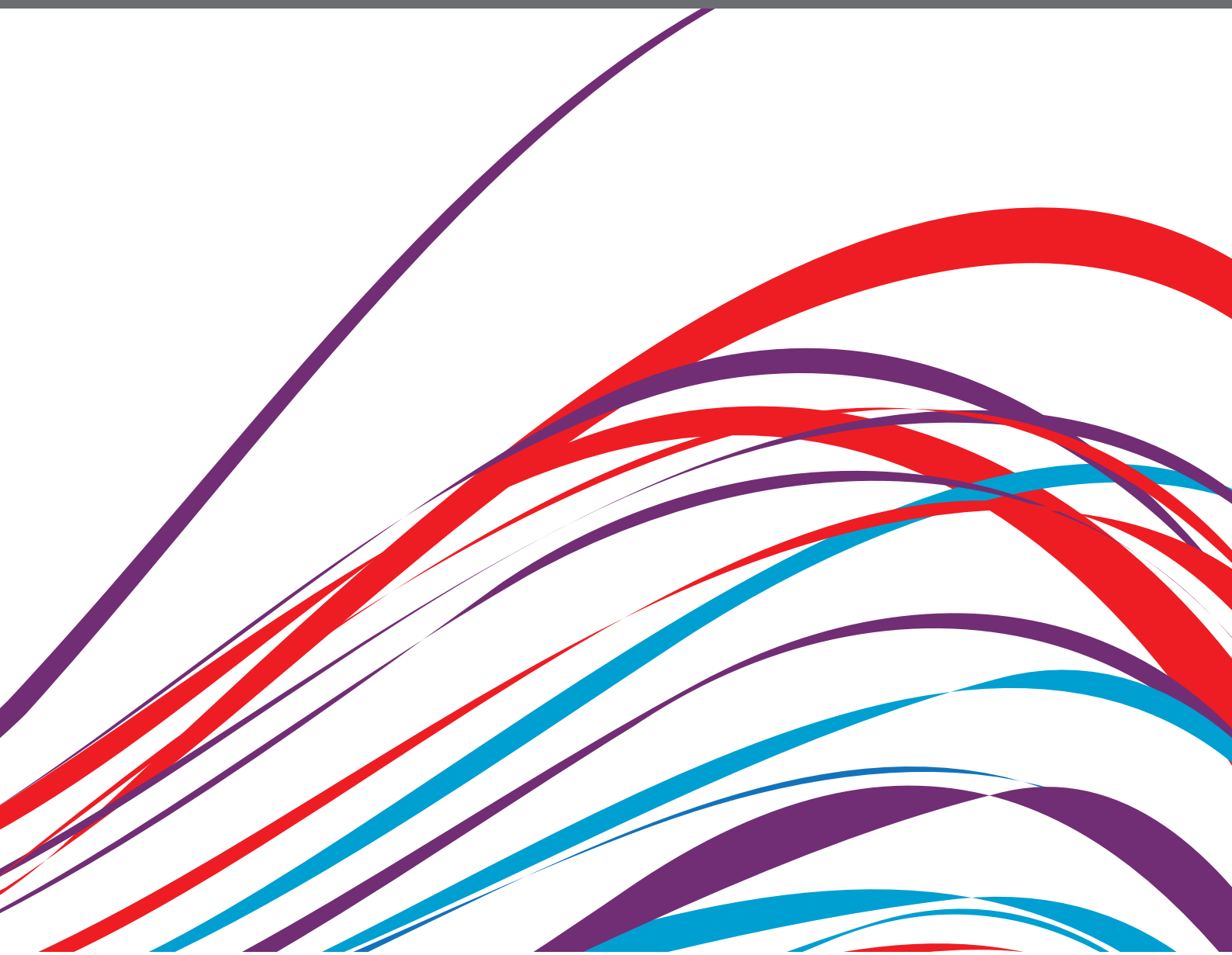
TRANSLATING ARTIFICIAL INTELLIGENCE INTO CLINICAL USE WITHIN CARDIOLOGY

EDITED BY: Paul Leeson, Shane Nanayakkara and Pablo Lamata
PUBLISHED IN: Frontiers in Cardiovascular Medicine and
Frontiers in Medical Technology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-735-9
DOI 10.3389/978-2-88976-735-9

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: frontiersin.org/about/contact

TRANSLATING ARTIFICIAL INTELLIGENCE INTO CLINICAL USE WITHIN CARDIOLOGY

Topic Editors:
Paul Leeson, University of Oxford, United Kingdom
Shane Nanayakkara, Alfred Hospital, Australia
Pablo Lamata, King's College London, United Kingdom

Citation: Leeson, P., Nanayakkara, S., Lamata, P., eds. (2022). Translating Artificial Intelligence Into Clinical Use Within Cardiology. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-735-9

Table of Contents

05 ***Editorial: Translating artificial intelligence into clinical use within cardiology***
Paul Leeson, Shane Nanayakkara and Pablo Lamata

08 ***Calibration-Free Cuffless Blood Pressure Estimation Based on a Population With a Diverse Range of Age and Blood Pressure***
Syunsuke Yamanaka, Koji Morikawa, Hiroshi Morita, Ji Young Huh and Osamu Yamamura

19 ***Machine Learning Augmented Echocardiography for Diastolic Function Assessment***
Andrew J. Fletcher, Winok Lapidaire and Paul Leeson

30 ***Improving Diuretic Response in Heart Failure by Implementing a Patient-Tailored Variability and Chronotherapy-Guided Algorithm***
Ariel Kenig, Yotam Kolben, Rabea Asleh, Offer Amir and Yaron Ilan

44 ***Modified GAN Augmentation Algorithms for the MRI-Classification of Myocardial Scar Tissue in Ischemic Cardiomyopathy***
Umesh C. Sharma, Kanhao Zhao, Kyle Mentkowski, Swati D. Sonkawade, Badri Karthikeyan, Jennifer K. Lang and Leslie Ying

55 ***Deep Learning for Classification and Selection of Cine CMR Images to Achieve Fully Automated Quality-Controlled CMR Analysis From Scanner to Report***
Vittoria Vergani, Reza Razavi, Esther Puyol-Antón and Bram Ruijsink

66 ***Deep Learning Predicts Heart Failure With Preserved, Mid-Range, and Reduced Left Ventricular Ejection Fraction From Patient Clinical Profiles***
Mohanad Alkhodari, Herbert F. Jelinek, Angelos Karlas, Stergios Soulaidopoulos, Petros Arsenos, Ioannis Doundoulakis, Konstantinos A. Gatzoulis, Konstantinos Tsioufis, Leontios J. Hadjileontiadis and Ahsan H. Khandoker

79 ***MOCOnet: Robust Motion Correction of Cardiovascular Magnetic Resonance T1 Mapping Using Convolutional Neural Networks***
Ricardo A. Gonzales, Qiang Zhang, Bartłomiej W. Papież, Konrad Werys, Elena Lukaschuk, Iulia A. Popescu, Matthew K. Burrage, Mayooran Shanmuganathan, Vanessa M. Ferreira and Stefan K. Piechnik

88 ***Machine Learning for the Prediction of Complications in Patients After Mitral Valve Surgery***
Haiye Jiang, Leping Liu, Yongjun Wang, Hongwen Ji, Xianjun Ma, Jingyi Wu, Yuanshuai Huang, Xinhua Wang, Rong Gui, Qinyu Zhao and Bingyu Chen

99 ***Corrigendum: Machine Learning for the Prediction of Complications in Patients After Mitral Valve Surgery***
Haiye Jiang, Leping Liu, Yongjun Wang, Hongwen Ji, Xianjun Ma, Jingyi Wu, Yuanshuai Huang, Xinhua Wang, Rong Gui, Qinyu Zhao and Bingyu Chen

100 ***Corrigendum: Machine Learning for the Prediction of Complications in Patients After Mitral Valve Surgery***
Haiye Jiang, Leping Liu, Yongjun Wang, Hongwen Ji, Xianjun Ma, Jingyi Wu, Yuanshuai Huang, Xinhua Wang, Rong Gui, Qinyu Zhao and Bingyu Chen

101 ***The Role of AI in Characterizing the DCM Phenotype***
Clint Asher, Esther Puyol-Antón, Maleeha Rizvi, Bram Ruijsink, Amedeo Chiribiri, Reza Razavi and Gerry Carr-White

121 ***Machine Learning for Clinical Decision-Making: Challenges and Opportunities in Cardiovascular Imaging***
Sergio Sanchez-Martinez, Oscar Camara, Gemma Piella, Maja Cikes, Miguel Ángel González-Ballester, Marius Miron, Alfredo Vellido, Emilia Gómez, Alan G. Fraser and Bart Bijnens

132 ***WSSNet: Aortic Wall Shear Stress Estimation Using Deep Learning on 4D Flow MRI***
Edward Ferdian, David J. Dubowitz, Charlene A. Mauger, Alan Wang and Alistair A. Young

151 ***EP-PINNs: Cardiac Electrophysiology Characterisation Using Physics-Informed Neural Networks***
Clara Herrero Martin, Alon Oved, Rasheda A. Chowdhury, Elisabeth Ullmann, Nicholas S. Peters, Anil A. Bharath and Marta Varela

166 ***A Collaborative Approach for the Development and Application of Machine Learning Solutions for CMR-Based Cardiac Disease Classification***
Markus Huellebrand, Matthias Ivantsits, Lennart Tautz, Sebastian Kelle and Anja Hennemuth

179 ***Predicting the Prognosis of Patients in the Coronary Care Unit: A Novel Multi-Category Machine Learning Model Using XGBoost***
Xingchen Wang, Tianqi Zhu, Minghong Xia, Yu Liu, Yao Wang, Xizhi Wang, Lenan Zhuang, Danfeng Zhong, Jun Zhu, Hong He, Shaoxiang Weng, Junhui Zhu and Dongwu Lai

192 ***Left Ventricular Pressure Estimation Using Machine Learning-Based Heart Sound Classification***
Philip Westphal, Hongxing Luo, Mehrdad Shahmohammadi, Luuk I. B. Heckman, Marion Kuiper, Frits W. Prinzen, Tammo Delhaas and Richard N. Cornelussen

203 ***Development of an Electronic Frailty Index for Predicting Mortality and Complications Analysis in Pulmonary Hypertension Using Random Survival Forest Model***
Jiandong Zhou, Oscar Hou In Chou, Ka Hei Gabriel Wong, Sharen Lee, Keith Sai Kit Leung, Tong Liu, Bernard Man Yung Cheung, Ian Chi Kei Wong, Gary Tse and Qingpeng Zhang

frontiers | Frontiers in Cardiovascular Medicine

TYPE Editorial
PUBLISHED 04 August 2022
DOI 10.3389/fcvm.2022.995234

OPEN ACCESS

EDITED AND REVIEWED BY
Ngan F. Huang,
Stanford University, United States

*CORRESPONDENCE
Paul Leeson
paul.leeson@cardiov.ox.ac.uk

SPECIALTY SECTION
This article was submitted to Cardiovascular Biologics and Regenerative Medicine, a section of the journal Frontiers in Cardiovascular Medicine

RECEIVED 15 July 2022
ACCEPTED 25 July 2022
PUBLISHED 04 August 2022

CITATION
Leeson P, Nanayakkara S and Lamata P (2022) Editorial: Translating artificial intelligence into clinical use within cardiology.
Front. Cardiovasc. Med. 9:995234.
doi: 10.3389/fcvm.2022.995234

Editorial: Translating artificial intelligence into clinical use within cardiology

Paul Leeson[1]*, Shane Nanayakkara[2,3] and Pablo Lamata[4]

[1]Cardiovascular Clinical Research Facility, RDM Division of Cardiovascular Medicine, University of Oxford, Oxford, United Kingdom, [2]Department of Cardiology, The Alfred, Melbourne, VIC, Australia, [3]Baker Heart and Diabetes Institute, Melbourne, VIC, Australia, [4]Department of Biomedical Engineering, School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom

KEYWORDS

artificial intelligence, cardiology, heart failure, cardiovascular imaging, disease prediction

Editorial on the Research Topic
Translating artificial intelligence into clinical use within cardiology

Introduction

In the foreword to "The Seeds of Artificial Intelligence," Gregory Freiherr sets out a vision for artificial intelligence in medicine (1): "*..the intelligent machine, a device that mimics the expert's reasoning power and can retain in retrievable power much of the knowledge currently available to experts in a given specialty.*" He also set out a timetable for when these systems could reach clinical practice: "*Most systems of this type are still immature. But some are already moving into the real world and others will make the transition within the next few years.*" The book was written in 1980 and the "*next few years*" has extended to "*40 years*" but we have now, arguably, reached that threshold, where artificial intelligence is at a stage of development when it could transition into routine application in medicine (2).

Throughout this period cardiology, and in particular cardiovascular imaging, has been at the forefront of understanding how to apply computational approaches to common medical tools such as an ECG recording or an ultrasound image (3). Now there is a final hurdle to overcome to ensure these solutions impact clinical practice. How do we translate them effectively into regular clinical use? This Research Topic was designed to provide an opportunity for authors to present their solutions to this problem. The topic was kept broad to attract submissions about both innovations and applications and, across the wide variety of papers submitted, several themes have emerged. These reflect the most active academic areas in the application of artificial intelligence to cardiovascular medicine.

Cardiovascular imaging and cardiovascular magnetic resonance

Cardiovascular imaging has been an exemplar area of cardiovascular research for adoption of artificial intelligence (3). The volume of data and natural collaborations between academic centers internationally have allowed testing and development of different approaches using large scale patient datasets. Research efforts such as UK Biobank, with hundreds of thousands of freely available cardiovascular magnetic resonance image datasets have also accelerated development of solutions (4). The analysis of images using artificial intelligence is also central to many of the secular applications of artificial intelligence during for example internet searches. Therefore, the technology solutions are well-advanced. In this Research Topic, Sanchez-Martinez et al. provide a review that brings together understanding of how machine learning can be applied to decision making in cardiac imaging. Furthermore, four new technical approaches to segment and analyse cardiovascular magnetic resonance scans for different disease states are presented. Approaching the problem from a more generalized angle, Vergani et al. present a deep learning approach for classification of images while Huellebrand et al. provide a holistic analysis of methods for disease classification. More focused on specific cardiovascular magnetic resonance sequences, Sharma et al. present a method for classification of myocardial scar identified with late gadolinium imaging, a stalwart of the modality for the last 20 years. In comparison Gonzales et al. present a new way to manage automation of T1 mapping, an emerging option for gadolinium-free scar imaging. Together these manuscripts describe innovations across the whole pipeline of image handling from image sorting and classification to disease identification.

Signal processing in cardiovascular medicine

Technological solutions using artificial intelligence, however, clearly extend beyond image handling. Intelligent signal processing is an area of rapid development and this has been effectively presented in several papers within the Research Topic. Electrophysiology is well-known for its analysis of ECG signals and Herrero Martin et al. present a framework for characterization of signals to define cardiac disease. Three others paper approach another well-known signal processing problem, the characterization of cardiac and vascular pressures. Yamanaka et al. have tackled a problem with a very large potential application footprint. Blood pressure measurement is relevant to both the assessment of health and for the identification or monitoring of disease. The ability to do this without the need of an inflatable cuff would transform daily practice and their early data shows new ways for cuffless blood pressure estimation. Westphal et al. take a different approach to understanding pressures within the cardiovascular system, with a similarly widespread potential clinical footprint. Heart sounds are readily accessible with a stethoscope but the ability to estimate left ventricular pressure based off heart sounds is a novel application of machine learning. The other region of the cardiovascular system in which pressures and stresses can have important clinical implications is the aorta. Increased aortic shear stress has been implicated in aortic remodeling as well as dissection and atheroma development. Ferdian et al. present new data on technological approaches to estimate aortic shear stress from simply acquired parameters. These solutions provide new approaches to extract information from clinical signals. Several of these measures are not in routine clinical use but if these methods increase availability they may encourage translation into practice.

Heart failure diagnosis and management

With regard to disease areas, heart failure has proved to be the most popular for application of new artificial intelligence management approaches within our Research Topic. Heart failure represents the end stage of most cardiac conditions and has a heterogenous pathology with numerous potential management approaches. Identifying the type of heart failure is critical and identifying the disease process early is important to prevent irreparable damage. Identification of heart failure has been addressed in three papers in this Research Topic. Two focus on imaging with Fletcher et al. having reviewed the possibility for artificial intelligence to augment the power of echocardiography for identification of heart failure with preserved ejection fraction, while, Asher et al. present a review covering the possibility for imaging classification of dilated cardiomyopathy phenotypes. In contrast, Alkhodari et al. have explored whether deep learning based on patient records is effective for classification of heart failure profiles. Finally, Kenig et al. have tackled the question of management approaches and present an algorithm for assessment and titration of diuretic treatment in heart failure. Heart failure is an increasing problem within healthcare and the reviews suggest there remains a huge untapped potential for applications of artificial intelligence in this space.

Disease characterization, prognosis, and prediction

In a final theme of the Research Topic four papers address the ongoing question of how artificial intelligence can be applied to characterize disease and predict outcomes. These

papers cover an extensive range of topics and provide some of the most challenging potential applications. The ability to predict the future and identify patients at high risk is hugely difficult but may be tractable with artificial intelligence processing techniques. Zhou et al. have simply modeled using existing data what happens to patients who have pulmonary hypertension to try and understand who may be most at risk of frailty-related problems. Wang et al. have also used patient-related data to look at outcomes from Coronary Care Units and Jiang et al. have adopted the same approach to look at outcomes after mitral valve surgery. These papers show promise for predictive algorithms based on patient acquired data but further clinical validation is required in independent datasets.

Conclusion

When we proposed this Research Topic we had a vision to present a broad range of papers promoting real world clinical applications of artificial intelligence in medicine. We were unsure of what would emerge from the clinical and academic community but the spread of topics and technical solutions is impressive. Interestingly, many are not yet ready for clinical application and there is clearly still some way to go to achieve our goal of clinically translatable artificial intelligence in cardiology. However, the first steps are being made and we look forward to a time when, in Eric Topol's words (2): "*Eventually, doctors will adopt AI and algorithms as their work partners. This leveling of the medical knowledge landscape will ultimately lead to a new premium: to find and train doctors who have the highest level of emotional intelligence.*"

Author contributions

PLe drafted the manuscript. SN and PLa edited for important intellectual content. All authors contributed to the article and approved the submitted version.

Funding

PLe was supported by the Oxford NIHR Biomedical Research Centre and Oxford BHF Centre for Research Excellence.

Conflict of interest

PLe is a founder and shareholder of Ultromics, an AI imaging company, and has patents related to clinical applications of AI.

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.

Publisher's note

References

1. Freiherr G. *The Seeds of Artificial Intelligence: SUMEX-AIM*. Bethesda, MD; Washington, DC: U.S. Dept. of Health, Education, and Welfare, Public Health Service; National Institutes of Health (1980).

2. Topol EJ. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. 1st Edn. New York, NY: Basic Books (2019). p. 378.

3. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP, et al. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *J Am Coll Cardiol*. (2019). 73:1317–35. doi: 10.1016/j.jacc.2018.12.054

4. Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat Commun*. (2020) 11:2624. doi: 10.1038/s41467-020-15948-9

frontiers in Medical Technology

ORIGINAL RESEARCH
published: 27 July 2021
doi: 10.3389/fmedt.2021.695356

Calibration-Free Cuffless Blood Pressure Estimation Based on a Population With a Diverse Range of Age and Blood Pressure

Syunsuke Yamanaka[1], *Koji Morikawa*[2], *Hiroshi Morita*[1], *Ji Young Huh*[3] and *Osamu Yamamura*[4]*

[1] Department of Emergency Medicine, General Internal Medicine, University of Fukui Hospital, Fukui, Japan, [2] Connect Inc., Tokyo, Japan, [3] Emergency and Critical Care Center, Kobe City Medical Center General Hospital, Kobe, Japan, [4] Second Department of Internal Medicine, University of Fukui Hospital, Fukui, Japan

This study presents a new blood pressure (BP) estimation algorithm utilizing machine learning (ML). A cuffless device that can measure BP without calibration would be precious for portability, continuous measurement, and comfortability, but unfortunately, it does not currently exist. Conventional BP measurement with a cuff is standard, but this method has various problems like inaccurate BP measurement, poor portability, and painful cuff pressure. To overcome these disadvantages, many researchers have developed cuffless BP estimation devices. However, these devices are not clinically applicable because they require advanced preparation before use, such as calibration, do not follow international standards (81060-1:2007), or have been designed using insufficient data sets. The present study was conducted to combat these issues. We recruited 127 participants and obtained 878 raw datasets. According to international standards, our diverse data set included participants from different age groups with a wide variety of blood pressures. We utilized ML to formulate a BP estimation method that did not require calibration. The present study also conformed to the method required by international standards while calculating the level of error in BP estimation. Two essential methods were applied in this study: (a) grouping the participants into five subsets based on the relationship between the pulse transit time and systolic BP by a support vector machine ensemble with bagging (b) applying the information from the wavelet transformation of the pulse wave and the electrocardiogram to the linear regression BP estimation model for each group. For systolic BP, the standard deviation of error for the proposed BP estimation results with cross-validation was 7.74 mmHg, which was an improvement from 17.05 mmHg, as estimated by the conventional pulse-transit-time-based methods. For diastolic BP, the standard deviation of error was 6.42 mmHg for the proposed BP estimation, which was an improvement from 14.05mmHg. The purpose of the present study was to demonstrate and evaluate the performance of the newly developed BP estimation ML method that meets the international standard for non-invasive sphygmomanometers in a population with a diverse range of age and BP.

Keywords: continuous blood pressure, electrocardiogram, pulse wave, wavelet transformation, machine learning, cuffless

OPEN ACCESS

Edited by:
Shane Nanayakkara, Alfred Hospital, Australia

Reviewed by:
Choon-Sik Jhun, The Pennsylvania State University, United States
Robert Gaul, Trinity College Dublin, Ireland

***Correspondence:**
Osamu Yamamura
kapi@u-fukui.ac.jp

Specialty section:
This article was submitted to Cardiovascular Medtech, a section of the journal Frontiers in Medical Technology

Received: *14 April 2021*
Accepted: *30 June 2021*
Published: *27 July 2021*

Citation:
Yamanaka S, Morikawa K, Morita H, Huh JY and Yamamura O (2021) Calibration-Free Cuffless Blood Pressure Estimation Based on a Population With a Diverse Range of Age and Blood Pressure. Front. Med. Technol. 3:695356. doi: 10.3389/fmedt.2021.695356

INTRODUCTION

Non-invasive blood pressure (BP) measurement with cuff-based devices is widely used, and these devices are necessary for various medical situations (1). However, there are some disadvantages of the cuff-based BP measurement methods: (i) a study showed that three out of 10 home cuff-based BP measurement devices were inaccurate (2); (ii) the measurement is usually intermittent and does not capture all the BP changes occurring throughout the measurement period; (iii) the current cuff-based BP devices are still bulky, and are not portable or practical for daily or long-term use (3, 4); (iv) cuff pressure can be painful for some patients; it can also interrupt their state of rest. The cuff pressure results in difficulty for measuring BP during sleep or everyday life or, even worse, may affect the BP measurement itself (5).

Many researchers have developed cuffless BP estimation devices to overcome these disadvantages, allowing patients to monitor BP continuously (6–8). Pulse transit time (PTT) is the pulse wave propagating time from two separate arterial sites on the same cardiac cycle (9, 10), which usually needs to be examined using a continuous electrocardiogram (ECG) (11). The PTT indirectly depends on blood pressure; the higher the pressure, the faster the PTT (11). This phenomenon has been used for non-invasive BP estimation. However, many conventional cuffless PTT-based BP estimation studies have some drawbacks, divided into four categories: (1) analysis of biased, (2) small datasets, (3) studies with devices that required calibration, and (4) insufficient accuracy as required by international standards. Some studies analyzed biased data that included only young participants with a narrow blood pressure range (12, 13), while others had insufficient participants (13–18). Other studies needed additional advanced preparations such as frequent calibrations (19–21), required additional parameters (22, 23), or used devices that needed to be anchored to the body, resulting in annoyance for some users (24). Besides, some studies either had a wide gap of means and standard deviations from the reference (25, 26), or a low regression coefficient (R^2) (27, 28), resulting in the need for more precise mathematical models (29). **Table 1** shows the previous studies for non-invasive cuffless BP estimation. Poon and Zhang's cuffless BP measurement was the only study that had handled a large variety of participants' blood pressures (39 with hypertension), range of age (57 $\pm$ 27 years old), and large participant number (85 participants) (21). The estimated SBP and DBP in the Poon and Zhang's method differed from the reference BP by 0.6 $\pm$ 9.8 mmHg and 0.9 $\pm$ 5.6 mmHg, respectively. However, the BP estimation method in Poon and Zhang's study had major shortcomings to need a calibration procedure for each participant, and the accuracy of BP estimation was not precise.

An international standard has already been formulated for cuffless sphygmomanometers (ISO 81060-1:2007) (31). ISO standard is an international standard that must be met when releasing the cuffless, non-invasive blood pressure estimation model as a medical device to the market in the future. However, to the best of our knowledge, a cuffless BP estimation model that meets the international standard does not currently exist. The purpose of the present study was to demonstrate and evaluate the performance of the newly developed ML method for BP estimation that meets the international standard for non-invasive sphygmomanometers in a population with a diverse range of age and BP.

METHODS

Participants

We recruited 127 participants (73 males and 54 females) at the University of Fukui Hospital and its affiliated institutions. The study was conducted with the approval of the Research Ethics Committee of the University of Fukui (Approval Number: 20148035). Written informed consent was obtained from all the participants. All participants were asked to fill out a medical form that included sex, date of birth, past medical history, and current medications before measuring participant BP. We excluded those participants who were either pregnant, <18 years old, or had a persistent arrhythmia. We recruited participants with a specified range of BP, as required by a protocol of the International Standard of Non-invasive Sphygmomanometers (ISO 81060-1:2007) (31).

Experimental System

A biopotential sensing system was developed to measure ECG and pulse wave simultaneously, as shown in **Figure 1**. The system consisted of (A) a biopotential amplification device and data transmitter, (B) an ECG electrical potential electrode, (C) a pulse wave sensor and the second ECG electrical potential electrode, (D) a receiving dongle, and (E) a personal computer for data recording. The personal computer recorded and analyzed the waveform of the ECG and the pulse wave. Participants held one of the two ECG sensor electrodes with their thumb and index finger of the left hand (B in **Figure 1**). Another sensor electrode that could sense the ECG and the pulse wave simultaneously was attached to the index finger of the right hand in the sitting position (C in **Figure 1**). The sampling frequencies and the resolution of these waveforms were 1024 Hz and 12 bits, respectively.

Data Collection Protocol

We collected the reference and analyzed datasets according to the protocol of ISO81060-1:2007 (31). We acquired the reference and analyzed datasets from the participants in the sitting position at room temperature without disturbing influences. The left arm of the participants was used for the reference measurement. The protocol for collecting data is shown in **Figure 2**. After a 5-min rest to stabilize the BP, the first reference BP measurement was taken. Then, the participants took a 1-min rest after the first reference measurement to avoid venous congestion. The measurement was then taken with the device being tested, after the first reference measurement but before the second reference measurement. Participants retook a 1-min rest after the test device data acquisition. After the rest, the second reference measurement was taken. We took the measurement using the test device in each participant at least three times.

TABLE 1 | Prior non-invasive cuffless blood pressure estimation studies.

References	Source	# of participants	Range of age	Methods	MAE SBP	MAE DBP
Gao et al. (30)	PPG	65	22–65	W SVM	5.1 ± 4.3	4.6 ± 4.3
Chen et al. (25)	BCG ECG	51	20–74	AS	9.0 ± 5.6	1.8 ± 1.3
Chan et al. (26)	ECG PPG PTT	/	/	AS	7.5 ± 8.8	4.1 ± 5.6
Ding et al. (13)	PTT PPG	27	21–29	AS	−0.4 ± 5.2	−0.1 ± 4.0
Chen et al. (15, 16)	PTT	23/26	19–60	AS	2.2 ± 6.2	−1.5 ± 6.5
Poon and Zhan (21)	PTT	85	/ (57 ± 27)	AS	0.6 ± 9.8	0.9 ± 5.6

MAE, mean absolute error; SBP, systolic blood pressure; DBP, diastolic blood pressure; PPG, photoplethysmogram; ECG, electrocardiogram; PTT, pulse transit time; BCG, ballistocardiograph; AS, analytical solution; SVM, support vector regression machine; W, wavelet.

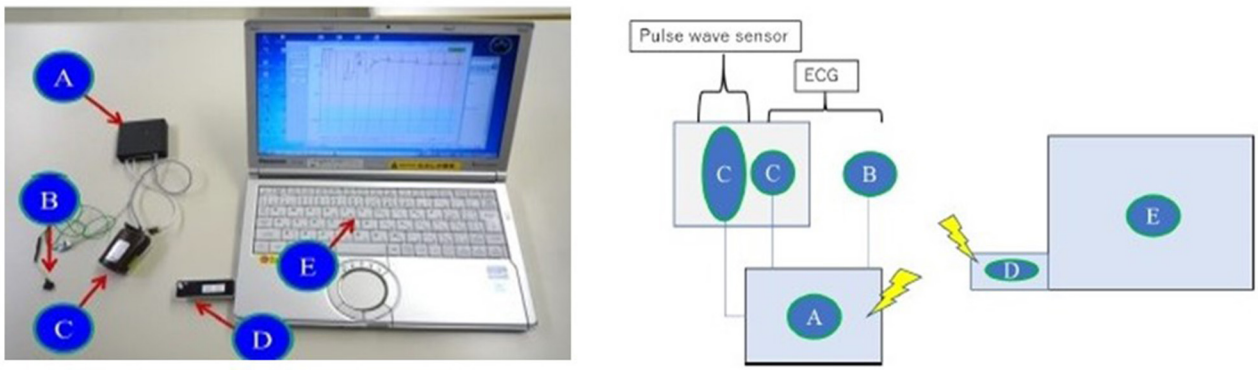


FIGURE 1 | Biopotential sensing system. (A) Biopotential amplification device and data transmitter, (B) First ECG electrical potential electrode, (C) Pulse wave sensor and the second ECG electrical electrode, (D) Receiving dongle, (E) Personal computer for data recording.

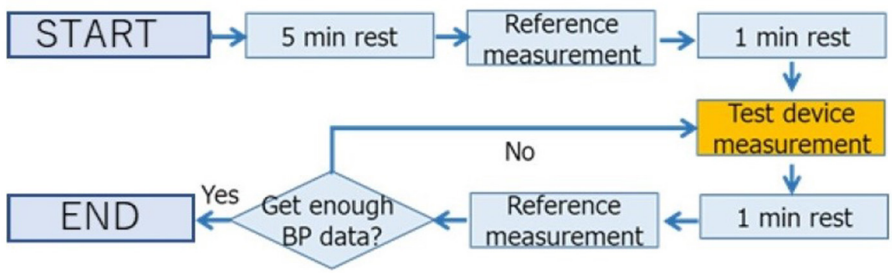


FIGURE 2 | Protocol for collecting data. We took the measurement using the test device in each participant at least three times.

Reference Data

Two medically trained observers measured the reference BP simultaneously with one reference mercury sphygmomanometer using a "Y" connector that lets two observers measure one participant's BP. The systolic blood pressure (SBP) and diastolic blood pressure (DBP) were determined by phase 1 and phase 5 of Korotkoff sounds, respectively. All measurements were recorded to the nearest 2 mmHg. If the values of SBP and DBP as measured by the two observers were <4 mmHg apart, the mean value of the BP was calculated from the observed values and used as the reference data. If the difference in the measured BP between the two observers was more significant than 4 mmHg, we excluded both the reference and the test device data. These reference BP acquisition methods were stipulated by the ISO standard (31).

Preparation of the Data Set for Machine Learning

Due to the quality of the measurement, some data were not suitable for data analysis, included either poor ECG, poor pulse wave signals, significant blood pressure changes between

the reference measurements taken before and after the test device BP measurement. We applied several structured criteria to the dataset for the preprocessing of machine learning, as shown below:

Exclusion Criteria Based on Reference Data

We excluded the data according to exclusion criteria, as designated by the ISO standard for the reference dataset (31); (a) if the difference between the reference measurements taken by the two observers was more than 4 mmHg, (b) if the difference between the reference measurements before and after the test device measurement was more than 12 mmHg for SBP or more than 8 mmHg for DBP, (c) if three valid datasets could not be acquired from one participant due to any reason (e.g., unstable BP).

Exclusion Criteria Based on Waveform

We excluded the following cases from the data analysis: (a) the amplitude of the ECG or the pulse waveform was too low due to the dryness of the participant's hands, (b) the crest of the ECG and the pulse waveform was not clear, (c) steady noise (mainly caused due to power supply noise) interfering with the waveform due to unstable electrode holding, and (d) a wavelet calculation error occurring either in the ECG or the pulse waveform mainly caused due to measurement errors. A specific cutoff value was set for each exclusion criteria so that data could be excluded objectively.

Data Selection to Satisfy the ISO Protocol

The proportion of participants with high blood pressure and those with low blood pressure was designated according to the protocol of ISO80601-1:2007 (31). ISO requests a wide range of blood pressure proportions to demonstrate the device's applicability to a wide range of participants with various blood pressures. Since we acquired many datasets within the normal range of blood pressure, we needed to exclude some of these datasets from the other datasets for an accurate analysis. The datasets containing large PTT fluctuations and low signal quality of ECG were excluded in descending order until the datasets met the ISO standard protocol.

Final Data Set for Training and Evaluation

One hundred and twenty-seven participants with 878 datasets remained in the study. We excluded 21 participants with 423 datasets because of the exclusion criteria based on reference BP data. Nineteen participants with 127 datasets were excluded because of the exclusion criteria based on waveform. Among the 85 patients with 328 datasets, 68 datasets were excluded (66 datasets were normal BP range and 2 datasets were SBP $\geq$140 mmHg) in descending order of PTT fluctuation and exclude one participant for meeting BP distribution of the ISO standard. Finally, we acquired 260 datasets from 84 participants. The ISO protocol requires more than 85 participants with more than 255 datasets. In this study, we prioritized the distribution of BP values with a certain proportion of participants with high and low BP participants because we aimed to confirm the ability of the device to apply to a wide variety of BP values. The participant's BP distributions are shown in **Table 2**. We also intended that the device should apply to diverse age groups. The ISO protocol requires the participants' age to range from 18 years old to 65 years old. We collected datasets from young and older participants, and our age range was more expansive than the requirements of the ISO protocols. **Table 3** shows the age distribution of the participants.

TABLE 2 | Blood pressure distribution of participants.

	ISO standard	Acquired data
Participant	85	84
Valid data set	255	260
Gender ratio	Each ≥30%	Female 54 (64.3%)
SBP ≤ 100 mmHg	≥13 data/5%	39 data (15.0%)
SBP ≥ 160 mmHg	≥13 data/5%	16 data (6.2%)
SBP ≥ 140 mmHg	≥52 data/20%	52 data (20.0%)
DBP ≤ 60 mmHg	≥13 data/5%	43 data (16.5%)
DBP ≤ 100 mmHg	≥13 data/5%	13 data (5.0%)
DBP ≤ 85 mmHg	≥52 data/20%	71 data (27.3%)

SBP, systolic blood pressure; DBP, diastolic blood pressure.

TABLE 3 | Age distribution of participants.

Age (years)	58.1 ± 16.1
20∼24	2 (2.4%)
25∼29	2 (2.4%)
30∼34	5 (6.0%)
35∼39	5 (6.0%)
40∼44	4 (4.8%)
45∼49	5 (6.0%)
50∼54	7 (8.3%)
55∼59	8 (10.0%)
60∼64	13 (15.5%)
65∼69	12 (14.3%)
70∼74	7 (13.0%)
75∼79	7 (13.0%)
80∼84	3 (3.6%)
85∼	4 (4.8%)

Implementation of the BP Estimation Algorithm

The proposed algorithm consists of four steps: (i) waveform preparation, (ii) participant group classification, (iii) feature extraction, and (iv) BP value estimation according to the selected participants' group.

Waveform Preparation

The ECG and the pulse waveforms were averaged to a single waveform in a 10-s window to decrease the difference between each pulse. Due to the heart rate fluctuations, the pulse intervals were normalized and extended to 1 s. Normalization was also applied to the amplitude of the wave due to the amplitude fluctuation for each pulse. By multiplying the amplitude peak

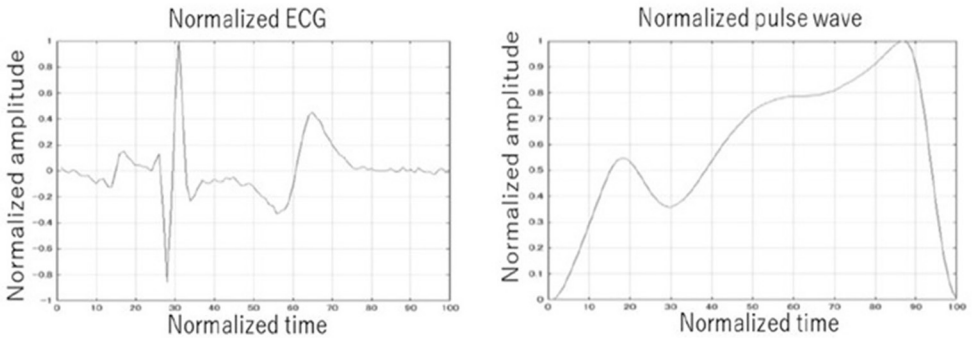


FIGURE 3 | Example of normalized ECG and pulse wave. The ECG and the pulse waveforms were averaged to a single waveform in a 10-second window. The pulse intervals were normalized and extended to one second. Normalization was also applied to the amplitude of the wave.

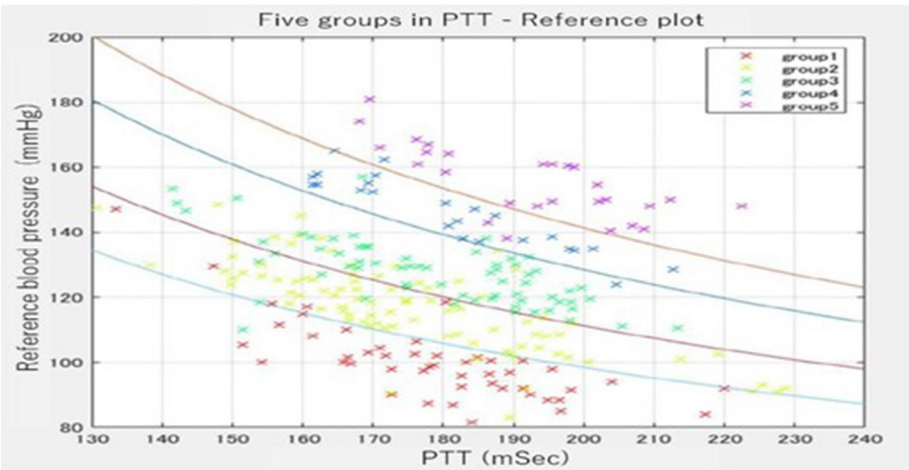


FIGURE 4 | Five groups were divided by group estimator in the PTT-reference plot. The participants were distributed 1:1:3:3:2 in the five areas surrounded by the four curves.

by the multiplication coefficient, the amplitude peak of the waveform is calculated. The R wave peak was defined as the largest peak of the ECG wave from the baseline before and after its generation. As the T wave in ECG is also a big positive wave and we needed to recognize the R wave and the T wave correctly, the threshold was set to 0.7 from the baseline in the normalized amplitude width. The wave that was bigger than 0.7 was recognized as the R wave, and the wave that was smaller than 0.7 just after R wave was recognized as the T wave. The similar procedure was also performed for the pulse waveform. Examples of the normalized ECG and the normalized pulse waveforms are shown in **Figure 3**.

Participants' Group Classification

We divided the entire data into five categories (group 1, group 2, group 3, group 4, and group 5.) at a ratio of 1:1:3:3:2 with four curved borders. **Figure 4** shows the distribution of sample data divided by the four curved borders. The four curved borders divide the entire data into five subgroups, such as group 1 for the very high BP subgroup, group 2 for the high BP subgroup, group 3 for the moderate BP subgroup, group 4 for the low BP subgroup, and group 5 for the very low BP subgroup at the ratio of 1:1:3:3:2. We noticed that even with almost the same PTT, the BP differed significantly among participants through data collection. However, BP was mainly stable in the same participant. The group classification formula was derived from the relationship between the BP tendency and the waveform features. We expected the classified groups to function as rough BP estimators from waveform features and PTT alone. The ratio of the number of samples was determined as 1:1:3:3:2 while shifting the coefficient term of the 1/PTT line for SBP and PTT. Since we drew four curved borders evenly spaced, subgroup 1(very high BP subgroup) contains very few participants. Hence, we drew the bottom three curves evenly spaced and the top curved border 50% nearly to the second one. As a result, the participants were distributed 1:1:3:3:2 in the five areas surrounded by the four curves.

A support vector machine ensemble with bagging was selected as a group estimator (32, 33). The support vector machine is one machine learning and constructs a hyperplane or set of

TABLE 4 | Selected feature values.

Selected features	Waveform	Normalized time**	Normalized frequency**
1 * 1/ PTT	-	-	-
2. The square root of body weight	-	-	-
3. Wavelet coefficients	Pulse wave	11	7
4. Wavelet coefficients	ECG	14	3
5. Bodyweight	-	-	-
6. Wavelet coefficients	Pulse wave	19	1
7. Heart rate	-	-	-
8. Wavelet coefficients	Pulse wave	1	2
9. Wavelet coefficients	ECG	1	2
10. Wavelet coefficients	Pulse wave	17	1
11. Pulse transit time	-	-	-
12. Wavelet coefficients	ECG	19	2
13. Wavelet coefficients	Pulse wave	13	1
14. Wavelet coefficients	Pulse wave	14	1
15. Wavelet coefficients	Pulse wave	20	1
16. Wavelet coefficients	Pulse wave	5	1
17. Wavelet coefficients	ECG	5	1
18. Wavelet coefficients	ECG	8	1
19. Wavelet coefficients	ECG	19	1
20. Peak of ECG	ECG	-	-
21. Wavelet coefficients	ECG	20	4
22. Wavelet coefficients	Pulse wave	10	1
23. Wavelet coefficients	Pulse wave	8	1
24. Wavelet coefficients	Pulse wave	8	2

PTT, pulse transit time, ECG, electrocardiogram.
**Numbers are assigned in order of the effectiveness of features.*
***Wavelet coefficients of ECG/pulse wave: The time indicates the number from the beginning of the normalized time axis divided into 20 equal parts. The frequency indicates the number from the beginning of the normalized frequency axis divided into 8 equal parts.*

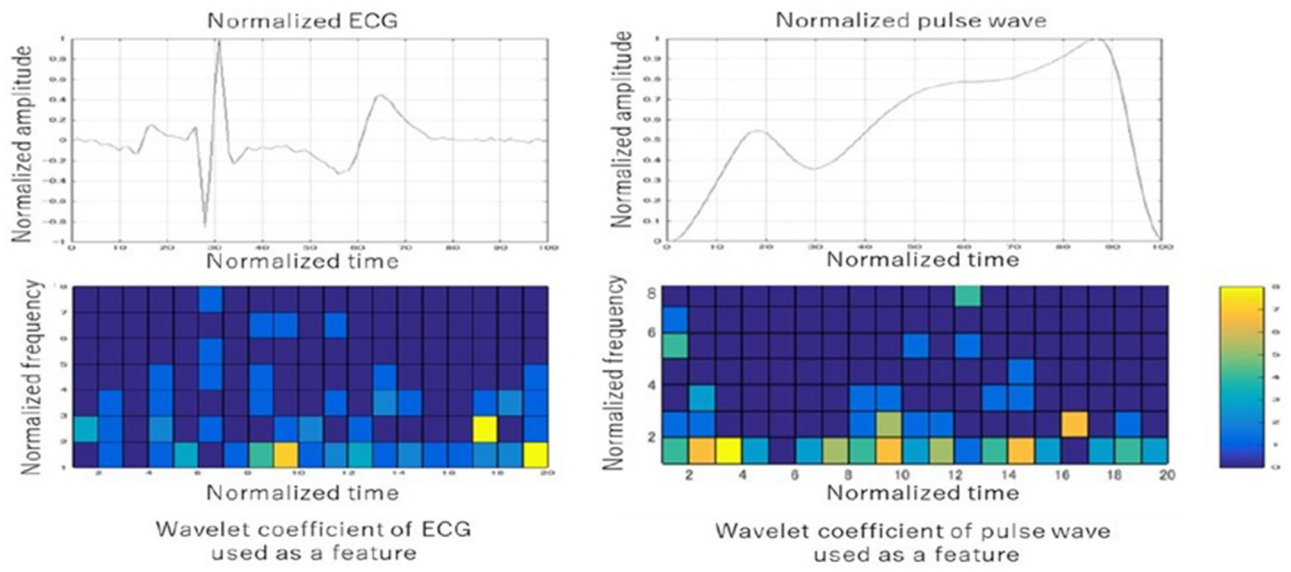


FIGURE 5 | Examples of the wavelet transformation of the ECG and pulse waveforms. The ST-segment and the baseline between the T wave and the P wave had frequently used the wavelet coefficient features in both the wavelet coefficients of the ECG and pulse waveforms.

hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outlier detection. We applied 10-fold cross-validation for learning. Cross-validation is a technique for assessing how the results of a statistical analysis generalize to an independent data set (34). Even the support vector machine (SVM) has been proposed to provide a good generalization performance; if we use only one SVM, the classification result of the practically implemented SVM is often far from the theoretically expected level. To improve the limited classification performance of the one SVM estimator, we prepared 35 support vector machine estimators with bagging (bootstrap technique). Each SVM was trained independently using the randomly chosen training samples via a bootstrap technique. Then, they were aggregated into to make a collective decision for participants' group clarification (32, 33).

Feature Extraction

The candidate features for machine learning consisted of the following features: participants basic information (age, body weight, height, etc.), PTT, basic information of second derivative photoplethysmogram (a, b, c, etc.), wavelet features from the ECG, and wavelet features from the pulse wave. The continuous wavelet transform was adopted for waveform analysis. We obtained the wavelet features from each cell by dividing the normalized wave of both the ECG and the pulse wave into 20 bands in the horizontal and eight bands in the vertical direction. In this way, we prepared 320 features (20 [the horizontal direction] $\times$ 8[the vertical direction] $\times$ 2[ECG, pulse wave]) as wavelet features from the ECG and the pulse wave. These features were used to create a BP estimation linear regression model for each group. **Table 4** shows the specific features that were used in machine learning in descending order. Although some traditional features such as 1/PTT, the square root of body weight, heart rate, PTT, and the peak of ECG were used, many newly-developed features such as wavelet coefficients of the ECG/pulse waveforms were also used in the models. The ST-segment and the baseline between the T wave and the P wave had frequently used the wavelet coefficient features in both the wavelet coefficients of the ECG and pulse waveforms. **Figure 5** shows one example of the wavelet transformation of the ECG and pulse waveforms.

BP Value Estimation From the Selected Participants' Group

The present study's model comprises two steps: grouping estimation by the SVM ensemble with bagging (mentioned above) and the BP estimation linear regression model for each group. The estimation formula for BP value was derived from the relationship between the feature values and the reference BP values. Since the BP estimation model was prepared for each group, five BP estimation models were learned from the 5-grouped datasets. Linear regression was used for the BP value estimator. The incremental feature value selection was applied to select useful features only. Feature value candidates consisted of the wavelet transformation of the ECG and the pulse waveforms, weight, height, body mass index (BMI), PTT, and 1/PTT.

TABLE 5 | Comparison of conventional PTT-based methods and the proposed method.

	Standard deviation of error in SBP	Standard deviation of error in DBP
*Measured BP	21.34 mmHg	14.65 mmHg
PTT-based methods	17.05 mmHg	14.05 mmHg
Proposed method	7.74 mmHg	6.42 mmHg

BP, blood pressure; SBP, systolic blood pressure; DBP, diastolic blood pressure; PTT, pulse transit time.
**Measured BP is the overall distribution of the BP data without any data treatment.*

Primary Outcome

The primary outcome in this study was the evaluation of the models formulated with the standard deviation of the error between the estimated BP value and the reference value. The 10-fold cross-validation was adopted for evaluating the model performance. The original sample was partitioned into 10 subsamples. Of the 10 subsamples, a single subsample was retrained as the validation data for testing the model, and the remaining nine subsamples were used as training data. The cross-validation process was repeated 10 times, with each of the 10 subsamples used exactly once as the validation data. The 10 standard deviation error results were averaged for the evaluation result (34).

RESULTS

Comparison of Conventional PTT-Based Method and the Proposed Method

Table 5 shows a comparison of the conventional PTT-based methods and the proposed method for the primary outcome. For SBP, the error of the standard deviation of the proposed BP estimation results with cross-validation was 7.74 mmHg, which was an improvement from 17.05 mmHg, as estimated by the conventional PTT-based methods. For DBP, the error of the standard deviation of the proposed BP estimation results with cross-validation was 6.42 mmHg, which was an improvement from 14.05 mmHg, as estimated by the conventional PTT-based methods. The association between the estimated BP by the proposed method and the measured reference BP for SBP and DBP are shown in **Figures 6**, **7**, respectively. The proposed method had small-range prediction values closer to the reference values compared to the PTT-based methods.

DISCUSSIONS

Comparison to Prior Work

This study aimed to propose a new BP estimation method that did not require calibration based on a wide range of age and BP distribution while assessing the participants according to the

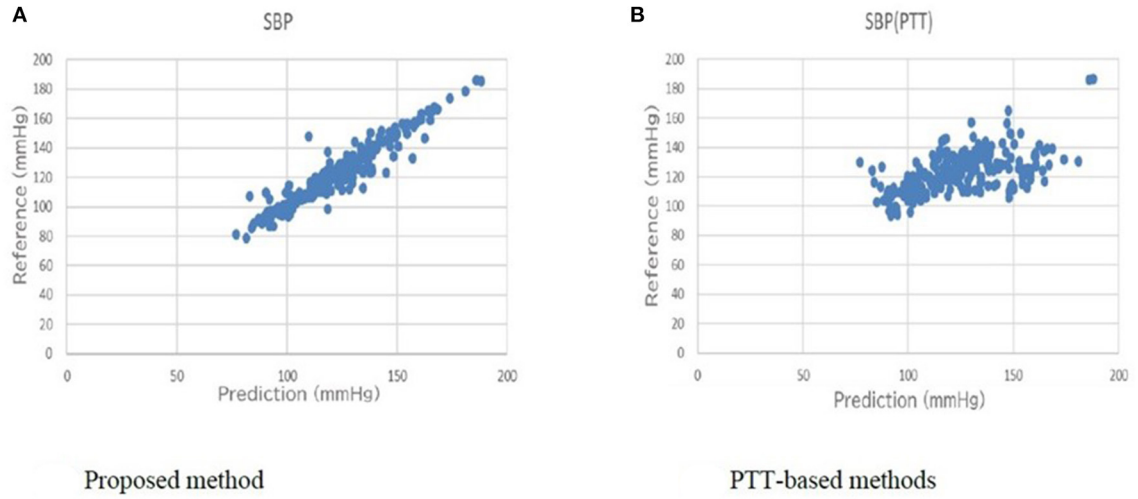


FIGURE 6 | (A) Proposed method reference-prediction plot (Systolic blood pressure). **(B)** PTT-based method reference-prediction plot (Systolic blood pressure). The proposed method had small-range prediction values closer to the reference values compared to the PTT-based methods.

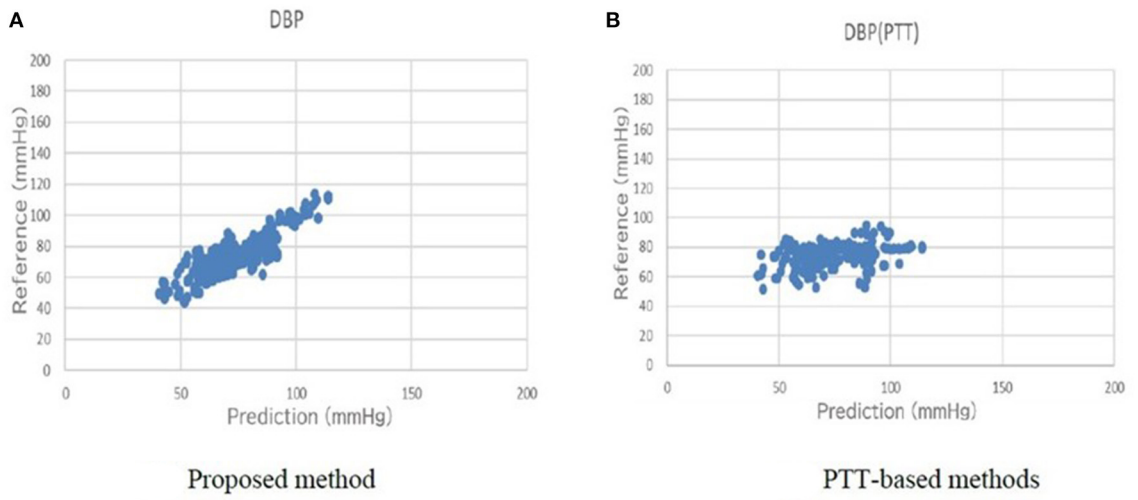


FIGURE 7 | (A) Proposed method reference-prediction plot (Diastolic blood pressure). **(B)** PTT-based method reference-prediction plot (Diastolic blood pressure). The proposed method had small-range prediction values closer to the reference values compared to the PTT-based methods.

ISO standard protocol. To the best of our knowledge, the present study is the only one that has met most of the protocols of the ISO standard in reference BP acquire method, the number of participants, BP distribution, and the standard deviation of the error for prediction model (<8 mm SBP). **Table 1** shows prior studies performed for BP estimation using PTT and ECG analysis. It presents a comparison of the signal source, number of participants, range of age, estimation methods, and error.

Implementing Wavelet Transformation, Grouping, and Used Features

As shown in **Table 4** and **Figure 5**, the characteristics of the features used in the present models were divided into two: traditional features and newly developed features. Traditional features such as 1/PTT, the square root of body weight, heart rate, and PTT were well-studied, and they were also effective in improving the accuracy in the present study (29). Besides, newly developed features such as wavelet transformation of the ECG and the pulse waveforms were as effective as traditional features (17, 18). The cells of frequently used wavelet coefficient features in both ECG and pulse wave were the ST-segment and the baseline between the T wave and P wave in the ECG (shown in **Figure 5**). The ST segment represents the interval between ventricular depolarization and repolarization and pumps blood to blood vessels during the ST segment (35). Since the P wave represents the depolarization of the left and right atria

and corresponds to atrial contraction, blood moves from the circulatory system to the heart during the baseline between the T wave and P wave (36). Because the arteriosclerosis of the blood vessel wall plays a significant role in determining blood pressure, it seems to be consistent with the many essential wavelet coefficients in the ST segment and the baseline between the T and P waves.

Another feature of the present study was the grouping of the datasets into 5 subgroups. Since the arterial walls are composed of the proteoglycans, endothelium, elastin, collagen, and smooth muscles in varying quantities depending on individuals and vessel size, grouping participants into five subgroups was considered effective (37). By grouping the datasets, the participants' BP estimation model can be accurately selected for each participant that matches the stiffness of the vessel wall.

The Merit of the Cuffless Sphygmomanometer

We adopted a data collection and evaluation method in line with the ISO standard for non-invasive sphygmomanometers (31). If the estimation accuracy is sufficiently improved, it is possible to implement a cuffless sphygmomanometer that predicts the trends of BP instead of using cuff-based BP devices in the near future. A cuffless sphygmomanometer has many potential advantages. Considerable fluctuations in BP trends can be used as a warning signal by medical practitioners. Most vital parameters in today's operating rooms are measured continuously, except for BP.

Physicians are always concerned about the deterioration of a patient's vital signs in the emergency departments, especially for BP. When we use the model in actual clinical settings in the future, we can use a watch-type pulse wave sensor for 24/7 monitoring. Blood pressure estimates are calculated within 1 s from pulse wave detection, and we can know blood pressure estimates for each pulse, resulting in improved patient safety. In the intensive care unit, invasive arterial blood pressure monitoring is routinely used to monitor seriously ill patients. For these patients, this is not only painful but, more seriously, can cause life-threatening infections or bleeding. However, since the device is non-invasive, the present BP estimation method has no risk of these complications. As it can record BP changes at every beat, it may also contribute to ensuring patient safety by retrospective investigation when incidents occur in a hospital.

A cuffless sphygmomanometer can bring various benefits to ordinary households as well. BP fluctuations are essential in monitoring systems for older patients, like the ones present in smart homes. By installing a cuffless sphygmomanometer in places where sudden changes in the BP of an older individual are likely to occur, such as in beds, chairs, toilets, and bathrooms, the healthcare provider can respond swiftly to these changes. Accumulation of daily blood pressure data can be utilized in outpatient treatment, leading to the improved prescription of appropriate antihypertensive medications and compliance. Furthermore, machine learning with other parameters can lead to the prediction of sudden events in daily life. Therefore, the development of a cuffless sphygmomanometer is expected to impact an aging society's social security system significantly.

Limitations of the Present Study

First of all, the present study results were data-dependent, and different datasets might create different BP estimation models. The characteristics used in the present study may differ depending on the datasets. In machine learning, 260 data sets in this study are relatively small, and it is desirable to perform sensitivity analysis with more data. However, to our knowledge, this study is one of the largest studies in the field of non-invasive BP estimation. This study also meets the requirements of ISO standards that demanding more than 255 datasets. We have a plan for a validation study with more participants in ICU/ER settings. Secondly, it can be challenging to interpret the created algorithm. As shown in **Figure 5**, it is challenging to determine precisely why this frequency during the specific period in the ECG and the pulse waveforms were related to the BP. However, these limitations are generally found in ML, and despite these challenges, applying the ML model to clinical practice is rapidly progressing (38–40). Third, we performed waveform measurements on motionless participants and excluded participants in arrhythmias in the present study. Since the current model averages pulse waveforms, patient's movement and arrhythmias can cause the poor performance of BP estimates. Finally, we did not validate the present model with a new dataset in the settings where BP change can be bigger, such as ICU/ER, further validation in ICU/ER settings is needed in the future.

CONCLUSIONS

Based on the participants with a wide age range and BP distribution, we proposed a novel cuffless BP estimation method by grouping participants and applying wavelet features. The standard deviation of error improved from 17.05 to 7.74 mmHg for SBP and from 14.05 to 6.42 mmHg for DBP compared to the PTT-only estimation methods. We plan to increase the number of datasets in ICU and ER settings and improve the accuracy of the estimation methods in future studies.

DATA AVAILABILITY STATEMENT

The original contributions generated for the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee of the University of Fukui (Approval Number: 20148035). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SY, KM, and OY contributed to conception and design of the study. KM organized the database and performed the statistical analysis. SY wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

The biopotential sensing system was used in joint research with Panasonic Corporation. The National Health Insurance Ota Hospital greatly supported us in enrolling the participants for BP measurements. Thanks to their participation and expertise, we were able to obtain good data sets for the present study.

REFERENCES

1. Tzourio C, Hanon O, Godin O, Soumaré A, Dufouil C. Impact of home blood pressure monitoring on blood pressure control in older individuals: a French randomized study. *J Hypertens.* (2017) 35:612–20. doi: 10.1097/HJH.0000000000001191
2. Leung AA, Nerenberg K, Daskalopoulou SS, McBrien K, Zarnke KB, Dasgupta K, et al. Hypertension Canada's 2016 Canadian hypertension education program guidelines for blood pressure measurement, diagnosis, assessment of risk, prevention, and treatment of hypertension. *Can J Cardiol.* (2016) 32:569–88. doi: 10.1016/j.cjca.2016.02.075
3. Parati G, Stergiou G, O'Brien E, Asmar R, Beilin L, Bilo G, et al. European Society of Hypertension practice guidelines for ambulatory blood pressure monitoring. *J Hypertens.* (2014) 32:1359–66. doi: 10.1097/HJH.0000000000000221
4. O'Brien E, Parati G, Stergiou G, Asmar R, Beilin L, Bilo G, et al. European Society of Hypertension position paper on ambulatory blood pressure monitoring. *J Hypertens.* (2013) 31:1731–68. doi: 10.1097/HJH.0b013e328363e964
5. Ward M, Langton JA. Blood pressure measurement. *Cont Educ Anaesthes Crit Care Pain.* (2007) 7:122–6. doi: 10.1093/bjaceaccp/mkm022
6. Kim J, Park J, Kim K, Chee Y, Lim Y, Park K. Development of a nonintrusive blood pressure estimation system for computer users. *Telemed J E Health.* (2007) 13:57–64. doi: 10.1089/tmj.2006.0034
7. Kim JS, Chee YJ, Park JW, Choi JW, Park KS. A new approach for non-intrusive monitoring of blood pressure on a toilet seat. *Physiol Meas.* (2006) 27:203. doi: 10.1088/0967-3334/27/2/010
8. Baek HJ, Lee HB, Kim JS, Choi JM, Kim KK, Park KS. Nonintrusive biological signal monitoring in a car to evaluate a driver's stress and health state. *Telemed e-Health.* (2009) 15:182–9. doi: 10.1089/tmj.2008.0090
9. Nye ER. The effect of blood pressure alteration on the pulse wave velocity. *Br Heart J.* (1964) 26:261–5. doi: 10.1136/hrt.26.2.261
10. Steptoe A, Smulyan H, Gribbin B. Pulse wave velocity and blood pressure change: calibration and applications. *Psychophysiology.* (1976) 13:488–93. doi: 10.1111/j.1469-8986.1976.tb00866.x
11. Geddes L, Voelz M, Babbs C, Bourland J, Tacker W. Pulse transit time as an indicator of arterial blood pressure. *Psychophysiology.* (1981) 18:71–4. doi: 10.1111/j.1469-8986.1981.tb01545.x
12. Teng XF, Zhang YT. An evaluation of a PTT-based method for noninvasive and cuffless estimation of arterial blood pressure. *Conf Proc IEEE Eng Med Biol Soc.* (2006) 1:6049–52. doi: 10.1109/IEMBS.2006.260823
13. Ding X-R, Zhang Y-T, Liu J, Dai W-X, Tsang HK. Continuous cuffless blood pressure estimation using pulse transit time and photoplethysmogram intensity ratio. *IEEE Trans Biomed Eng.* (2016) 63:964–72. doi: 10.1109/TBME.2015.2480679
14. Jeong I, Wood J, Finkelstein J. Using individualized pulse transit time calibration to monitor blood pressure during exercise. *Stud Health Technol Inform.* (2013) 190:39–41. doi: 10.3233/978-1-61499-276-9-39
15. Chen Y, Wen C, Tao G, Bi M. Continuous and noninvasive measurement of systolic and diastolic blood pressure by one mathematical model with the same model parameters and two separate pulse wave velocities. *Ann Biomed Eng.* (2012) 40:871–82. doi: 10.1007/s10439-011-0467-2
16. Chen Y, Wen C, Tao G, Bi M, Li G. Continuous and noninvasive blood pressure measurement: a novel modeling methodology of the relationship between blood pressure and pulse wave velocity. *Ann Biomed Eng.* (2009) 37:2222–33. doi: 10.1007/s10439-009-9759-1
17. Lee C, Zhang Y. Cuffless and noninvasive estimation of blood pressure based on a wavelet transform approach. In: *IEEE EMBS Asian-Pacific Conference on Biomedical Engineering*. Kyoto (2003). p. 148-9.
18. Sahoo A, Manimegalai P, Thanushkodi K. Wavelet based pulse rate and blood pressure estimation system from ECG and PPG signals. In: International Conference on Computer, Communication and Electrical Technology (ICCCET). Tirunelveli (2011). p. 285-9. doi: 10.1109/ICCCET.2011.5762486
19. McCarthy B, O'Flynn B, Mathewson A. An investigation of pulse transit time as a non-invasive blood pressure measurement method. *J Phys.* (2011) 307:012060. doi: 10.1088/1742-6596/307/1/012060
20. Cattivelli FS, Garudadri H. Noninvasive cuffless estimation of blood pressure from pulse arrival time and heart rate with adaptive calibration. In: Sixth International Workshop on Wearable and Implantable Body Sensor Networks Conference. Berkeley, CA: IEEE (2009). p. 114–9. doi: 10.1109/BSN.2009.35
21. Poon CC, Zhang YT. Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time. *Conf Proc IEEE Eng Med Biol Soc.* (2005) 6:5877–80. doi: 10.1109/IEMBS.2005.1615827
22. Singh RB, Cornelissen G, Weydahl A, Schwartzkopff O, Katinas G, Otsuka K, et al. Circadian heart rate and blood pressure variability considered for research and patient care. *Int J Cardiol.* (2003) 87:9–28; discussion 29–30. doi: 10.1016/S0167-5273(02)00308-X
23. Liu Q, Yan BP, Yu CM, Zhang YT, Poon CC. Attenuation of systolic blood pressure and pulse transit time hysteresis during exercise and recovery in cardiovascular patients. *IEEE Trans Biomed Eng.* (2014) 61:346–52. doi: 10.1109/TBME.2013.2286998
24. Lobodzinski SS, Laks MM. New devices for very long-term ECG monitoring. *Cardiol J.* (2012) 19:210–4. doi: 10.5603/CJ.2012.0039
25. Chen Z, Yang X, Teo JT, Ng SH. Noninvasive monitoring of blood pressure using optical ballistocardiography and photoplethysmograph approaches. In: *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Osaka: IEEE (2013). p. 2425–8.
26. Chan K, Hung K, Zhang Y. Noninvasive and cuffless measurements of blood pressure for telemedicine. In: *Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Istanbul (2001). p. 3592–3.
27. Wong Y, Zhang Y. The effects of exercises on the relationship between pulse transit time and arterial blood pressure. *Conf Proc IEEE Eng Med Biol Soc.* (2005) 2005:5576–8. doi: 10.1109/IEMBS.2005.1615748
28. Wong MY-M, Pickwell-MacPherson E, Zhang Y-T. The acute effects of running on blood pressure estimation using pulse transit time in normotensive subjects. *Eur J Appl Physiol.* (2009) 107:169–75. doi: 10.1007/s00421-009-1112-8
29. Sharma M, Barbosa K, Ho V, Griggs D, Ghirmai T, Krishnan SK, et al. Cuff-less and continuous blood pressure monitoring: a methodological review. *Technologies.* (2017) 5:21. doi: 10.3390/technologies5020021
30. Gao SC, Wittek P, Zhao L, Jiang WJ. Data-driven estimation of blood pressure using photoplethysmographic signals. *Annu Int Conf IEEE Eng Med Biol Soc.* (2016) 2016:766–9. doi: 10.1109/EMBC.2016.7590814
31. Non-invasive sphygmomanometers-Part 1: Requirements and test methods for non automated measurement type. ANSI/AAMI/ISO 81060-1. Association for the Advancement of Medical Instrumentation (2007).
32. Cortes C, Vapnik V. Support-vector networks. *Mach Learn.* (1995) 20:273–97. doi: 10.1007/BF00994018
33. Kim H-C, Pang S, Je H-M, Kim D, Bang S-Y. Support vector machine ensemble with bagging. In: *International Workshop on*

Support Vector Machines. Niagara Falls: Springer (2002). p. 397–408. doi: 10.1007/3-540-45665-1_31
34. Fushiki T. Estimation of prediction error by using K-fold cross-validation. *Stat Comput*. (2011) 21:137–46. doi: 10.1007/s11222-009-9153-8
35. Kashou AH, Basit H, Malik A. *ST Segment*. Treasure Island, FL: StatPearls Publishing LLC (2020).
36. Douedi S, Douedi H. *P wave*. Treasure Island, FL: StatPearls Publishing LLC (2020).
37. Burton AC. Relation of structure to function of the tissues of the wall of blood vessels. *Physiol Rev*. (1954) 34:619–42. doi: 10.1152/physrev.1954.34.4.619
38. Tunthanathip T, Sae-Heng S, Oearsakul T, Sakarunchai I, Kaewborisutsakul A, Taweesomboonyat C. Machine learning applications for the prediction of surgical site infection in neurological operations. *Neurosurg Focus*. (2019) 47:E7. doi: 10.3171/2019.5.FOCUS19241
39. Kang SY, Cha WC, Yoo J, Kim T, Park JH, Yoon H, et al. Predicting 30-day mortality of patients with pneumonia in an emergency department setting using machine-learning models. *Clin Exp Emerg Med*. (2020) 7:197–205. doi: 10.15441/ceem.19.052
40. Fickling SD, Smith AM, Pawlowski G, Ghosh Hajra S, Liu CC, Farrell K, et al. Brain vital signs detect concussion-related neurophysiological impairments in ice hockey. *Brain*. (2019) 142:255–62. doi: 10.1093/brain/awy317

Conflict of Interest: KM is employed Connect Inc.

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cardiovascular Medicine

REVIEW
published: 04 August 2021
doi: 10.3389/fcvm.2021.711611

Machine Learning Augmented Echocardiography for Diastolic Function Assessment

Andrew J. Fletcher[1,2], *Winok Lapidaire*[1] *and Paul Leeson*[1]*

[1] *Oxford Cardiovascular Clinical Research Facility, Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom,* [2] *Department of Cardiac Physiology, Royal Papworth Hospital National Health Service Foundation Trust, Cambridge, United Kingdom*

Cardiac diastolic dysfunction is prevalent and is a diagnostic criterion for heart failure with preserved ejection fraction—a burgeoning global health issue. As gold-standard invasive haemodynamic assessment of diastolic function is not routinely performed, clinical guidelines advise using echocardiography measures to determine the grade of diastolic function. However, the current process has suboptimal accuracy, regular indeterminate classifications and is susceptible to confounding from comorbidities. Advances in artificial intelligence in recent years have created revolutionary ways to evaluate and integrate large quantities of cardiology data. Imaging is an area of particular strength for the sub-field of machine-learning, with evidence that trained algorithms can accurately discern cardiac structures, reliably estimate chamber volumes, and output systolic function metrics from echocardiographic images. In this review, we present the emerging field of machine-learning based echocardiographic diastolic function assessment. We summarise how machine-learning has made use of diastolic parameters to accurately differentiate pathology, to identify novel phenotypes within diastolic disease, and to grade diastolic function. Perspectives are given about how these innovations could be used to augment clinical practice, whilst areas for future investigation are identified.

Keywords: artificial inteligence, echocardiogaphy, diastolic dysfunction, machine learning, heart failure preserved ejection fraction

OPEN ACCESS

Edited by:
Rene M. Botnar, King's College London, United Kingdom

Reviewed by:
Haikun Qi, ShanghaiTech University, China
Gen-Min Lin, Hualien Armed Forces General Hospital, Taiwan

***Correspondence:**
Paul Leeson
paul.leeson@cardiov.ox.ac.uk

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

Received: *18 May 2021*
Accepted: *13 July 2021*
Published: *04 August 2021*

Citation:
Fletcher AJ, Lapidaire W and Leeson P (2021) Machine Learning Augmented Echocardiography for Diastolic Function Assessment. Front. Cardiovasc. Med. 8:711611. doi: 10.3389/fcvm.2021.711611

INTRODUCTION—DIASTOLIC ASSESSMENT IN CLINICAL PRACTICE

Left-sided cardiac diastolic dysfunction can lead to patients developing debilitating symptoms such as dyspnoea and fatigue, as well as conferring worse survival and increased morbidity (1–3). Prevalence estimates of diastolic dysfunction vary widely depending upon the population studied and the definition used, but a recent review of community studies suggests it is in the range of 20-30% in the general population (4). In studies of thousands of patients in clinical settings, prevalences of 5.0% (5) to 9.2% (3) have been reported when accompanied by a normal left ventricular (LV) ejection fraction (EF) > 50%. Diastolic dysfunction is a criterion for the diagnosis of heart failure (HF) with preserved EF (HFpEF) (6), which represents about a third of all hospitalised heart failure in the United Kingdom (7) and about half of community heart failure in North America (8).

The gold standard for assessing diastolic function is invasive pressure-volume loop analysis, which directly measures ventricular compliance and relaxation, but is seldomly performed

clinically as the process and analyses are technically challenging. The next best assessment technique is direct measurement of cardiac pressures with catheterisation, because diastolic pressures rise with advanced dysfunction. Cardiac catherisation can either be performed via left heart catheterisation (LHC), where an end diastolic pressure (LV-EDP) or pre-A wave pressure are the common benchmarks, or by right heart catheterisation (RHC), where a pulmonary capillary wedge pressure (PCWP) is recorded which approximates the LA pressure from across the pulmonary capillary bed. These invasive procedures are also not routinely performed in most patient cohorts as they are resource intensive whilst exposing the patient to radiation and possible discomfort. Therefore, echocardiography is often preferred for diastolic function assessment.

Echocardiography is the primary imaging tool used for assessing diastology and HFpEF in routine clinical practice because it is non-invasive and widely available. There are over 20 different variables of diastolic relevance that can be measured using routine transthoracic echocardiography (9), as well as variables emerging from the research domain such as speckle-tracking strain parameters (10). Unfortunately, no single echocardiographic parameter adequately captures the complexity of diastolic function, given the different structural and functional changes which can manifest at different time points during the cardiac cycle. Routine variables, commonly derived from pulse-wave Doppler and tissue Doppler techniques, at best modestly correlate with invasively measured diastolic pressures (11–13).

Diastolic function is therefore classified echocardiographically by combining multiple parameters. The most widely adopted method for this comes from the American Society of Echocardiography and European Association of Cardiovascular Imaging (ASE/EACVI) guideline (9). The method uses a series of decision steps in the form of two algorithms, one screening for the presence of diastolic dysfunction and the other to grade diastolic dysfunction if it is found to exist (or is assumed to exist based upon defined clinical and structural observations), however there are a number of caveats which complicate matters. Classification outcomes are either that the guideline cannot be applied due to insufficient requisite information, that the diastolic function is indeterminate or normal, or that diastolic function is graded as mild, moderate, or severely impaired. Filling-pressure, which refers to left atrial (LA) and/or LV diastolic pressures, is often dichotomously described as "normal" or "raised." Moderate and severe ASE/EACVI guideline graded diastolic dysfunction correspond to raised filling-pressures (9).

LIMITATIONS OF ROUTINE DIASTOLIC ASSESSMENT

There are a number of barriers to widespread and robust diastolic evaluation with echocardiography, particularly with following guideline suggestions, which include: accuracy uncertainties, unclassifiable and indeterminate situations, and confounding from comorbidities. These introduce clinical uncertainty, which can lead to inappropriate treatment decisions, and are explored in more detail below.

Clinical Guideline Accuracy

Dual echocardiographic and invasive-catheterisation validation studies show suboptimal accuracy of the current ASE/EACVI guideline method to identify patients with raised filling-pressures. Sato et al. demonstrated that guideline classified moderate or severe diastolic dysfunction predicted raised invasively measured filling-pressures with an accuracy of 66% in an all-comers clinical population receiving echocardiography and LHC within 24 h (1). Lancellotti et al. found an accuracy of 56% of the guideline for predicting raised filling-pressures in patients receiving LHC for known or suspected coronary artery disease (14). Balaney et al. reported an accuracy of 68% in patients attending for LHC for a variety of clinical indications (15). Andersen et al. demonstrated accuracy of 87% in patients having either LHC or RHC for any valid clinical reason in a multi-institutional study (16).

It is worth noting that all of these studies excluded patients with confounding factors before recruitment or analysis, so they may not represent "real-world" accuracies in all-comer populations. Unfortunately, varying methodologies of researchers also limits our ability to compare these results. For the invasive validation studies above, different definitions of raised filling-pressures were used [LV-EDP > 14 mmHg (14), LV-EDP > 15 mmHg (17), LV-EDP > 16 mmHg (1), pre-A pressure > 12 mmHg (15, 16), PCWP > 12 mmHg (16), and PCWP > 15 mmHg (17)]. Furthermore, whilst the ASE/EACVI guideline recommends using an average of medial and lateral mitral annular e' values to calculate an average E/e' ratio in the majority of pathologies (9), some institutions have historically only acquired one or the other, or only report one in publications, again limiting clinical applicability of results (18, 19).

Unclassifiable and Indeterminate Diastolic Grading

Unclassifiable diastolic function can arise from key parameters, needed to follow the guideline decision steps, being missing. A frequent cause of this is a suboptimal acoustic window which precludes measurements. One author reported unclassifiable diastolic function in 8% of consecutive echocardiograms due to poor image quality and missing data (20), whilst another study found this in 22% of scans (3), showing that this situation occurs regularly.

Indeterminate diastolic grading also creates uncertainty and may result in additional resource intensive or higher risk investigations, like exercise echocardiography or cardiac catheterisation. In the guideline approach, it occurs because parameters required in the decision-steps are contradictory/inconclusive. This too is frequent—a report from the National Echocardiography Database of Australian (3) found 27% of 344,646 scans were labelled as indeterminate, whilst a study of consecutive Canadian tertiary centre echocardiograms found 36% indeterminate (5).

Further real-world clinical data from Europe (1,000 individuals), Britain (189 individuals), Asia (57,630 individuals), America (866 individuals), and Canada (71,727 individuals),

show that even after excluding patients with diastology-confounding factors, between 11 and 22% of scans are labelled indeterminate (5, 18, 19, 21, 22). A pertinent limitation of the literature base collectively is that a breakdown of reasons for indeterminate grading is often not presented.

For data acquired under stricter research study protocols or during dual invasive validation studies, the indeterminate proportion is not too dissimilar at 7-24% (1, 14–16, 23, 24). However, in the setting of pulmonary hypertension, indeterminate classification may be as high as 53% (17) suggesting that confounding factors may magnify the indeterminate issue.

Factors Confounding Diastolic Assessment

Pulmonary hypertension (PH), arrhythmias, tachycardia, and valvular pathologies can all complicate the assessment of diastolic function by confounding associations between individual diastolic parameters and filling-pressures, thus reducing diagnostic accuracy and sometimes even precluding the measurement of parameters altogether. Such situations are common and were found to occur in 52% of consecutive echocardiograms (20).

Atrial fibrillation not only causes technical problems concerning parameter measurement, due to variability in cardiac cycle length and potentially misleading LA dilation, but it also prevents any meaningful late-diastolic atrial pumping of blood into the LV. This removes key diastolic parameters like the E/A ratio. Between 48 and 57% of HFpEF patients have confounding atrial fibrillation (25, 26), highlighting the pressing need to overcome this obstacle. The routine diastolic parameter E/e' has suboptimal association with invasively measured filling-pressures in patients with atrial fibrillation (27), whilst a range of common diastolic parameters are known to be altered in atrial fibrillation when compared in the same patients to sinus rhythm (28), showing how difficult assessment is in the presence of arrhythmias.

Pre-capillary PH, that is PH not of a left-sided aetiology, results in left heart preload reduction, which creates a disconnect between the intrinsic left-sided diastolic state, its filling pressures and hence its echocardiographic parameters. It is not always clear at the time of echocardiography whether the PH aetiology is pre-capillary, post-capillary, or mixed, and hence diastolic assessment is often confounded. This situation may arise regularly given that PH occurs in 50-80% of HFpEF patients (29, 30). Leung et al. investigated patients referred for suspected PH with echocardiography and both LHC/RHC (17). The ASE/EACVI algorithm accuracy for identifying raised or normal filling-pressures was only 29 and 23%, respectively, although it must be noted that the guideline recommends emphasis of different parameters in situations such as PH, which are not applied in the same algorithmic way.

Left sided valvular disease is also known to confound traditional diastolic assessment due to variable influences upon individual parameters. Mitral stenosis and mitral annular calcification both reduce the e' velocity and uncouple the measure (and therefore E/e') from the underlying diastolic state and filling-pressure (31, 32). Significant mitral regurgitation raises the e' velocity, again uncoupling it from the intrinsic state and rendering it unreliable (31). Analysis of 161,468 echocardiograms excluded from diastolic classification due to confounding factors showed that mild or more mitral stenosis was present in 1.6% of these scans, moderate or greater mitral annular calcification in 1.2% and more than moderate mitral regurgitation in 3.3% (33). Given the huge quantity of echocardiograms performed annually globally, this represents confounding of a significant number of scans.

Confounding factors tend to be become more prevalent as diastolic function deteriorates and often coexist with HFpEF. For example, atrial fibrillation may develop within 4 years in a third of HFpEF patients who originally present in sinus rhythm (34). Echocardiographers are hence in need of techniques to better assess diastolic function in the presence of confounding factors, and more robust tools which reduce indeterminate and unclassifiable situations.

MACHINE LEARNING APPLICATION TO DIASTOLIC ASSESSMENT

Machine-Learning in Echocardiography

Machine-learning (ML), a domain of artificial intelligence born from advanced computer science, mathematical and statistical techniques, holds huge potential for improving echocardiographic analysis in terms of streamlining workflow, automating feature quantification and accurately identifying pathology (35). The term "supervised ML" refers to algorithms that are trained using data labelled with an important feature, outcome, or diagnosis. Supervised ML aims to perform a task which could be of a regression type, such as predicting an exact value of left atrial pressure, but in cardiology is often a classification type, for example to state whether a particular disease is present or not. Support vector machines, random forests, and artificial neural networks (deep-learning) are examples of supervised ML, where heart-failure hospitalisation or invasively measured left heart pressures could be suitable training labels.

Unsupervised ML refers to analyses which learn from unlabelled input data to perform the required task. Cluster analysis is such an algorithm, which for example can group patients with similar echocardiographic variable values to create novel stages of a disease process or can unearth homogeneous subgroups within a larger heterogeneous cohort. Another class of unsupervised ML is dimensionality reduction, where the input data is projected onto a lesser number of new variables, thus reducing complexity, increasing interpretability and visualisation, and making the dataset better primed for other ML techniques. Principle component analysis is an example of this, where original variables are mapped to a smaller number of new "principal component" variables, which retain as much of the original data variance as possible.

ML is thus a powerful tool able to process large high-dimensional datasets, such as those obtained with genomics, metabolomics, and imaging. It can learn not only from the acquired variables presented to it, but also by discerning novel

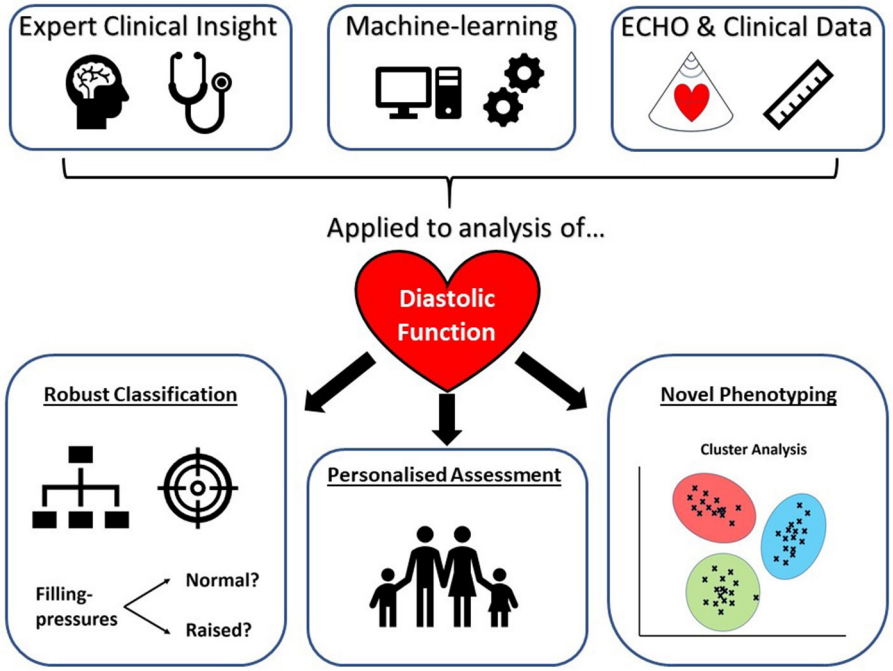


FIGURE 1 | Machine learning augmented diastolic function assessment with echocardiography.

features and latent data relationships (35, 36). Echocardiographic data, with its many parameters of diastolic function, combined with detailed clinical and demographic features is therefore well-suited as training material for ML models which could augment traditional diastolic assessment techniques (**Figure 1**). ML is particularly well-suited to detect and describe non-linear relationships (37), which is pertinent to diastolic assessment where, for example, the E/A ratio has a *U*-shaped relationship with dysfunction.

Much of the application of ML to echocardiography thus far has concerned automation and accuracy improvement of tasks such as image structure segmentation, left-sided chamber size calculation and estimation of systolic function metrics like ejection fraction (38–40). Prediction of the development of relevant pathology from images, such as clinically significant coronary artery disease, is also a focus of efforts given the prognostic implications (41). Attention is now increasingly being afforded to diastolic function, given the aforementioned challenges of contemporary assessment methodology, diastolic dysfunction prevalence, and suitability of echocardiographic data for training ML models.

ML Integration of Diastolic Parameters

There are various applications of ML to echocardiographic diastolic function assessment—the key illustrative studies of which are summarised in **Table 1**. The first concerns the integration of diastolic parameters, whether collectively or alongside non-diastolic and/or non-echocardiographic variables, with ML to assist with disease diagnosis. Hubert et al. (42) report a new method for assessing diastolic function using strain-volume loops (SVL) derived from speckle-tracking strain imaging. SVL area differentiated between amyloidosis and HFpEF with an area under the receiver operator curve (AUC) of 0.76. However, when supervised linear discriminant analysis ML was applied to integrate the SVL with routine diastolic related echocardiographic parameters, the AUC increased to 0.91, showing the added value of individual diastolic variables.

ML's ability to effectively combine diastolic information to improve disease identification is also demonstrated by Sengupta et al. (43) who used an associative memory classifier–based ML algorithm to differentiate constrictive pericarditis from restrictive cardiomyopathy. The AUC of 0.89 for speckle-tracking strain variables increased to 0.96 when just four routine diastolic pertinent variables were included in their model (septal thickness, posterior wall thickness, e' and E/e').

Further evidence comes from Choi et al. (44) who tested a range of different ML algorithms for their ability to diagnose both HFpEF and systolic HF with a range of clinical, blood, electrocardiographic and echocardiographic variables. The diastolic pertinent variables combined included EF, indexed mass, septal E/e' and tricuspid regurgitation maximum velocity (TR-Vmax). When compared to physician diagnosis, ML diagnosed HFpEF with 99.6% concordance.

Novel parameters of diastolic dysfunction, relevant to breathless patients, hypertensives and those with HFpEF, have been elucidated through ML. Sanchez-Martinez et al. (36) obtained tissue Doppler data during exercise echocardiography,

TABLE 1 | Summary of key publications demonstrating application of machine-learning to the echocardiographic assessment of diastolic function.

References	Application category	ML technique(s)	Training data types	Key finding(s)
Choi et al., (44)	Integration of diastolic parameters	Five techniques: "classification and regression tree" performed best.	1. Echocardiographic routine variables 2. Electrocardiogram variables 3. Clinical/haematological variables	When compared to physician diagnosis, ML diagnosed HFpEF with 99.6% concordance
Sanchez-Martinez et al., (36)	Novel diastolic variable discovery	Unsupervised agglomerative hierarchical clustering	1. Exercise echocardiography tissue-doppler variables	Premature fusion of early and late diastolic waves, increased variability in the onset of atrial contraction (a' wave) and a blunted response in atrial velocities during exercise were novel diastolic variables for assessing HFpEF with ML.
Segar et al., (49)	Phenotyping, prognostication	Unsupervised penalized finite mixture model-based clustering	1. Echocardiographic routine variables 2. Electrocardiogram variables 3. Clinical/Haematological variables	ML derived three phenogroups of HFpEF which varied in diastolic dysfunction, hospitalisation, and mortality.
Omar et al., (56)	Diastolic function grading	Random forest, artificial neural network, and support vector machine	1. Echocardiographic strain variables	ML predicted invasively measured PCWP $\geq$ 18 mmHg with AUC = 0.88. ML correctly identified 80% of patients with raised PCWP, with no indeterminate classifications.
Lancaster et al., (19)	Diastolic function grading, prognostication	Unsupervised hierarchical clustering	1. Echocardiographic routine variables	ML found two distinct clusters in those who would normally be 'screened' for diastolic dysfunction with the guideline (9); one cluster was mostly (72%) guideline-normal whilst the other cluster was mostly (also 72%) guideline-defined dysfunction or indeterminate grading.
Tokodi et al., (57)	Diastolic function grading, prognostication	Unsupervised topological data analysis and clustering	1. Echocardiographic routine variables 2. Major adverse cardiac event hospitalisation records	Continuous 'patient similarity networks', derived with ML and later split into segments, vary in diastolic function, mortality, and morbidity, with no indeterminate classifications.
Cho et al., (59)	Diastolic function grading, phenotyping	Unsupervised topological data analysis, supervised decision tree, ensemble and deep neural network	1. Echocardiographic routine diastolic variables 2. Echocardiographic strain variables 3. Vector flow mapping variables	ML produced a patient similarity network with four regions and no indeterminate classifications—regions linearly progressed in terms of diastolic variables, heart failure stages A-D and New York Heart Association functional classes.
Pandey et al., (60)	Diastolic function grading, phenotyping, prognostication	Unsupervised topological data analysis, agglomerative hierarchical clustering and supervised deep neural network	1. Echocardiographic routine diastolic variables	ML was superior to ASE 2016 diastolic guideline grades for predicting invasively-measured elevated left ventricular filling pressure (AUC = 0.88 vs. 0.67). Two clusters of patients were found—the high-risk phenogroup showed higher rates of heart failure hospitalization and/or death than the low-risk phenogroup in multiple external validation cohorts.

ML, machine-learning; HFpEF, heart failure with preserved ejection fraction; AUC, area under the receiver operator curve; PCWP, pulmonary capillary wedge pressure.

and *via* unsupervised agglomerative hierarchical clustering ML identified premature fusion of early and late diastolic waves, increased variability in the onset of atrial contraction (a' wave) and a blunted response in atrial velocities (a' wave peak) during exercise as novel diastolic variables for assessing those with, or at risk of, HFpEF. The importance of these in other populations and their potential clinical utility remains unknown.

ML for Diastolic Phenotyping and Prognostication

Over the last 5 years or so a rapidly growing body of evidence has accumulated where ML has been applied to phenotype patient's diastology based upon clinical and echocardiographic data. Most often for this purpose, unsupervised ML cluster analysis has been used, which groups patients in a potentially

novel way based upon input data similarities. Commonly the disease of interest has been HF, with different authors finding two (45, 46), three (47–49), four (50) and even six (51) phenogroups of HFpEF when applying ML clustering to echocardiographic variables.

For example, Nouraei et al. (51) found six clusters of HFpEF which varied in diastolic dysfunction, endpoints and clinical features. Echocardiographic diastolic variables they used included indexed LA volume, indexed mass, E/A ratio, average E/e', tricuspid regurgitation maximum velocity and grade of diastolic dysfunction. Segar et al. (49) found three phenogroups of HFpEF which varied in diastolic dysfunction, BNP, comorbidities, mortality and hospitalisations. E/A ratio and LA area were among the optimal 20 variables for predicting phenogroup membership.

Hedman et al. (45) found two HFpEF clusters based upon 32 echocardiographic variables, which was possible despite >50% of the patients being in atrial fibrillation at the time of the scan. This supports the notion that ML does not necessarily need individual parameters like E/A ratio all of the time to grade diastolic function, hence offering a more flexible approach than the current clinical methodology.

Among acute HF admissions Horiuchi et al. (52) undertook cluster analysis which yielded three phenogroups that varied in diastolic function, risk of death and subsequent hospitalisation. This highlights the ability of ML to create novel groupings, or classifications, that are not only diagnostically important, but that are also prognostically important. ML can provide clinically useful results in conditions other than HF as shown by Mishra et al. (53), who performed cluster analysis of clinical and echocardiographic data from stable coronary artery disease patients with seven years mean follow-up. Four phenogroups resulted, which varied in diastolic dysfunction and hospitalisation risk.

Unique insights about non-cardiac chronic conditions are also possible with ML of diastolic variables. The relationship between diastolic function and renal function, in the setting of type 2 diabetes, was the subject of investigation by Pecková et al. (54). Unsupervised ML clustering found two subgroups of those with impaired renal function: when there was an early diastolic tissue velocity e' $\leq$ 7.1 cm/s there was a significant correlation between the echocardiographic ratio E/e' and renal function, whereas when e' > 7.1 cm/s then there was no significant correlation. This highlights the aforementioned potential for ML to unearth associations which may not be obvious to the human eye nor with traditional statistics.

Grading Diastolic Function With ML

Another application of ML is for grading of diastolic function. Evidence exists suggesting that ML can overcome some of the confounding factor, indeterminate classification and accuracy barriers surrounding the routine clinical guideline grading of diastolic function with echocardiography. Omar et al. (55) investigated whether ML of LV and LA speckle-tracking strain (STS) variables could assess LV diastolic function independently of routine Doppler parameters. They undertook cluster analysis using nine STS variables from 130 patients with heart failure symptoms. This produced three clusters which varied concordantly in diastolic Doppler indices and LA maximal volume, with no indeterminate classifications. The clusters were invasively validated in a further 44 patients, where PCWP and LV-EDP increased concordantly across their three pre-identified clusters.

These findings show that ML can identify discrete phenotypes of diastolic function which vary in severity, much like the current grading system, but which do not rely upon the acquisition of standard diastolic variables. This could greatly assist healthcare professionals in identifying those with diastolic dysfunction when technical limitations or missing Doppler variables may preclude following the guideline grading algorithms. How these novel ML results relate to clinical markers of diastolic dysfunction and heart failure, like symptoms and b-type natriuretic peptides, is unclear.

In an extension of their work, Omar et al. published further analysis from the same patient cohort (56). Using 14 STS parameters they trained three separate ML algorithms (random forest, artificial neural network, and support vector machine) in a supervised fashion to diagnose raised filling-pressures, with E/e' or PCWP as a label and a majority voting system to decide the outcome. Taking the best 11 STS parameters, an AUC = 0.85 was obtained for predicting E/e' $\geq$ 13 in the derivation group, with AUC = 0.88 for predicting PCWP $\geq$ 18 mmHg in the invasive validation group. ML of the echocardiographic parameters correctly identified 80% of patients with raised PCWP, with no indeterminate classifications, exceeding the performance of the guideline grading algorithm in most invasive studies of its accuracy (as per section Clinical guideline accuracy). A limitation of this work is that referral for heart failure symptoms created a biased population in terms of diastolic function.

Lancaster et al. (19) analysed routine diastolic parameters in a retrospective analysis of 866 consecutive patients referred for myocardial function assessment. Scans were grouped according to which ASE/EACVI guideline algorithm (9) would be applied clinically: screening or grading. Unsupervised ML cluster analysis in the screening group found two clusters with no indeterminate classifications. The first larger cluster (n = 460, 82% of 559) contained mostly guideline classified "normal" diastolic function (72%), whilst the second smaller cluster (n = 99, 18%) contained mostly guideline classified diastolic dysfunction or "indeterminate" (72% of 99). Agreement between ML and guideline classification of diastolic dysfunction was poor (kappa = 0.41).

In the grading group, ML again found two clusters: one (n = 236, 61%) comprised mainly guideline graded "mild" (44%) and "moderate" (50%), with little "severe" (6%). In contrast, the second (n = 151, 39%) contained mostly "moderate" (78%) with some "severe" (15%) and infrequently "mild" (7%). Using binary classification of mild vs. moderate/severe, the agreement between ML and guideline grading was better than for screening (kappa = 0.62). Given the known guideline limitations, a modest agreement metric should not discourage the notion of a more effective ML grading system.

This evidence leads to some interesting questions. There was a lot of overlap between their two clusters for those who

the guideline grades as having moderate diastolic dysfunction. Does the traditional "moderate" category perhaps contain two phenotypes, one with normal filling-pressures and one with raised? Furthermore, given the convention of associating moderate/severe with raised filling-pressures, is binary grading (normal/raised filling-pressures) clinically more useful than the historic four grades (normal/mild/moderate/severe)?

Additionally, the authors found their ML clusters to show improved prognostication compared to the ASE/EACVI guideline. Their ML diastolic dysfunction screening algorithm better predicted event-free survival and rehospitalisations whilst their grading algorithm was better at predicting mortality but was indifferent to the guideline for rehospitalisations. Again, this hints that a dichotomous classification system may better correspond to patient outcomes.

Placing patients onto a continuous spectrum from normality to disease, rather than aggregated into categorical boxes, represents an achievable aspiration for ML and would facilitate a major leap towards personalised medicine for diastology. Tokodi et al. (57) applied "Topological Data Analysis" and unsupervised clustering ML to the same cohort of 866 patients used by Lancaster et al. (19). The TDA technique is well-suited for detecting subtle geometric patterns in high dimensional data (57, 58). The authors studied nine echocardiographic parameters used for diastolic classification (EF, indexed LV mass, E, A, E/A, e', E/e', LA indexed volume and TR velocity). Their initial analyses created a "patient similarity network" (57) which cluster analysis divided into four regions. These varied in cardiac structure, LV systolic/diastolic function, mortality, morbidity, and diastolic dysfunction risk-factors, with no indeterminate classifications.

The authors also tested the prognostic ability of their loop in $n = 96$ completely new patients with two serial echocardiograms each. A supervised random-forest ML algorithm was trained with the derivation cohort data with loop region as a label. In the unseen validation cohort, the ML loop region was associated with major adverse cardiac event hospitalisation (MACE-h). Upon comparing the first and second echocardiograms, an improvement to (or remaining within) a lower risk region was associated with lower MACE-h rates. These results support the notion that ML can identify patients with different stages of diastolic dysfunction in a personalised fashion, with fidelity for linking diastolic changes to outcomes. How their loop-regions correspond to guideline diastolic grades, and to invasive filling-pressures, would be of interest.

Whilst the other pieces of evidence concerning ML classification of diastolic dysfunction have used either STS or routine Doppler diastolic parameters exclusively, added benefit may be realised by simultaneously applying ML to novel variables emerging from the research setting. Cho et al. (59) prospectively recruited $n = 247$ consecutive patients and $n = 50$ healthy control participants. All routine parameters needed to perform guideline classification of LV diastolic function were obtained, apart from lateral e'. LV and LA deformation, plus vector flow mapping (VFM) parameters were also measured.

Topological Data Analysis produced a patient similarity network with four regions and no indeterminate classifications—regions linearly progressed in terms of diastolic parameters, heart failure stages A-D and New York Heart Association functional classes I-IV. Three supervised ML techniques were then individually trained with the label of network region. A Deepnet neural network performed best at classifying scans into regions when given all 42 variables (AUC between 0.83-0.99 for the four regions). Of the 25 most important variables, 44% were VFM, 40% routine and only 16% STS. A high dependency of the model upon niche VFM variables, which most cardiologists and echocardiographers are not skilled in measuring, perhaps limits the potential for adoption of these findings into current clinical practice. Interestingly and reassuringly in this study though, the regions of the patient similarity network appeared to mimic the four-grade system of the guidelines.

Pandey et al. (60) are the first to have applied deep learning to the assessment of diastolic function from echocardiography data. As in the work of Tokodi et al. (57) and Cho et al. (59), a patient similarity network was firstly derived with Topological Data Analysis of the routine parameters needed to follow guideline diastolic assessment. Two network regions were defined with the help of clustering; high-risk and low-risk for heart-failure hospitalisation or cardiac death. Supervised deep learning with a neural network model then classified phenogroup membership before further evaluations were undertaken in multiple independent datasets with invasive haemodynamic, outcome, cardiac biomarker, and exercise performance metrics.

The model of Pandey et al. (60) showed better prediction than the guideline diastolic grades for elevated LV filling-pressure > 15 mmHg (AUC $= 0.88$ vs. 0.67; $p < 0.01$). Tellingly, most of the outperformance was driven by guideline "indeterminate" subjects, as when removed the AUC were similar between the deep-learning model and guideline method. Furthermore, the high-risk phenogroup showed higher rates of heart failure hospitalisation and/or death than the low-risk group in multiple HFpEF trial datasets. As such, these results further support a clinically meaningful augmentation of diastolic assessment with ML of echocardiographic variables.

Requisites for ML Translation to Clinical Practice

Several important steps are required to effectively use ML for echocardiographic diastolic function analysis, and subsequently translate the results into clinical practice (**Figure 2**). Firstly, high quality structured datasets, with sufficient granularity to describe the system of interest, are required. These are traditionally obtained through clinical trials, however, clinical archives are now often mined due to the lower resource costs and need for large quantities of training data that reflects real life practice (38). Care must be taken to ensure that the raw data is still acquired following high-quality protocols, is in a suitable format for ML ingestion (e.g., normalised, standardised), and consideration of missing data and outliers is paramount.

Secondly, a choice of suitable ML algorithm(s) to train should be made by considering the research aims. For classification, where labelled data is available and the so called "ground-truth" is known, a supervised algorithm such as a support vector machine or decision tree would be considered. If the

FIGURE 2 | Key steps required for using machine-learning for echocardiographic diastolic analysis.

TABLE 2 | Challenges and potential benefits of machine-learning augmented echocardiography for diastolic analysis.

Machine-Learning Augmented Echocardiography for Diastolic Analysis	
Potential benefits	**Challenges**
• Improved diagnostic accuracy • Enhanced prognostication • Reduced indeterminate classifications • Assessment in presence of confounding factors • Personalised diastolic assessment • Use all scan information collected • Discover novel diastolic parameters • Automated assessment • Track serial changes	• Lack of echocardiographic data with simultaneous invasive haemodynamics • Quality of clinical datasets for training • Quantity of records in training datasets • To be robust to changes in pre-load / after-load • "Black-box" perception • Regulatory approval • How to best integrate into clinical practice?

aim is to derive novel groupings or phenotypes, and the data is unlabelled, then unsupervised methods such as clustering algorithms or principal component analysis are better suited. Some aims may require several algorithms in combination, i.e., an ensemble.

Training of the ML model with the data, tuning of model hyperparameters, and performance testing form the next stage of the process. Validation with independent external datasets is recommended to assess performance of the model, however investigators often do not have access to such data and hence frequently use cross-fold validation. This entails splitting data into sections (known as folds), using some folds to train and some folds to test the model, before rotating the folds around the whole dataset (61). Traditional statistical analysis methodologies are normally applied to the resulting outputs where, for example, well-known metrics such as sensitivity, specificity, and accuracy, allow appreciation of performance and a comparison of models for a classification task. These hence permit identification of the optimal model or ensemble.

Variations in equipment, protocols, staff, and patients between datasets may influence the performance of a model. Validation with independent external datasets hence allows appreciation of the real-world applicability or generalisability of the findings, which is especially important if the training dataset was from a clinical trial. Overfitting, where the performance of the model is high in the training dataset but lower in other data, suggests that the model has leant the training data well but lacks generalisability. Underfitting can be found where the training data was too small and/or of too low a dimensionality to allow the model to accurately perform the desired task. Validation thus helps to check for over/under-fitting of the ML model, which is particularly relevant if the data came from a specific cohort, demographic, or location for example.

Once a ML algorithm has been trained, tested, and validated, and there is confidence that the model can positively impact clinical care, regulatory approval should be sought to enable widespread adoption or commercialisation. Regulatory approval by the Food and Drug Administration in America, or with the Conformité Européene (CE) mark in Europe has been increasingly sought in recent years for devices incorporating ML (62). Given the potential impact upon patient care, ML algorithms should be certified as a medical device to make sure that they are safe and fit for purpose, and to provide reassurance of quality to purchasers, users, and patients. The risk category being assessed under by the regulatory body, and transparency of the submitted information, are the subject of debate (63).

The final step of introducing a produced ML tool into clinical care then requires some more considerations. Regular audit must be undertaken to ensure safety and effectiveness of decisions made as a result of the tool introduction. The outcome data produced by the ML tool, such as a predicted diagnosis, should be regularly compared to a reference standard or clinical diagnosis. The ML output(s) could then feedback into the entire development cycle of another ML tool, because of this potential alteration in patient care.

POTENTIAL BENEFITS, CHALLENGES, AND FUTURE DIRECTIONS

With the applications of ML augmented echocardiography now stepping out of the shadows for diastolic assessment, a number of potential benefits and challenges can be seen (**Table 2**). From the healthcare system perspective, improved prognostication of cardiology patients with ML may facilitate efficient resource allocation, meaning that the right care is available to the right patient at the right time.

In the clinical echocardiographic laboratory, healthcare staff may benefit from automated assessment tools which could save them time, reduce inter- and intra-rater variability and allow for subtle serial changes to be monitored. ML may be able to negate the influence of confounding factors and reduce the impact of a missing routine diastolic parameter, creating a more robust and widely applicable technique. Use could also be made of all the rich detail collected in an echocardiogram, which alongside detection of latent data relationships and novel parameters could vastly improve diagnostic accuracy for diastolic dysfunction. All of these aspects are likely to be the focus of future research.

From the perspective of the physician, a personalised diastolic assessment with ML would allow tailored investigations and patient management. A reduction or complete elimination of indeterminate diastolic grading would reduce clinical uncertainty and lessen the need for additional complex or invasive investigations, such as cardiac catheterisation. This would also facilitate better decision making in high-risk groups where diastolic status may alter risk-benefit balances of interventions.

Several challenges are apparent though for ML augmented diastolic assessment. Firstly, the thirst of ML algorithms for high-quality, large-volume, high-dimensionality training data is a problem in a world where healthcare systems are still struggling to digitalise and integrate electronic systems. Significant progress has been achieved thus far with modest sample sizes, but a truly "big-data" approach to diastolic assessment seems warranted to validate existing findings, to unlock new insights, and to increase the clinical applicability of results. Limitations of the current evidence are that often patients with missing data, or indeterminate guideline classification, are excluded from studies. Where these patients fit on the spectrum of diastology would be of great interest given their prevalence.

Secondly, a lack of direct invasively measured diastolic data, to use as a "ground truth," also limits ML research about diastolic function in many populations, as patients simply do not receive such a test routinely. Given ethical and resource considerations, it is unlikely that this can be overcome. A third challenge is that a ML diastolic assessment tool would require resilience to changes in preload, afterload, and heart rate, etc. Given oscillatory patterns of, for example, de-compensated heart failure and subsequent treatment, the field should aim to not only categorise diastolic dysfunction, but also to assign a personalised "live" diastolic status, which could then be used to track temporal changes in diastolic function. This would greatly advance the field towards precise and personalised medicine.

Fourthly, a perception that ML represents a "black box" technology, where healthcare professionals do not understand how it is arriving at a decision, is also a problem for ML augmented echocardiography. Input parameter feature weightings, and heat maps of image areas being used by the machine, offer ways for researchers to dispel this perception. Finally, regulatory approval also acts as a potential obstacle for ML innovators.

The field of echocardiography is moving towards a more automated, data-driven, and analytical approach. Diastolic function is not something that can be readily eyeballed—it needs expert clinical insight to meet rigorous science to improve its assessment. How to best integrate a ML diastolic assessment tool into clinical practice should form the basis of future debate.

CONCLUSIONS

Evidence shows that ML can use diastolic parameters to differentiate diseases, improve the accuracy of disease diagnoses, and identify diastolic phenotypes within heterogeneous conditions such as HFpEF. There is also evidence to suggest that ML can improve identification of raised filling-pressures, classify or grade diastolic function in novel ways, and improve upon the prognostic ability of the current diastolic clinical standard. Although there are numerous potential benefits, many challenges stand in the way of progress for the field. ML augmented echocardiography for diastolic assessment is here, but real-world applicability and its relationship to clinical decision making remains to be seen.

AUTHOR CONTRIBUTIONS

AF and PL conceptualised the review topic. AF performed the literature search and drafted the article. All authors edited and reviewed the article and approved the manuscript for submission.

REFERENCES

1. Sato K, Grant A, Negishi K, Cremer P, Negishi T, Kumar A, et al. Reliability of updated left ventricular diastolic function recommendations in predicting elevated left ventricular filling pressure and prognosis. *Am Heart J.* (2017) 189:28-39. doi: 10.1016/j.ahj.2017.03.022
2. Kuznetsova T, Thijs L, Knez J, Herbots L, Zhang Z, Staessen JA. Prognostic value of left ventricular diastolic dysfunction in a general population. *J Am Heart Assoc.* (2014) 3:e000789. doi: 10.1161/JAHA.114.000789
3. Playford D, Strange G, Celermajer DS, Evans G, Scalia GM, Stewart S, et al. Diastolic dysfunction and mortality in 436 360 men and women: the National Echo Database Australia (NEDA). *Eur Heart J Cardiovasc Imaging.* (2020) 22:505-15. doi: 10.1093/ehjci/jeaa253
4. Chetrit M, Cremer PC, Klein AL. Imaging of diastolic dysfunction in community-based epidemiological studies and randomized controlled trials of HFpEF. *JACC Cardiovasc Imaging.* (2020) 13:310-26. doi: 10.1016/j.jcmg.2019.10.022
5. Yeung DF, Jiang R, Behnami D, Jue J, Sharma R, Turaga M, et al. Impact of the updated diastolic function guidelines in the real world. *Int J Cardiol.* (2021) 326:124-30. doi: 10.1016/j.ijcard.2020.10.068
6. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur J Heart Fail.* (2016) 18:891-975. doi: 10.1002/ejhf.592
7. National-Institute-for-Cardiovascular-Outcomes-Research. *UK National Heart Failure Audit Online: NICOR.* (2020). Available online at: https://www.nicor.org.uk/national-cardiac-audit-programme/heart-failure-heart-failure-audit/ (accessed December 18, 2020).
8. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. *N Engl J Med.* (2006) 355:251-9. doi: 10.1056/NEJMoa052256

9. Nagueh SF, Smiseth OA, Appleton CP, Byrd BF, Dokainish H, Edvardsen T, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr.* (2016) 29:277-314. doi: 10.1016/j.echo.2016.01.011
10. Nagueh SF. Left ventricular diastolic function: understanding pathophysiology, diagnosis, and prognosis with echocardiography. *JACC Cardiovasc Imaging.* (2020) 13:228-44. doi: 10.1016/j.jcmg.2018.10.038
11. Sharifov OF, Schiros CG, Aban I, Denney TS, Gupta H. Diagnostic accuracy of tissue doppler index E/è for evaluating left ventricular filling pressure and diastolic dysfunction/heart failure with preserved ejection fraction: a systematic review and meta-analysis. *J Am Heart Assoc.* (2016) 5:e002530. doi: 10.1161/JAHA.115.002530
12. Jones R, Varian F, Alabed S, Morris P, Rothman A, Swift AJ, et al. Meta-analysis of echocardiographic quantification of left ventricular filling pressure. *ESC Heart Fail.* (2021) 8:566-76. doi: 10.1002/ehf2.13119
13. Nauta JF, Hummel YM, Van Der Meer P, Lam CSP, Voors AA, Van Melle JP. Correlation with invasive left ventricular filling pressures and prognostic relevance of the echocardiographic diastolic parameters used in the 2016 ESC heart failure guidelines and in the 2016 ASE/EACVI recommendations: a systematic review in patients with heart failure with preserved ejection fraction. *Eur J Heart Fail.* (2018) 20:1303-11. doi: 10.1002/ejhf.1220
14. Lancellotti P, Galderisi M, Edvardsen T, Donal E, Goliasch G, Cardim N, et al. Echo-Doppler estimation of left ventricular filling pressure: results of the multicentre EACVI Euro-Filling study. *Eur Heart J Cardiovasc Imaging.* (2017) 18:961-8. doi: 10.1093/ehjci/jex067
15. Balaney B, Medvedofsky D, Mediratta A, Singh A, Ciszek B, Kruse E, et al. Invasive validation of the echocardiographic assessment of left ventricular filling pressures using the 2016 diastolic guidelines: head-to-head comparison with the 2009 guidelines. *J Am Soc Echocardiogr.* (2018) 31:79-88. doi: 10.1016/j.echo.2017.09.002
16. Andersen OS, Smiseth OA, Dokainish H, Abudiab MM, Schutt RC, Kumar A, et al. Estimating left ventricular filling pressure by echocardiography. *J Am Coll Cardiol.* (2017) 69:1937-48. doi: 10.1016/j.jacc.2017.01.058
17. Leung EC, Swiston JR, Alahmari L, Alahmari T, Huckell VF, Brunner NW. Validity of algorithm for estimating left sided filling pressures on echocardiography in a population referred for pulmonary arterial hypertension. *Pulm Circ.* (2018) 8:1-7. doi: 10.1177/2045893217740471
18. Almeida JG, Fontes-Carvalho R, Sampaio F, Ribeiro J, Bettencourt P, Flachskampf FA, et al. Impact of the 2016 ASE/EACVI recommendations on the prevalence of diastolic dysfunction in the general population. *Eur Heart J Cardiovasc Imaging.* (2018) 19:380-6. doi: 10.1093/ehjci/jex252
19. Lancaster MC, Salem Omar AM, Narula S, Kulkarni H, Narula J, Sengupta PP. Phenotypic clustering of left ventricular diastolic function parameters. *JACC Cardiovasc Imaging.* (2019) 12:1149-61. doi: 10.1016/j.jcmg.2018.02.005
20. Gottbrecht M, Salerno M, Aurigemma G. Evolution of diastolic function algorithms: implications for clinical practice. *Echocardiography.* (2018) 35:39-46. doi: 10.1111/echo.13746
21. Luke P, Eggett C, Spyridopoulos I, Irvine T. A comparative analysis of British and American Society of Echocardiography recommendations for the assessment of left ventricular diastolic function. *Echo Res Pract.* (2018) 5:139-47. doi: 10.1530/ERP-18-0024
22. Liang HY, Lo YC, Chiang HY, Chen MF, Kuo CC. Validation and comparison of the 2003 and 2016 diastolic functional assessments for cardiovascular mortality in a large single-center cohort. *J Am Soc Echocardiogr.* (2020) 33:469-80. doi: 10.1016/j.echo.2019.11.013
23. Obokata M, Kane GC, Reddy YNV, Olson TP, Melenovsky V, Borlaug BA. Role of diastolic stress testing in the evaluation for heart failure with preserved ejection fraction. *Circulation.* (2017) 135:825-38. doi: 10.1161/CIRCULATIONAHA.116.024822
24. Chubuchny V, Pugliese NR, Taddei C, Poggianti E, Spini V, Barison A, et al. A novel echocardiographic method for estimation of pulmonary artery wedge pressure and pulmonary vascular resistance. *ESC Heart Fail.* (2021) 8:1216-29. doi: 10.1002/ehf2.13183
25. Eapen ZJ, Greiner MA, Fonarow GC, Yuan Z, Mills RM, Hernandez AF, et al. Associations between atrial fibrillation and early outcomes of patients with heart failure and reduced or preserved ejection fraction. *Am Heart J.* (2014) 167:369-75. doi: 10.1016/j.ahj.2013.12.001
26. Santhanakrishnan R, Wang N, Larson MG, Magnani JW, McManus DD, Lubitz SA, et al. Atrial fibrillation begets heart failure and vice versa. *Circulation.* (2016) 133:484-92. doi: 10.1161/CIRCULATIONAHA.115.018614
27. Ma G, Fang L, Gao P, Cheng Z, Chen T, Lin X, et al. Association between the ratio of early diastolic transmitral velocity to early diastolic mitral annular velocity and invasive measured left atrial pressure in patients with atrial fibrillation and preserved left ventricular ejection fraction. *Zhonghua xin xue guan bing za zhi.* (2018) 46:292-7. doi: 10.3760/cma.j.issn.0253-3758.2018.04.008
28. Wijesurendra RS, Liu A, Eichhorn C, Ariga R, Levelt E, Clarke WT, et al. Lone atrial fibrillation is associated with impaired left ventricular energetics that persists despite successful catheter ablation. *Circulation.* (2016) 134:1068-81. doi: 10.1161/CIRCULATIONAHA.116.022931
29. Thenappan T, Prins KW, Cogswell R, Shah SJ. Pulmonary hypertension secondary to heart failure with preserved ejection fraction. *Can J Cardiol.* (2015) 31:430-9. doi: 10.1016/j.cjca.2014.12.028
30. Guazzi M. Pulmonary hypertension in heart failure preserved ejection fraction. *Circ Heart Fail.* (2014) 7:367-77. doi: 10.1161/CIRCHEARTFAILURE.113.000823
31. Diwan A, McCulloch M, Lawrie GM, Reardon MJ, Nagueh SF. Doppler estimation of left ventricular filling pressures in patients with mitral valve disease. *Circulation.* (2005) 111:3281-9. doi: 10.1161/CIRCULATIONAHA.104.508812
32. Abudiab MM, Chebrolu LH, Schutt RC, Nagueh SF, Zoghbi WA. Doppler echocardiography for the estimation of LV filling pressure in patients with mitral annular calcification. *JACC Cardiovasc Imaging.* (2017) 10:1411-20. doi: 10.1016/j.jcmg.2016.10.017
33. Hammond MM, Shen C, Li S, Kazi DS, Sabe MA, Garan AR, et al. Retrospective evaluation of echocardiographic variables for prediction of heart failure hospitalization in heart failure with preserved versus reduced ejection fraction: a single center experience. *PLoS ONE.* (2020) 15:e0244379. doi: 10.1371/journal.pone.0244379
34. Zakeri R, Chamberlain AM, Roger VL, Redfield MM. Temporal relationship and prognostic significance of atrial fibrillation in heart failure patients with preserved ejection fraction. *Circulation.* (2013) 128:1085-93. doi: 10.1161/CIRCULATIONAHA.113.001475
35. Alsharqi M, Woodward WJ, Mumith JA, Markham DC, Upton R, Leeson P. Artificial intelligence and echocardiography. *Echo Res Pract.* (2018) 5:R115-25. doi: 10.1530/ERP-18-0056
36. 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. *Circulation Cardiovasc Imaging.* (2018) 11:e007138. doi: 10.1161/CIRCIMAGING.117.007138
37. Qutrio Baloch Z, Raza SA, Pathak R, Marone L, Ali A. Machine learning confirms nonlinear relationship between severity of peripheral arterial disease, functional limitation and symptom severity. *Diagnostics.* (2020) 10:515-26. doi: 10.3390/diagnostics10080515
38. 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
39. Ghorbani A, Ouyang D, Abid A, He B, Chen JH, Harrington RA, et al. Deep learning interpretation of echocardiograms. *NPJ Digit Med.* (2020) 3:10-20. doi: 10.1038/s41746-019-0216-8
40. Asch FM, Poilvert N, Abraham T, Jankowski M, Cleve J, Adams M, et al. Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. *Circulation Cardiovasc Imaging.* (2019) 12:e009303. doi: 10.1161/CIRCIMAGING.119.009303
41. Amarbayasgalan T, Park KH, Lee JY, Ryu KH. Reconstruction error based deep neural networks for coronary heart disease risk prediction. *PLoS ONE.* (2019) 14:e0225991. doi: 10.1371/journal.pone.0225991
42. Hubert A, Le Rolle V, Galli E, Bidaud A, Hernandez A, Donal E. New expectations for diastolic function assessment in transthoracic echocardiography based on a semi-automated computing of strain–volume loops. *Eur Heart J Cardiovasc Imaging.* (2020) 21:1366-71. doi: 10.1093/ehjci/jeaa123

43. Sengupta PP, Huang Y-M, Bansal M, Ashrafi A, Fisher M, Shameer K, et al. Cognitive machine-learning algorithm for cardiac imaging. *Circ Cardiovasc Imaging.* (2016) 9:e004330. doi: 10.1161/CIRCIMAGING.115.004330
44. Choi DJ, Park JJ, Ali T, Lee S. Artificial intelligence for the diagnosis of heart failure. *NPJ Digit Med.* (2020) 3:54-60. doi: 10.1038/s41746-020-0261-3
45. Hedman ÅK, Hage C, Sharma A, Brosnan MJ, Buckbinder L, Gan L-M, et al. Identification of novel pheno-groups in heart failure with preserved ejection fraction using machine learning. *Heart.* (2020) 106:342-9. doi: 10.1136/heartjnl-2019-315481
46. Przewlocka-Kosmala M, Marwick TH, Dabrowski A, Kosmala W. Contribution of cardiovascular reserve to prognostic categories of heart failure with preserved ejection fraction: a classification based on machine learning. *J Am Soc Echocardiogr.* (2019) 32:604-15. doi: 10.1016/j.echo.2018.12.002
47. 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
48. Kaptein YE, Karagodin I, Zuo H, Lu Y, Zhang J, Kaptein JS, et al. Identifying phenogroups in patients with subclinical diastolic dysfunction using unsupervised statistical learning. *BMC Cardiovasc Disord.* (2020) 20:367-82. doi: 10.1186/s12872-020-01620-z
49. Segar MW, Patel KV, Ayers C, Basit M, Tang WHW, Willett D, et al. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail.* (2020) 22:148-58. doi: 10.1002/ejhf.1621
50. Harada D, Asanoi H, Noto T, Takagawa J. Different pathophysiology and outcomes of heart failure with preserved ejection fraction stratified by K-Means clustering. *Front Cardiovasc Med.* (2020) 7:e607760. doi: 10.3389/fcvm.2020.607760
51. Nouraei H, Rabkin SW. A new approach to the clinical subclassification of heart failure with preserved ejection fraction. *Int J Cardiol.* (2021) 331:138-43. doi: 10.1016/j.ijcard.2021.01.052
52. Horiuchi Y, Tanimoto S, Latif A, Urayama KY, Aoki J, Yahagi K, et al. Identifying novel phenotypes of acute heart failure using cluster analysis of clinical variables. *Int J Cardiol.* (2018) 262:57-63. doi: 10.1016/j.ijcard.2018.03.098
53. Mishra RK, Tison GH, Fang Q, Scherzer R, Whooley MA, Schiller NB. Association of machine learning-derived phenogroupings of echocardiographic variables with heart failure in stable coronary artery disease: the heart and soul study. *J Am Soc Echocardiogr.* (2020) 33:322-31. doi: 10.1016/j.echo.2019.09.010
54. Pecková M, Charvat J, Schuck O, Hill M, Svab P, Horackova M. The association between left ventricular diastolic function and a mild-to-moderate decrease in glomerular filtration rate in patients with type 2 diabetes mellitus. *J Int Med Res.* (2011) 39:2178-86. doi: 10.1177/147323001103900614
55. Omar AMS, Narula S, Abdel Rahman MA, Pedrizzetti G, Raslan H, Rifaie O, et al. Precision phenotyping in heart failure and pattern clustering of ultrasound data for the assessment of diastolic dysfunction. *JACC Cardiovasc Imaging.* (2017) 10:1291-303. doi: 10.1016/j.jcmg.2016.10.012
56. Omar AMS, Shameer K, Narula S, Abdel Rahman MA, Rifaie O, Narula J, et al. Artificial intelligence-based assessment of left ventricular filling pressures from 2-dimensional cardiac ultrasound images. *JACC Cardiovasc Imaging.* (2018) 11:509-10. doi: 10.1016/j.jcmg.2017.05.003
57. Tokodi M, Shrestha S, Bianco C, Kagiyama N, Casaclang-Verzosa G, Narula J, et al. Interpatient similarities in cardiac function: a platform for personalized cardiovascular medicine. *JACC Cardiovasc Imaging.* (2020) 13:1119-32. doi: 10.1016/j.jcmg.2019.12.018
58. Carlsson G. Topology and data. *Bull Am Math Soc.* (2009) 46:255-308. doi: 10.1090/S0273-0979-09-01249-X
59. Cho JS, Shrestha S, Kagiyama N, Hu L, Ghaffar YA, Casaclang-Verzosa G, et al. A network-based "Phenomics" approach for discovering patient subtypes from high-throughput cardiac imaging data. *JACC Cardiovasc Imaging.* (2020) 13:1655-70. doi: 10.1016/j.jcmg.2020.02.008
60. Pandey A, Kagiyama N, Yanamala N, Segar MW, Cho JS, Tokodi M, et al. Deep-learning models for the echocardiographic assessment of diastolic dysfunction. *JACC Cardiovasc Imaging.* (2021). doi: 10.1016/j.jcmg.2021.04.010. [Epub ahead of print].
61. Stone M. Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Ser B.* (1974) 36:111-33. doi: 10.1111/j.2517-6161.1974.tb00994.x
62. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Health.* (2021) 3:e195-203. doi: 10.1016/S2589-7500(20)30292-2
63. Fraser AG, Butchart EG, Szymański P, Caiani EG, Crosby S, Kearney P, et al. The need for transparency of clinical evidence for medical devices in Europe. *Lancet.* (2018) 392:521-30. doi: 10.1016/S0140-6736(18)31270-4

Conflict of Interest: PL acknowledges current grant support related to medical imaging from the British Heart Foundation, Wellcome Trust, National Institute of Health Research and Lantheus Medical Imaging. PL is a stockholder, non-executive director and co-founder of Ultromics: a medical imaging artificial intelligence company.

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cardiovascular Medicine

REVIEW
published: 11 August 2021
doi: 10.3389/fcvm.2021.695547

Improving Diuretic Response in Heart Failure by Implementing a Patient-Tailored Variability and Chronotherapy-Guided Algorithm

Ariel Kenig[1†], *Yotam Kolben*[1†], *Rabea Asleh*[2], *Offer Amir*[2,3] *and Yaron Ilan*[1*]

[1] *Department of Medicine, Hebrew University-Hadassah Medical Center, Jerusalem, Israel,* [2] *Department of Cardiology, Hebrew University-Hadassah Medical Center, Jerusalem, Israel,* [3] *The Azrieli Faculty of Medicine in the Galilee, Bar-Ilan University, Safed, Israel*

OPEN ACCESS

Edited by:
Paul Leeson,
University of Oxford, United Kingdom

Reviewed by:
Julio Nunez,
Hospital Clínico Universitario de Valencia, Spain
Abhijit Chakraborty,
Baylor College of Medicine, United States
Zhexue Qin,
Xinqiao Hospital, China
Frank Davis,
University of Michigan, United States

***Correspondence:**
Yaron Ilan
ilan@hadassah.org.il

[†]*These authors share first authorship*

Specialty section:
This article was submitted to Cardiovascular Therapeutics, a section of the journal Frontiers in Cardiovascular Medicine

Received: *15 April 2021*
Accepted: *21 July 2021*
Published: *11 August 2021*

Citation:
Kenig A, Kolben Y, Asleh R, Amir O and Ilan Y (2021) Improving Diuretic Response in Heart Failure by Implementing a Patient-Tailored Variability and Chronotherapy-Guided Algorithm. Front. Cardiovasc. Med. 8:695547. doi: 10.3389/fcvm.2021.695547

Heart failure is a major public health problem, which is associated with significant mortality, morbidity, and healthcare expenditures. A substantial amount of the morbidity is attributed to volume overload, for which loop diuretics are a mandatory treatment. However, the variability in response to diuretics and development of diuretic resistance adversely affect the clinical outcomes. Morevoer, there exists a marked intra- and inter-patient variability in response to diuretics that affects the clinical course and related adverse outcomes. In the present article, we review the mechanisms underlying the development of diuretic resistance. The role of the autonomic nervous system and chronobiology in the pathogenesis of congestive heart failure and response to therapy are also discussed. Establishing a novel model for overcoming diuretic resistance is presented based on a patient-tailored variability and chronotherapy-guided machine learning algorithm that comprises clinical, laboratory, and sensor-derived inputs, including inputs from pulmonary artery measurements. Inter- and intra-patient signatures of variabilities, alterations of biological clock, and autonomic nervous system responses are embedded into the algorithm; thus, it may enable a tailored dose regimen in a continuous manner that accommodates the highly dynamic complex system.

Keywords: heart failure, diuretic resistance, chronobiology, variability, digital systems

INTRODUCTION

Heart failure (HF) is a staggering clinical and public health problem with high morbidity and mortality burden, affecting more than 6 million individuals in the United States where ∼670,000 individuals are diagnosed with HF each year (1). Patients with HF are frequently hospitalized with HF exacerbation and have reduced quality of life and increased mortality rates. HF exacerbation is the leading cause of hospitalization among patients aged >65 years (2).

HF is associated with high rates of morbidity and reduced quality of life which are related to symptoms of volume overload resulting from sodium retention and volume overload. Loop diuretics are commopnly used for the teratment of volume overload resulting from HF. There is a marked inter and intra-patient variability that characterizes the response to diuretics at different stages. These variations may alter patients' clinical course and adversely affect their clinical outcomes (3, 4). A diminished response to loop diuretics is a well-recognized clinical challenge,

which limits their clinical use (5, 6). Current methods for dose adjustment are somewhat arbitrary and do not take into consideration patients' resistance to diuretics (6). A continuous increase in the diuretic doses may further worsen diuretic resistance.

In the present article, we review the common mechanisms underlying the development of diuretic resistance and discuss current methods to overcome this devastating phenomenon. The role of the autonomic nervous system (ANS) and chronobiology in the pathogenesis of congestive heart failure (CHF) are discussed. The implementation of a patient-tailored variability and chronotherapy-guided algorithm for the treatment of these patients is presented as a potential strategy for alleviating drug resistance and improving their long-term beneficial effects.

DIURETIC RESISTANCE IS A MAJOR DIFFICULTY IN THE TREATMENT OF HF

The effects of loop diuretics on systemic and renal hemodynamics are driven by multiple factors, including the dose, route of administration, concomitant diseases, and medications and chronicity of their use (7). Furosemide is a potent prototypic loop diuretic that exerts its effect by binding to the translocation pocket at the extracellular surface of sodium-potassium-chloride symporters (NKCCs). It blocks ion transport directly by inhibiting NKCCs at the apical surface of the thick ascending loop of Henle (7, 8). The half-life of loop diuretics is generally shorter than typical dosing intervals of twice daily. Additionally, as they inhibit solute transport at only a single segment out of the numerous sodium-reabsorbing nephron sections, their impacts on extracellular fluid volume are multifaceted. Loop diuretics cause renal vasodilation through direct vascular dilation of the afferent arterioles and inhibition of the tubuloglomerular feedback (9).

Definition of Diuretic Resistance

Diuretic resistance is defined as the failure to achieve effective decongestion with low urine sodium concentration despite the use of appropriate or escalating diuretic doses (10). When diuretic resistance develops, the response to diretics is reduced before the achieving the goal of treatment, leading to poor prognosis. Resistance (or tolerance) to diuretic therapy can develop over time, making volume reduction in HF more challenging. There are two forms of diuretic tolerance, namely, short- and long-term resistance. Short-term tolerance occurs when the effect of the diuretic is weakened after the first dose. It may be prevented by reestablishing diuretic-induced loss of volume. Long-term tolerance is observed following administration of a loop diuretic for prolonged periods of time, which is associated with sodium reabsorption at the distal sites (10).

Factor That Are Associated With Diuretic Resistance

Multiple factors can explain non-responsiveness to diuretics, including inadequate doses, lack of adherence, advanced age, high sodium intake, impaired secretion into the tubule lumen, chronic kidney disease (CKD), gut edema, use of non-steroidal anti-inflammatory agents, hypoproteinemia, hypotension, nephrotic syndrome, reduced renal blood flow, and neurohormonal activation (7, 11).

Tolerance at the receptor- or post-receptor points may be associated with diuretic tolerance or resistance (11). In the circulation, loop diuretics are bound to various proteins (mainly albumin) and secreted into the tubules by the organic anion transporters (OAT1 and OAT2) located at the basolateral membrane and multidrug resistant protein-4 located at the apical membrane. As diuretics compete with chloride for binding to NKCC2, increased salt absorption in the proximal tubules limits the diuretic-sensitive transport (7).

In patients with HF, RAAS activation results in sodium retention. These patients may manifet with renal dysfunction assciated with additional activation of neurohormones. As HF progresses, persistent activation of these neurohormonal systems enhances sodium retention and contributes to the development of diuretic resistance (12–14). Worsening renal function in HF reduces the usefulness of loop diuretics by reducing their secretion into the renal tubules, a process mediated by increased organic ions competing for organic ion transporter binding (15).

Compensatory Responses to Diuretics

Compensatory responses to diuretics may increase drug resistance. The diuretic-associated increase of the plasma renin activity promotes angiotensin II while blocking the tubuloglomerular feedback for increasing the glomerular filtration rate (16). Angiotensin II stimulates proximal tubular sodium reabsorption, thereby reducing distal sodium delivery and leading to diuretic resistance (11). Elevated plasma levels of both angiotensin II and aldosterone activate sodium transporters in the distal nephron (17). Post-diuretic sodium retaining may occur once the loop diuretic concentration drops below a certain threshold in the renal tubules. Sodium reabsorption is increased in the distal tubules and collecting ducts. This effect counters the loop diuretic effects (14). Adding of a second or third daily dose of loop diuretic may overcome this effect but may be associated with further long-term tolerance.

Abbreviations: CHF, congestive heart failure; HF, heart failure; ANS, autonomic nervous system; NKCC, sodium-potassium-chloride symporter; RASS, renin-angiotensin-aldosterone system; CKD, chronic kidney disease; OAT, organic anion transporter; SNS, sympathetic nervous system; ADHF, acute decompensated heart failure; PA, pulmonary artery; QALY, quality-adjusted life years; SCN, suprachiasmatic nucleus; CLOCK, circadian locomotor output cycle kaput; BMAL1, brain and muscle ARNT-like 1; Per, period; Cry, cryptochrome; BP, blood pressure; HR, heart rate; L/D, light/dark; HFrEF. heart failure with reduced ejection fraction; MI, myocardial infarction; SCD, sudden cardiac death; HRV, heart rate variability; MAP, mean arterial pressure; ACEi, angiotensin converting enzyme inhibitors; Klf15, kruppel-like factor 15; SNS, sympathetic nervous system; LF, low frequency; HIF, high frequency; BRS, baroreflex sensitivity; OSA, obstructive sleep apnea; CVHR, cyclic variation of heart rate; FMD, flow-mediated dilation; SA, sleep apnea; LVAD, left ventricular assist device; SDANN, standard deviation of averages of normal R-R intervals; AT1R, angiotensin II type 1 receptor; RA, renin activity; Una, fe, fractional excretion of sodium in the urine; UK, fe, fractional excretion of potassium in the urine; CANA, cardiac autonomous nervous activity; CSA, central sleep apnea.

Distal Nephron Remodeling

An additional contributing factor to diuretic resistance is the "braking phenomenon." Chronic use of high dosages of loop diuretics leads to distal nephron remodeling, which involves hypertrophy and hyperplasia of the distal convoluted and connecting tubules and the collecting duct. This process augments the reabsorption capacity of the distal nephron and results in enhanced sodium reabsorption that negates the anticipated beneficial effects of these diuretics (18–20). This process is partially mediated by the RAAS (21) and activation of the sympathetic nervous system (SNS) (22). Adding a thiazide diuretic that blocks sodium reabsorption in the distal tubules This occurrence can overcome this effect; thus, it augments the net loss of sodium. Activation of baroreceptors in the arterial system triggers the SNS and RAAS, nephron remodeling, and extracellular volume depletion, thereby resulting in sodium retention (23). Extracellular volume contraction, which occurs with prolonged diuretic use in the setting of persistent congestion, contributes to the development of drug resistance (7).

CURRENT MEASURES FOR OVERCOMING DIURETIC RESISTANCE

Several pharmacologic and non-pharmacologic interventions are being used in an attempt to improve the diuretic response, including intravenous administration, increasing diuretic doses, and changing diuretic agents (7, 11). Several of these are directly or indirectly related to continuous daily administration of similar or higher dosages of diuretics. However, despite the prevalence of diuretic resistance, there remains a paucity of clinical trials to provide evidence on how to mitigate the resistance and guide therapy when patients develop this phenomenon (24).

Increasing dosages of loop diuretics may lead to a plateau in their effect, suggesting that raising doses beyond a "ceiling" will not further augment response (7). High diuretic doses, which stimulate the RAAS and SNS, are associated with worse outcomes, raising the possibility that they should be avoided in patients with severe decompensated HF (25). However, the Diuretic Optimization Strategies Evaluation study demonstrated that patients with acute decompensated HF on higher diuretic doses showed more favorable outcomes, including significant relief of dyspnea, reduction in weight, and higher net fluid loss, although they had a higher incidence of worsening renal function. Notably, the initial elevation in serum creatinine levels was associated with improved clinical outcomes (26).

Combined diuretic therapy is increasingly used to overcome high sodium retention. A common way to overcome diuretic resistance is to combine a diuretic with other medications including mineralocorticoid receptor antagonists, acetazolamide, or metolazone (6). The sequential nephron blockade by combining loop and thiazide diuretics in patients with inadequate response to optimal doses of loop diuretics was reportedly effective in some studies (27). Dopamine can improve renal perfusion and exert a diuretic effect (28). Hypertonic saline transiently increases the intravascular volume and improves sodium delivery to the tubules of the nephron (28). Additionally, metalazone (zaroxylin) in combination with loop diuretics or ultrafiltration is also being used in patients with resistance (29, 30).

The CardioMEMS HF System was designed for monitoring and measuring pulmonary artery (PA) pressure in patients with CHF. It transmits daily sensor readings from patients to their healthcare providers, allowing for personalized diuretic management in order to reduce the likelihood of hospitalization. In a retrospective analysis, PA pressure-guided management reduced HF hospitalization by 43% and mortality by 57% (31–33). The CHAMPION trial was a controlled, single-blind study of 550 patients with New York Heart Association class III HF and an HF hospitalization in the previous year. After 6 months of follow-up, the PA sensor-actively monitored patients who experienced an increased frequency of medication adjustments, higher dosages of diuretics, and diuretics intensification (34). Compared with the standard of care, its use was cost-effective, with an incremental cost-effectiveness ratio per quality-adjusted life year (31). However, dose adjusting is marginally arbitrary and does not consider patients developing resistance to diuretics.

Taken together, similar mechanisms are responsible for both the resistance and adaptation to diuretics. Several measures taken for overcoming diuretic resistance may further enhance the vicious cycle of inducing resistance and jeopardizing the clinical condition. Continuous administration on a regular basis of the same or higher dosages of diuretics is associated with a vicious cycle of actually augmenting the resistance.

DISRUPTION OF THE CIRCADIAN RHYTHM IN HF

Chronobiology describes the control of multiple biological functions by the circadian rhythm. Many cellular and physiological processes exhibit a circadian rhythm, oscillating approximately once in 24 h. These endogenous cycles enable the organism to optimally arrange its patterns of behavior in synchronization with the predictable changes in environmental conditions (35, 36). The central pacemaker is loated in the suprachiasmatic nuclei (SCN) and is synchronized with multiple peripheral clocks in various cells (37). In the cellular level, four genes—*circadian locomotor output cycle kaput* (*clock), brain and muscle ARNT-like 1 (bmal1), period 1 and 2 (Per1,2),* and *Cryptochrome (Cry)*—form a transcriptions-translation feedback loop that cycles approximately every 24 h, providing the periodicity of the cellular circadian rhythm.

Daily Oscillations in Heart and Blood Vessels

The heart and blood vessels are characterized by marked daily oscillations in gene expression, metabolism, growth, and remodeling. The circadian clocks within the cardiomyocytes are linked to the regulation of myocardial function and metabolism (38). A proper response of the heart to its diurnal environment is mandatory for survival. It involves response to changes in workload, nutrients, neurohumoral stimuli, and metabolic alterations. Blood pressure (BP), heart rate (HR), coagulation

activity, and endothelial function manifest in a day/night pattern (39).

Chrono Disruption of the Circadian Rhythm

Chronodisruption is an alteration of circadian rhythms associated with diverse diseases, including cardiovascular diseases and malignancy. Night-shift workers, who are exposed to artificial light disrupting the endogenous circadian rhythm, have a higher risk of all-cause and cardiovascular mortality (40). Obesity, high triglycerides, and low high-density lipoprotein cholesterol levels are more common in night-shift workers (41). Brachial artery endothelial function is determined as the reaction to reactive hyperemia (flow-mediated dilation [FMD]) and was irregular in residents following a 24-h shift, including night duty. A marked decrease in FMD was noted in physicians with a history of night-shift duty and in those reporting fewer sleeping hours during the shift (42).

Chronodisruption was described with acute and chronic cardiovascular events. Intramyocellular circadian clock and diurnal variations in fatty acid responsiveness were noted in the rat cardiomyocytes. Reversal of the 12-h/12-h light/dark (L/D) cycle led to re-entrainment of the circadian clock in the heart. Disruption of the circadian clock within the heart via the expression of a dominant negative CLOCK mutant lowered the promotion of myocardial fatty acid-responsive genes during fasting (43). Under pathologic conditions, the endogenous circadian phase corresponds to 10 a.m., the peak time for adverse cardiac events. Diurnal rhythm disruption after myocardial infarction (MI) hinders cardiac healing and exacerbates pathological cardiac remodeling. In a mouse model of MI, a short-term diurnal disruption adversely affected metabolism while altering innate immune responses shown by differences in cytokines, cardiac myeloperoxidase, and macrophage and neutrophil infiltration. Clock mutant mice showed changed infiltration after MI, which is linked to the innate immune responses required for scar formation and associated with reduced blood vessel formation in the infarct region, increased left ventricular dilation, and infarct expansion (44).

In an animal study, the likelihood of ventricular fibrillation (VF) was assessed by the time of the day during which acute coronary failure occurred. A coronary failure between 15.30 and 18.00 led to irreversible VF and death while modeling a similar condition from 11.00 to 15.00 was not associated with VF (45). In a study of 268 consecutive healthy subjects, FMD was diminished in early compared to late morning post-waking hours and predicted long-term cardiovascular events in healthy subjects with no known heart disease (46). FMD of the brachial artery analyzed three times a day (6:30 a.m., 11:30 a.m., and 9 p.m.) in patients with idiopathic dilated cardiomyopathy was lower during the day, manifesting reduction of the normal circadian variation in endothelial function (47).

A meta-analysis encompassing more than 2 million people, described a moderately increased risk for MI and ischemic stroke among shift workers (48). Another chronobiological characteristic of cardiovascular diseases is exhibited in the varying prevalence of adverse cardiovascular events throughout the day. MI, ischemic strokes, and sudden cardiac death (SCD) have higher incidences in the morning hours, corresponding with an increase in the sympathetic tone, peripheral arteries resistance, platelet aggregability, decreased parasympathetic tone, and endothelial function (39).

Chrono Disruption Is Associated With Heart Failure

Disruption of the circadian rhythm contributes to the pathogenesis of HF (49). In HF, the dynamics are associated with an increased parameter of the scaling exponent of the inter-beat interval. In these patients, the peak in the scaling exponent at the circadian phase corresponds to the time of increased heart vulnerability. The endogenous circadian-mediated effects on cardiac regulation are linked to a day/night pattern of adverse cardiac events (50). The heart rate variability (HRV) and mean arterial pressure (MAP) variability were decreased in rats with HF; this decrease was accompanied by disturbances in the normal circadian pattern of HR and BP (51). Differences were demonstrated in the circadian outcome of angiotensin converting enzyme inhibitors (ACEis) on BP in rats with HF (52). The circadian clock contributes to regulation of the mitochondrial metabolism and maintaining the cardiac function. Ablation of the Bmal1circadian clock gene is associated with mitochondrial defects in the heart including morphological changes and reduced enzymatic activities within the respiratory complex, reduced expression of cardiac genes associated with the tricarboxylic acid cycle, fatty acid oxidative pathway, and mitochondrial respiratory chain. These changes were associated with the development of HF. Similar changes were noted in mice exposed to the chronic reversal of the L/D cycle and disrupted circadian rhythmicity (53). In a study of 1,401 asymptomatic subjects in the Cardiovascular Health Study with interpretable 24-h baseline Holter recordings, irregular HRV parameters were related with CHF. Combined with higher incidence of PVCs, HRV enhanced the predictive power of the Health ABC score (54). Autonomic dysfunction quantified HRV parameters characterized subjects with no benefits from cardiac resynchronization treatment, suggesting that pre-implant HRV study helps in improving qualifications for this treatment. In a study of 719 subjects with normal sinus rhythm enrolled in MADIT-CRT (Multicenter Automatic Defibrillator Implantation Trial-Cardiac Resynchronization Therapy), followed for over 3 years, 124 patients reached the primary end point of heart failure or death and 47 died. In multivariate analysis, low SDNN or low VLF was related to a significantly increased risk of HF and mortality (55).

A blunt BP circadian rhythm in HF was documented in humans and linked to the disappearance of circadian variation in atrial natriuretic peptide (56) and in normotensive patients, with the degree of left ventricle function impairment (57). A narrower decrease in the nocturnal systolic BP correlated with a lower ejection fraction. Respiratory function also varies between the light and dark periods, and these L/D variations were shown in HF. In an animal model, HF was associated with blunted L/D differences in resting and chemoreflex breathing, suggesting

that the HF disrupts cardiovascular and respiratory circadian rhythms. Cyclic variation of heart rate (CVHR) is linked with sleep-disordered breathing reflecting cardiac autonomic responses to apneic/hypoxic stress. Blunted CVHR in a night-time Holter ECG predicts increased mortality risk in patients with post-MI, end-stage renal disease, and HF. In a study of 717 patients after MI, decreased cyclic variation amplitude was a predictor of mortality (58). Diurnal variations in respiration and arterial BP were abolished in HF with reduced ejection fraction (HFrEF) (59). Circadian oscillations in calcineurin-dependent activities in the left ventricle of normal mice showed that calcineurin-dependent transcript levels and nuclear occupancy of the nuclear factor of activated T-cells fluctuate over the course of a day, peaking in the morning when mice enter a period of rest (60).

SCD, a prominent cause of death in patients with HF, exhibits diurnal variation and is linked to irregularities in the duration (short or long QT syndromes and HF) or pattern (Brugada's syndrome) of myocardial repolarization. Cardiac ion-channel expression and QT-interval duration, which reflects myocardial repolarization, manifest endogenous circadian rhythmicity controlled by a clock-dependent oscillator, Krüppel-like factor 15 (Klf15). Klf15 transcriptionally regulates a rhythmic expression of the Kv channel-interacting protein 2, which is necessary for the transient outward potassium current (61).

Overall, these data suggest a possible association between HF and alteration in the normal circadian rhythms, implying that novel treatment strategies for HF must take into consideration the timing of treatment administration, adjusting it to the individualized circadian rhythm of patients.

THE AUTONOMIC NERVOUS SYSTEM REGULATES THE CIRCADIAN DISRUPTION IN HF

The ANS underlie some mechanisms associated with CHF. The sympathetic nerve fibers travel alongside the coronary arteries and terminate in the sub-epicardium. Its activation results in accelerated HR, augmented contractility, and increased afterload. In contrast, the parasympathetic system, which branches from the vagal nerve to the sub-endocardium, slows the HR (62). Chronic sympathetic overactivation was described in HF and diuretic resistance and is related to components of the metabolic syndrome, such as obesity, dyslipidemia, BP elevation, and reduced fasting glucose with hyperinsulinemia (63, 64).

Figure 1 describes the interfaces between the circadian rhythm and the autonomic nervous system, contributing to the pathogenesis of congestive heart failure.

The Role of Sodium in Circadian Rhythm

Dietary sodium affects the circadian oscillators downstream of the master light-dark-adjusted pacemaker in the SCN. The chronobiology of renin activity (RA), BP, and fractional excretion of renal sodium (UFENa) and potassium (UFEK) handling in relation to meal timing was studied in dogs. Data showed that RA, UFENa, UFEK, diastolic and systolic BP fluctuate with a circadian periodicity in dogs fed at 07:00 h with a regular diet. Modeling using a fixed 24-h period reflected the variations of UFEK, RA, and BP, and cyclic changes in UFENa suggested a postprandial sodium excretion and a monotonous decay. A delay in the feeding schedule by 6- or 12-h was associated with a shift of comparable magnitude in the rhythm of these biomarkers (65).

Diurnal Rhythm of the ANS Affect Cardiac Parameters

The diurnal rhythm in the standard deviation of the averages of normal R-R intervals (SDANN) and LF/HIF ratio is disrupted in MI. L/D differences in the LF/HIF ratio change from negative to positive values along with a decrease in SDANN, HR, LF/HIF ratio in the dark phase, and elevated plasma norepinephrine levels (66). These data suggest that the timing of the disturbance of diurnal rhythm in SDANN and the LF/HIF ratio are different from those in HR and in the plasma norepinephrine levels (67). In patients with HF, arterial underfilling caused by decreased cardiac output or peripheral arterial vasodilation activates the SNS, RAAS, and non-osmotic vasopressin release (68). Excess activation of the SNS, which aims to maintain cardiac output, has deleterious effects on the heart in the long run. Activation of the SNS in HF results from interactions between the appropriate compensatory reflexes to pathologic excitatory stimuli associated with the depressed systolic function and additional comorbidities. These interactions elicit adrenergic activation in excess to homeostatic requirements (62). To counteract the increased SNS activation, vagal nerve stimulation was suggested as a novel strategy. However, although profound anti-arrhythmic effects were exerted and cardiac function in HF models was improved (69), this method failed to reduce mortality and disease progression in HFrEF patients in the INOVATE-HF trail (70).

The Circadian Rhythm of ANS Affects Heart Rate and Blood Pressure

The circadian pattern of HR and BP, which are affected by the ANS, is disrupted in patients with HF. In an animal model of MI, the association between vagal nerve activity and stellate ganglion nerve activity was documented. A circadian variation following MI reached a peak at a time when sympathetic activation was the uppermost and vagal activity was the lowest (71). The circadian and short-term regulation of BP and HR were shown to be preserved in young, non-failing beta1-transgenic mice, suggesting that the loss of blood pressure and HRV in HF cannot be attributed to over activity of the sympathetic system. However, it reveals loss of adrenergic responsiveness to changes in the activity of the ANS (72). Angiotensin II participates in abnormal autonomic cardiovascular control, which occurs during HF progression. In a model of post-MI, HR increased with the severity of HF, loss of circadian HR, MAP, and BRS rhythms were noted, along with an upregulation in the central angiotensin II type 1 receptors (AT1R) and gp91 proteins in the brainstem. Losartan reduced AT1R expression in daytime but failed to restore circadian variability (73). HF is associated with

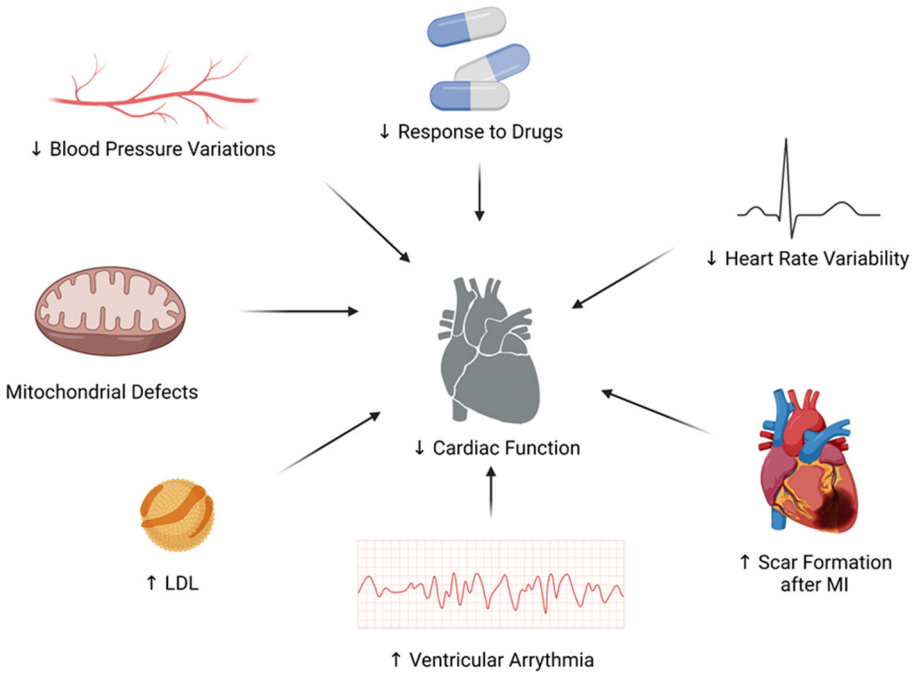


FIGURE 1 | The interfaces between the circadian rhythm and the autonomic nervous system, contributing to the pathogenesis of congestive heart failure.

increased LF constituents of diastolic BP variability, an index of sympathetic tone, during the awakening period compared to during the sleeping period. Amiodarone suppressed this transient increase in LF power during the awakening period (74). Circadian changes in ANS function were also observed at the molecular level. Rats with coronary artery ligation-induced HF showed a shift in the adrenoreceptor beta 1:beta 2 ratio toward beta 2 and decreased beta 1 adrenergic stimulation by adenylyl cyclase. These findings were in association with disturbed circadian patterns in BP and HR (75).

Heart Rate Variability as a Measure of the Association Between ANS and Heart Failure

HRV is a method to visualize the link between ANS and circadian dysfunction to the development of HF. It measures irregularity of intervals between adjacent heartbeats, representing a neurocardiac function and ANS regulation state (76). Normal HRV presents a circadian pattern with parasympathetic parameters peaking at night-time and shows sleep-stage dependence (77, 78). HRV alteration is associated with numerous disease states and correlates with cardiovascular and all-cause mortality, mainly in patients with established cardiovascular disease (79). A decrease in the HRV was associated with increased mortality and was a better predictor of death than conventional clinical management (80). HRV parameters are altered in both animal models of HF and in humans. In a study determining autonomic parameters 3 and 7 weeks after left coronary artery ligation in rats, LF and HF parameters of HRV were reduced in CHF 3 weeks after infarction, in addition to the prolonged loss of baroreflex sensitivity (BRS). Correlation between HRV and MAP variability in the LF domain was reduced in HF (81). HRV and BRS are severely affected in patients with ADHF and improve with clinical stabilization (77–80, 82). In addition, HRV parameters are abnormal in stable patients with HF. HRV parameters were linked to the incidence of CHF in 1,401 asymptomatic, older adults (54). Increased sympathetic activity in CHF is associated with obstructive and central sleep apnea (SA). In a study of patients with CHF and SA, a lowered cardiac autonomic modulation across the 24-h period was documented. The RR variance, LF, HIF parameters of HRV, and spontaneous BRS were reduced in subjects with SA. The HIF power, a marker of vagal activity, elevated during sleep in patients without SA, whereas it did not alter across the 24-h period in subjects with SA. In a study of 167 patients with CHF with central sleep apnea (CSA) and obstructive sleep apnea (OSA), morning premature ventricular contractions and non-sustained VT were more frequent in CSA. CSA was linked to the occurrence of VT irrespective of sleep/wake status (83–85).

Patients with CHF manifested with lower RR interval complexity and loss of its circadian rhythm, along with reduced frequency-domain RR interval variability and its irregular circadian rhythm (86). The circadian variability of RR and QT intervals is altered in CHF due to neurohumoral activation and functional and structural remodeling of the heart. In a study of 121 patients with HF in the sinus rhythm, all subjects showed marked circadian rhythms in the QT, RR, and QTc intervals and the QT/RR slope by cosine-curve fitting. The increased HR was associated with longer QT interval, and steeper QT/RR slope,

lowered circadian variability of QT interval and later maximum RR interval were related to increased cardiac mortality (87).

As sympathetic activation underlies the pathogenesis of HF and diuretic resistance, the selective reduction of the renal afferent and sympathetic efferent nerves was proposed to improve diuretic resistance, CHF, and cardio renal disorders (63). Treatments directed against neurohormonal compensatory mechanisms, such as losartan, spironolactone, and beta-blockers, resulted in an improved HRV profile (88, 89). The favorable effect of spironolactone on the ANS showed a circadian pattern and was limited to the morning hours. Patients with decompensated HF on beta-blockers had a higher power in all HRV frequency bands but were significantly associated with low circulating norepinephrine levels (90). Altering effect of these drugs on the circadian pattern of HRV is known in ischemic heart disease (91).

Implantation of a left ventricular assist device (LVAD) can increase the reduced HRV, which correlates with the degree of ventricular dysfunction. HRV was lower in LVAD patients early after device implantation, suggesting that cardiac dysautonomia is still evident in the first 2 months (92). The algorithm provided by LVADs shows circadian variation. In a study of 18 patients with fixed rotation speed continuous-flow LVAD support, the motor current was lower during the night and higher during the day (93).

Despite these supportive data, other studies failed to support circadian variation in patients with HF. Analysis of data from the PRAISE trial showed that SCD in HF did not show an a.m. peak, signifying that circadian sympathetic activation did not affect these events (94). A study of eight patients with mild-to-moderate HF showed that circadian variation of the cardiac autonomic nervous activity (CANA) was preserved. The CANA was affected by patient position and the time of day. Circadian variation was noted in the recumbent position. The HIF power was lower and LF/HIF was higher in the morning than at night (95).

Overall, the currently available data support the role of the ANS in regulating the heart function under circadian rhythm in health and disease. As many of the effective treatments in HF aim to antagonize the neurohormonal compensatory mechanisms, a personalized approach based on understanding the ANS characteristics and circadian rhythm changes may confer clinical benefits.

CHRONOTHERAPY-BASED THERAPY FOR CHF

Chronobiology impacts the effectiveness and toxicity of drugs and is linked to the pharmacodynamics and pharmacokinetics of medications. Chronotherapy is based on linking the absorption, metabolism, and elimination of medications to the circadian patterns (96) and involves the use of a specific drug at the most optimal time, pattern, and dose possible, with the goal of maximizing efficacy and minimizing toxicity. The ideal timing is based on circadian-rhythm-dependent biological systems, which alter drugs bioavailability, pharmacokinetics, and pharmacodynamics (97, 98). Identification of a rhythmic marker for selecting the dosing time has been suggested to improve the response to drugs (99).

Table 1 summarizes examples of the linkage between administration time and effect in cardiovascular drugs.

Chronotherapeutic benefits of antihypertensive medications have been demonstrated (107). A study of 19,084 hypertensive patients treated with BP-lowering drugs at bedtime, as opposed to upon wakening, reported improved BP control and cardiovascular outcomes, including MI, HF, and cardiovascular death (101). Low-dose aspirin administered at bedtime compared with during the awakening hours can improve platelet inhibition during the critical morning hours (102). Several studies have supported the chronotherapeutic attributes of loop diuretics. Administration of furosemide to rats at 10 a.m. showed greater urine volume and urinary excretion of sodium and furosemide than that at 10 p.m. A correlation between the urinary output of furosemide and urinary volume of sodium, representing the known mechanism of furosemide, has been described (104).

Pretreatment with clorgyline, a monoamine-oxidase inhibitor that alters circadian variations in the SNS, diminished the variability in furosemide effect (108). Pretreatment with propranolol also caused a loss of variability, while atenolol did not cause the same effect, suggesting the involvement of the beta2-adrenergic receptor (109). In two additional studies, continuous infusion of norepinephrine and denervation of the kidneys caused the same influence of diminished variability (110, 111). In humans, the efficacy of furosemide was improved by bedtime dosing as compared to dosing upon awakening. The fraction of patients with controlled ambulatory BP following therapy was also increased after bedtime treatment (105).

Chronotherapy can control the RAAS overactivation in HF. The efficacy of ACEi in HF depends on the time of administration. ACEis were more effective in preventing heart remodeling when administered during sleep (103). Dosing of ACEi in the evening, at an inactive period, was associated with an improved protective effect against heart hypertrophy in hypertensive rats and also proposed to reduce the severity of the drug-related dry cough of hypertensive subjects treated in the morning (100). Bedtime administration of valsartan normalizes the circadian rhythm and protects the kidneys and heart in patients with CKD (106). An integrated pharmacokinetic-pharmacodynamics model was designed based on the kinetics of the drug and time-varying changes of RAAS, showing that the optimal efficacy of ACEis is achieved with bedtime dosing (112).

Melatonin determines the circadian physiology, and its circulating levels vary on a daily cycle, allowing the regulation of circadian rhythms of multiple biological systems. It was shown to exert cardioprotective effects impacting the clinical course of HF. In a study of 32 patients with CHF, nocturnal melatonin secretion negatively associated with N-terminal pro-brain natriuretic peptide (NT-proBNP). Higher melatonin levels were noted at 02:00 than at 07:00 (113).

Rev-erb alpha, which is the nuclear receptor 1D1, is a circadian rhythm regulator that controls inflammation and glucose and lipid metabolism. Administration of SR9009, a Rev-erb, in a mouse model of MI and HF improved the survival rates and reduced the left ventricular function.

TABLE 1 | Examples of the linkage between administration time and effect in cardiovascular drugs.

Drug[a]	Timing	Effect	References
ACEi	Evening vs. morning	Reduction in treatment-related cough	(100)
Antihypertensives	Bedtime vs. awakening	Improved blood pressure control and decrease in major cardiovascular events	(101)
Aspirin	Bedtime vs. awakening	Decrease in early morning platelet activity	(102)
Captopril (rats)	Sleep vs. wake time	Effect on cardiovascular remodeling was achieved only when administered during sleep	(103)
Furosemide (rats)	10 a.m. vs. 10 p.m.	Increased urinary volume and sodium exertion	(104)
Torsemide	Bedtime vs. awakening	Improved blood pressure control and 24-hour therapeutic duration	(105)
Valsartan	Bedtime vs. awakening	Greater reduction in proteinuria, delayed decline in GFR and reduced risk of MI	(106)

[a]*Administered to human subjects unless otherwise stated. ACEi, angiotensin converting enzyme inhibitors; GFR, glomerular filtration rate; MI, myocardial infarction.*

These effects were associated with reduced BNP levels and reduced expression of inflammatory-related molecules, such as matrix metallopeptidase 9, Ly6g, Cd11b, IL-6, Mcp1, and phosphorylated NF-kappaB p65, phosphorylated extracellular signal-regulated ligand, and phosphorylated p38. This treatment also reduced neutrophil and pro-inflammatory macrophage infiltration (114). SR9009 improved the long-term cardiac repair post-myocardial ischemia-reperfusion injury in animals. A single therapy decreased cardiac NLRP3 inflammasome, fibroblast, and immunocyte infiltration, in turn augmenting the MI healing process (115).

Taken together, these studies suggest that chronotherapy, including the targeting of ANS activation, exerts a beneficial effect on HF and diuretic resistance-related parameters.

INTER- AND INTRA-PATIENT VARIABILITY AFFECT THE RESPONSE TO MEDICATIONS

Variability is inherent to biological systems and has been proposed to be part of the normal function of cells and organs (116–119). Variability at the genome and cellular level (118, 120), HRV (121), respiratory variability (122, 123), and gate variability (124) are some examples. The loss or alteration in the physiologic variability is associated with disease states and poor prognoses (125–127). At the cellular level, the cell is packed with a non-uniform spreading of macromolecules that interact specifically and nonspecifically with a drug. This leads to a high degree of variability between individual cells in the response to drugs (128). Many of these variabilities do not follow specific patterns or rules and are characterized by marked inherent intra- and inter-patients unpredictability.

Variability in the heart function was the basis for the development of the heart-slowing medication ivabradine. The sinus node is the central cardiac pacemaker, and its function is associated with several ionic transporter circuits that can produce a rhythm. If one apparatus fails, another one can take over the task. Eliminating a transporter that could carry as much as 80% of the ionic current required for producing the rhythm alters the incidence by only around 10–15%, and this is due to a substitution mechanism. This supports the concept of a high degree of molecular-level variations as a protective mechanism (129, 130).

High degrees of both inter- and intra-patient variability are described for the response to many drugs. There are marked intra- and inter-patient variabilities in drug pharmacodynamics associated with the loss of an effective response to drugs (119, 131–136). For example, high tacrolimus intra-patient variability was associated with graft rejection (134), and intra-patient variability in many antiepileptic circulatory levels in stable patients was described with observed differences of tens of percentages in the serum levels (137).

Regular dosing regimens for diuretics are not compatible with physiological variability and may further increase drug resistance. Taking a constant dose of diuretics on a regular basis, or continuously increasing the dose, may increase the serum drug levels reaching a peak level and its subsequent gradual decrease. This process is observed to repeat on a daily basis. It is a monotonic cyclic pattern associated with adaptation to the effect and results in the partial or complete loss of the response to the diuretic (12–14). Part of the loss of the effect is associated with augmentation of the compensatory responses to diuretics as a result of a continous increase in doses, which further reduces their effect. For several drugs, a regular dosing regimen or continuous increase of a dose was proposed to contribute to drug resistance, compared with the irregular consumption of the same dose or with altering the daily dose (117, 138).

A model for heartbeat was developed built on a noise titration assay, which provides a time-resolved and quantitative degree of the chaos. It uses the HIF and LF parameters of HRV for determining whether they reflect stochastic or chaotic phenomenon (139). Noise titration of the running short-segment Holter tachograms from healthy people showed circadian sleep/wake-dependent heartbeat chaos that linked with the HF parameter (respiratory sinus arrhythmia). In contrast, in patients with HF, a near-regular heartbeat was non-chaotic and resulted in transient chaotic rhythms. HRV alterations in HF were accompanied by little changes in approximate entropy, a measure of signal irregularity, which reflects an autonomic, cardiac, respiratory, and circadian/sleep-wake changes (139).

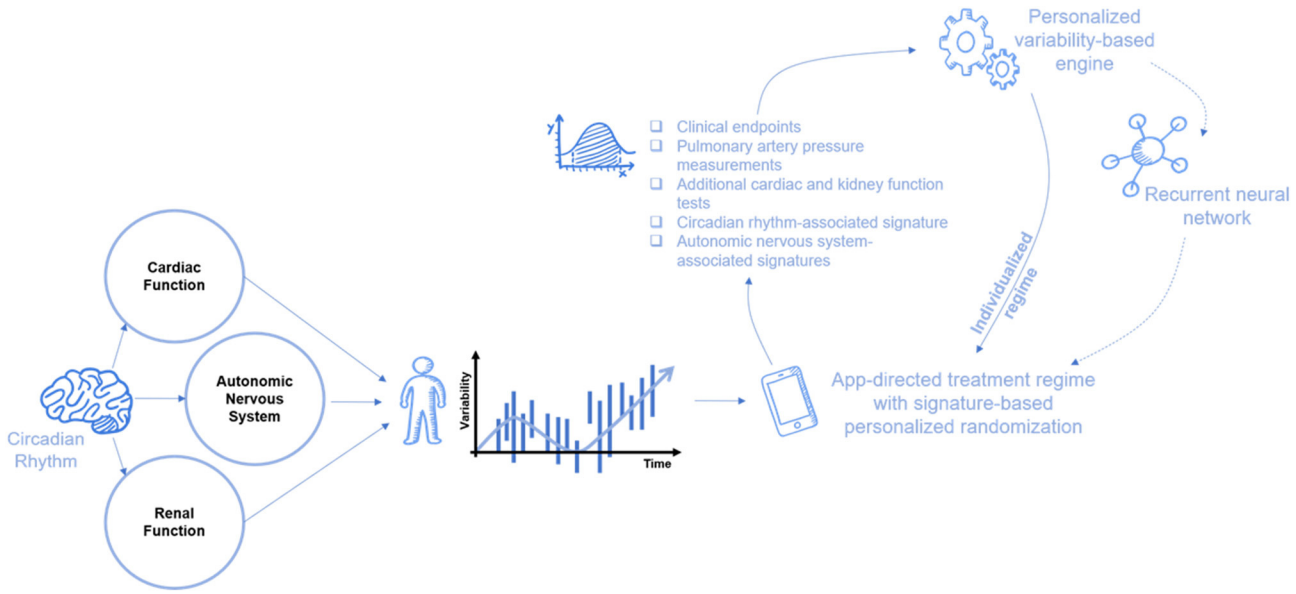


FIGURE 2 | A suggested scheme of the closed loop machine learning system for improving the response to diurectis in patients with HF. Inputs from various sensors are being incorporated in a dynamic manner for determining the timings of administration and dosages in a personalized approach. In a dynamic changing system with multiple alternating parameters, a continuously changing dosing regimen is generated based on the individualized-variability pattern that comprises data from the ANS (e.g., HRV), heart, and kidney. Cardiac, renal, and ANS, along with numerous other physiological and biological processes, exhibit intrinsic randomness. These variations in their activity are also contributed from the generalized circadian rhythm and from intrinsic biological rhythms. Randomized treatment administration within the preset limits augmented with personalized signature from the clinical status, echocardiographic, HRV, pulmonary artery pressure monitoring, biomarkers, and more can improve outcomes of the diuretic treatment. In addition, randomization might prevent renal adaptivity involved in the development of diuretic resistance. ANS, autonomic nervous system; HRV, heart rate variability.

A model using time irreversibility analysis was developed for the analysis of the short heart period sequences derived from 24 h Holter recordings. Irreversible dynamics over short time scales were noted, while the irreversibility of longer time scales was marginal. In healthy subjects, the percentage of irreversible dynamics was higher during the daytime than during the night-time, suggesting augmented non-linear dynamics during the daytime. In healthy subjects, the non-linear behavior reflected that bradycardic ones run shorter than the tachycardic ones during the daytime. In contrast, patients with HF demonstrated an increased percentage of irreversible series along with a reverse pattern, showing the tachycardic ones run shorter than the bradycardic ones (140).

ESTABLISHING A CHRONOTHERAPY AND VARIABILITY-BASED ALGORITHM FOR IMPROVING THE RESPONSE TO EFFECTIVENSS OF DIURETICS

Using treatment regimens based on aperiodic intervals and at irregular strengths has been suggested to improve response to diuretics. For maximum benefits, irregular, pulsed, or multiple intervals-based administration of a chronic drug at continually changing dosage strengths may improve the overall effect, thereby reducing the likelihood of drug resistance (117, 119, 141). Manipulating the conditions of living organs using rhythmic administration of altered feeding schedules or several drugs appears successful (99, 142).

Establishing a novel model for overcoming diuretic resistance was proposed based on a closed loop system, which comprises clinical, laboratory, and sensors-derived inputs, including PA pressure sensors. Alteration of rhythmicity may adversely affect homeostatic regulation and lead to deleterious effects. These changes in therapeutic regimens should take into account the inter- and intra-patient variability, alterations of the biological clock, and ANS responses.

This model has been developed at four levels. In the first step, random alterations in the dosing and timing of administration are introduced into the regimens. The implementation of irregular regimens is expected to improve the response to diuretics by lowering the harmful effects of compensatory mechanisms (117, 138, 143, 144). The physician is asked to register the patient and to enter into the systme the ranges for the dosages and for the times of administration. Patients receive daily reminders for taking the drug. The algorithm randomly alters the dosages and times of administration on a daily basis (143, 144).

In the second step, an increase in the efficiency of diuretics is achieved by administering drugs that are at times the best tolerated by synchronizing medication concentrations to rhythms in disease activity. Technologies for drug delivery precisely in a time-modulated mode by bedside or ambulatory pumps can improve both the safety and efficacy of the chronic use of diuretics (99).

In the third stage, machine learning is introduced to regulate the variability and chronotherapy-based regimens using closed loop systems that control the dosing and timing in an individualized manner. This enables the overcoming of intra- and inter-patient variability, which prevents the "single bullet for all" concept from being applicable to the majority of patients. This concept is being introduced for overcoming drug resistance in several chronic diseases (142, 143, 145–156). In the last step, signatures based on parameters relevant to the pathogenesis of HF, mechanisms of action, and the cardio renal axis are introduced in a personalized approach for tailoring the appropriate dosing in a continuous and consistent manner (143, 144) (**Figure 2**).

In summary, CHF remains a major clinical problem with an enormous morbidity and mortality burden. Diuretic resistance is a major obastacle for the effective long-term treatment of HF with considerable inter- and intra-patient variability, in addition to the implication of both the ANS and chronobiology in the pathogenesis and progression of HF. A personalized machine learing algorithm, which comprises continously changing parameters derived from clinical, laboratory, and sensors-derived inputs, including inputs from PA measurements, is suggested as an effective tool for improving the resposne to duretics. Ongoing trials will determine if these platforms are efficacious in reducing the adverse clinical outcomes and improving long-term prognoses.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Tran HA, Lin F, Greenberg BH. Potential new drug treatments for congestive heart failure. *Expert Opin Investig Drugs.* (2016) 25:811–26. doi: 10.1080/13543784.2016.1181749
2. Chioncel O, Lainscak M, Seferovic PM, Anker SD, Crespo-Leiro MG, Harjola VP, et al. Epidemiology and one-year outcomes in patients with chronic heart failure and preserved, mid-range and reduced ejection fraction: an analysis of the ESC heart failure long-term registry. *Eur J Heart Fail.* (2017) 19:1574–85. doi: 10.1002/ejhf.813
3. Zepeda P, Rain C, Sepulveda P. Continuous infusion or bolus injection of loop diuretics for congestive heart failure? *Medwave.* (2016) 16(Suppl. 2):e6426. doi: 10.5867/medwave.2016.6426
4. Graziani G, Pini D, Oldani S, Cucchiari D, Podesta MA, Badalamenti S. Renal dysfunction in acute congestive heart failure: a common problem for cardiologists and nephrologists. *Heart Fail Rev.* (2014) 19:699–708. doi: 10.1007/s10741-013-9416-5
5. Wilcox CS, Testani JM, Pitt B. Pathophysiology of diuretic resistance and its implications for the management of chronic heart failure. *Hypertension.* (2020) 76:1045–54. doi: 10.1161/HYPERTENSIONAHA.120.15205
6. Felker GM, Ellison DH, Mullens W, Cox ZL, Testani JM. Diuretic therapy for patients with heart failure: JACC state-of-the-art review. *J Am Coll Cardiol.* (2020) 75:1178–95. doi: 10.1016/j.jacc.2019.12.059
7. Ellison DH, Felker GM. Diuretic treatment in heart failure. *N Engl J Med.* (2017) 377:1964–75. doi: 10.1056/NEJMra1703100
8. Somasekharan S, Tanis J, Forbush B. Loop diuretic and ion-binding residues revealed by scanning mutagenesis of transmembrane helix 3 (TM3) of Na-K-Cl cotransporter (NKCC1). *J Biol Chem.* (2012) 287:17308–17. doi: 10.1074/jbc.M112.356014
9. Oppermann M, Hansen PB, Castrop H, Schnermann J. Vasodilatation of afferent arterioles and paradoxical increase of renal vascular resistance by furosemide in mice. *Am J Physiol Renal Physiol.* (2007) 293:F279–87. doi: 10.1152/ajprenal.00073.2007
10. Jardim SI, Ramos dos santos L, Araujo I, Marques F, Branco P, Gaspar A, Fonseca C. A 2018 overview of diuretic resistance in heart failure. *Rev Port Cardiol.* (2018) 37:935–45. doi: 10.1016/j.repc.2018.03.014
11. Hoorn EJ, Ellison DH. Diuretic resistance. *Am J Kidney Dis.* (2017) 69:136–42. doi: 10.1053/j.ajkd.2016.08.027
12. Robinette B. *Strategies to overcome diuretic resistance in congestive HF.* Cardiology Today (2015).
13. Ter Maaten JM, Rao VS, Hanberg JS, Perry Wilson F, Bellumkonda L, Assefa M, et al. Renal tubular resistance is the primary driver for loop diuretic resistance in acute heart failure. *Eur J Heart Fail.* (2017) 19:1014–22. doi: 10.1002/ejhf.757
14. Shah N, Madanieh R, Alkan M, Dogar MU, Kosmas CE, Vittorio TJ. A perspective on diuretic resistance in chronic congestive heart failure. *Ther Adv Cardiovasc Dis.* (2017) 11:271–8. doi: 10.1177/1753944717718717
15. Oh SW, Han SY. Loop diuretics in clinical practice. *Electrolyte Blood Press.* (2015) 13:17–21. doi: 10.5049/EBP.2015.13.1.17
16. Harrison-Bernard LM. The renal renin-angiotensin system. *Adv Physiol Educ.* (2009) 33:270–4. doi: 10.1152/advan.00049.2009
17. Seva Pessoa B, Van der Lubbe N, Verdonk K, Roks AJ, Hoorn EJ, Danser AH. Key developments in renin-angiotensin-aldosterone system inhibition. *Nat Rev Nephrol.* (2013) 9:26–36. doi: 10.1038/nrneph.2012.249
18. De Bruyne LK. Mechanisms and management of diuretic resistance in congestive heart failure. *Postgrad Med J.* (2003) 79:268–71. doi: 10.1136/pmj.79.931.268
19. Ellison DH. Diuretic resistance: physiology and therapeutics. *Semin Nephrol.* (1999) 19:581–97.
20. Kaissling B, Bachmann S, Kriz W. Structural adaptation of the distal convoluted tubule to prolonged furosemide treatment. *Am J Physiol.* (1985) 248:F374–81. doi: 10.1152/ajprenal.1985.248.3.F374
21. Abdallah JG, Schrier RW, Edelstein C, Jennings SD, Wyse B, Ellison DH. Loop diuretic infusion increases thiazide-sensitive na(+)/Cl(-)-cotransporter abundance: role of aldosterone. *J Am Soc Nephrol.* (2001) 12:1335–41. doi: 10.1681/ASN.V1271335
22. Ellison DH. The physiologic basis of diuretic synergism: its role in treating diuretic resistance. *Ann Intern Med.* (1991) 114:886–94. doi: 10.7326/0003-4819-114-10-886
23. Gordan R, Gwathmey JK, Xie LH. Autonomic and endocrine control of cardiovascular function. *World J Cardiol.* (2015) 7:204–14. doi: 10.4330/wjc.v7.i4.204
24. Bowman BN, Nawarskas JJ, Anderson JR. Treating diuretic resistance: an overview. *Cardiol Rev.* (2016) 24:256–60. doi: 10.1097/CRD.0000000000000116
25. Felker GM, O'Connor CM, Braunwald E, Heart failure clinical research network Investigators. Loop diuretics in acute decompensated heart failure: necessary? Evil? A necessary evil? *Circ Heart Fail.* (2009) 2:56–62. doi: 10.1161/CIRCHEARTFAILURE.108.821785
26. Brisco MA, Zile MR, Hanberg JS, Wilson FP, Parikh CR, Coca SG, et al. Relevance of changes in serum creatinine during a heart failure trial of decongestive strategies: insights from the DOSE trial. *J Card Fail.* (2016) 22:753–60. doi: 10.1016/j.cardfail.2016.06.423
27. Jentzer JC, DeWald TA, Hernandez AF. Combination of loop diuretics with thiazide-type diuretics in heart failure. *J Am Coll Cardiol.* (2010) 56:1527–34. doi: 10.1016/j.jacc.2010.06.034
28. Verbrugge FH. Editor's choice-diuretic resistance in acute heart failure. *Eur Heart J Acute Cardiovasc Care.* (2018) 7:379–89. doi: 10.1177/2048872618768488
29. Rosenberg J, Gustafsson F, Galatius S, Hildebrandt PR. Combination therapy with metolazone and loop diuretics in outpatients with refractory heart

failure: an observational study and review of the literature. *Cardiovasc Drugs Ther.* (2005) 19:301–6. doi: 10.1007/s10557-005-3350-2
30. Shi X, Bao J, Zhang H, Wang H, Li L, Zhang Y. Patients with high-dose diuretics should get ultrafiltration in the management of decompensated heart failure: a meta-analysis. *Heart Fail Rev.* (2019) 24:927–40. doi: 10.1007/s10741-019-09812-2
31. Schmier JK, Ong KL, Fonarow GC. Cost-effectiveness of remote cardiac monitoring with the cardioMEMS heart failure system. *Clin Cardiol.* (2017) 40:430–6. doi: 10.1002/clc.22696
32. Adamson PB, Abraham WT, Stevenson LW, Desai AS, Lindenfeld J, Bourge RC, et al. Pulmonary artery pressure-Guided heart failure management reduces 30-day readmissions. *Circ Heart Fail.* (2016) 9:e002600. doi: 10.1161/CIRCHEARTFAILURE.115.002600
33. Benza RL, Raina A, Abraham WT, Adamson PB, Lindenfeld J, Miller AB, et al. Pulmonary hypertension related to left heart disease: insight from a wireless implantable hemodynamic monitor. *J Heart Lung Transplant.* (2015) 34:329–7. doi: 10.1016/j.healun.2014.04.014
34. Costanzo MR, Stevenson LW, Adamson PB, Desai AS, Heywood JT, Bourge RC, et al. Interventions linked to decreased heart failure hospitalizations during ambulatory pulmonary artery pressure monitoring. *JACC Heart Fail.* (2016) 4:333–44. doi: 10.1016/j.jchf.2015.11.011
35. Morris CJ, Aeschbach D, Scheer FA. Circadian system, sleep and endocrinology. *Mol Cell Endocrinol.* (2012) 349:91–104. doi: 10.1016/j.mce.2011.09.003
36. Panda S. Circadian physiology of metabolism. *Science.* (2016) 354:1008–15. doi: 10.1126/science.aah4967
37. Honma S. The mammalian circadian system: a hierarchical multi-oscillator structure for generating circadian rhythm. *J Physiol Sci.* (2018) 68:207–19. doi: 10.1007/s12576-018-0597-5
38. Durgan DJ, Young ME. Linking the cardiomyocyte circadian clock to myocardial metabolism. *Cardiovasc Drugs Ther.* (2008) 22:115–24. doi: 10.1007/s10557-008-6086-y
39. Thosar SS, Butler MP, Shea SA. Role of the circadian system in cardiovascular disease. *J Clin Invest.* (2018) 128:2157–67. doi: 10.1172/JCI80590
40. Gu F, Han J, Laden F, Pan A, Caporaso NE, Stampfer MJ, et al. Total and cause-specific mortality of U.S. nurses working rotating night shifts. *Am J Prev Med.* (2015) 48:241–52. doi: 10.1016/j.amepre.2014.10.018
41. Karlsson B, Knutsson A, Lindahl B. Is there an association between shift work and having a metabolic syndrome? Results from a population based study of 27,485 people. *Occup Environ Med.* (2001) 58:747–52. doi: 10.1136/oem.58.11.747
42. Amir O, Alroy S, Schliamser JE, Asmir I, Shiran A, Flugelman MY, et al. Brachial artery endothelial function in residents and fellows working night shifts. *Am J Cardiol.* (2004) 93:947–9. doi: 10.1016/j.amjcard.2003.12.032
43. Durgan DJ, Trexler NA, Egbejimi O, McElfresh TA, Suk HY, Petterson LE, et al. The circadian clock within the cardiomyocyte is essential for responsiveness of the heart to fatty acids. *J Biol Chem.* (2006) 281:24254–69. doi: 10.1074/jbc.M601704200
44. Alibhai FJ, Tsimakouridze EV, Chinnappareddy N, Wright DC, Billia F, O'Sullivan ML, et al. Short-term disruption of diurnal rhythms after murine myocardial infarction adversely affects long-term myocardial structure and function. *Circ Res.* (2014) 114:1713–22. doi: 10.1161/CIRCRESAHA.114.302995
45. Blagonravov ML, Azova MM, Frolov VA. Chronobiology of cardiac ventricular fibrillation development in experimental acute coronary failure. *Bull Exp Biol Med.* (2010) 149:559–61. doi: 10.1007/s10517-010-0992-2
46. Hirsch L, Shechter A, Feinberg MS, Koren-Morag N, Shechter M. The impact of early compared to late morning hours on brachial endothelial function and long-term cardiovascular events in healthy subjects with no apparent coronary heart disease. *Int J Cardiol.* (2011) 151:342–7. doi: 10.1016/j.ijcard.2010.08.069
47. Maruo T, Nakatani S, Kanzaki H, Kakuchi H, Yamagishi M, Kitakaze M, et al. Circadian variation of endothelial function in idiopathic dilated cardiomyopathy. *Am J Cardiol.* (2006) 97:699–702. doi: 10.1016/j.amjcard.2005.09.118
48. Vyas MV, Garg AX, Iansavichus AV, Costella J, Donner A, Laugsand LE, et al. Shift work and vascular events: systematic review and meta-analysis. *BMJ.* (2012) 345:e4800. doi: 10.1136/bmj.e4800
49. Sole MJ, Martino TA. Diurnal physiology: core principles with application to the pathogenesis, diagnosis, prevention, and treatment of myocardial hypertrophy and failure. *J Appl Physiol 1985.* (2009) 107:1318–27. doi: 10.1152/japplphysiol.00426.2009
50. Hu K, Ivanov P, Hilton MF, Chen Z, Ayers RT, Stanley HE, et al. Endogenous circadian rhythm in an index of cardiac vulnerability independent of changes in behavior. *Proc Natl Acad Sci USA.* (2004) 101:18223–7. doi: 10.1073/pnas.0408243101
51. Teerlink JR, Clozel JP. Hemodynamic variability and circadian rhythm in rats with heart failure: role of locomotor activity. *Am J Physiol.* (1993) 264:H2111–8. doi: 10.1152/ajpheart.1993.264.6.H2111
52. Giles TD, Roffidal L, Quiroz A, Sander G, Tresznewsky O. Circadian variation in blood pressure and heart rate in nonhypertensive congestive heart failure. *J Cardiovasc Pharmacol.* (1996) 28:733–40. doi: 10.1097/00005344-199612000-00001
53. Kohsaka A, Das P, Hashimoto I, Nakao T, Deguchi Y, Gouraud SS, et al. The circadian clock maintains cardiac function by regulating mitochondrial metabolism in mice. *PLoS ONE.* (2014) 9:e112811. doi: 10.1371/journal.pone.0112811
54. Patel VN, Pierce BR, Bodapati RK, Brown DL, Ives DG, Stein PK. Association of holter-derived heart rate variability parameters with the development of congestive heart failure in the cardiovascular health study. *JACC Heart Fail.* (2017) 5:423–31. doi: 10.1016/j.jchf.2016.12.015
55. Sherazi S, Kutyifa V, McNitt S, Aktas MK, Couderc JP, Peterson B, et al. Prognostic significance of heart rate variability among patients treated with cardiac resynchronization therapy: MADIT-CRT (Multicenter automatic defibrillator implantation trial-cardiac resynchronization therapy). *JACC Clin Electrophysiol.* (2015) 1:74–80. doi: 10.1016/j.jacep.2015.03.004
56. Yoshino F, Sakuma N, Date T, Unoki T, Fukagawa K, Miyamoto T, et al. Diurnal change of plasma atrial natriuretic peptide concentrations in patients with congestive heart failure. *Am Heart J.* (1989) 117:1316–9. doi: 10.1016/0002-8703(89)90412-2
57. Caruana MP, Lahiri A, Cashman PM, Altman DG, Raftery EB. Effects of chronic congestive heart failure secondary to coronary artery disease on the circadian rhythm of blood pressure and heart rate. *Am J Cardiol.* (1988) 62:755–9. doi: 10.1016/0002-9149(88)91217-9
58. Hayano J, Yasuma F, Watanabe E, Carney RM, Stein PK, Blumenthal JA, et al. Blunted cyclic variation of heart rate predicts mortality risk in post-myocardial infarction, end-stage renal disease, and chronic heart failure patients. *Europace.* (2017) 19:1392–400. doi: 10.1093/europace/euw222
59. Lewis R, Hackfort BT, Schultz HD. Chronic heart failure abolishes circadian rhythms in resting and chemoreflex breathing. *Adv Exp Med Biol.* (2018) 1071:129–36. doi: 10.1007/978-3-319-91137-3_16
60. Sachan N, Dey A, Rotter D, Grinsfelder DB, Battiprolu PK, Sikder D, et al. Sustained hemodynamic stress disrupts normal circadian rhythms in calcineurin-dependent signaling and protein phosphorylation in the heart. *Circ Res.* (2011) 108:437–45. doi: 10.1161/CIRCRESAHA.110.235309
61. Jeyaraj D, Haldar SM, Wan X, McCauley MD, Ripperger JA, Hu K, et al. Circadian rhythms govern cardiac repolarization and arrhythmogenesis. *Nature.* (2012) 483:96–9. doi: 10.1038/nature10852
62. Kishi T. Heart failure as an autonomic nervous system dysfunction. *J Cardiol.* (2012) 59:117–22. doi: 10.1016/j.jjcc.2011.12.006
63. Sobotka PA, Mahfoud F, Schlaich MP, Hoppe UC, Bohm M, Krum H. Sympatho-renal axis in chronic disease. *Clin Res Cardiol.* (2011) 100:1049–57. doi: 10.1007/s00392-011-0335-y
64. Dillier R, Zuber M, Arand P, Erne S, Erne P. Assessment of systolic and diastolic function in heart failure using ambulatory monitoring with acoustic cardiography. *Ann Med.* (2011) 43:403–11. doi: 10.3109/07853890.2010.550309
65. Mochel JP, Fink M, Bon C, Peyrou M, Bieth B, Desevaux C, et al. Influence of feeding schedules on the chronobiology of renin activity, urinary electrolytes and blood pressure in dogs. *Chronobiol Int.* (2014) 31:715–30. doi: 10.3109/07420528.2014.897711
66. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. *Front Public Health.* (2017) 5:258. doi: 10.3389/fpubh.2017.00258

67. Lee HW, Han TH, Yi KJ, Choi MC, Lee SY, Ryu PD. Time course of diurnal rhythm disturbances in autonomic function of rats with myocardial infarction. *Auton Neurosci.* (2013) 179:28–36. doi: 10.1016/j.autneu.2013.06.007
68. Schrier RW, Gurevich AK, Cadnapaphornchai MA. Pathogenesis and management of sodium and water retention in cardiac failure and cirrhosis. *Semin Nephrol.* (2001) 21:157–72. doi: 10.1053/snep.2001.20933
69. De Ferrari GM. Vagal stimulation in heart failure. *J Cardiovasc Transl Res.* (2014) 7:310–20. doi: 10.1007/s12265-014-9540-1
70. Gold MR, Van Veldhuisen DJ, Hauptman PJ, Borggrefe M, Kubo SH, Lieberman RA, et al. Vagus nerve stimulation for the treatment of heart failure: the INOVATE-HF trial. *J Am College Cardiol.* (2016) 68:149–58. doi: 10.1016/j.jacc.2016.03.525
71. Piccirillo G, Moscucci F, D'Alessandro G, Pascucci M, Rossi P, Han S, et al. Myocardial repolarization dispersion and autonomic nerve activity in a canine experimental acute myocardial infarction model. *Heart Rhythm.* (2014) 11:110–8. doi: 10.1016/j.hrthm.2013.10.022
72. Witte K, Engelhardt S, Janssen BJ, Lohse M, Lemmer B. Circadian and short-term regulation of blood pressure and heart rate in transgenic mice with cardiac overexpression of the beta1-adrenoceptor. *Chronobiol Int.* (2004) 21:205–16. doi: 10.1081/CBI-120037801
73. Mousa TM, Schiller AM, Zucker IH. Disruption of cardiovascular circadian rhythms in mice post myocardial infarction: relationship with central angiotensin II receptor expression. *Physiol Rep.* (2014) 2: doi: 10.14814/phy2.12210
74. Ohori T, Hirai T, Joho S, Kameyama T, Nozawa T, Asanoi H, et al. Circadian changes in autonomic function in conscious rats with heart failure: effects of amiodarone on sympathetic surge. *Auton Neurosci.* (2011) 159:20–5. doi: 10.1016/j.autneu.2010.07.001
75. Witte K, Hu K, Swiatek J, Müssig C, Ertl G, Lemmer B. Experimental heart failure in rats: effects on cardiovascular circadian rhythms and on myocardial beta-adrenergic signaling. *Cardiovasc Res.* (2000) 47:350–58. doi: 10.1016/S0008-6363(00)00099-7
76. Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Front Psychol.* (2014) 5:1040. doi: 10.3389/fpsyg.2014.01040
77. Bonnemeier H, Richardt G, Potratz J, Wiegand UK, Brandes A, Kluge N, et al. Circadian profile of cardiac autonomic nervous modulation in healthy subjects: differing effects of aging and gender on heart rate variability. *J Cardiovasc Electrophysiol.* (2003) 14:791–9. doi: 10.1046/j.1540-8167.2003.03078.x
78. Boudreau P, Yeh WH, Dumont GA, Boivin DB. Circadian variation of heart rate variability across sleep stages. *Sleep.* (2013) 36:1919–28. doi: 10.5665/sleep.3230
79. Gerritsen J, Dekker JM, TenVoorde BJ, Kostense PJ, Heine RJ, Bouter LM, et al. Impaired autonomic function is associated with increased mortality, especially in subjects with diabetes, hypertension, or a history of cardiovascular disease: the Hoorn study. *Diabetes Care.* (2001) 24:1793–8. doi: 10.2337/diacare.24.10.1793
80. Nolan J, Batin PD, Andrews R, Lindsay SJ, Brooksby P, Mullen M, et al. Prospective study of heart rate variability and mortality in chronic heart failure: results of the United Kingdom heart failure evaluation and assessment of risk trial (UK-heart). *Circulation.* (1998) 98:1510–6. doi: 10.1161/01.CIR.98.15.1510
81. Henze M, Hart D, Samarel A, Barakat J, Eckert L, Scrogin K. Persistent alterations in heart rate variability, baroreflex sensitivity, and anxiety-like behaviors during development of heart failure in the rat. *Am J Physiol Heart Circ Physiol.* (2008) 295:H29–38. doi: 10.1152/ajpheart.01373.2007
82. Rydlewska A, Jankowska EA, Ponikowska B, Borodulin-Nadzieja L, Banasiak W, Ponikowski P. Changes in autonomic balance in patients with decompensated chronic heart failure. *Clin Auton Res.* (2011) 21:47–54. doi: 10.1007/s10286-010-0089-z
83. Omran H, Bitter T, Horstkotte D, Oldenburg O, Fox H. Characteristics and circadian distribution of cardiac arrhythmias in patients with heart failure and sleep-disordered breathing. *Clin Res Cardiol.* (2018) 107:965–74. doi: 10.1007/s00392-018-1269-4
84. Beckers F, Verheyden B, Couckuyt K, Aubert AE. Fractal dimension in health and heart failure. *Biomed Technik.* (2006) 51:194–7. doi: 10.1515/BMT.2006.035
85. Ueno LM, Drager LF, Rodrigues ACT, Rondon MUPB, Mathias W, Jr, Krieger EM, et al. Day-night pattern of autonomic nervous system modulation in patients with heart failure with and without sleep apnea. *Int J Cardiol.* (2011) 148:53–8. doi: 10.1016/j.ijcard.2009.10.023
86. Kim SG, Yum MK. Decreased RR interval complexity and loss of circadian rhythm in patients with congestive heart failure. *Jpn Circ J.* (2000) 64:39–45. doi: 10.1253/jcj.64.39
87. Watanabe E, Arakawa T, Uchiyama T, Tong M, Yasui K, Takeuchi H, et al. Prognostic significance of circadian variability of RR and GT intervals and GT dynamicity in patients with chronic heart failure. *Heart Rhythm.* (2007) 4:999–1005. doi: 10.1016/j.hrthm.2007.04.019
88. Jandackova VK, Scholes S, Britton A, Steptoe A. Are changes in heart rate variability in middle-Aged and older people normative or caused by pathological conditions? Findings from a large population-based longitudinal cohort study. *J Am Heart Assoc.* (2016) 5:e002365. doi: 10.1161/JAHA.115.002365
89. Shehab A, Elnour AA, Struthers AD. A randomised, controlled, double-blind, cross-over pilot study assessing the effects of spironolactone, losartan and their combination on heart rate variability and GT dispersion in patients with chronic heart failure. *Cardiovasc J Afr.* (2008) 19:292–6.
90. Aronson D, Burger AJ. Effect of beta-blockade on autonomic modulation of heart rate and neurohormonal profile in decompensated heart failure. *Ann Noninvas Electrocardiol.* (2001) 6:98–106. doi: 10.1111/j.1542-474X.2001.tb00093.x
91. Burger AJ, Kamalesh M. Effect of beta-adrenergic blocker therapy on the circadian rhythm of heart rate variability in patients with chronic stable angina pectoris. *Am J Cardiol.* (1999) 83:596–A8. doi: 10.1016/S0002-9149(98)00921-7
92. Compostella L, Russo N, Setzu T, Tursi V, Bottio T, Tarzia V, et al. Cardiac autonomic dysfunction in the early phase after left ventricular assist device implant: implications for surgery and follow-up. *Int J Artif Organs.* (2013) 36:410–8. doi: 10.5301/ijao.5000210
93. Suzuki K, Nishinaka T, Miyamoto T, Ichihara Y, Yamazaki K. Circadian variation of motor current observed in fixed rotation speed continuous-flow left ventricular assist device support. *J Artif Organs.* (2014) 17:157–61. doi: 10.1007/s10047-014-0762-4
94. Carson PA, O'Connor CM, Miller AB, Anderson S, Belkin R, Neuberg GW, et al. Circadian rhythm and sudden death in heart failure: results from prospective randomized amlodipine survival trial. *J Am Coll Cardiol.* (2000) 36:541–6. doi: 10.1016/S0735-1097(00)00728-2
95. Miyamoto S, Fujita M, Tambara K, Sekiguchi H, Eiho S, Hasegawa K, et al. Circadian variation of cardiac autonomic nervous activity is well preserved in patients with mild to moderate chronic heart failure: effect of patient position. *Int J Cardiol.* (2004) 93:247–52. doi: 10.1016/S0167-5273(03)00190-6
96. Dallmann R, Brown SA, Gachon F. Chronopharmacology: new insights and therapeutic implications. *Annu Rev Pharmacol Toxicol.* (2014) 54:339–61. doi: 10.1146/annurev-pharmtox-011613-135923
97. Buttgereit F, Smolen JS, Coogan AN, Cajochen C. Clocking in: chronobiology in rheumatoid arthritis. *Nat Rev Rheumatol.* (2015) 11:349–56. doi: 10.1038/nrrheum.2015.31
98. Smolensky MH, Peppas NA. Chronobiology, drug delivery, and chronotherapeutics. *Adv Drug Deliv Rev.* (2007) 59:828–51. doi: 10.1016/j.addr.2007.07.001
99. Ohdo S. Chronotherapeutic strategy: rhythm monitoring, manipulation and disruption. *Adv Drug Deliv Rev.* (2010) 62:859–75. doi: 10.1016/j.addr.2010.01.006
100. Ohmori M, Fujimura A. ACE inhibitors and chronotherapy. *Clin Exp Hypertens.* (2005) 27:179–85. doi: 10.1081/CEH-200048762
101. Hermida RC, Crespo JJ, Domínguez-Sardiña M, Otero A, Moyá A, Ríos MT, et al. Bedtime hypertension treatment improves cardiovascular risk reduction: the hygia chronotherapy trial. *Eur Heart J.* (2019) 41:4565–76. doi: 10.1093/eurheartj/ehz754
102. Buurma M, Van diemen JJK, Thijs A, Numans ME, Bonten TN. Circadian rhythm of cardiovascular disease: the potential of chronotherapy with aspirin. *Front Cardiovasc Med.* (2019) 6:84. doi: 10.3389/fcvm.2019.00084
103. Martino TA, Tata N, Simpson JA, Vanderlaan R, Dawood F, Kabir MG, et al. The primary benefits of angiotensin-converting enzyme inhibition on cardiac remodeling occur during sleep time in murine

pressure overload hypertrophy. *J Am Coll Cardiol.* (2011) 57:2020–8. doi: 10.1016/j.jacc.2010.11.022
104. Fujimura A, Ebihara A. Chronopharmacological study of furosemide in rats. *Life Sci.* (1986) 38:1215–20. doi: 10.1016/0024-3205(86)90176-1
105. Calvo C, Hermida RC, Ayala DE, Lopez JE, Rodriguez M, Chayan L, et al. [Chronotherapy with torasemide in hypertensive patients: increased efficacy and therapeutic coverage with bedtime administration]. *Med Clin.* (2006) 127:721–9. doi: 10.1157/13095521
106. Wang C, Zhang J, Liu X, Li CC, Ye ZC, Peng H, et al. Effect of valsartan with bedtime dosing on chronic kidney disease patients with nondipping blood pressure pattern. *J Clin Hypertens.* (2013) 15:48–54. doi: 10.1111/jch.12021
107. Tsimakouridze EV, Alibhai FJ, Martino TA. Therapeutic applications of circadian rhythms for the cardiovascular system. *Front Pharmacol.* (2015) 6:77. doi: 10.3389/fphar.2015.00077
108. Fujimura A, Sudoh T, Shiga T, Ohashi K, Ebihara A. Influence of clorgyline treatment on chronopharmacology of furosemide in rats. *Life Sci.* (1993) 52:819–24. doi: 10.1016/0024-3205(93)90080-M
109. Fujimura A, Ebihara A. Chronopharmacological study of furosemide in rats: (II). Influence of beta-adrenoceptor blockade. *Life Sci.* (1988) 42:1431–7. doi: 10.1016/0024-3205(88)90053-7
110. Fujimura A, Ohashi K, Ebihara A. Chronopharmacological study of furosemide; (IX). Influence of continuous norepinephrine infusion. *Life Sci.* (1992) 50:449–55. doi: 10.1016/0024-3205(92)90380-8
111. Fujimura A, Shiga T, Sudoh T, Ohashi K, Ebihara A. Influence of renal denervation on chronopharmacology of furosemide in rats. *Life Sci.* (1992) 51:1811–6. doi: 10.1016/0024-3205(92)90052-Q
112. Mochel JP, Danhof M. Chronobiology and pharmacologic modulation of the renin-angiotensin-aldosterone system in dogs: what have we learned? *Rev Physiol Biochem Pharmacol.* (2015) 169:43–69. doi: 10.1007/112_2015_27
113. Dzida G, Prystupa A, Lachowska-Kotowska P, Kadas T, Kamienski P, Kimak E, et al. Alteration in diurnal and nocturnal melatonin serum level in patients with chronic heart failure. *Ann Agric Environ Med.* (2013) 20:745–8.
114. Stujanna EN, Murakoshi N, Tajiri K, Xu D, Kimura T, Qin R, et al. Rev-erb agonist improves adverse cardiac remodeling and survival in myocardial infarction through an anti-inflammatory mechanism. *PLoS ONE.* (2017) 12:e0189330. doi: 10.1371/journal.pone.0189330
115. Reitz CJ, Alibhai FJ, Khatua TN, Rasouli M, Bridle BW, Burris TP, et al. SR9009 administered for one day after myocardial ischemia-reperfusion prevents heart failure in mice by targeting the cardiac inflammasome. *Commun Biol.* (2019) 2:353. doi: 10.1038/s42003-019-0595-z
116. Goldberger AL. Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. *Lancet.* (1996) 347:1312–4. doi: 10.1016/S0140-6736(96)90948-4
117. Ilan Y. Advanced tailored randomness: a novel approach for improving the efficacy of biological systems. *J Comput Biol.* (2019) 27:20–29. doi: 10.1089/cmb.2019.0231
118. Ilan Y. Randomness in microtubule dynamics: an error that requires correction or an inherent plasticity required for normal cellular function? *Cell Biol Int.* (2019) 43:739–48. doi: 10.1002/cbin.11157
119. Ilan Y. Generating randomness: making the most out of disordering a false order into a real one. *J Transl Med.* (2019) 17:49. doi: 10.1186/s12967-019-1798-2
120. Finn EH, Misteli T. Molecular basis and biological function of variability in spatial genome organization. *Science.* (2019) 365:eaaw9498. doi: 10.1126/science.aaw9498
121. Singh N, Moneghetti KJ, Christle JW, Hadley D, Plews D, Froelicher V. Heart rate variability: an old metric with new meaning in the era of using mHealth technologies for health and exercise training guidance. Part One: Physiology and Methods. *Arrhythm Electrophysiol Rev.* (2018) 7:193–8. doi: 10.15420/aer.2018.27.2
122. Shields RW, Jr. Heart rate variability with deep breathing as a clinical test of cardiovagal function. *Cleve Clin J Med.* (2009) 76(Suppl 2):S37–40. doi: 10.3949/ccjm.76.s2.08
123. Kox M, Pompe JC, Van der hoeven JG, Hoedemaekers CW, Pickkers P. Influence of different breathing patterns on heart rate variability indices and reproducibility during experimental endotoxaemia in human subjects. *Clin Sci.* (2011) 121:215–22. doi: 10.1042/CS20110027
124. Konig N, Singh NB, Baumann CR, Taylor WR. Can gait signatures provide quantitative measures for aiding clinical decision-making? A systematic meta-analysis of gait variability behavior in patients with Parkinson's disease. *Front Hum Neurosci.* (2016) 10:319. doi: 10.3389/fnhum.2016.00319
125. Nayyar S, Hasan MA, Roberts-Thomson KC, Sullivan T, Baumert M. Effect of loss of heart rate variability on T-wave heterogeneity and QT variability in heart failure patients: implications in ventricular arrhythmogenesis. *Cardiovasc Eng Technol.* (2017) 8:219–228. doi: 10.1007/s13239-017-0299-9
126. Avolio A. Heart rate variability and stroke: strange attractors with loss of complexity. *J Hypertens.* (2013) 31:1529–31. doi: 10.1097/HJH.0b013e328362ff5e
127. Moon Y, Sung J, An R, Hernandez ME, Sosnoff JJ. Gait variability in people with neurological disorders: a systematic review and meta-analysis. *Hum Mov Sci.* (2016) 47:197–208. doi: 10.1016/j.humov.2016.03.010
128. Elgart V, Lin JR, Loscalzo J. Determinants of drug-target interactions at the single cell level. *PLoS Comput Biol.* (2018) 14:e1006601. doi: 10.1371/journal.pcbi.1006601
129. Noble D. Evolution viewed from physics, physiology and medicine. *Interface Focus.* (2017) 7:20160159. doi: 10.1098/rsfs.2016.0159
130. DiFrancesco D, Camm JA. Heart rate lowering by specific and selective i(f) current inhibition with ivabradine: a new therapeutic perspective in cardiovascular disease. *Drugs.* (2004) 64:1757–65. doi: 10.2165/00003495-200464160-00003
131. Leino AD, King EC, Jiang W, Vinks AA, Klawitter J, Christians U, et al. Assessment of tacrolimus intrapatient variability in stable adherent transplant recipients: establishing baseline values. *Am J Transplant.* (2018) 19:1410–20. doi: 10.1111/ajt.15199
132. Gueta I, Markovits N, Yarden-Bilavsky H, Raichlin E, Freimark D, Lavee J, et al. High tacrolimus trough level variability is associated with rejections after heart transplant. *Am J Transplant.* (2018) 18:2571–78. doi: 10.1111/ajt.15016
133. Gueta I, Markovits N, Yarden-Bilavsky H, Raichlin E, Freimark D, Lavee J, et al. Intrapatient variability in tacrolimus trough levels after solid organ transplantation varies at different postoperative time periods. *Am J Transplant.* (2018) 19:611. doi: 10.1111/ajt.15134
134. Del Bello, Congy-Jolivet N, Danjoux M, Muscari F, Lavayssiere L, Esposito L, et al. High tacrolimus intra-patient variability is associated with graft rejection, and *de novo* donor-specific antibodies occurrence after liver transplantation. *World J Gastroenterol.* (2018) 24:1795–802. doi: 10.3748/wjg.v24.i16.1795
135. Ilan Y. Why targeting the microbiome is not so successful: can randomness overcome the adaptation that occurs following gut manipulation? *Clin Exp Gastroenterol.* (2019) 12:209–17. doi: 10.2147/CEG.S203823
136. Ilan Y. Beta-glycosphingolipids as mediators of both inflammation and immune tolerance: a manifestation of randomness in biological systems. *Front Immunol.* (2019) 10:1143. doi: 10.3389/fimmu.2019.01143
137. Contin M, Alberghini L, Candela C, Benini G, Riva R. Intrapatient variation in antiepileptic drug plasma concentration after generic substitution vs stable brand-name drug regimens. *Epilepsy Res.* (2016) 122:79–83. doi: 10.1016/j.eplepsyres.2016.02.012
138. Strik AB, Mould S, Mathôt D, Ponsioen R, van den Brande C, Jansen J, et al. Dashboard driven vs. conventional dosing of infliximab in inflammatory bowel disease patients: the PRECISION trial. *J Crohns Colitis.* (2019) 13:S063. doi: 10.1093/ecco-jcc/jjy222.090
139. Wu GQ, Arzeno NM, Shen LL, Tang DK, Zheng DA, Zhao NQ, et al. Chaotic signatures of heart rate variability and its power spectrum in health, aging and heart failure. *PLoS ONE.* (2009) 4:e4323. doi: 10.1371/journal.pone.0004323
140. Porta A, D'Addio G, Corbi G, Maestri R, Pinna GD. Circadian variations of short-term heart period irreversibility in healthy and chronic heart failure patients. *Conf Proc IEEE Eng Med Biol Soc.* (2008) 2008:2116–9. doi: 10.1109/IEMBS.2008.4649611
141. Kyriazis M. Practical applications of chaos theory to the modulation of human ageing: nature prefers chaos to regularity. *Biogerontology.* (2003) 4:75–90. doi: 10.1023/A:1023306419861
142. Ilan Y. Overcoming compensatory mechanisms toward chronic drug administration to ensure long-term, sustainable beneficial effects. *Mol Ther Methods Clin Dev.* (2020) 18:335–44. doi: 10.1016/j.omtm.2020.06.006

143. Ilan Y. Improving global healthcare and reducing costs using second-generation artificial intelligence-based digital pills: a market disruptor. *Int J Environ Res Public Health*. (2021) 18:811. doi: 10.3390/ijerph18020811
144. Ilan Y. Second-generation digital health platforms: placing the patient at the center and focusing on clinically meaningful endpoints title: second-generation artificial intelligence algorithms. *Front Digit Health.* (2020) 2:569178. doi: 10.3389/fdgth.2020.569178
145. Kenig A, Ilan Y. A personalized signature and chronotherapy-Based platform for improving the efficacy of sepsis treatment. *Front Physiol.* (2019) 10:1542. doi: 10.3389/fphys.2019.01542
146. Khoury T, Ilan Y. Introducing patterns of variability for overcoming compensatory adaptation of the immune system to immunomodulatory agents: a Novel method for improving clinical response to anti-TNF therapies. *Front Immunol.* (2019) 10:2726. doi: 10.3389/fimmu.2019.02726
147. El-Haj M, Kanovitch D, Ilan Y. Personalized inherent randomness of the immune system is manifested by an individualized response to immune triggers and immunomodulatory therapies: a novel platform for designing personalized immunotherapies. *Immunol Res.* (2019) 67:337–47. doi: 10.1007/s12026-019-09101-y
148. Ilan Y. Advanced tailored randomness: a novel approach for improving the efficacy of biological systems. *J Comput Biol*. (2020) 27:20–9.
149. Khoury T, Ilan Y. Platform introducing individually tailored variability in nerve stimulations and dietary regimen to prevent weight regain following weight loss in patients with obesity. *Obes Res Clin Pract.* (2021) 15:114–23. doi: 10.1016/j.orcp.2021.02.003
150. Kolben Y, Weksler-Zangen S, Ilan Y. Adropin as a potential mediator of the metabolic system-autonomic nervous system-chronobiology axis: implementing a personalized signature-based platform for chronotherapy. *Obes Rev.* (2021) 22:e13108. doi: 10.1111/obr.13108
151. Ilan Y, Spigelman Z. Establishing patient-tailored variability-based paradigms for anti-cancer therapy: using the inherent trajectories which underlie cancer for overcoming drug resistance. *Cancer Treat Res Commun.* (2020) 25:100240. doi: 10.1016/j.ctarc.2020.100240
152. Kessler A, Weksler-Zangen S, Ilan Y. Role of the immune system and the circadian rhythm in the pathogenesis of chrnic pancreatitis: establishing a personalized signature for improving the effect of immunotherapies for chronic pancreatitis. *Pancreas.* (2020) 49:1024–32. doi: 10.1097/MPA.0000000000001626
153. Potruch A, Khoury ST, Ilan Y. The role of chronobiology in drug-resistance epilepsy: the potential use of a variability and chronotherapy-based individualized platform for improving the response to anti-seizure drugs. *Seizure.* (2020) 80:201–11. doi: 10.1016/j.seizure.2020.06.032
154. Forkosh E, Kenig A, Ilan Y. Introducing variability in targeting the microtubules: review of current mechanisms and future directions in colchicine therapy. *Pharmacol Res Perspect.* (2020) 8:e00616. doi: 10.1002/prp2.616
155. Gelman R, Bayatra A, Kessler A, Schwartz A, Ilan Y. Targeting SARS-CoV-2 receptors as a means for reducing infectivity and improving antiviral and immune response: an algorithm-based method for overcoming resistance to antiviral agents. *Emerg Microbes Infect.* (2020) 9:1397–406. doi: 10.1080/22221751.2020.1776161
156. Ilan Y. Order through disorder: the characteristic variability of systems. *Front Cell Dev Biol.* (2020) 8:186. doi: 10.3389/fcell.2020.00186

Conflict of Interest: YI was the founder of Oberon Sciences.

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 13 September 2021
doi: 10.3389/fcvm.2021.726943

Modified GAN Augmentation Algorithms for the MRI-Classification of Myocardial Scar Tissue in Ischemic Cardiomyopathy

Umesh C. Sharma[1,2*†], *Kanhao Zhao*[3†], *Kyle Mentkowski*[1,3], *Swati D. Sonkawade*[1], *Badri Karthikeyan*[1], *Jennifer K. Lang*[1,2,3,4] *and Leslie Ying*[3]

[1] Department of Medicine, Division of Cardiology, Jacobs School of Medicine and Biomedical Sciences, Buffalo, NY, United States, [2] Department of Pharmacology and Toxicology, University at Buffalo, Buffalo, NY, United States, [3] Department of Biomedical Engineering, University at Buffalo, Buffalo, NY, United States, [4] Veterans Affairs Western New York Healthcare System, Buffalo, NY, United States

Contrast-enhanced cardiac magnetic resonance imaging (MRI) is routinely used to determine myocardial scar burden and make therapeutic decisions for coronary revascularization. Currently, there are no optimized deep-learning algorithms for the automated classification of scarred vs. normal myocardium. We report a modified Generative Adversarial Network (GAN) augmentation method to improve the binary classification of myocardial scar using both pre-clinical and clinical approaches. For the initial training of the MobileNetV2 platform, we used the images generated from a high-field (9.4T) cardiac MRI of a mouse model of acute myocardial infarction (MI). Once the system showed 100% accuracy for the classification of acute MI in mice, we tested the translational significance of this approach in 91 patients with an ischemic myocardial scar, and 31 control subjects without evidence of myocardial scarring. To obtain a comparable augmentation dataset, we rotated scar images 8-times and control images 72-times, generating a total of 6,684 scar images and 7,451 control images. In humans, the use of Progressive Growing GAN (PGGAN)-based augmentation showed 93% classification accuracy, which is far superior to conventional automated modules. The use of other attention modules in our CNN further improved the classification accuracy by up to 5%. These data are of high translational significance and warrant larger multicenter studies in the future to validate the clinical implications.

Keywords: cardiac MRI, generative adversarial networks, data augmentation, myocardial scarring, deep learning

OPEN ACCESS

Edited by:
Paul Leeson,
University of Oxford, United Kingdom

Reviewed by:
Vivek Muthurangu,
University College London, United Kingdom
Ming-Jie Wang,
Fudan University, China

****Correspondence:***
Umesh C. Sharma
sharmau@buffalo.edu

† These authors have contributed equally to this work

Specialty section:
This article was submitted to General Cardiovascular Medicine, a section of the journal Frontiers in Cardiovascular Medicine

Received: *17 June 2021*
Accepted: *18 August 2021*
Published: *13 September 2021*

Citation:
Sharma UC, Zhao K, Mentkowski K, Sonkawade SD, Karthikeyan B, Lang JK and Ying L (2021) Modified GAN Augmentation Algorithms for the MRI-Classification of Myocardial Scar Tissue in Ischemic Cardiomyopathy. Front. Cardiovasc. Med. 8:726943. doi: 10.3389/fcvm.2021.726943

INTRODUCTION

Acute myocardial infarction (MI), commonly known as a heart attack, is an unpredictable complication of coronary artery disease (CAD). The location, size, density and heterogeneity of myocardial scarring provides both diagnostic and prognostic information for patient management. Such information is critical to manage patients at risk for heart failure (HF) and lethal cardiac arrhythmias (1, 2). HF and cardiac arrhythmias usually result from diseased myocardium and electrically unstable scars (3–5). Therefore, myocardial scar classification using emerging data augmentation methods is of great clinical significance.

Deep learning and artificial intelligence are rapidly gaining importance in the field of medical imaging. The development of newer generation cardiac MRI scanners with a higher signal-to-noise ratio and better edge definition has enabled us to better characterize myocardial scar tissue. However, we still lack smart decision-making tools to accurately classify the "*scar tissue*" in an objective and reproducible manner. Fortunately, there are major enhancements in the Graphics Processing Unit (GPU) development that enable us to train a large dataset in a relatively short time span.

A promising approach to improve the accuracy and consistency of myocardial scar detection lies in artificial intelligence – the use of a machine to perform deep cognitive analysis based on data input. The availability of such platforms has the potential to improve clinical workflow, enhance diagnostic accuracy, and offer options for early interventions. Furthermore, integrating preclinical and clinical analytical algorithms will allow us to directly examine the clinical implication of scar tissue after acute MI in humans, which has the potential to have an immediate impact on patient management.

Since its inception in 2012, the ImageNet classification platform with convolutional neural networks (CNNs) has been developed as the most efficient data analytical platform (6). Well-designed CNN models, such as VGG, Inception v3, and Resnet 50 have shown exceptional performance for image classification, and now deep learning applications are frequently used in medical image analysis (7–9). To enhance accuracy, multiple applications including feature detection, segmentation, classification, and image reconstruction are being integrated into various data platforms (10, 11).

In this translational study, we generated the first proof-of-concept data in a pre-clinical model of acute MI induced by ligating the left anterior descending coronary artery. The mice then underwent contrast-induced cardiac MRI and confirmatory histology analysis of the infarct and remote regions. After preclinical testing, we studied over 7,000 augmented images generated from patients with a history of CAD and known myocardial scar development.

METHODS

Mouse Model of Acute MI

All preclinical procedures and protocols conformed to institutional guidelines for the care and use of animals in research and were reviewed and approved by the University at Buffalo Institutional Animal Care and Use Committee (IACUC). Acute MI was induced in mice (age 14–15 weeks, C57B1/6 background) by using our study protocol described previously (12–14). Mice underwent permanent ligation of the left anterior descending (LAD) coronary artery producing an infarct in the anterior/anteroseptal walls of the LV. Concisely, mice were anesthetized with ketamine (1 mg/kg intramuscular) and xylazine (5 mg/kg subcutaneous) and were intubated to undergo a ligation procedure (9-0 nylon) of the LAD. Our laboratory performs AMI experiments on a routine basis with 70–80% post-MI survival. We studied 12 survivor mice (6 MI and 6 controls) for 2 weeks. On day 14, mice underwent cardiac MRI with gadolinium contrast infusion. To minimize pain and distress, all studies were performed on anesthetized (1.5% of isoflurane) animals. Upon completion of cardiac MRI, mice were sacrificed using the CO2 euthanasia protocol approved by the IACUC. The euthanasia procedure conformed to the guidelines from the Panel on Euthanasia of the American Veterinary Medical Association.

Myocardial Histology

Myocardial histology was performed to provide a gold-standard (tissue) validation of myocardial infarction. Since Human subjects are not required to undergo cardiac biopsy for tissue validation of myocardial scar, pre-clinical studies were performed for the conclusive evidence of myocardial infarction, along with cross-validation with cardiac MRI in mice.

To visualize the extent of MI in mice, an Evans Blue/tetrazolium chloride (TTC) method was employed. For the TTC assays, 0.5 mL of a 2% Evans blue solution (Sigma) devoid of bubbles was slowly perfused, turning the heart blue except for the risk regions. The heart was then removed, rinsed with KCl and PBS, and chilled at -20°C for 5 min prior to sectioning the LV into 7–8 transverse rings of 1 mm thickness using a heart slicer matrix (Zivic Instruments). Sections were subsequently incubated in a 1% TTC solution (Sigma) in a 37°C incubator for 15 min until a red stain developed to assess infarct size. The sections were placed between two clamped pieces of plexiglass with a 2 mm spacer and digital images were taken of both sides of each slice as described previously by our group (15). The total area and left ventricular (LV) area were calculated using Fiji. Using a color thresholding technique, we classified the blue regions of the heart as viable and the bright red as the risk regions.

Hematoxylin and eosin (H&E)-stained myocardial tissue sections were used to examine the infarct zone using the whole heart tissue, covering both ventricles and interventricular septum. Whole heart images were obtained from Leica Aperio VERSA whole slide imaging System at 63× magnifications (Multispectral Imaging suite, University at Buffalo). The extent of total myocardial fibrosis was visualized by trichrome staining (Thermo Scientific™ Richard-Allan Scientific™ Masson Trichrome Kit/22110648). The total myocardial area and the area of positive staining for fibrosis were quantified using color deconvolution algorithms as described previously (16).

Preclinical Cardiac MRI

Based on our study protocol explained previously (14), we used a 20 cm diameter horizontal-bore 9.4 Tesla magnet (Biospec 94/20 USR, Bruker Biospin) equipped with a gradient coil supporting 440 mT/m gradient strength and 3,440 T/m/s maximum linear slew rate (BGA-12S HP; Bruker Biospin). A series of three orthogonal gradient echo (GRE) scout images of the heart were acquired. For the tagged images, we acquired ECG and respiration-gated axial views of the heart in cine mode with 2D SPAMM tagging (0.1 mm thickness; 0.5 mm grid distance) using a single-slice fast low-angle shot (FLASH) sequence with the following parameters: 2 ms Gaussian pulse for slice selection; 30 flip angle; TE/TR = 2.2/15 ms; 50 kHz readout bandwidth; fat

suppression; 1 mm slice thickness; 30 × 30 mm^2 field of view; 256 × 256 matrix; 8 averages; 8 cardiac movie frames. The scan time was between 10 and 15 min, depending on respiration and heart rate.

For late gadolinium enhancement (LGE), we discharged the syringe and acquired late contrast enhancement data at 20 min after the injection using an ECG and respiration-gated inversion-recovery T_1-weighted FLASH sequence with the following parameters: 60° flip angle; TE/TR = 2.1/1200 ms; 65 kHz readout bandwidth; TI = 200 ms; 8 axial slices with 0.8 mm thickness and 0.2 mm gap; 25 × 25 mm^2 field of view; 256 × 256 matrix; no averages. The scan time was between 3 and 5 min, depending on respiration and heart rate. Cardiac MRI images were taken both before and after the MI induction procedure (median time = 2 weeks), which were denoted as pre-MI and post-MI images, respectively, as described previously by our group (14). We selected 272 images from the MI group and 383 images from the normal control mice. After an initial quality review, 392 images were chosen as the training dataset.

Contrast-Enhanced Cardiac MRI Protocol in Patients

Human experimental protocols were approved by an institutional review board (IRB) committee from University at Buffalo and all methods involving human/human data were performed in accordance with the relevant guidelines and regulations. Since the MRI database was accessed retrospectively and no direct patient contact was involved, the informed consent was waived by the IRB committee. Patient identifiers were securely processed using our existing Health Insurance Portability and Accountability Act (HIPAA) guidelines. A GE 1.5-T scanner with technical parameters recommended by the manufacturer was used. LGE sequence was obtained after intravenous (IV) gadolinium injection with an inversion recovery prepared T1 gradient echo and manually adapted inversion time. Typically, the images taken after 7–10 min after gadolinium injection were used for the current data analysis algorithms. Further details on cardiac MRI protocols were reported previously (17).

Clinical MRI Dataset

We first obtained 1,447 images from 91 patients with a history of coronary artery disease. Most of these patients were referred for contrast-enhanced cardiac MRI after visualization of coronary artery disease on the invasive coronary angiogram (44% had a prior history of stent placement and 13% had previously undergone surgical revascularization). Patients underwent a comprehensive MRI protocol including gadolinium contrast injection. The presence of abnormal LGE signal after optimal inversion recovery in the contrast-enhanced MRI was considered as the presence of myocardial scar. Additionally, 313 MR images from 31 control subjects were used for comparison. The controls included age-matched subjects with identical myocardial function, but no evidence of myocardial scar. Only 660 images with abnormal LGE, and 207 control images passed the initial image quality review. The DCM image format was then transformed into JPG format using MicroDicom viewer before further processing of the datasets. Representative examples are shown in **Supplementary Figure 1**. The model was first tested with 206 images (103 randomly split from each category) to precisely evaluate the model classification accuracy. Next, 557 images from patients with ischemic scars and 104 control images were used for training. Since there were discrepancies between the number of MI and control images, we augmented the control data size to match the myocardial scar data size. To ensure a balanced comparative dataset, we adjusted the training sample size from both classes as demonstrated in data **Supplementary Figure 2**.

MobileNetV2

MobileNet/MobileNetV1 was first proposed by a Google researcher team in 2017 (18). In this modified CNN model, a convolutional layer is replaced by a depthwise-separable convolution layer to reduce the parameters and speed-up the training process. The main refinement of MobileNetV2 is to improve the precision by the introduction of inverted residual blocks. The basic idea of residual blocks is derived from Resnet (9, 19). The inverted residual in MobileNetV2 reverses the residual block sequence in Resnet. Considering the accuracy and training speed, we used MobileNetV2 as the fundamental model in our experiment. To improve this model performance, *Finetune* (initialization by a pre-trained classification network, and then training for a different task) was used. Since pretrained parameters provide an excellent initiation point, *Finetuning* is widely employed in medical image analysis for a faster convergence of the model (10, 20, 21). The comparison of random initialization and pretrained parameter initialization is shown in **Supplementary Figure 3**.

Attentional Units

To further enhance the MobileNetV2 classification accuracy, squeeze-and-excitation block (22) is considered as a channel attention (CA) unit to embed into MobileNetV2. The first layer in the squeeze-and-excitation block is a convolutional step. The remaining structure is similar to the residual block. First, a global average pooling is used to obtain individual channel information U. The key formulation is defined as $S = \sigma(g(z, W)) = \sigma(W_2\delta(W_1 z))$, where σ is a sigmoid activation and δ refers to a ReLU activation. These two activation layers learn non-linear interactions between the channels and generate a mask (S) of multiple channels to emphasize features. The output is the channel-wise multiplication from mask S and feature map U. Spatial attention (SA) is derived from convolution block attention module (23). The difference is that we only use the average pooling layer followed by a convolutional layer and a sigmoid layer, and mix attention (MA) adds these two attentions together to shift the parameters. Because of the highest performance of CA, in a later experiment, we only focused on the performance of MobileNetV2 with or without CA.

Traditional Data Augmentation

A single flip shift scale is used for the images in the training set. To ensure comparable dataset size, we rotated scar images 8 times and control images 72 times. Finally, 6,684 MI augmentation images and 7,451 control augmentation images were generated. These augmented images formed an image pool, which was

sampled randomly in different TA scales. In these experiments, we used 4x augmentation. We first sampled 453 control images to match the MI image size of 557. Upon sampling randomly from the image pool, we generated 2,228 MI and 2,228 control images as the training dataset.

GAN Augmentation

PGGAN, a generative adversarial network (GAN) variant, has greater power in generating 1,024 × 1,024 high-resolution image platform from a random noise (24). Different from the traditional GAN training process, PGGAN trains with progressive growth resolution. The training starts with a low spatial resolution of 4 × 4 pixels. As the training advances, the generator and discriminator layers increase to match the spatial image resolutions, and hyperparameter α continues to update for a smooth transition to a higher resolution. In our experiments, we generated 256 × 256 scar and control MR images. The overview of PGGAN architecture is shown in **Supplementary Figure 4**. To quantitatively analyze the shift effect to PGGAN, ImageNet pretrained MobileNetV2 is used again to calculate Frechet inception distance (FID) in equation 1:

$$FID = \parallel \mu_r - \mu_g \parallel^2 + Tr(\sum_r + \sum_g - 2(\sum_r * \sum_g)^{\frac{1}{2}}) \quad (1)$$

where μ_r, μ_g are feature mean of real images and generated images, and $\sum_r$, $\sum_g$, are feature covariance matrix of real images and generated images. Low FID means a low distance from real image distribution to generated image distribution.

Different from PGGAN, CycleGAN generates a fake image from the real image. In CycleGAN, two-cycle consistency losses are introduced to enforce the generated image (G) and reconstructed image (F) from G to be consistent with each other (25). Since CycleGAN uses actual images as an input, we expect the stable generated images to have better classification accuracy with CycleGAN. Because GAN needs a large dataset for training, TA image pool is used to train a GAN. For better GAN training results, we generate images from each category separately.

Filtration of the Unusual-Looking Images

Principal component analysis (PCA) is a typical technique to reduce data dimensions and visualize data structure. K-means is a classical unsupervised algorithm to cluster data by calculating individual sample distance. We used MobileNetV2 pretraining by ImageNet to extract 1000-dimension features of each image. Next, we used PCA to map the 1000-dimension embedding features into one-dimensional latent space. We assumed that nearly 30% of the values, one standard deviation σ away from the mean, could be considered as unusual images (outliers), which were filtered out from the algorithm.

Gradient-Weighted Class Activation Mapping

Because the last convolutional layers contain detailed spatial information, Grad-CAM uses the gradient information flowing into the last convolutional layer of the CNN to look for semantic class-specific information in the images (26). We generated a heat map of all the test images by this method to visualize the region of interest for CNN and implement the quantitative analysis.

Results

Cardiac MRI Showed Anterior Wall Thinning and Scarring in Mice With Coronary Artery Ligation

Acute MI led to regional changes in myocardial morphology as demonstrated by histology and cardiac MRI. Evans Blue/TTC method was employed to assess infarct size as shown in **Figure 1A**. The severity of myocardial injury following LAD ligation was calculated as the ratio of the area at risk (AAR) to left ventricular area (LV) (AAR/LV: 32.18% ± 13.32, $n = 6$), assessed across the heart at 1 mm intervals from base to apex. Representative images illustrating the TTC-based histological confirmation of myocardial infarction, and an abnormal gadolinium enhancement on contrast-enhanced cardiac MRI are shown in **Figure 1**. Most of the anterior wall contained the myocardial scar. The remaining regions that had no scar were considered viable remote regions.

The Principal Component Analysis Showed a Gaussian Distribution of the Data Dimensions

We used the PCA method to reduce the dimensionality of large data sets by transforming data variables into a smaller one, but preserving most of the information of the larger set. The PCA analysis of feature distribution is shown in **Figure 2**. **Figures 2A,B** show the normal distribution of the features generated from PGGAN. Therefore, we used k-means of one to calculate the Euclidean metric between each data point with a mean. **Figures 2C,D** demonstrate that the data features are subjected to binary-variate Gaussian distribution. This algorithm enabled us easier data exploration, and thus making data analysis much easier and faster for k-means without extraneous variables to process.

MobileNetV2 With Heatmap Generation Had 100% Accuracy to Classify Acute MI in Mice

After multiple epochs in the training dataset with a stabilized training model, we generated the learning curves to determine the suitability fit to the training dataset. The MobileNetV2 classification accuracy curve demonstrated 100% accuracy in mice as illustrated in **Figure 3**. Because the accuracy was so high, further data augmentation algorithm was not applied. The heatmap of MobileNetV2 was focused on the whole thoracic cavity, which also includes the ventricles with acute MI.

Compared to the Spatial or Mixed Attention Modules, the Channel Attention Module Had the Highest Accuracy

Different attention modules were tested based on 4x traditional augmentation. Compared to spatial or mixed attention training modules, CA showed the highest accuracy with the most stable system, as illustrated in **Figure 4**. The combined module might have been less efficient than the original CA, since the

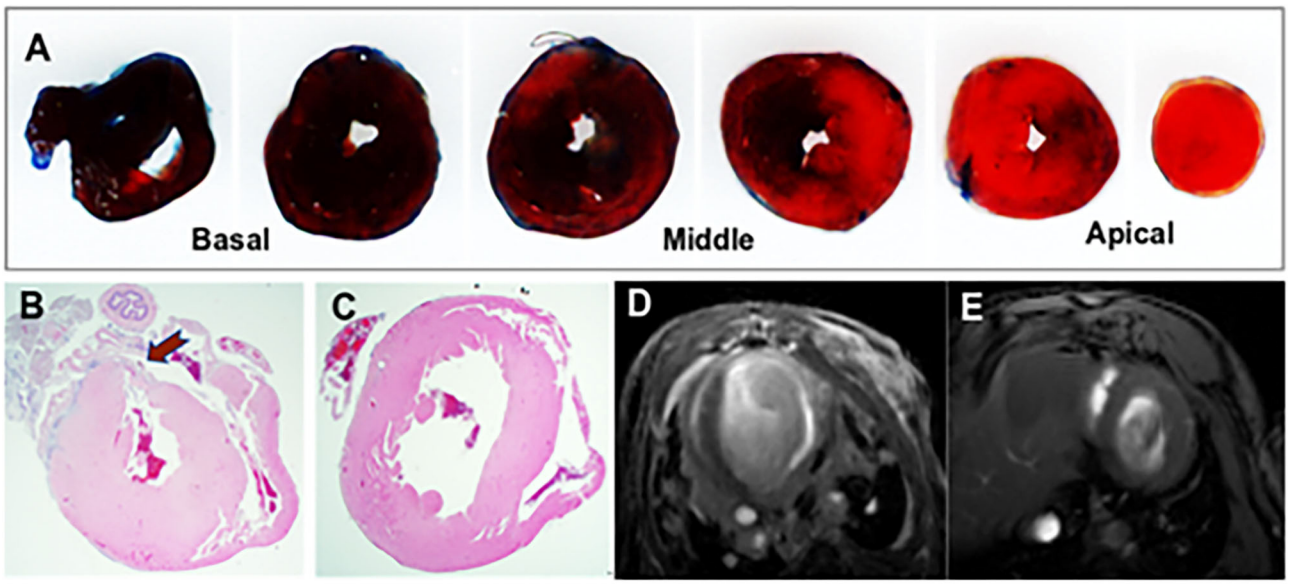


FIGURE 1 | Representative images from the mouse model of acute myocardial infarction (MI). **(A)** TTC stained mouse heart sections at 1 mm after myocardial infraction. The viable myocardium appeared in blue and the area at risk (AAR) in bright red. **(B)** Histological demonstration of a thinned myocardium (shown by an arrow) in a mouse that underwent left anterior descending artery ligation. **(C)** Comparative histology from a normal control mouse. **(D)** Contrast-enhanced cardiac MRI showing anterior wall scar and wall thinning in a mouse model. **(E)** Cardiac MRI showing normal left ventricular morphology.

summation of the attention modules in different axes could partly counterbalance the original features.

PGGAN Shifted by K-Means Removed the Data Outliers, CycleGAN Improved the Data Outline

Compared to the original images (**Figures 5A,F**), Many PGGAN generated images are of unusual shape as demonstrated in **Figures 5B,G**. Since the original training dataset is relatively small, a small raw dataset does not provide sufficiently high variance information for stable image generation. **Table 1** shows that k-means selection reduces FID, and therefore should be an effective method to remove unusually shaped images. Some typical PGGAN generated images with k-means are shown in **Figures 5C,H**. Although some images are still of unusual shape, k-means filters the outliers and reduces the number of unusually shaped images.

Figures 5D–J show CycleGAN generated images. CycleGAN translates images from the real images in different source domains. Unlike the images generated by PGGAN, all CycleGAN-generated images represent a clearer tissue outline. The main problem associated with CycleGAN is that scar tissue images are imperfectly translated from the normal image domain. In our study, this resulted in the overestimation of the scar size (green circle, **Figures 5D,E**) with a hollow (black) core. This semantic difference would be hard to be shifted by k-means. The classification results also showed that k-means is more effective for PGGAN generation shift than for CycleGAN generation shift.

Combination of GAN Augmentation and Kmeans Selection Was a Reliable Method to Improve CNN Performance

All classification results of different augmentation methods are presented in **Table 2**. The performance accuracy (high-to-low) is determined to be PGGAN augmentation, TA and CycleGAN. In particular, 4x PGGAN-k-means based augmentation shows the best accuracy (92.7%). We infer that perhaps the inferior performance of CycleGAN was due to unthorough translation from source image domain. The GAN augmentation classification results without k-means are shown in **Table 3**. Taken together, k-means selection enhances PGGAN augmentation accuracy removing the unusually-shaped images. Images generated by CycleGAN have a better edge definition. Additionally, the ROC (receiver operating characteristic) curves are shown in **Supplementary Figure 5**.

Higher Accuracy Was Correlated With a Smaller Region of Interest

As shown in **Figures 6D,G**, the myocardial scar was correctly localized through Grad-CAM. The ROI (marked red) represents the scar-bearing myocardial segment. However, Grad-CAM was not sensitive enough to locate the myocardial scar at the border zones. One possible explanation for this caveat could be that without pixel-level labeling of myocardial scar, the self-localization of CNN from a small Grad-CAM training set is limited. Further augmentation results with smaller ROI and higher accuracy as shown in **Figures 6A–F** and **Table 2**.

Higher accuracy is likely derived from a precise location of the ROI. A similar heatmap generated from TA (**Figure 6D**)

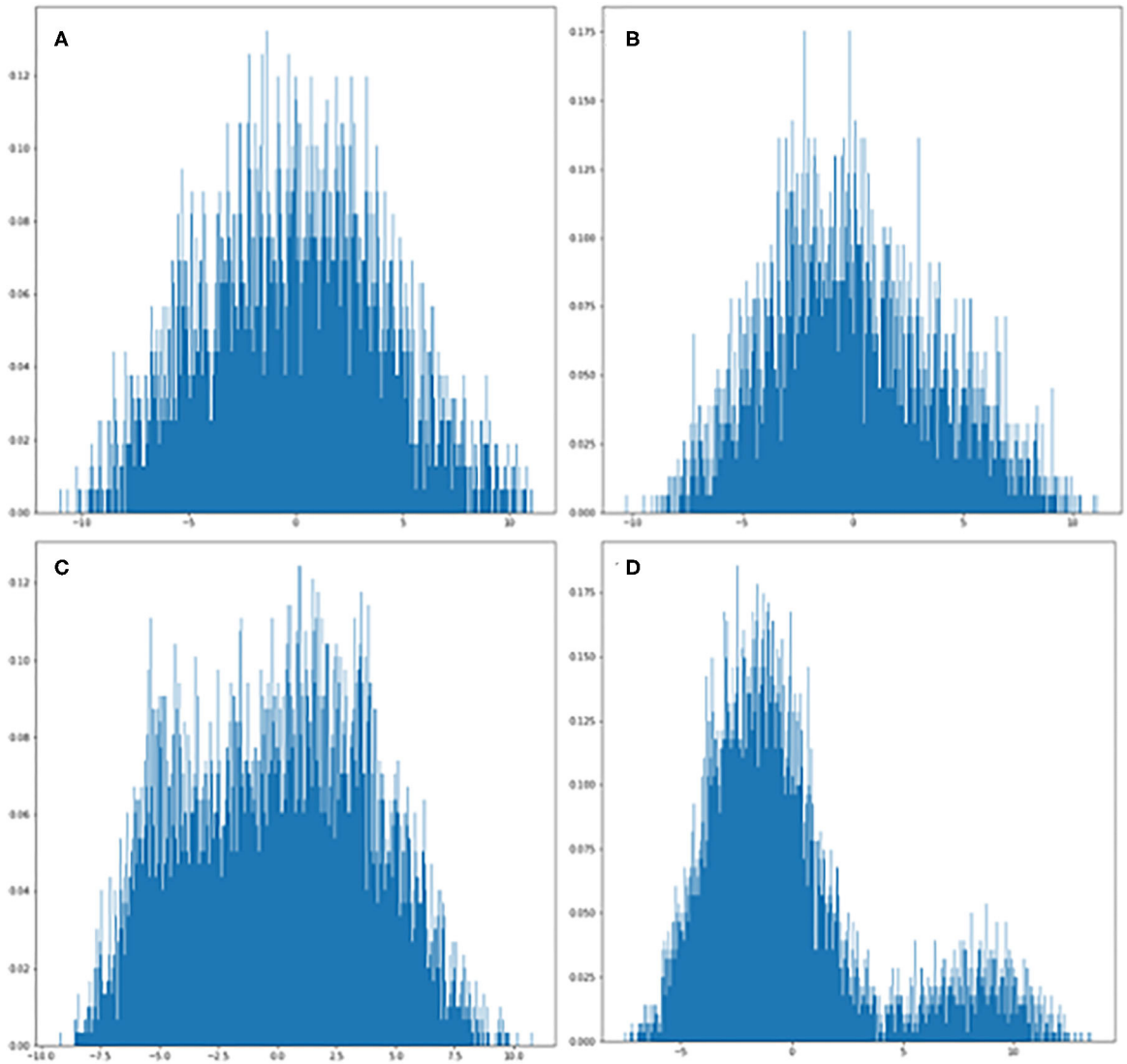


FIGURE 2 | PCA plots of GAN-generated images from each category. **(A)** Visualization of ischemic scar image set generated from PGGAN. **(B)** Visualization of control image set generated from PGGAN. **(C)** Visualization of ischemic scar image set generated from CycleGAN. **(D)** Visualization of control image set generated from CycleGAN.

and from PGGAN augmentation (**Figure 6E**) reveals that ROI generated from PGGAN+k-means is the same as in TA, and the classification accuracies from PGGAN+k-means augmentation (92.7) and TA (92.2) are similar. This also suggests that attention shift could be the reason for inferior training results from CycleGAN (91.7). Since CycleGAN (**Figure 6F**) provides higher emphasis to the subdominant features around the ventricles, this can lead to attention shift.

DISCUSSION

Despite recent advances in the field of deep learning to predict cardiovascular outcomes (27, 28), there are limited data examining the accuracy of myocardial scar classification. This is the first multidisciplinary study that combines the preclinical and clinical approaches to develop a tissue-validated classification and augmentation algorithms in subjects with an ischemic myocardial scar. Our CNN model introduces an attentional block-based data processing approach to improve MobileNetV2 classification of myocardial scar. We also report that GAN is an effective method to mitigate the data imbalance and present a comparative data analysis algorithm to show which GAN-type is must suitable to augment myocardial scar imaging. Finally, we combine k-means and PCA to identify abnormal images with the goal of improving the augmentation effects in advance.

Quantitative interpretation of myocardial scar has remained a challenging task despite the use of

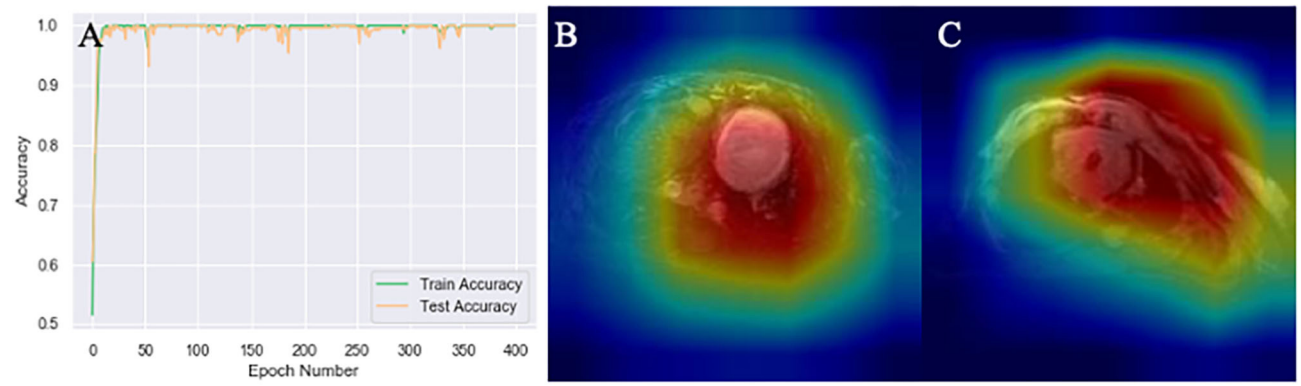


FIGURE 3 | Mouse MRI heatmaps and training curves generated by MobileNetV2. **(A)** Classification accuracy curve. The test accuracy was at 100% after several epochs that stabilized the model. **(B,C)** Heatmaps generated by Gcam from acute myocardial infarction (MI) and control mice, respectively.

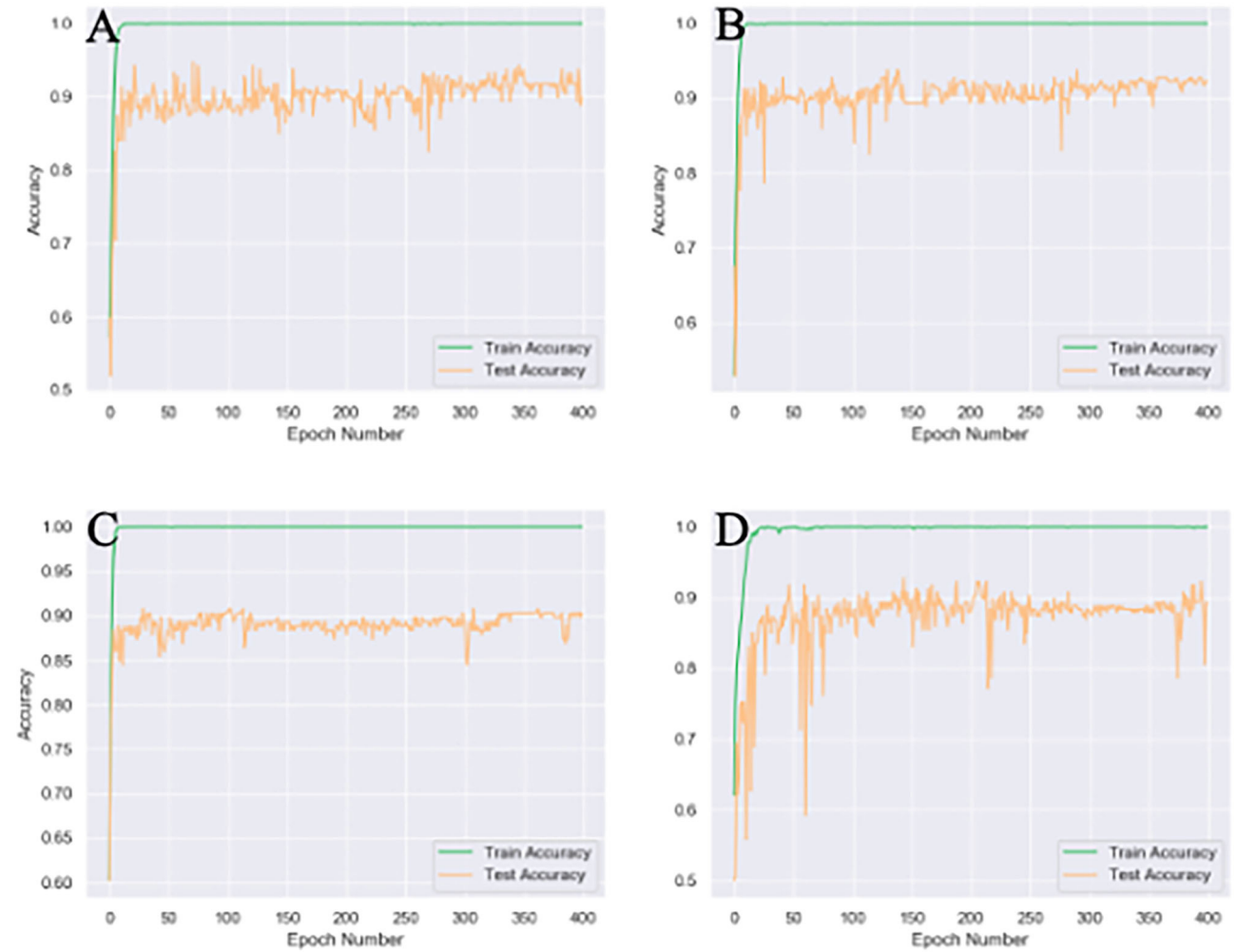


FIGURE 4 | Different attention unit MobileNetV2 training results of 4x traditional augmentation. **(A)** Spatial attention (SA) accuracy curve (89.8% testset accuracy). **(B)** Channel attention (CA) accuracy curve (92.2% testset accuracy). **(C)** Accuracy curve without any attention (89.8% testset accuracy). **(D)**, Mix attention (MA) accuracy curve (89.3% testset accuracy). The accuracy curve fluctuate more violently without any attention module and the final accuracy is the lowest.

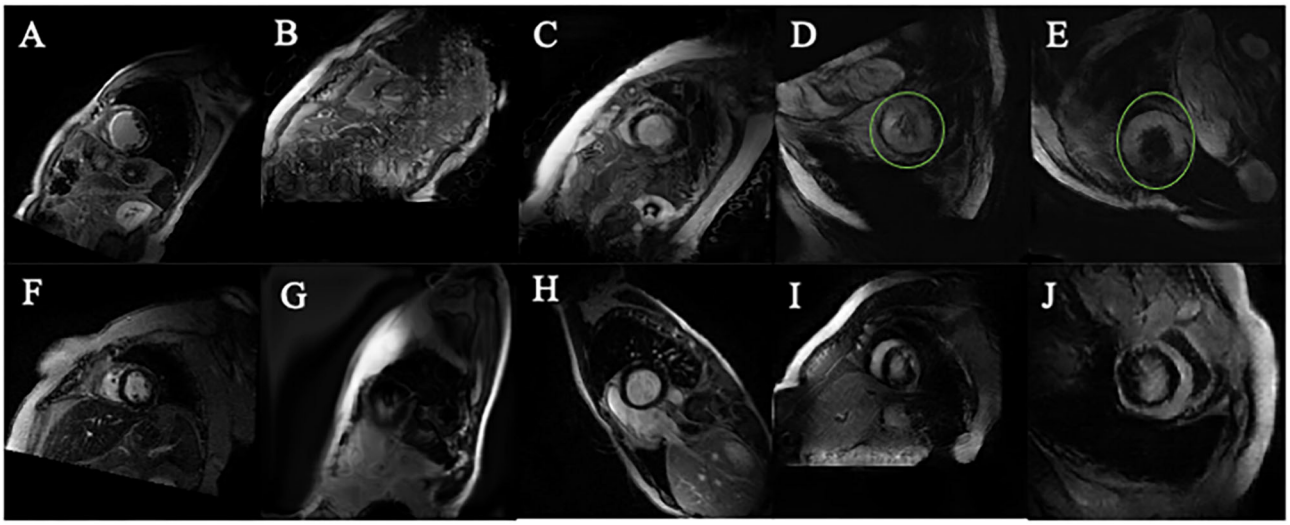


FIGURE 5 | Images generated by PGGAN and CycleGAN. **(A)** MI image from original data. **(B)** MI image generated form PGGAN. **(C)** MI image shifted by kmeans after PGGAN generation. **(D)** MI image generated from CycleGAN. **(E)** MI image shifted by kmeans after CycleGAN generation. **(F)** Control image from original data. **(G)** Control image generated form PGGAN. **(H)** Control image image shifted by kmeans after PGGAN generation. **(I)** Control image generated from CycleGAN. **(J)** Control image shifted by kmeans after CycleGAN generation. Green circle in **(D,E)** is manually added so as to mark the abnormal morphology of the generated image.

TABLE 1 | Generation of FID score from PGGAN.

	FID
PGGAN (MI)	289.208
PGGAN + kmeans (MI)	285.746
PGGAN (control)	323.129
PGGAN + kmeans (control)	322.873

PGGAN, the FID score of PGGAN; PGGAN+kmeans, FID of kmeans sift of original PGGAN generation; MI, myocardial infarction; Control, normal images from the control group.

TABLE 2 | Comparative analysis of various classification approaches.

	Acc (%)
MobilenetV2 (raw)	55.8
MobilenetV2 (raw) (CA)	50.5
TA base MobilenetV2	83.5
TA base MobilenetV2 (CA)	88.8
4x TA MobilenetV2	90.3
4x TA MobilenetV2 (CA)	92.2
PGGAN+kmeans base MobilenetV2	84.5
PGGAN+kmeans base MobilenetV2 (CA)	88.3
4x PGGAN+kmeans MobilenetV2	89.8
4x PGGAN+kmeans MobilenetV2 (CA)	**92.7**
CycleGAN+kmeans base MobilenetV2	83.5
CycleGAN+kmeans base MobilenetV2 (CA)	88.3
4x CycleGAN+kmeans MobilenetV2	87.4
4x CycleGAN+kmeans MobilenetV2 (CA)	91.7

Acc (accuracy), kmeans (k = 1, removes PGGAN generative images. K = 2, removes CycleGAN generative images), base (augmentation applied to increase normal set size same as acute set size so as to mitigate the imbalance). TA, traditional augmentation; CA, channel attention. Bold means highest accuracy.

automated edge-detection techniques (29, 30). The manual segmentation is time-consuming and subjective, which can lead to high intra- and inter-observer variations. To date, there were no studies that attempted the gold standard approaches of myocardial scar classification using a histological validation in an acute MI model, and clinical validation in patients with chronic ischemic myocardial remodeling.

Our initial data classification algorithm tested in a mouse model of acute MI demonstrated 100% accuracy for the classification of acute MI. Although this approach was highly promising and did not require additional data filtration or augmentation algorithms, there can be limitations of fully extrapolating the mouse data into the clinical scenario. First, the mouse model of acute MI was developed by irreversible occlusion of the left anterior descending coronary artery, which is not entirely representative of our patient population with chronic CAD. Second, the mouse model of acute MI was studied within the first two weeks of MI induction. Nevertheless, mice MRI data showed smaller variance and higher interpretation accuracy compared to the human data. This preclinical training model provided additional intimations to advance classification performance and improve accuracy.

Generally, our results show the classification accuracy obtained from combined PGGAN and k-means is comparable to traditional data augmentation. The inferior data augmentation from CycleGAN is likely due to unsupervised image-to-image translation generated without prior restriction. For image-to-image augmentation, training from more powerful translation

GAN like UGATIT platforms may enhance performance and improve accuracy (31). As for noise-to-image augmentation, newer approaches with layer normalization instead of pixel norm in PGGAN could be attempted to improve the semantic understanding.

Although channel attention has shown the highest model stability in our experiments, other attention modules could also be utilized to fulfill additional tasks. Powerful spatial attention modules such as the non-local layer or graph reasoning layer could be optimized and then applied to our system for the accurate classification of myocardial scar in different coronary artery territories (32, 33). Since the image-shifting is a commonly encountered challenge, newer outlier detection algorithms could be utilized to shift the generated images more effectively. The unusual-looking images pose a semantic difference with well-looking images. In particular, for CycleGAN generated images, an unthorough translation of the myocardial scar leads to a high-level of semantic difference.

In contrast with CycleGAN-based image generation, k-means outlier selection is based on the Gaussian distribution rule. Although the feature extraction step from the pretrained MobileNetV2 involves image semantic information, this has a limited utility for our imaging processing. This advantage of PGGAN over the CycleGAN could be the reason for the higher accuracy of PGGAN data augmentation after k-means selection. Our data suggest that the use of an algorithm with an integrated medical image semantic extraction module can extract outliers, and enhance classification performance.

TABLE 3 | GAN augmentation classification results without kmeans.

	Acc (%)
4x PGGAN MobilenetV2	81.1
4x PGGAN MobilenetV2 (CA)	87.9
4x CycleGAN MobilenetV2	82.5
4x CycleGAN MobilenetV2 (CA)	90.7

LIMITATIONS

Our study has a few limitations which can be overcome with future research. This study has a small image dataset size as described above. However, we have, at least in part, addressed this issue by rotating the scar images 8 times and control images 72 times, followed by scaling, shifting and flipping images one time, so that ∼7,000 images with balanced augmentation were obtained. In addition, this study was not designed to study the in-depth mechanisms of ischemic remodeling in mice. Since the image classification was our main goal, the clinical data were not tailored to study the long-term cardiovascular outcomes. Nevertheless, this is the first step toward the training a validation of the myocardial scar classification algorithm using a multidisciplinary approach.

CONCLUSIONS AND FUTURE IMPLICATIONS

We have shown that the channel attention is the most effective attention unit to improve CNN performance. We conclude that the data performance can be improved by utilizing a min-max contrast between the discriminator and generator models of GAN. K-means has a strong ability to remove unusually-shaped

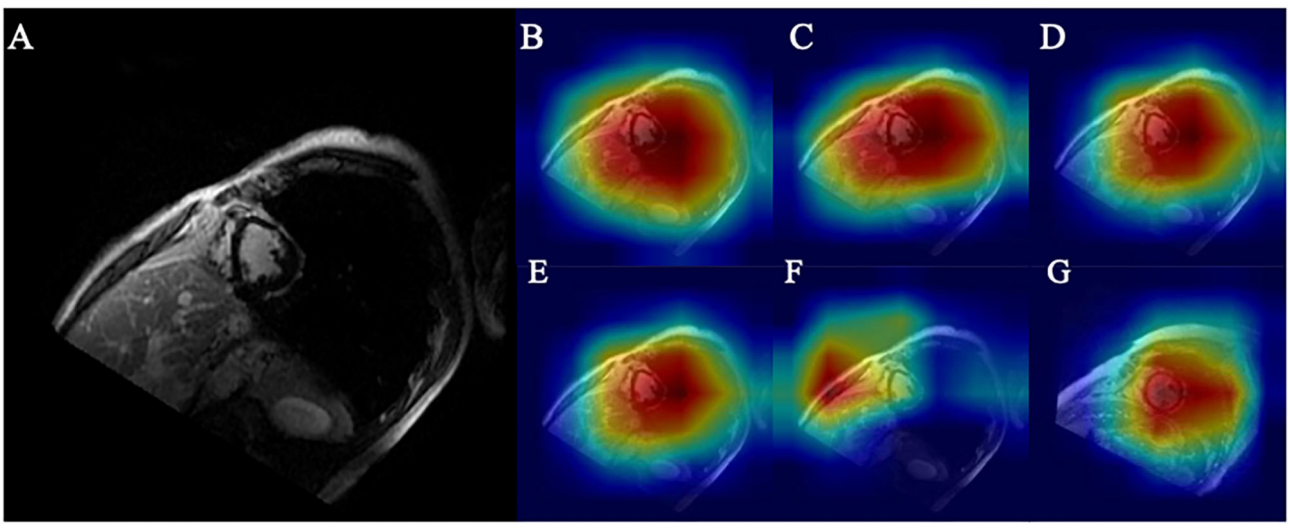


FIGURE 6 | (A) Original ischemic scar image. **(B)** Heatmap of base model. **(C)** Heatmap of CA embedded to base model. **(D)** Heatmap of CA embedded to base model with 4x traditional augmentation. **(E)** Heatmap of CA embedded to base model with 4x PGGAN augmentation. **(F)** Heatmap of CA embedded to base model with 4x CycleGAN augmentation. **(G)** Heatmap of CA embedded to base model with 4x traditional augmentation of image in the control group.

generated images and thus amplifying the accuracy training from PGGAN augmentation (17).

In the future, besides image interpretation, this model could be applied in several medical applications, including a GAN-based data augmentation anonymization tool for large-scale data sharing, and a clinical training tool to educate medical practitioners. One promising approach is to use this method to augment the size and variability of myocardial scar, which can predict clinical outcomes, including heart failure and sudden arrhythmic events.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Human experimental protocols were approved by a institutional review board (IRB) committee from University at Buffalo and all methods involving human/human data were performed in accordance with the relevant guidelines and regulations. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. The animal study was reviewed and approved by all preclinical procedures and protocols conformed to institutional guidelines for the care and use of animals in research and were reviewed and approved by the University at Buffalo Institutional Animal Care and Use Committee (IACUC).

AUTHOR CONTRIBUTIONS

US supervised the clinical aspects of this study and contributed to the manuscript writing. KZ performed experiments, analyzed data, and contributed to the manuscript writing. BK contributed to the manuscript writing. JL, KM, and SS contributed to pre-clinical experiments and data analysis. LY was responsible for the overall supervision of machine learning techniques and critiqued the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (UL1TR001412) to the University at Buffalo. US received support from NIH/NHLBI (K08HL131987 and R01HL152090) and JL received support from VA/United States (IK2 BX004097/BX/BLRD).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2021.726943/full#supplementary-material

REFERENCES

1. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, et al. Heart disease and stroke statistics-2014 update: a report from the American heart association. *Circulation*. (2014). 129:e28–292. doi: 10.1161/01.cir.0000441139.02102.80
2. Goldberger JJ, Cain ME, Hohnloser SH, Kadish AH, Knight BP, Lauer MS, et al. American heart association/American college of cardiology foundation/heart rhythm society scientific statement on noninvasive risk stratification techniques for identifying patients at risk for sudden cardiac death: a scientific statement from the American heart association council on clinical cardiology committee on electrocardiography and arrhythmias and council on epidemiology and prevention. *Circulation.* (2008) 118:1497–518. doi: 10.1161/CIRCULATIONAHA.107.189375
3. Schmidt A, Azevedo CF, Cheng A, Gupta SN, Bluemke DA, Foo TK, et al. Infarct tissue heterogeneity by magnetic resonance imaging identifies enhanced cardiac arrhythmia susceptibility in patients with left ventricular dysfunction. *Circulation.* (2007) 115:2006–14. doi: 10.1161/CIRCULATIONAHA.106.653568
4. Klocke FJ, Wu E, Lee DC. "Shades of gray" in cardiac magnetic resonance images of infarcted myocardium: can they tell us what we'd like them to? *Circulation.* (2006) 114:8–10. doi: 10.1161/CIRCULATIONAHA.106.636456
5. Heydari B, Kwong RY. Cardiac magnetic resonance infarct heterogeneity: is it ready to be used on patients for the prevention of sudden cardiac death? *Eur Heart J Cardiovasc Imag.* (2014) 15:108–9. doi: 10.1093/ehjci/jet188
6. Krizhevsky A, Sutskever I, Hinton GE. *In Advances in Neural Information Processing Systems*. (2012). p. 1097–105.
7. Simonyan K, Zisserman A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. (2014).
8. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. *In Proceedings of the IEEE Conference on Computer Vision and pattern Recognition*. (2016). p. 2818–26.
9. He K, Zhang X, Ren S, Sun J. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2016) 770–8.
10. Litjens G, Kooi T, Bejnordi BE, Setio AA, 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
11. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Ann Rev Biomed Eng.* (2017) 19:221–48. doi: 10.1146/annurev-bioeng-071516-044442
12. Mosleh W, Chaudhari MR, Sonkawade S, Mahajan S, Khalil C, Frodey K, et al. The therapeutic potential of blocking galectin-3 expression in acute myocardial infarction and mitigating inflammation of infarct region: a clinical outcome-based translational study. *Biomark Insights.* (2018) 13:1177271918771969. doi: 10.1177/1177271918771969
13. Aartsen WM, Schuijt MP, Danser AH, Daemen MJ, Smits JF. The role of locally expressed angiotensin converting enzyme in cardiac remodeling after myocardial infarction in mice. *Cardiovasc Res.* (2002) 56:205–13. doi: 10.1016/S0008-6363(02)00516-3
14. Karthikeyan B, Sonkawade SD, Pokharel S, Preda M, Schweser F, Zivadinov R, et al. Tagged cine magnetic resonance imaging to quantify regional mechanical changes after acute myocardial infarction. *Magn Reson Imag.* (2020) 66:208–18. doi: 10.1016/j.mri.2019.09.010
15. Mentkowski KI, Lang JK. Exosomes engineered to express a cardiomyocyte binding peptide demonstrate improved cardiac retention *in Vivo*. *Sci Rep.* (2019) 9:10041. doi: 10.1038/s41598-019-46407-1
16. Teman CJ, Wilson AR, Perkins SL, Hickman K, Prchal JT, Salama ME. Quantification of fibrosis and osteosclerosis in myeloproliferative neoplasms: a computer-assisted image study. *Leukemia Res.* (2010) 34:871–6. doi: 10.1016/j.leukres.2010.01.005

17. Burnett HW Jr, Steinberg I, Dotter CT. Seven-year survival after radiation therapy in a patient with inoperable cancer of the lung. *N Engl J Med.* (1952) 247:850–1. doi: 10.1056/NEJM195211272472207
18. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. *Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. (2017).
19. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. *In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* (2018). p. 4510–20.
20. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. *A Comprehensive Survey on Transfer Learning*. (2019).
21. Wang Y, Yao Q, Kwok J, Ni LM. in arXiv: 1904.05046 (2019).
22. Hu J, Shen L, Sun G. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2018). p. 7132–41.
23. Woo S, Park J, Lee J-Y, So Kweon I. *In Proceedings of the European Conference on Computer Vision (ECCV).* (2018). p. 3–19.
24. Karras T, Aila T, Laine S, Lehtinen J. *Progressive Growing of Gans for Improved Quality, Stability, and Variation*. (2017).
25. Zhu J-Y, Park T, Isola P, Efros AA. *In Proceedings of the IEEE International Conference on Computer Vision*. (2017). 2223–32.
26. Selvaraju AR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, et al. In *Proceedings of the IEEE International Conference on Computer Vision.* (2017). p. 618–26.
27. Kotu LP, Engan K, Borhani R, Katsaggelos AK, Ørn S, Woie L, et al. Cardiac magnetic resonance image-based classification of the risk of arrhythmias in post-myocardial infarction patients. *Artif Intellig Med.* (2015) 64:205–15. doi: 10.1016/j.artmed.2015.06.001
28. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circ Res.* (2017) 121:1092–101. doi: 10.1161/CIRCRESAHA.117.311312
29. Petitjean C, Dacher J-N. 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
30. 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. *Mag Reson Mat Phys Biol Med.* (2016) 29:155–95. doi: 10.1007/s10334-015-0521-4
31. Kim J, Kim M, Kang H, Lee K. *U-GAT-IT: Unsupervised Generative Attentional Networks With Adaptive Layer-Instance Normalization for Image-to-Image Translation.* (2019).
32. Li X, Lin C, Li C, Sun M, Wu W, Yan J, et al. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 13836–13845.
33. Wang X, Girshick R, Gupta A, He K. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2018) 7794–803.

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 14 October 2021
doi: 10.3389/fcvm.2021.742640

Deep Learning for Classification and Selection of Cine CMR Images to Achieve Fully Automated Quality-Controlled CMR Analysis From Scanner to Report

Vittoria Vergani[1]*, *Reza Razavi*[1,2], *Esther Puyol-Antón*[1†] and *Bram Ruijsink*[1,3*†]

[1] School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom, [2] Department of Adult and Paediatric Cardiology, Guy's and St. Thomas' NHS Foundation Trust, London, United Kingdom, [3] Department of Cardiology, Division of Heart and Lungs, University Medical Centre Utrecht, Utrecht, Netherlands

OPEN ACCESS

Edited by:
Paul Leeson,
University of Oxford, United Kingdom

Reviewed by:
Richard Nethononda,
Chris Hani Baragwanath Hospital, South Africa
Emmanuel Androulakis,
Royal Brompton and Harefield NHS Foundation Trust, United Kingdom

***Correspondence:**
Bram Ruijsink
jacobus.ruijsink@kcl.ac.uk
Vittoria Vergani
vergani.vittoria@gmail.com

† These authors have contributed equally to this work and share last authorship

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

Received: 16 July 2021
Accepted: 09 September 2021
Published: 14 October 2021

Citation:
Vergani V, Razavi R, Puyol-Antón E and Ruijsink B (2021) Deep Learning for Classification and Selection of Cine CMR Images to Achieve Fully Automated Quality-Controlled CMR Analysis From Scanner to Report. Front. Cardiovasc. Med. 8:742640. doi: 10.3389/fcvm.2021.742640

Introduction: Deep learning demonstrates great promise for automated analysis of CMR. However, existing limitations, such as insufficient quality control and selection of target acquisitions from the full CMR exam, are holding back the introduction of deep learning tools in the clinical environment. This study aimed to develop a framework for automated detection and quality-controlled selection of standard cine sequences images from clinical CMR exams, prior to analysis of cardiac function.

Materials and Methods: Retrospective study of 3,827 subjects that underwent CMR imaging. We used a total of 119,285 CMR acquisitions, acquired with scanners of different magnetic field strengths and from different vendors (1.5T Siemens and 1.5T and 3.0T Phillips). We developed a framework to select one good acquisition for each conventional cine class. The framework consisted of a first pre-processing step to exclude still acquisitions; two sequential convolutional neural networks (CNN), the first (CNN_{class}) to classify acquisitions in standard cine views (2/3/4-chamber and short axis), the second (CNN_{QC}) to classify acquisitions according to image quality and orientation; a final algorithm to select one good acquisition of each class. For each CNN component, 7 state-of-the-art architectures were trained for 200 epochs, with cross entropy loss and data augmentation. Data were divided into 80% for training, 10% for validation, and 10% for testing.

Results: CNN_{class} selected cine CMR acquisitions with accuracy ranging from 0.989 to 0.998. Accuracy of CNN_{QC} reached 0.861 for 2-chamber, 0.806 for 3-chamber, and 0.859 for 4-chamber. The complete framework was presented with 379 new full CMR studies, not used for CNN training/validation/testing, and selected one good 2-, 3-, and 4-chamber acquisition from each study with sensitivity to detect erroneous cases of 89.7, 93.2, and 93.9%, respectively.

Conclusions: We developed an accurate quality-controlled framework for automated selection of cine acquisitions prior to image analysis. This framework is robust and generalizable as it was developed on multivendor data and could be used at the beginning of a pipeline for automated cine CMR analysis to obtain full automatization from scanner to report.

Keywords: cardiac magnetic resonance, deep learning, quality control, cardiac function, view-selection, multivendor

INTRODUCTION

Cardiac magnetic resonance (CMR) is the state-of-the-art clinical tool to assess cardiac morphology, function, and tissue characterization (1), and both European and American guidelines advocate its use to diagnose and monitor a large number of cardiovascular diseases (2, 3). The role of CMR continues to grow due to the technical advances that allow increasingly detailed analysis of the cardiovascular system.

However, systematic manual analysis of the different CMR sequences which are acquired during a typical CMR exam is highly time consuming, where the bulk of the time is taken up by repetitive tasks, such as image identification, selection, and segmentation, which are at the basis of CMR post-processing.

Deep learning (DL), a branch of artificial intelligence (AI), is securing an emergent role in the field of CMR, as it provides for automatization of repetitive tasks, significantly reducing the time required for image analysis, while maintaining a high degree of accuracy (4, 5). Physicians' time can thus be optimized and targeted for critical review of clinical and imaging information to reach a correct diagnosis. Moreover, automated analysis allows access to biomarkers of cardiac function that would normally be too labor intensive to obtain, such as peak ejection and filling rates from ventricular volume curves (6, 7) or atrioventricular valve planar motion (8) from long-axis segmentations.

Several groups have shown promising results on the implementation of DL in the analysis of CMR, including segmentation of cine images to derive cardiac function (5), analysis of perfusion defects to detect inducible ischemia (9), and assessment of late gadolinium enhancement and T1 mapping to aid tissue characterization (10, 11).

However, some limitations still need to be addressed before a widespread clinical adoption of DL tools, such as steps to perform automated selection of target images from the full CMR exam, as well as robust systems to flag inadequate quality of image or of analysis, which are the necessary steps that precede analysis of CMR sequences in the clinical setting. We have previously shown that comprehensive quality-control can be introduced into a DL pipeline for accurate, automated analysis of cine CMR images that adheres to clinical safety standards (5). On the other hand, DL has not yet been systematically implemented for image recognition and selection prior to analysis in CMR.

This study aimed at developing a framework for automated identification and quality-controlled (QC) selection of cine images used for cardiac function analysis from routine clinical CMR exams. This framework was then implemented as the first step of a larger pipeline for QC CMR analysis of cine images previously developed by our group (5).

Abbreviations: 2Ch, 2-Chamber view; 3Ch, 3-Chamber view; 4Ch, 4-Chamber view; ACHD, adult congenital heart disease; AI, artificial intelligence; CMR, cardiac magnetic resonance; CNN, convolutional neural network; DL, deep learning; GE, General Electrics; GSTFT, Guy's and St. Thomas' NHS Foundation Trust; LVOT, left ventricular outflow tract; QC, quality control; SAX, short axis.

MATERIALS AND METHODS

The framework we present is composed of a set of algorithms combined with two convolutional neural networks (CNN) aimed at identifying conventional cine views (CNN_{class}) and at sorting these images according to quality into "correct" and "wrong" (CNN_{QC}). This construction allows for the framework, presented with a full CMR exam, to perform a selection of one good quality acquisition for each conventional cine class, which is then used for analysis of cardiac function, and to flag exams when no image of sufficient quality could be identified (see **Figure 1**).

The developed pipeline was implemented in Python using standard libraries such as Numpy and Pandas as a dedicated deep learning library Pytorch.

Study Population

This is a retrospective multivendor study conducted on a large CMR dataset. 3,445 CMR exams were included: 1,510 from UK Biobank and 1,935 from Guy's and St. Thomas' NHS Foundation Trust (GSTFT), London, comprising of a total of 119,285 individual CMR acquisitions. Images were acquired on 1.5T Siemens and 1.5T and 3.0T Phillips CMR scanners using a large variety of protocols, with variable voxel- and image-size, acquisition techniques and under-sampling factors.

Our dataset was randomly selected from the pool of available studies in the UK Biobank and GSTFT databases. All exams were acquired between 2004 and 2020. The random selection was used to obtain a heterogeneous population, including both healthy and pathological hearts, with a variety of cardiac pathologies (ischemic heart disease, dilated and hypertrophic cardiomyopathy, valvular heart disease, adult congenital heart disease (ACHD), and others). In the case of grossly disruptive artifacts (e.g., device artifacts covering the majority of the chambers), or grossly distorted anatomy (e.g., patients with single-ventricle morphology or Ebstein's disease), CMR exams were excluded from the dataset.

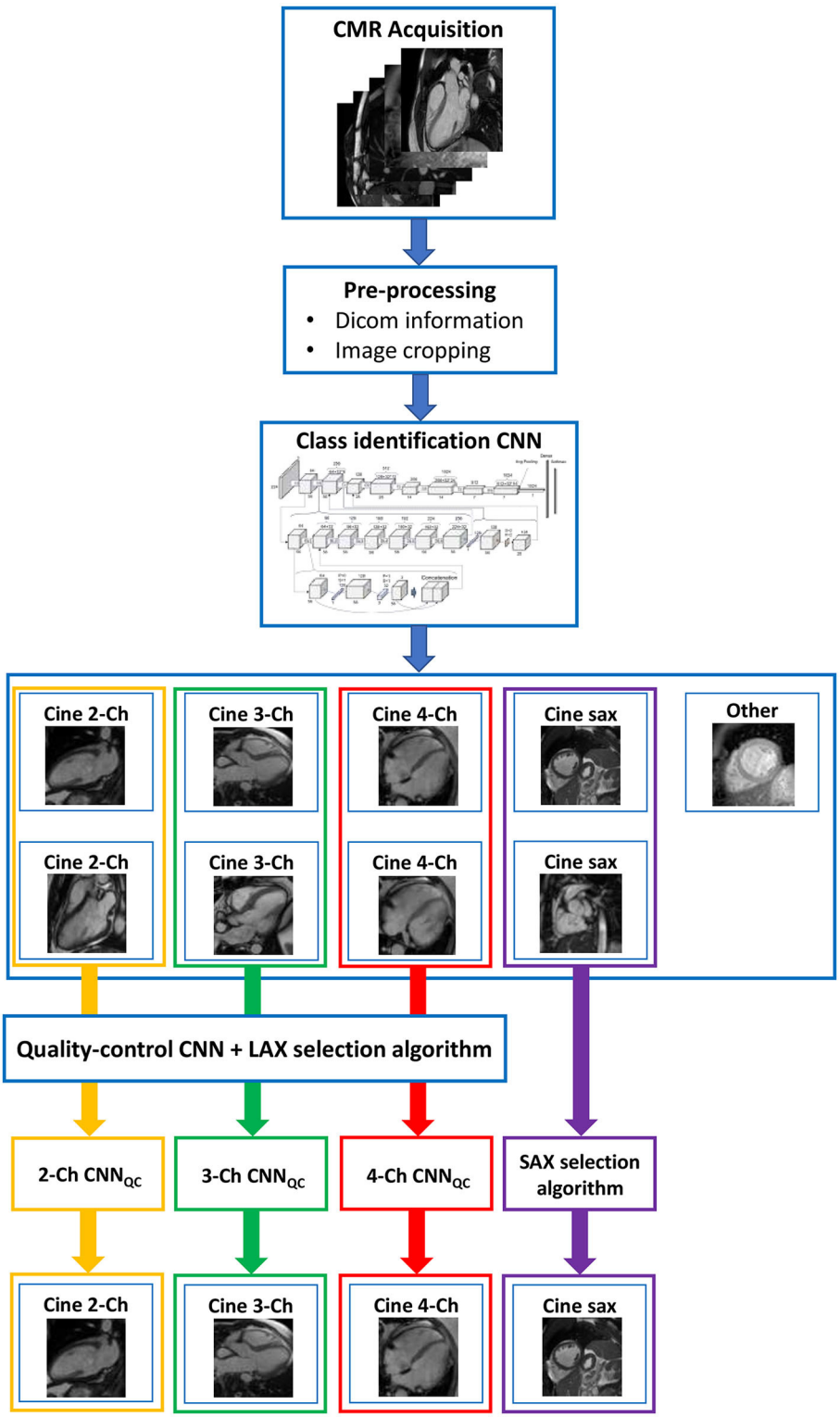


FIGURE 1 | Complete framework. The framework consisted of a first pre-processing step to exclude still images; two sequential convolutional neural networks (CNN), the first to classify images in standard cine views (2/3/4-chamber and short axis), the second to classify images according to image quality and orientation; a final algorithm to select one good image of each class. This construction allows for the framework, presented with a full CMR exam, to perform a quality-controlled selection of one good image for each conventional cine class, which is then used for analysis of cardiac function. Ch, chamber; CNN, convolutional neural network; LAX, long axis; SAX, short axis; QC, quality control.

Identification of Conventional Cine Classes

The first part of our framework aimed to identify the standard cine views used for analysis of cardiac function from the complete exam. These were the 2-Chamber view (2Ch), 3-Chamber view (3Ch), 4-Chamber view (4Ch), and the short-axis stack (SAX).

First, all single-frame acquisitions were excluded from the exams based on the dicom-header information. This allowed to reduce the bulk of acquisitions available in a complete CMR exam to a set of multi-frame acquisitions. The 2Ch, 3Ch, 4Ch, and SAX sequences were subsequently selected from the remaining acquisitions using our view selection algorithm. To identify the class of the cine acquisitions, only a single frame of the acquisition is needed. Therefore, the first frames of each of the remaining data was used for training. All data was manually classified into the conventional cine classes by an expert physician (2,905 2Ch, 1,171 3Ch, 2,963 4Ch, 9,112 SAX images) and a class of "other" (4,043). In case of doubt, a second opinion was sought, and a decision was made by consensus. Prior to CNN training, all images were cropped to a standard size of 256 × 256 pixels and converted to numpy arrays.

The manually classified data were divided as follows: 80% was used for training, 10% was used for validation, and 10% for testing. The training, validation and test data cohorts had a mutually exclusive subject pool, i.e., acquisitions from the same subject could only be used in one of the three cohorts. We trained seven state-of-the-art CNN architectures: AlexNet (12), DenseNet (13), MobileNet (14), ResNet (15), ShuffleNet (16), SqueezeNet (17), and VGG (18). Each network was trained for 200 epochs with cross entropy loss, to classify end-diastolic images into the five classes described. For training data, data augmentation was performed on-the-fly using random translations (±30 pixels), rotations (±90°), flips (50% probability) and scalings (up to 20%) to each mini-batch of images before feeding them to the network. The probability of augmentation for each of the parameters was 50%. Augmentation is the only technique we use to prevent over-fitting, as other techniques were not found to improve performance and their omission contributed to a simpler network architecture.

An additional algorithm was used after CNN_{class} to check complete classification of the short axis acquisition. An image classified as SAX was confirmed to belong to a short axis acquisition if the following two criteria, screened by the algorithm, were met: (1) the image belonged to a stack composed of a minimum of 8 slices, (2) at least 2 out the 3 central images of the stack were classified as short axis by CNN_{class}.

Quality Control of Selected Images

The second part of our framework aimed to scrutinize the quality of the identified long-axis cine images. QC of short axis acquisitions was not performed in this step, as our downstream pipeline for automated CMR analysis already includes short-axis QC (5).

To train the networks for QC (CNN_{QC}), a set of 2Ch (1,937), 3Ch (1,591), 4Ch (2,003) images from our database were reviewed by an expert physician and classified as "correct" or "wrong" based on whether image quality was satisfactory for subsequent analysis. Images that included mis-triggering, breathing, implant or fold-over artifacts were classified as "wrong" if the detection of myocardial borders was hindered. Moreover, all off-axis orientations (i.e., presence of left ventricular outflow tract (LVOT) in 4Ch, fore-shortening of the apex, absence of any of the valves in 3Ch) were also classified as "wrong." In case of doubt, a second opinion was sought, and a decision was made by consensus. The resulting database consisted of 1,444 "correct" and 493 "wrong" 2Ch, 1,098 "correct" and 493 "wrong" 3Ch images and 1,393 "correct" and 610 "wrong" 4Ch images.

The manually classified data were used to train QC networks for each class (2Ch-CNN_{QC}, 3Ch-CNN_{QC}, 4Ch-CNN_{QC}). The data were divided as follows: 80% training, 10% for validation, 10% for testing. We trained the same seven CNN architectures as described in the previous section. We used the same training process as described for training of CNN_{class}, with the difference that CNN_{QC} was trained as binary classifiers, i.e., a two-class classification problem as opposed to five, and therefore used binary cross entropy with a logit loss function. Additionally, we implemented an adaptive learning rate scheduler, which decreases the learning rate by a constant factor of 0.1 after 5 epochs stopping on plateau on the validation/test set (commonly known as ReduceLROnPlateau). This step was added as it improves CNN training when presented with unbalanced datasets. The CNN_{QC}'s output was a binary classification ("correct" vs. "wrong"), as well as the probabilities (i.e., certainty) associated with the classification for each case.

Complete Framework: From Full Study to Selection

To complete the framework, the CNN_{class} and CNN_{QC} were combined with a final selection algorithm. This algorithm selected one good quality acquisition of each standard cine view for image analysis, when multiple acquisitions of a single class were present in the exam. For long axis data, it did so by identifying the case with the highest probability of being scored "correct" by the CNN_{QC}. For short axis, the stack with the highest probability of belonging to SAX (obtained from the output of the CNN_{class}) was selected. If any of the classes was absent in an exam, or the framework did not identify an image of sufficient quality, the case was flagged for clinician review.

As the individual components act in series in the complete framework, their sequential action will yield an overall performance that is different from the sum of the individual ones. To find the best combination of components, each possible combination of the trained CNN architectures trained in the previous steps was tested using an additional test-set of 379 scans randomly selected from the database, not previously used for CNN training. For each exam, a manual operator selected the best cine long and short axis acquisitions. To determine the intra- and inter-observer variability present in the manual analysis, 100 randomly selected scans were re-analyzed by the same operator and by a second operator.

TABLE 1 | Population characteristics.

		CNN training	Complete framework
Number		3,445	379
Age (years)		57 ± 16	49 ± 19
Sex (males)		1,911 (55)	228 (60)
Height (cm)		176 ± 32	171 ± 18
Weight (kg)		79 ± 18	80 ± 19
BMI (kg/m^2)		27 ± 5	27 ± 7
Ethnicity	Caucasian	2,401 (69.7)	231 (60.9)
	Afro-Caribbean	172 (5.0)	54 (14.2)
	Asian	85 (2.5)	10 (2.6)
	Other	21 (0.6)	7 (1.8)
	Not stated	766 (22.2)	77 (20.3)
Cardiac pathology	Healthy	1,886 (54.7)	68 (17.9)
	IHD	315 (9.1)	43 (11.3)
	DCM	167 (4.8)	27 (7.1)
	HCM	77 (2.2)	16 (4.2)
	ACHD	185 (5.4)	59 (15.6)
	Valvular	133 (3.9)	37 (9.8)
	Vascular	104 (3.0)	32 (8.4)
	Arrhythmic	159 (4.6)	26 (6.9)
	Other	419 (12.2)	71 (18.7)

Age, sex, height, weight, and cardiac pathology of subjects used for training of CNNs, framework and full pipeline validation. All continuous values are reported as mean ± standard deviation, while categorical variables are reported as number (percentage). ACHD, adult congenital heart disease (excluding valvular and vascular abnormalities); CNN, convolutional neural network; HCM, hypertrophic cardiomyopathy; IHD, ischaemic heart disease; SD, standard deviation.

Statistics

Class Identification CNN

Precision, recall, and F1-score of each class ("4Ch," "3Ch," "2Ch," "SAX," "other") and overall accuracy were computed at test-time to evaluate the performance of each trained CNN_{class}.

Quality Control CNNs

Precision, recall, and F1-score of each class ("correct," "wrong") and overall accuracy were assessed to evaluate the performance at test-time of each trained 2Ch/3Ch/4Ch-CNN_{QC}.

Complete Framework

Sensitivity (defined as: the percentage of incorrect cases identified as incorrect), specificity (defined as: the percentage of correct cases identified as correct), and balanced accuracy were computed for each framework. Cohen kappa coefficient was used to assess intra- and inter-observer variability.

Full Pipeline: From Scanner to Report

Finally, we added the complete framework as the first step of our previously validated pipeline for quality-controlled AI-based analysis of cardiac function from CMR (5). Broadly, this pipeline consists of quality-controlled image segmentation and analysis of cine images to obtain LV and RV volumes and mass, LV ejection and filling dynamics, and longitudinal, radial and circumferential strain.

TABLE 2 | Class identification CNN_{class} performance.

	AlexNet			DenseNet			MobileNet			ResNet			ShuffleNet			SqueezeNet			VGG		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
2-chamber	0.998	0.993	0.996	0.998	1.000	0.999	0.997	1.000	0.998	0.998	1.000	0.999	0.993	1.000	0.997	0.986	0.991	0.989	0.993	1.000	0.997
3-chamber	0.991	0.953	0.972	1.000	1.000	1.000	1.000	0.996	0.998	0.996	0.996	0.996	1.000	0.983	0.991	0.987	0.979	0.983	0.987	0.987	0.987
4-chamber	1.000	0.990	0.995	1.000	0.998	0.999	0.990	0.995	0.992	0.998	0.995	0.997	0.997	1.000	0.998	0.987	0.992	0.989	0.993	0.997	0.995
Short axis	0.995	0.999	0.997	0.996	0.999	0.998	0.997	1.000	0.998	0.999	0.999	0.999	0.996	1.000	0.998	0.992	0.997	0.994	0.996	1.000	0.998
Other	0.973	0.985	0.979	0.998	0.990	0.994	0.996	0.984	0.990	0.994	0.995	0.994	0.996	0.984	0.990	0.984	0.969	0.976	0.995	0.978	0.986
Accuracy			**0.992**			**0.998**			**0.996**			**0.998**			**0.996**			**0.989**			**0.994**

Precision, recall, and F1 score for each CNN_{class}-determined class, and accuracy of each trained CNN_{class}.

TABLE 3 | Class identification CNN_{QC} performance.

	AlexNet			DenseNet			MobileNet			ResNet			Shufflenet			SqueezeNet			VGG		
	Precision	Recall	F1-score	precision	recall	F1-score	precision	recall	F1-score	Precision	recall	F1-score	precision	recall	F1-score	precision	recall	F1-score	precision	recall	F1-score
2Ch-CNN_{QC}																					
Correct	0.856	0.844	0.850	0.931	0.841	0.884	0.889	0.917	0.903	0.877	0.910	0.893	0.881	0.869	0.875	0.755	0.993	0.858	0.873	0.955	0.912
Wrong	0.536	0.559	0.547	0.620	0.806	0.701	0.714	0.645	0.678	0.683	0.602	0.640	0.608	0.634	0.621	0.000	0.000	0.000	0.803	0.570	0.667
Accuracy			**0.775**			**0.832**			**0.851**			**0.835**			**0.812**			**0.751**			**0.861**
3Ch-CNN_{QC}																					
Correct	0.788	0.795	0.792	0.815	0.923	0.866	0.850	0.873	0.861	0.830	0.868	0.849	0.880	0.800	0.838	0.690	1.000	0.816	0.824	0.873	0.848
Wrong	0.536	0.525	0.531	0.757	0.535	0.627	0.699	0.657	0.677	0.674	0.606	0.638	0.630	0.758	0.688	0.000	0.000	0.000	0.674	0.586	0.627
Accuracy			**0.712**			**0.803**			**0.806**			**0.787**			**0.786**			**0.690**			**0.784**
4Ch-CNN_{QC}																					
Correct	0.877	0.821	0.848	0.900	0.900	0.900	0.891	0.878	0.884	0.878	0.900	0.888	0.906	0.832	0.867	0.705	1.000	0.827	0.891	0.878	0.884
Wrong	0.630	0.726	0.675	0.761	0.761	0.761	0.719	0.744	0.731	0.745	0.701	0.722	0.664	0.795	0.724	0.000	0.000	0.000	0.719	0.744	0.731
Accuracy			**0.793**			**0.859**			**0.838**			**0.841**			**0.821**			**0.705**			**0.838**

Precision, recall, and F1 score for "correct" and "wrong" cases for 2Ch-CNN_{QC}, 3Ch-CNN_{QC}, 4Ch-CNN_{QC}, and accuracy of each trained CNN_{QC}.

We present the feasibility and importance of a fully automated multi-step QC pipeline by (1) running 700 new CMR exams cases (not earlier seen during training) through the pipeline, and (2) presenting a video (**Supplementary Video 1**) of the full analysis for a good-quality case. For the 700 cases we ran through the pipeline we report the average time for selection and complete cine analysis from a full CMR study, and again report sensitivity, specificity and balanced accuracy of error detection.

RESULTS

Study Population

Of the 3,827 CMR exams used for this study 3,448 were used for the training and validation of CNN_{class} and CN_{QC} [1,026 acquisition of 16,151 (6.4%) were excluded because of grossly disruptive artifacts or grossly distorted anatomy]. These included patients undergoing clinical scans at GSTFT as well as subjects voluntarily enrolling onto the UK BioBank project, yielding a heterogeneous population in terms of sex (55% male) and clinical presentation (43% healthy, the remaining displaying a wide variety of cardiovascular pathologies, as shown in **Table 1**).

The remaining 379 CMR scans were used to test the complete framework. These were all selected from the GSTFT database to obtain a population representative of routine clinical practice. Demographic characteristics are comparable to the training population, but clinical presentation was more variable, with only 18% of patients having no cardiovascular pathology.

Population characteristics are summarized in **Table 1**.

Class Identification CNN

Precision, recall, F1-score, accuracy for all CNN_{class} are presented in **Table 2**. All trained architectures showed excellence performance, with accuracy ranging from 0.989 to 0.998. DenseNet and ResNet reached highest accuracy, i.e., 0.988. DenseNet showed best precision, recall and F1-score for conventional cine classes: 0.998, 1.00, 0.999 for 2Ch, 1.00, 1.00, 1.00 for 3Ch, 1.00, 0.998, 1.00 for 4Ch, and 0.996, 0.999, 0.998 for SAX.

Quality Control CNN

Precision, recall, F1-score, and accuracy for all CNN_{QC} are shown in **Table 3**. Accuracy was variable for different architectures and ranged from 0.751 to 0.861 for 2Ch, from 0.690 to 0.806 for 3Ch, and from 0.705 to 0.859 for 4Ch. Precision, recall and F1-score were consistently lower for the "wrong" class compared to the "correct" class for all trained architectures and across the 3 different chamber views.

Complete Framework

Sensitivity, specificity, and balanced accuracy of each constructed framework to identify and select one good quality 2Ch, one good quality 3Ch, and one good quality 4Ch image for each exam are shown in **Table 4**. For the sake of brevity, we present the results of one CNN_{class}, i.e., DenseNet, given the very high and similar performance of all different architectures, combined with all possible CNN_{QC}'s.

TABLE 4 | Framework validation.

	Framework for image identification and selection								
	DenseNet CNN_{class}+								
	2Ch-CNN_{QC}			3Ch-CNN_{QC}			4Ch-CNN_{QC}		
Network	**SEN**	**SPE**	**BACC**	**SEN**	**SPE**	**BACC**	**SPE**	**SEN**	**BACC**
AlexNet	79.4	89.8	84.6	84.4	85.8	85.1	70.3	90.9	80.6
DenseNet	75.0	91.1	83.0	**93.2**	**85.3**	**89.2**	88.2	90.9	89.5
MobileNet	78.1	92.4	85.2	88.9	85.9	87.4	72.2	91.9	82.1
ResNet	77.8	91.4	84.6	88.9	85.1	87.0	83.3	90.9	87.1
ShuffleNet	**89.7**	**91.5**	**90.6**	88.9	85.9	87.4	**93.9**	**89.2**	**91.6**
SqueezeNet	55.3	88.2	71.8	51.8	91.0	71.4	23.3	95.5	59.4
VGG	84.4	91.8	88.1	84.8	85.1	84.9	79.4	91.4	85.4

SEN, Sensitivity; SPE, Specificity; and BACC, balanced accuracy of each constructed framework for image identification and selection. Best performing framework per class is highlighted in blue.

The best performing framework was: DenseNet CNN_{class} + ShuffleNet 2Ch-CNN_{QC} (sensitivity = 89.7%, specificity = 91.5%, balanced accuracy = 90.6%), DenseNet 3Ch-CNN_{QC} (sensitivity = 93.2%, specificity = 85.3%, balanced accuracy = 89.2%), ShuffleNet 4Ch-CNN_{QC} (sensitivity = 93.9%, specificity = 89.2%, balanced accuracy = 91.6%).

Cohen's *k* for intra- and inter-observer agreement for the same manual operator and between the two different operators were 0.79 and 0.60, respectively.

Full Pipeline: From Scanner to Report

The sensitivity, specificity and balanced accuracy of the integrated view selection and quality-controlled cardiac analysis pipelines was 96.3, 85.0, and 90.6%, respectively. Performance was also assessed for healthy and pathological cases separately. Results are summarized in **Table 5**.

The average time for selection and complete cine analysis from a full CMR study was between 4 and 7 min for a clinical CMR exam.

Supplementary Video 1 portrays how implementing the new framework prior to segmentation results in good quality analysis.

TABLE 5 | Full pipeline performance.

	Full pipeline		
	BACC	**SEN**	**SPE**
Healthy	92.3	96.5	88.1
Pathological	87.3	95.7	78.9
Global	90.6	96.3	85.0

SEN, Sensitivity; SPE, Specificity; and BACC, balanced accuracy of full pipeline, global, and for left (LV) and right ventricle (RV), and healthy and pathological hearts.

DISCUSSION

In this study we present a DL-based framework to identify all conventional cine views from a full CMR exam, and subsequently select one image per class of good quality for further automated image analysis. To the best of our knowledge, this is the first automated framework developed for this purpose.

The framework was trained on a large database, a prerequisite to develop DL tools of good quality. It was also trained on multivendor and clinically heterogeneous data, which makes it generalizable to be implemented as the first step for other existing tools for image analysis. Moreover, the framework was developed through training and testing of 7 state-of the art CNN architectures for each step. In DL, several network variants are available, each exhibiting different strengths and weaknesses. Studies often focus on a single highly individualized network, tailored for a task through multiple trial-and-error experiments. This makes reproduction of the methods and appreciation of its performance in the context of other datasets challenging. In our work, we present the data of all trained CNN architectures, thus displaying our selection process in a reproducible, fair, and meaningful way.

Finally, we integrated the new framework as the first step of a larger pipeline we had previously developed (8), and we demonstrated that it could produce highly accurate, rapid, and fully-automated cine analysis from a complete collection of images routinely acquired during a clinical study.

Class Identification CNN

Class identification is the first necessary step for image analysis, making algorithmic classification of standard views a fundamental step for true automatization of analysis (19). DL has been used to meet this need for automated analysis of echo images (20, 21), but not in the field of CMR.

Identification of conventional cine classes can be seen as a trivial task. Nonetheless it is time consuming, especially in long CMR studies, where a multitude of sequences are acquired. Moreover, view recognition cannot rely on the name of the sequences, as these are not replicated across groups and misnaming is common, especially when images are repeated due to insufficient quality or slight errors in view-planning, or added during acquisition. These characteristics make the

problem of class identification well-suited for automation. In computer vision tasks in particular, DL has shown excellent performance (22). This is reflected by the high performance of all trained CNN_{class} architectures, which had an accuracy approaching 100%. Precision, recall and F1-scores were mostly between 0.99 and 1 for all classes.

Quality Control CNN

The second DL component of our framework was trained to add a quality-control step to our framework by identifying images of insufficient quality or inadequate planning to inform the automated image analysis process.

Quality control is crucial to transfer DL research tools to the clinical reality in a safe manner, and its importance is increasingly recognized (5, 10, 23, 24). The performance of our CNN_{QC} was lower compared to the CNN_{class} with highest recorded accuracy of 0.86 for 2Ch and 4Ch, and 0.80 for 3Ch. This is explained by several reasons. First, the input data were highly unbalanced, which is a natural consequence of the fact that radiographers aim to acquire good quality images, resulting in the poor quality class being significantly underrepresented. This is reflected by the significantly lower precision, recall and f1-scores for the identification of "wrong" images compared to that of "correct" ones. To reduce the bias of unbalanced data, we used cross entropy loss, adaptive learning rate scheduler, and balanced accuracy, but such bias can never be fully controlled. Second, images to be considered of insufficient quality have a wide range of problems, from motion artifacts to off-axis planning of different types, making their grouping into one class difficult for the CNN. In particular, when evaluating 3Ch views, the quality of both the cardiac chambers and the aorta were considered, which might explain the lower performance compared to 2Ch and 4Ch views. On the other hand, separation into different classes would have resulted in further imbalance of the data, with insufficient numbers in each hypothetical poor-quality class. Therefore, we decided to group them together. Last, there is a degree of subjectivity in this task, as the same problem can be present to a varying degree of severity; for example, off-axis planning of the 4Ch view can result in the presence of a clear LVOT, or just a small disruption of the basal septum. The subjectivity of QC is reflected in the intra- and inter-observer variability during manual assessment. Consequently, it is both unlikely and unnecessary for any algorithm to reach 100% accuracy in this task. The cases with low degree of severity were the most likely to be misclassified, as well as the most frequent source of inter- and intra-observer disagreement, as displayed in **Figure 2**.

Complete Framework

For the complete framework, we selected the CNN_{class} and CNN_{QC} that performed best in combination with a selection algorithm, rather than the networks that performed best in the validation of the individual steps. This was done for two reasons. First, a sequential process can leverage individual strengths and weaknesses to obtain the best combined result. Second, the addition of the selection algorithm after the two CNNs aimed at achieving a more complex and possibly more clinically relevant task: selection of images for further analysis, which can be highly time-consuming, especially in long exams containing several acquisitions. It was therefore our intent to test the integrated framework and select the one with the best overall performance.

The integration of the three steps of the framework yielded an accurate and rapid system to select images of interest for analysis. The best combination was DenseNet CNN_{class}, ShuffleNet 2Ch-CNN_{QC}, DenseNet 2Ch-CNN_{QC}, ShuffleNet 4Ch-CNN_{QC}, which had a 90% sensitivity for 2Ch, 93% for 3Ch, and 94% for 4Ch acquisitions. This is achieved at the cost of a small proportion of good quality images being mistakenly labeled as erroneous, thus requiring clinician review. However, we believe this is a reasonable compromise to ensure clinical safety within an automated process. Moreover, the process of review is fast in case of data falsely labeled as erroneous, which only requires a visual check from the clinician to accept the analysis results.

Full Pipeline: From Scanner to Report

Using the image-processing steps developed in this paper, we were able to present the first pipeline for analysis of cardiac function from cine CMR that automates the complete process from scanner to report, offering an automated system that reproduces manual analysis in current clinical practice. This pipeline is characterized by a high degree of QC (one step in the new framework, two steps in the previously published one). Sequential QC steps focusing on different quality problems ensures a "Swiss cheese" framework, where if a poor-quality image slips through a first barrier, it will likely be flagged up in a later stage. **Supplementary Video 1** displays how the new quality control step aids in selecting images where segmentation can be performed at high standards by subsequent pipeline steps.

Moreover, the addition of the new framework offers automated selection of all standard cine views, which can be further exploited for analysis of parameters beyond conventional ones, such as longitudinal strain and atrioventricular valve systolic excursion, expanding the role of CMR for the assessment of systolic and diastolic function.

The full pipeline is highly accurate, with a focus on high sensitivity, showing an improvement compared to our previously published work (5). The pipeline is also significantly time-efficient, producing outcome measures in about 4 min for standard scans, and in up to 7 min for longer research scans. This is faster than the time reported for the initial pipeline (5), due to changes in the previously developed code.

In the future, we aim to extend our framework to identify and analyze other CMR sequences, including late gadolinium enhancement, flow and T2 mapping. Moreover, with an expanding data set we will be able to train quality control CNNs to recognize specific types of quality and planning errors. In order to decrease subjectivity of this task, a collaborative initiative to build a consensus across a vast number of operators, similar to a recent one developed in the field of echo and AI (25), would be of great value. Lastly, our method now requires post-processing and is separated from the CMR scanner. In the future, effort should be made for direct implementation on the scanner console. In particular, the implementation of CNN_{QC} at the time of image acquisition would aid radiographers by

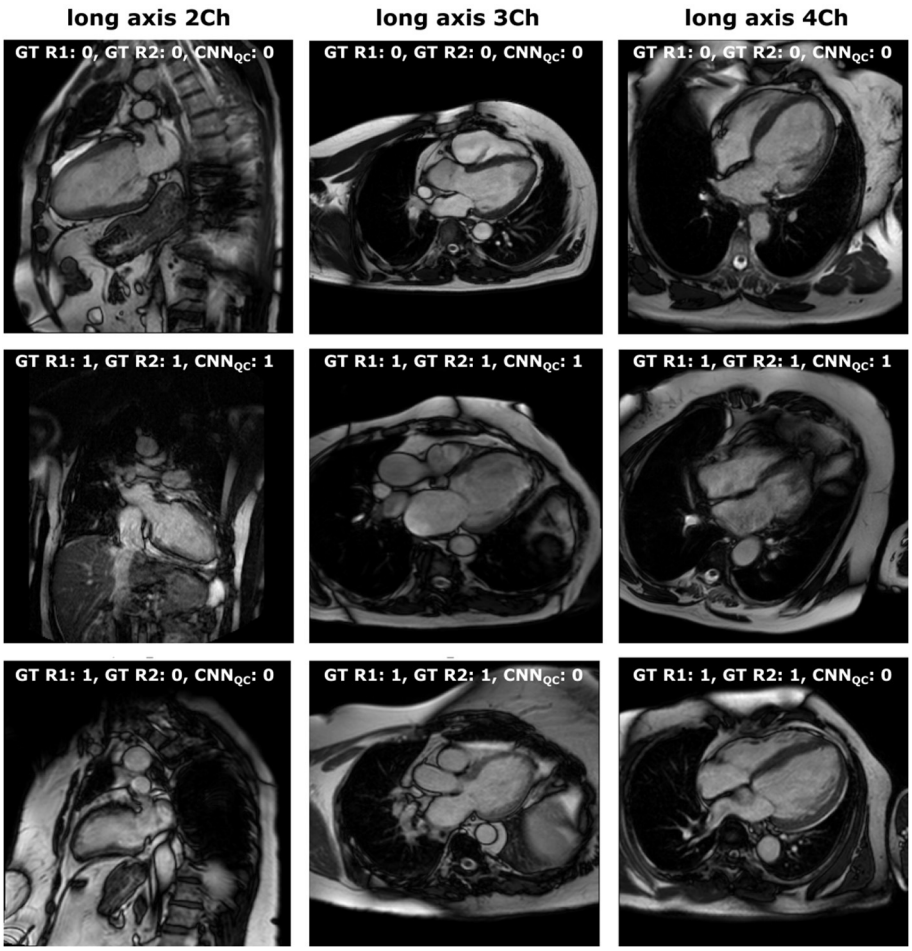


FIGURE 2 | Manual vs. automated classification. Visual representation of: (first line) cases classified as "correct" both by manual assessment (both operators) and CCN_{QC}; (second line) cases classified as "wrong" both by manual assessment (both operators) and CCN_{QC}; (third line) cases classified as "wrong" by manual assessment (with disagreement between operators for 2-chamber) and as "correct" by CCN_{QC}. 0, correct; 1, wrong; Ch, chamber; GT, ground truth; R1, first operator; R2, second operator.

promptly detecting images of unsatisfactory quality. This would improve image quality upstream and yield a greater accuracy of downstream image analysis (26).

Study Limitations

Our dataset did not include CMR studies acquired with General Electrics (GE), thus limiting our framework's generalizability. However, using Philips and Siemens granted a high degree of variability, which would facilitate further training with GE data.

DL algorithms inherently are black boxes. Therefore, interpretation of decisions remains challenging. In this paper we used a stepwise approach of classification and QC algorithms instead of a fused algorithm to allow at least some interpretation of the DL-based decisions.

Moreover, although we included a large number of patients with ACHD to train and test the model, exclusion of grossly distorted anatomy limits the use of this framework in patients with severe ACHD.

The framework presented in this paper performs limited quality control for short axis. We decided not to train a further CNN for this view as already present in the previously validated pipeline, and it would have therefore been redundant.

CONCLUSIONS

We developed and validated a framework to select cine acquisitions and perform QC of the selected images prior to automated cine CMR image analysis. We show that our network is able to select cine CMR from a full clinical CMR exam accurately and screen for image quality with a high rate of detecting erroneous acquisitions. We implemented our developed framework as the first step of a wider quality-controlled pipeline to obtain automated, quality-controlled

analysis of cardiac function from short and long axis cine images from complete CMR clinical studies.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: The UK Biobank data set is publicly available for approved research projects from https://www.ukbiobank.ac.uk/. The GSTFT data set cannot be made publicly available due to restricted access under hospital ethics and because informed consent from participants did not cover public deposition of data.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the North West Research Ethics Committee (REC 11/NW/0382; for UK Biobank data) and the South London Research Ethics Committee (REC 15/NS/0030; for GSTFT data) and was conducted according to the Declaration of Helsinki and Good Clinical Practice guidelines. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

VV, EP-A, and BR: conception, study design, and data analysis. EP-A: development of algorithms and analysis software. VV and EP-A: data pre-processing. VV, BR, and RR: clinical advice. VV and BR: manual image annotation. EP-A, BR, and RR: interpretation of data and results. VV, EP-A, BR, and RR: drafting and revising. All authors have read and approved the final manuscript.

FUNDING

This work was supported by the Wellcome/EPSRC Center for Medical Engineering at Kings College London (WT203148/Z/16/Z). EP-A was supported by the EPSRC (EP/R005516/1 and EP/P001009/1) and by core funding from the Wellcome/EPSRC Center for Medical Engineering (WT203148/Z/16/Z) and BR was supported by the NIHR Cardiovascular MedTech Co-operative award to the Guy's and St. Thomas' NHS Foundation Trust. This research was funded in whole, or in part, by the Wellcome Trust WT203148/Z/16/Z. For the purpose of open access, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission. The authors acknowledge financial support from the Department of Health through the National Institute for Health Research (NIHR) comprehensive Biomedical Research Center award to Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London.

ACKNOWLEDGMENTS

This research has been conducted using the UK Biobank Resource under Application Number 17806. The authors wish to thank all participants and staff from UK Biobank and GSTFT.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2021.742640/full#supplementary-material

Supplementary Video 1 | Full pipeline: from scanner to report. It displays the fully automated multi-step QC pipeline from scanner to image analysis and how implementing the new framework prior to segmentation results in good quality analysis.

REFERENCES

1. American College of Cardiology Foundation Task Force on Expert Consensus Documents, Hundley WG, Bluemke DA, Finn JP, Flamm SD, Fogel MA, et al. ACCF/ACR/AHA/NASCI/SCMR 2010 expert consensus document on cardiovascular magnetic resonance: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents. *J Am Coll Cardiol.* (2010) 55:2614–62. doi: 10.1016/j.jacc.2009.11.011
2. 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
3. Leiner T, Bogaert J, Friedrich MG, Mohiaddin R, Muthurangu V, Myerson S, et al. SCMR Position Paper (2020) on clinical indications for cardiovascular magnetic resonance. *J Cardiovasc Magn Reson.* (2020) 22:76. doi: 10.1186/s12968-020-00682-4
4. Bernard O, Lalande 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? *IEEE Trans Med Imaging.* (2018) 37:2514–25. doi: 10.1109/TMI.2018.2837502
5. 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. *JACC Cardiovasc Imaging.* (2020) 13:684–95. doi: 10.1016/j.jcmg.2019.05.030
6. Komi S, Inoue Y, Hata H, Nakajima A, Miyatake H. Cardiovascular magnetic resonance evaluation of left ventricular peak filling rate using steady-state free precession and phase contrast sequences. *Springerplus.* (2016) 5:1163. doi: 10.1186/s40064-016-2878-x
7. Mendoza D, Codella N, Wang Y, Prince MR, Sethi S, Manoushiagian SJ, et al. Impact of diastolic dysfunction severity on global left ventricular volumetric filling - assessment by automated segmentation of routine cine cardiovascular magnetic resonance. *J Cardiovasc Magn Reson.* (2010) 12:46. doi: 10.1186/1532-429X-12-S1-P86
8. Seemann F, Pahlm U, Steding-Ehrenborg K, Ostenfeld E, Erlinge D. Dubois-Rande JC. Time-resolved tracking of the atrioventricular plane displacement in Cardiovascular Magnetic Resonance (CMR) images. *BMC Med Imaging.* (2017) 17:19. doi: 10.1186/s12880-017-0189-5
9. Scannell CM, Veta M, Villa ADM, Sammut EC, Lee J, Breeuwer M, et al. Deep-learning-based preprocessing for quantitative myocardial perfusion MRI. *J Magn Reson Imaging.* (2020) 51:1689–96. doi: 10.1002/jmri.26983
10. 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
11. Puyol-Antón E, Ruijsink B, Baumgartner CF, Masci PG, Sinclair M, Konukoglu E, et al. Automated quantification of myocardial tissue characteristics from native T1 mapping using neural networks with

uncertainty-based quality-control. *J Cardiovasc Magn Reson.* (2020) 22:60. doi: 10.1186/s12968-020-00650-y
12. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst.* (2012) 25:1097–105.
13. Huang G, Liu Z, Weinberger KQ. Densely connected convolutional networks. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI (2017). p. 4700–8.
14. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. *ArXiv*. (2017) 1704:04861. Available online at: https://arxiv.org/abs/1704.04861
15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *ArXiv*. (2015) 1512:03385. doi: 10.1109/CVPR.2016.90
16. Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT (2018). p. 6848–56. doi: 10.1109/CVPR.2018.00716
17. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *ArXiv*. (2016) 1602:07360. Available online at: https://arxiv.org/abs/1602.07360
18. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv*. (2014) 409:1556. Available online at: https://arxiv.org/abs/1409.1556
19. Siegersma KR, Leiner T, Chew DP, Appelman Y, Hofstra L. Artificial intelligence in cardiovascular imaging: state of the art and implications for the imaging cardiologist. *Neth Heart J.* (2019) 27:403–13. doi: 10.1007/s12471-019-01311-1
20. Khamis H, Zurakhov G, Azar V, Raz A, Friedman Z, Adam D. Automatic apical view classification of echocardiograms using a discriminative learning dictionary. *Med Image Anal.* (2017) 36:15–21. doi: 10.1016/j.media.2016.10.007
21. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit Med.* (2018) 1:1–8. doi: 10.1038/s41746-017-0013-1
22. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: a brief review. *Comput Intell Neurosci.* (2018) 2018:7068349. doi: 10.1155/2018/7068349
23. 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
24. 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
25. Howard JP, Stowell CC, Cole GD, Ananthan K, Demetrescu CD, Pearce K, et al. Automated left ventricular dimension assessment using artificial intelligence developed and validated by a UK-wide collaborative. *Circ Cardiovasc Imaging.* (2021) 14:951. doi: 10.1161/CIRCIMAGING.120.011951
26. Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med.* (2020) 7:25. doi: 10.3389/fcvm.2020.00025

Author Disclaimer: The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, EPSRC, or the Department of Health.

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 22 November 2021
doi: 10.3389/fcvm.2021.755968

Deep Learning Predicts Heart Failure With Preserved, Mid-Range, and Reduced Left Ventricular Ejection Fraction From Patient Clinical Profiles

Mohanad Alkhodari[1], Herbert F. Jelinek[1,2], Angelos Karlas[3,4,5,6], Stergios Soulaidopoulos[7], Petros Arsenos[7], Ioannis Doundoulakis[7], Konstantinos A. Gatzoulis[7], Konstantinos Tsioufis[7], Leontios J. Hadjileontiadis[1,8,9] and Ahsan H. Khandoker[1]**

[1] Department of Biomedical Engineering, Healthcare Engineering Innovation Center (HEIC), Khalifa University, Abu Dhabi, United Arab Emirates, [2] Department of Biomedical Engineering, Biotechnology Center (BTC), Khalifa University, Abu Dhabi, United Arab Emirates, [3] Chair of Biological Imaging, Center for Translational Cancer Research (TranslaTUM), Technical University of Munich, Munich, Germany, [4] Institute of Biological and Medical Imaging, Helmholtz Zentrum München, Neuherberg, Germany, [5] Department for Vascular and Endovascular Surgery, Rechts der Isar University Hospital, Technical University of Munich, Munich, Germany, [6] DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany, [7] First Cardiology Department, School of Medicine, "Hippokration" General Hospital, National and Kapodistrian University of Athens, Athens, Greece, [8] Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates, [9] Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

OPEN ACCESS

Edited by:
Pablo Lamata,
King's College London,
United Kingdom

Reviewed by:
Bochra Tourki,
Morsani College of Medicine, USF Health, United States
Alexander E. Berezin,
Zaporizhia State Medical University, Ukraine

***Correspondence:**
Mohanad Alkhodari
mohanad.alkhodari@ku.ac.ae
Ahsan H. Khandoker
ahsan.khandoker@ku.ac.ae

Specialty section:
This article was submitted to Heart Failure and Transplantation, a section of the journal Frontiers in Cardiovascular Medicine

Received: 09 August 2021
Accepted: 19 October 2021
Published: 22 November 2021

Citation:
Alkhodari M, Jelinek HF, Karlas A, Soulaidopoulos S, Arsenos P, Doundoulakis I, Gatzoulis KA, Tsioufis K, Hadjileontiadis LJ and Khandoker AH (2021) Deep Learning Predicts Heart Failure With Preserved, Mid-Range, and Reduced Left Ventricular Ejection Fraction From Patient Clinical Profiles. Front. Cardiovasc. Med. 8:755968. doi: 10.3389/fcvm.2021.755968

Background: Left ventricular ejection fraction (LVEF) is the gold standard for evaluating heart failure (HF) in coronary artery disease (CAD) patients. It is an essential metric in categorizing HF patients as preserved (HFpEF), mid-range (HFmEF), and reduced (HFrEF) ejection fraction but differs, depending on whether the ASE/EACVI or ESC guidelines are used to classify HF.

Objectives: We sought to investigate the effectiveness of using deep learning as an automated tool to predict LVEF from patient clinical profiles using regression and classification trained models. We further investigate the effect of utilizing other LVEF-based thresholds to examine the discrimination ability of deep learning between HF categories grouped with narrower ranges.

Methods: Data from 303 CAD patients were obtained from American and Greek patient databases and categorized based on the American Society of Echocardiography and the European Association of Cardiovascular Imaging (ASE/EACVI) guidelines into HFpEF (EF $>$ 55%), HFmEF (50% $\leq$ EF $\leq$ 55%), and HFrEF (EF $<$ 50%). Clinical profiles included 13 demographical and clinical markers grouped as cardiovascular risk factors, medication, and history. The most significant and important markers were determined using linear regression fitting and Chi-squared test combined with a novel dimensionality reduction algorithm based on arc radial visualization (ArcViz). Two deep learning-based models were then developed and trained using convolutional neural networks (CNN) to estimate LVEF levels from the clinical information and for classification into one of three LVEF-based HF categories.

Results: A total of seven clinical markers were found important for discriminating between the three HF categories. Using statistical analysis, diabetes, diuretics medication, and prior myocardial infarction were found statistically significant ($p < 0.001$). Furthermore, age, body mass index (BMI), anti-arrhythmics medication, and previous ventricular tachycardia were found important after projections on the ArcViz convex hull with an average nearest centroid (NC) accuracy of 94%. The regression model estimated LVEF levels successfully with an overall accuracy of 90%, average root mean square error (RMSE) of 4.13, and correlation coefficient of 0.85. A significant improvement was then obtained with the classification model, which predicted HF categories with an accuracy ≥93%, sensitivity ≥89%, 1-specificity < 5%, and average area under the receiver operating characteristics curve (AUROC) of 0.98.

Conclusions: Our study suggests the potential of implementing deep learning-based models clinically to ensure faster, yet accurate, automatic prediction of HF based on the ASE/EACVI LVEF guidelines with only clinical profiles and corresponding information as input to the models. Invasive, expensive, and time-consuming clinical testing could thus be avoided, enabling reduced stress in patients and simpler triage for further intervention.

Keywords: heart failure, coronary artery disease, left ventricular ejection fraction, clinical profiles, demographical and clinical information, radial visualization, machine and deep learning

INTRODUCTION

Heart failure (HF) is a chronic and progressive pathologic state characterized by the inability of the heart to pump an adequate amount of blood to supply tissues with nutrients via the systemic circulation (1). Several conditions, such as coronary artery disease (CAD) and arterial hypertension, are considered major causes of HF progression (2, 3). According to the European Society of Cardiology (ESC), more than 26 million people around the world suffer from HF caused by CAD (4). Furthermore, the World Health Organization (WHO) estimates that HF accounts for more than 7.2 million deaths annually worldwide (3).

The systolic function of the heart, as indicated by the left ventricular ejection fraction (LVEF), is significantly decreased in HF. LVEF refers to the amount (%) of oxygenated blood pumped out of the left ventricle at each contraction of the heart (5, 6). It is considered an important diagnostic metric in evaluating the progression of HF, especially at early stages. Based on the LVEF, HF can be classified according to the American Society of Echocardiography and the European Association of Cardiovascular Imaging (ASE/EACVI) (7–9) into three main categories: heart failure with preserved ejection fraction (HFpEF) with an EF above 55%, heart failure with mid-range ejection fraction (HFmEF) with an EF between 50 and 55%, and heart failure with reduced ejection fraction (HFrEF) with an EF below 50%. The narrower range for the HFmEF category is considered as a variable criteria for this group in accordance to the etiology of HF. Other guidelines including the ESC (10) recommend different cut-off values for classification of HF, with a cut-off for HFrEF as low as 40%. The literature suggests that there are no strict rules and that the treatment is loosely associated with LVEF and clinical presentation. However, patients in the mid-range group between 40 and 49% based on the ESC guidelines show that 90% of patients either improved or deteriorated, whilst only 10% of cases remained unchanged (11).

Accurate LVEF-based assessment of HF therefore poses substantial challenges to clinicians (8, 9, 12). HFpEF, despite covering half of all patients with HF, is not yet well-understood and remains frequently undetected due to similarities in symptoms and adverse outcomes with HFrEF and, to a lesser extent, HFmEF (12, 13). Furthermore, HFmEF represents one-fifth of the HF population and remains ambiguous, as its pathogenesis was observed to be more similar to that of HFrEF and rather different from HFpEF depending on the guidelines applied. This raises the question of whether it should be considered a transient entity between HFpEF and HFrEF or a distinct entity on its own (14–16). Therefore, additional research is needed to investigate the effectiveness of LVEF-based categorization of HF patients. According to recently published studies, clinical profiles of patients allow for the discrimination between the three HF categories, especially the presence of comorbidities and quality of life based on the ESC guidelines (16–21). Based on these clinical results, HFmEF patients were found to fall between HFpEF and HFrEF while more closely resembling HFpEF (22, 23). Additionally, they were more likely to be younger than HFpEF and more prone to diabetes and hypertension than HFrEF (20, 24). Thus, further studies on a larger cohort of patients are still required to understand how demographical and clinical characteristics are associated with each HF category defined by clinically measured ejection fraction, especially in terms of optimizing treatment options to improve stratification and risk management of patients.

Most recently, machine learning has been widely implemented in medical research to assist in HF assessment through clinical

information (25–30). In addition, several studies have employed machine learning, including unsupervised clustering, to identify and characterize sub-groups of HFpEF from patient clinical profiles (31–34). However, there is still a limited knowledge on the complex relationship between demographical and clinical information and the three LVEF-based HF categories. In addition, it would be highly appreciated to offer a promising alternative tool to echocardiography for LVEF assessment which does not require the highly specialized knowledge and expensive equipment. In this vein, machine learning, including deep learning, can be essential in understanding the complicated clinical characteristics included in patient records leading to a better HF assessment. Therefore, in this study, we sought to investigate the ability of deep learning-based trained models in estimating LVEF levels as well as predicting HF categories from patient demographic and clinical information only in line with the ASE/EACVI guidelines. No previous studies have employed deep learning for analyzing HF categories associated with clinical profiles and LVEF. Thus, we developed trained models that could be capable of automatically providing assistance in clinical decision making in HF assessment based on LVEF levels. To prevent training the models using arbitrary or biased clinical variables, we ensured the following two steps: first, we investigated the statistical significance of each variable in discriminating between the three categories, and second, we followed a novel dimensionality reduction technique based on radial visualization to observe the best variables in characterizing and separating each LVEF-based HF category. We report the performance of the developed models that were trained based on the most important clinical variables to discuss the importance of deep learning in HF analysis based on LVEF as well as to elaborate on the significance of these clinical variables within patient profiles in differentiating between the three HF categories.

MATERIALS AND METHODS

Dataset and Patients Enrollment

Two datasets that contain clinical information of American and Greek patient cohorts were included in this study. Both datasets included patients with HF, more specifically CAD, with ages between 33 and 88 years ($n = 303$). These patients were divided into 129 HFpEF, 92 HFmEF, and 82 HFrEF according to the ASE/EACVI guidelines.

The American patient cohort was obtained from the archives of the Intercity Digital Electrocardiography (ECG) Alliance (IDEAL) study of the University of Rochester Medical Center Telemetric and Holter ECG Warehouse (THEW) (35). The database enrollment protocol was conducted according to Title 45, U.S. Code of Federal Regulations, Part 46, protection of human subjects (revised: November 13, 2001–effective: December 13, 2001) and in accordance with the Declaration of Helsinki. Furthermore, the research subject review board of the University of Rochester approved the IDEAL protocol (36). All patients provided a signed consent before participating in the study. The eligibility criteria to enroll in the IDEAL study included: (1) having either an evidence of previous MI or an exercise induced ischemia; (2) being in stable phase of ischemic heart disease at least 2 months after the last event; (3) not diagnosed with a congenital heart failure; and (4) being in sinus rhythm. Furthermore, all patients with dilated cardiomyopathy (left ventricular diameter (LVD) > 60 mm and EF $< 40\%$), congenital heart failure (CHF), coronary artery bypass grafting (CABG) surgery, non-sinus rhythm, and any cerebral, severe hepatic, or malignancy diseases were excluded from the study. A total of 199 patients were included from the IDEAL study. Out of these patients, HFpEF ($n = 106$), HFmEF ($n = 46$), and HFrEF ($n = 47$) categories were grouped based on the aforementioned ASE/EACVI guidelines.

The Greek patient cohort was obtained from the PRESERVE EF study with patients enrolled across seven cardiology departments in Greece (37). The protocol of the study was approved by the ethics committee at each cardiology department and was endorsed by the Hellenic Society of Cardiology. A database was created and is maintained by the Hellenic Society of Cardiology (38). All patients signed a consent form prior to enrollment in the study at each cardiology department. The eligibility criteria for patient enrollment included: (1) having a post-angiographically proven MI of at least 40 days after the event or 90 days after any CABG surgeries, if applicable; (2) being revascularized; (3) being not revascularized but without evidence of any active ischemia in previous the 6 months; and (4) following optimal and tolerated medical therapy. Furthermore, any patient with a secondary prevention indication for implantable cardioverter defibrillator (ICD) implantation, permanent pacemaker, persistent, long-standing persistent, and permanent atrial fibrillation, any neurological symptoms of syncope or pre-syncope within the last 6 months, and presence of any systemic illnesses such as liver failure, renal diseases, rheumatic diseases, thyroid dysfunction, and cancer was excluded from the study. Overall, a total of 104 patients were obtained from the PRESERVE EF study. These patients were distributed as 23 HFpEF, 46 HFmEF, and 35 HFrEF based on the ASE/EACVI guidelines.

Demographic and Clinical Markers

Both datasets included demographic and clinical information. Provided information was initially grouped into cardiovascular risk factors, cardiovascular medication, and cardiovascular history. As cardiovascular risk factors, age (years), sex (male—female), body mass index (BMI, kg/m^2), smoking (yes—no), diabetes (yes–no), and hypertension (yes—no) were the recorded markers. As cardiovascular medication, beta-blockers (yes—no), angiotensin-converting enzyme inhibitors (ACE-inhibitors, yes—no), anti-arrhythmics (yes—no), and diuretics (yes—no) were selected. Lastly, cardiovascular history included the presence of any previous angina pectoris (AP, yes—no), ventricular tachycardia (VT, yes—no), and myocardial infractions (Prior MI, yes—no).

Statistical Data Analysis

The statistical analysis was carried out using Student's t-test based on linear regression fitting (39), where the significance of each variable was evaluated based on the corresponding p-value measurement, with a p-value below 0.05 indicating significance.

A chi-squared (χ^2) test (40) was applied to examine which clinical variables were most important and highly dependent on individual LVEF categories. In this test, an important feature with a high score indicates a significant difference ($p < 0.05$) in discriminating between the three LVEF categories.

Multivariate Data Visualization

Patient information, being high-dimensional data, requires further projections into a low-dimensional space (dimensionality reduction) for multivariate visual inspection, and for clustering and pattern recognition purposes. A modified version of the original radial visualization (RadViz) algorithm (41, 42) proposed by Van Long (43), based on arc representation of variables (ArcViz) rather than point or line representations, was utilized. In ArcViz, a non-linear mapping into a two-dimensional plane is performed on the high-dimensional data (clinical information) by considering variables as arcs. Each multi-dimensional data point that includes clinical information of each patient is mapped as a point inside a circular convex hull. The arcs of this circle represent each variable, and new dimensional anchors (points) are calculated between these arcs to determine the location of the mapped point as well as the covering area of each arc. All points are normalized on the axes between the center [(x, y) = (0, 0)] and each calculated anchor point that is located on the arcs. The projection of the clinical variables was then optimized using a genetic algorithm based on linear discriminant analysis (LDA) fitting and the nearest centroid (NC) accuracy of the fitting was calculated accordingly.

Three properties are associated with the mapping process in ArcViz: (1) the larger the value of a variable inside the multi-dimensional data point, the closer the mapped point will be toward the anchor point located on the arc representing this variable; (2) the mapped point gets closer to the center if its data point values across the variables are similar; and (3) the mapped point is determined from a combination of anchor points calculated on the arcs and mapped within their convex hull.

Deep Learning Models

To provide a complete prediction approach (**Figure 1**), two deep learning-based models for regression (level estimations) and classification (category labels) of LVEF were developed. The input of these models was 303 patient clinical information (**Figure 1A**) including the previously mentioned demographic and clinical markers (**Figure 1B**). The results of the statistical analysis as well as dimensionality reduction based on ArcViz were used as feature selection approaches to assist in determining the most important markers for a maximized performance within the proposed deep learning models (**Figure 1C**). Both models for regression and classification were structured as a deep learning network (**Figure 1D**) with convolutional neural networks (CNN). Two convolutional layers were utilized, each followed by batch normalization (BN) and rectified linear unite (ReLu), to extract characteristics contaminated within patients' clinical markers of every LVEF category. The two consecutive convolutional layers were with kernel sizes of (1, 3) and (1, 2), respectively, and with 32 filters and 64 filters, respectively. The development of the models included training and prediction phases (**Figure 1E**). In the training phase, both models were trained for 300 epochs with a mini-batch size of 64. The adaptive moment estimation (ADAM) was selected as the optimizer with a learning rate of 0.001, L2-regularization of 0.0001, and decay rate of 0.90. For the prediction phase, a leave-one-out scheme, where each subject is held out as the testing subject on each training iteration, was adopted. This scheme provides a prediction for every subject in the dataset, while at the same time maximizing the amount of data included within the trained models. It allows for treating each patient as a completely hidden testing set to the trained models, thus, slightly addressing any issues on the generality in the training and testing phases due to the lack of any external patient testing sets.

The performance of the regression model was evaluated based on the overall accuracy level, which was calculated as the agreement between the estimated and original LVEF with an accepted error of ±5%. Furthermore, the average root mean square error (RMSE) and correlation coefficient, alongside the Bland-Altman (44) (with mean ± 2 std) and correlation plots of the estimation process were determined. To evaluate the performance of the classification model, analysis of the confusion matrix of predictions as well as the receiver operating characteristic (ROC) curves and the corresponding area under the ROC (AUROC) was applied. Additional performance evaluation metrics including accuracy, sensitivity, specificity, precision, and F1-score.

RESULTS

Clinical Characteristics of Patients

Patients included in this study had a median age of 58 years with an interquartile range of 50–65 years. Two hundred and fifty-eight patients were male (85.15%). Diabetes, diuretics medication, and prior MI showed significant differences in discriminating between the three LVEF categories ($p < 0.001$). Furthermore, for patients with diabetes, a significant difference was observed between HFpEF and HFrEF, whereas for diuretics medication and prior MI the significant differences were observed for HFpEF vs. HFrEF and HFpEF vs. HFmEF ($p < 0.001$). The complete clinical characteristics of the patient cohort is shown in **Table 1** alongside the *p*-value calculations using linear regression fitting. The three aforementioned markers had the highest normalized importance scores using the Chi-squared (χ^2) test as illustrated in **Figure 2** (diuretics: 1.0, Prior MI: 0.63, and diabetes: 0.32). Additionally, VT and AP had relatively high scores with 0.24 and 0.23, respectively, with the remaining clinical markers being below 0.1.

ArcViz Representations of Clinical Markers

The projection of clinical markers on the ArcViz convex hull (**Figure 3**) yielded an average NC accuracy of 93.73%, distributed as 99.01, 90.43, and 91.75% for cardiovascular risk factors, cardiovascular medication, and cardiovascular history categories, respectively. For cardiovascular risk factors (**Figure 3A**), the three LVEF categories were perfectly separated with a large arc area for diabetes. This indicates the strong impact of diabetes on

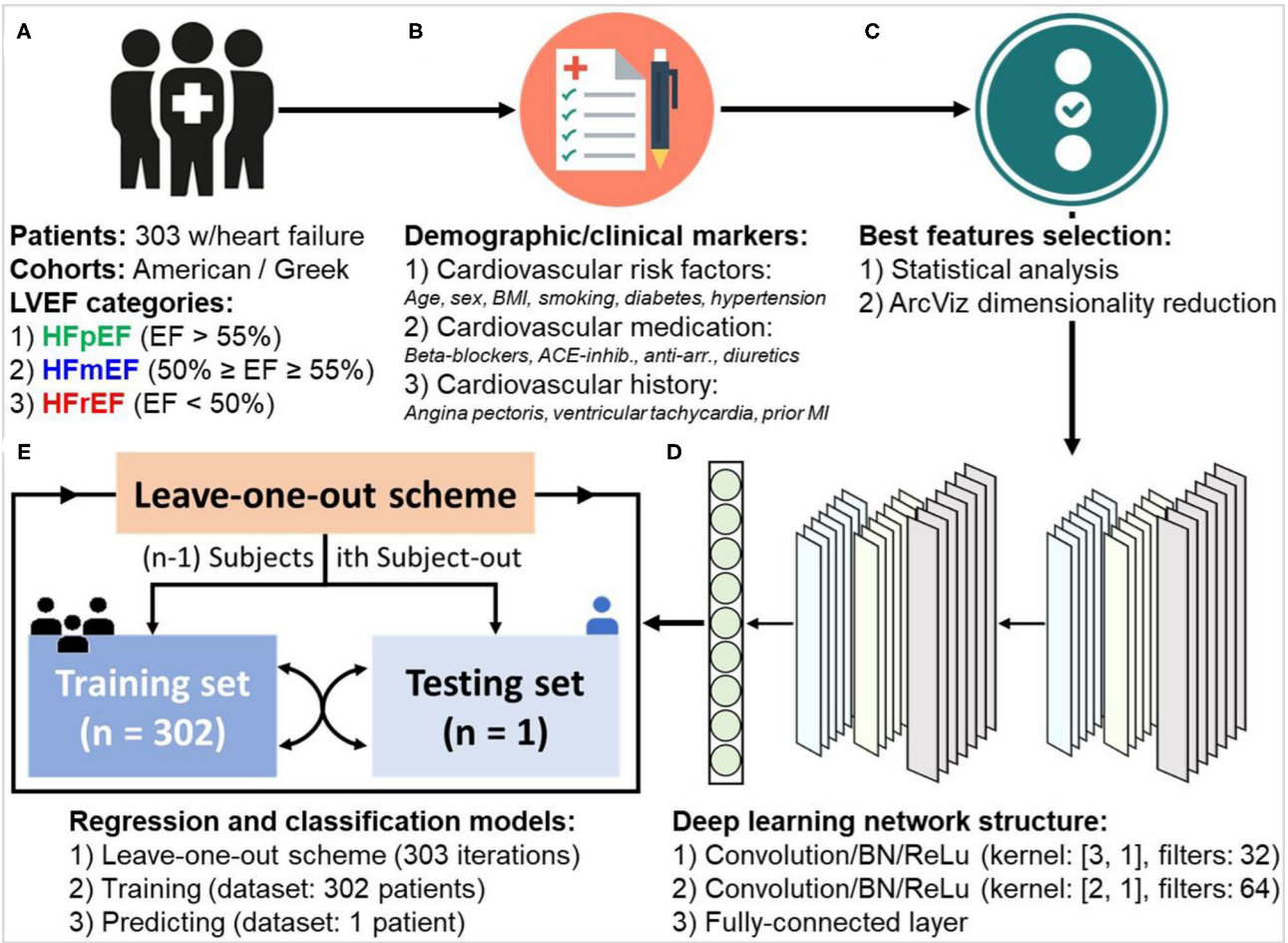


FIGURE 1 | Workflow of developing deep learning-based regression and classification models for the prediction of left ventricular ejection fraction (LVEF) levels and categories in heart failure patients. The procedure goes through: **(A)** dataset collection ($n = 303$ patients), **(B)** patient information and demographic/clinical markers categorization, **(C)** statistical analysis of markers and dimensionality reduction using ArcViz (feature selection), **(D)** designing deep learning network structure, and **(E)** development of regression and classification models with leave-one-out training and predicting scheme.

discriminating the three categories. Furthermore, although the centroids of HFpEF and HFmEF were located within diabetes, the centroid for HFrEF was located in the BMI region, which matches with the p-value observations of HFpEF vs. HFrEF in diabetes (**Table 1**—$p < 0.001$). It is worth noting that age had a greater effect on some HFpEF and therefore it was found to be significantly different for HFpEF vs. HFmEF as shown in **Table 1** ($p = 0.034$). For cardiovascular medication (**Figure 3B**), a fair separation was obtained between the three LVEF categories associated with anti-arrhythmics and diuretics medication use. Both centroids of HFmEF and HFrEF were located within the diuretics arc region, with $p < 0.001$ (**Table 1**) observed between HFmEF and HFrEF compared to HFpEF, which was located mostly within the anti-arrhythmics arc region. Lastly, for cardiovascular history (**Figure 3C**), the stronger impact was due to the prior MI marker that had the lowest ($p < 0.001$; **Table 1**) for differences between HFpEF and HFmEF as well as HFpEF and HFrEF. The centroid of the HFpEF was located within the VT arc region, whereas both centroids of HFmEF and HFrEF were located within the prior MI arc region. The slight shift of the HFrEF category toward the VT biomarker arc area is reflected by the low ($p = 0.005$; **Table 1**) when compared to patients in the HFmEF group.

Deep Learning Prediction of LVEF

Both deep learning models (regression and classification) were trained on the NVIDIA GeForce GTX 1070 graphics processing unit (GPU) of 8 GB display memory (VRAM). Training of each model required <1 min, while the prediction per-patient took <3 s. Both models were trained using the most important clinical markers (age, BMI, diabetes, anti-arrhythmics, diuretics, VT, and Prior MI) based on statistical significance and location of ArcViz centroids. The developed regression model (**Figure 4A**) successfully estimated patient LVEF levels with an overall accuracy of 90.43% (error: ±5%). Furthermore, the estimated LVEF levels had an average RMSE of 4.13 relative to the original LVEF levels. The Bland-Altman plot (**Figure 4B**) had a mean

TABLE 1 | Clinical characteristics of the heart failure patients based on their left ventricular ejection fraction categories.

Clinical variables	Overall subjects (n = 303)	LVEF categories			*p*-value			
		HFpEF (n = 129)	HFmEF (n = 92)	HFrEF (n = 82)	HFpEF HFrEF	HFpEF HFmEF	HFmEF HFrEF	HFpEF HFmEF HFrEF
LVEF, %	55 (46.5–63)	63 (60–70)	52.5 (50–55)	45 (40–47)	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Cardiovascular risk factors								
Age, years	58 (50–65)	57 (38–64.5)	58.5 (52–68)	60.5 (50–66)	0.155	**0.034**	0.533	0.110
Male	258 (85.15)	108 (83.72)	76 (82.61)	74 (90.24)	0.181	0.828	0.147	0.237
BMI, kg/m^2	27.28 (24.91–29.41)	27.12 (24.39–28.95)	27.22 (25.35–29.92)	27.68 (25.31–29.74)	0.159	0.166	0.945	0.133
Smoking	203 (67.00)	87 (67.44)	62 (67.39)	54 (65.85)	0.752	0.928	0.831	0.757
Diabetes	43 (14.19)	10 (7.75)	13 (14.13)	20 (24.39)	**<0.001**	0.127	0.086	**<0.001**
Hypertension	154 (50.83)	64 (49.61)	46 (50.00)	44 (53.66)	0.569	0.955	0.632	0.587
Cardiovascular medication								
Beta-Blockers	245 (80.86)	101 (78.30)	77 (83.70)	67 (81.71)	0.551	0.320	0.731	0.478
ACE-Inhibitors	113 (37.29)	47 (36.43)	33 (35.87)	33 (40.24)	0.555	0.932	0.580	0.611
Anti-Arrhythmics	12 (3.96)	3 (2.33)	4 (4.35)	5 (6.10)	0.164	0.400	0.605	0.166
Diuretics	114 (37.62)	24 (18.61)	51 (55.44)	39 (47.56)	**<0.001**	**<0.001**	0.302	**<0.001**
Cardiovascular history								
AP	186 (61.39)	89 (68.99)	46 (50.00)	51 (62.20)	0.311	**0.004**	0.107	0.189
VT	21 (6.93)	8 (6.20)	2 (2.17)	11 (13.42)	0.078	0.158	**0.005**	0.088
Prior MI	223 (73.60)	77 (59.69)	76 (82.61)	70 (85.37)	**<0.001**	**<0.001**	0.624	**<0.001**

All values are represented as median (interquartile range) or n (%). Bold p-values show statistically significant differences ($p < 0.050$) amongst the selected categories. LVEF, Left ventricular ejection fraction; HFpEF, Heart failure with preserved ejection fraction; HFmEF, Heart failure with mid-range ejection fraction; HFrEF, Heart failure with reduced ejection fraction; BMI, Body mass index; ACE, Angiotensin-converting enzyme; AP, Angina pectoris; VT, Ventricular tachycardia; MI, Myocardial infarction.

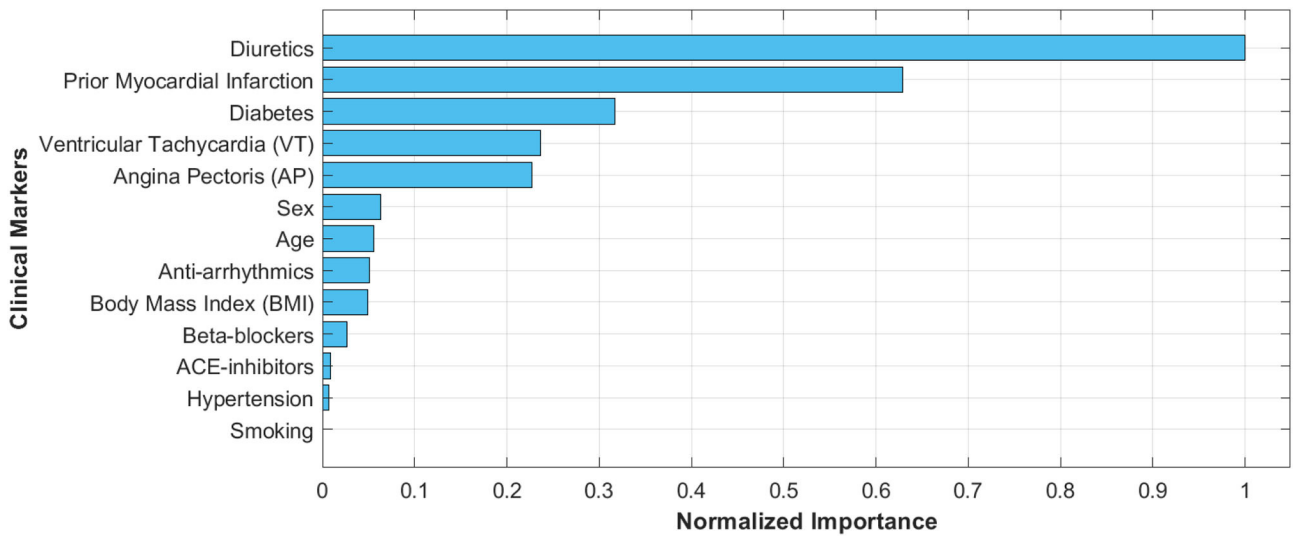


FIGURE 2 | Normalized importance scores for the clinical markers used in the study in differentiating between the three LVEF categories. Importance scores were calculated using the Chi-squared (χ^2) statistical test.

difference of 0.39 ± 11.61 between the estimated and original LVEF levels. Additionally, the correlation plot was skewed positively with an overall coefficient of 0.85 (**Figure 4C**). The classification model developed for this project (**Figure 5A**) efficiently predicted each LVEF category with a precision level of 93.00, 89.10, and 95.10% for HFpEF, HFmEF, and HFrEF, respectively.

The model resulted in an average AUROC of 0.975 (**Figure 5B**) distributed as 0.986 for HFpEF, 0.955 for HFmEF, and 0.983 for HFrEF. Furthermore, the model achieved high

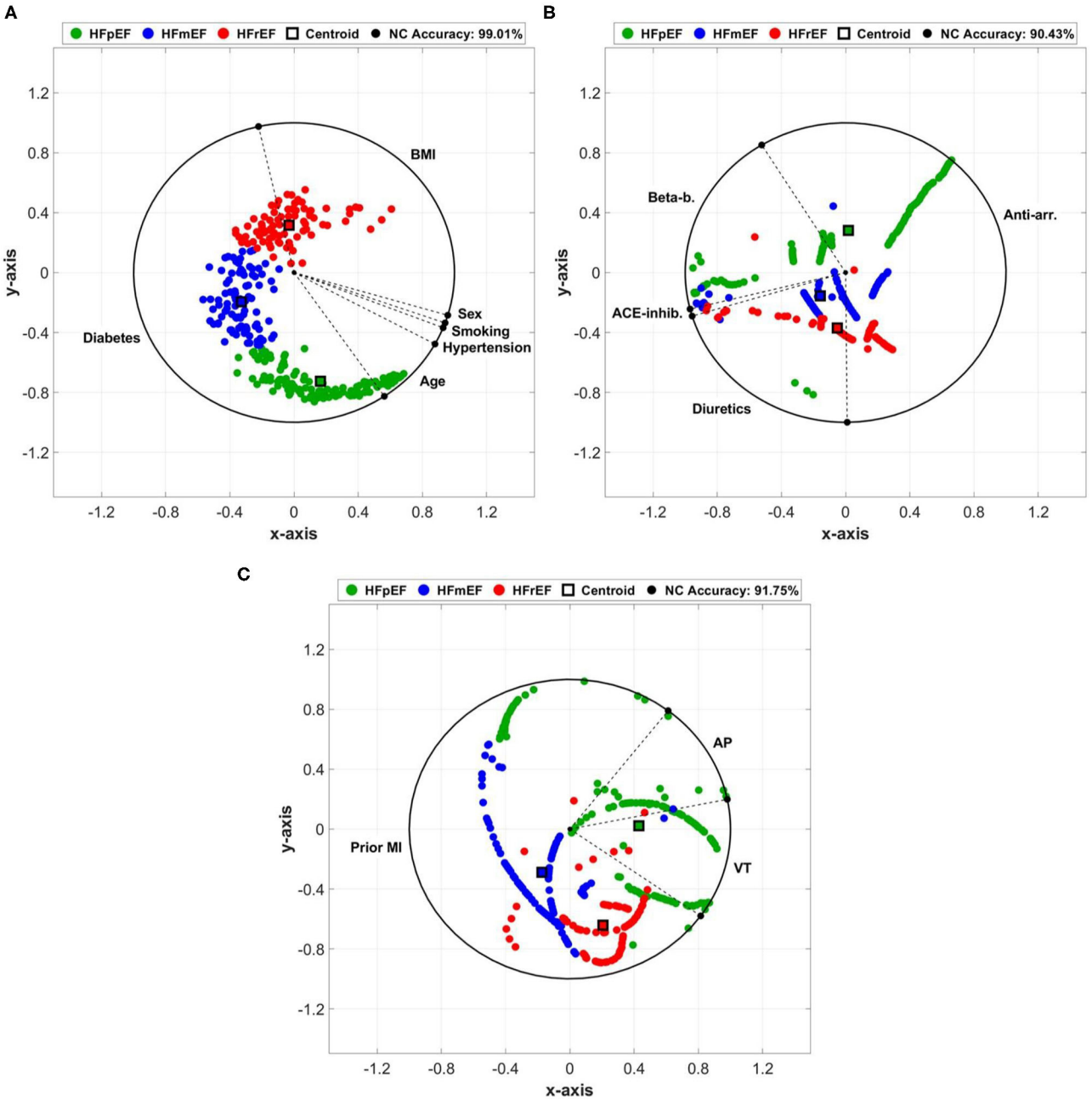


FIGURE 3 | Arc visualization (ArcViz) and dimensionality reduction for: **(A)** cardiovascular risk factors, **(B)** cardiovascular medication, and **(C)** cardiovascular history clinical markers. The nearest centroid (NC) accuracy was calculated after optimizing linear discriminant analysis (LDA) fitting.

levels of performance (**Figure 5C**) in accuracy, sensitivity, specificity, precision, and F1-score (more than 89%).

To elaborate on the generality and performance of the proposed deep learning models, support vector machines (SVM) based on a radial basis function (RBF) kernel and generalized linear model (GLM) were used to estimate LVEF (regression) and predict HF categories (classification). The performance of both models are compared with the aforementioned deep learning results in **Table 2**. LVEF estimation accuracies using SVM and GLM models have reached 87.46% and 84.82%, respectively, which was outperformed by deep learning (90.43%). In addition, the RMSE had its lowest levels for deep learning (4.13) compared with SVM (4.38) and GLM (5.11). In predicting HF categories, the overall accuracy reached 88.45 and 84.14% for SVM and GLM, respectively, whereas it reached 90.10% in deep learning. It is worth noting that both models had high precision levels in HFrEF prediction with a 94.12% using SVM and 96.34% using GLM. However, they both had lower performance metrics

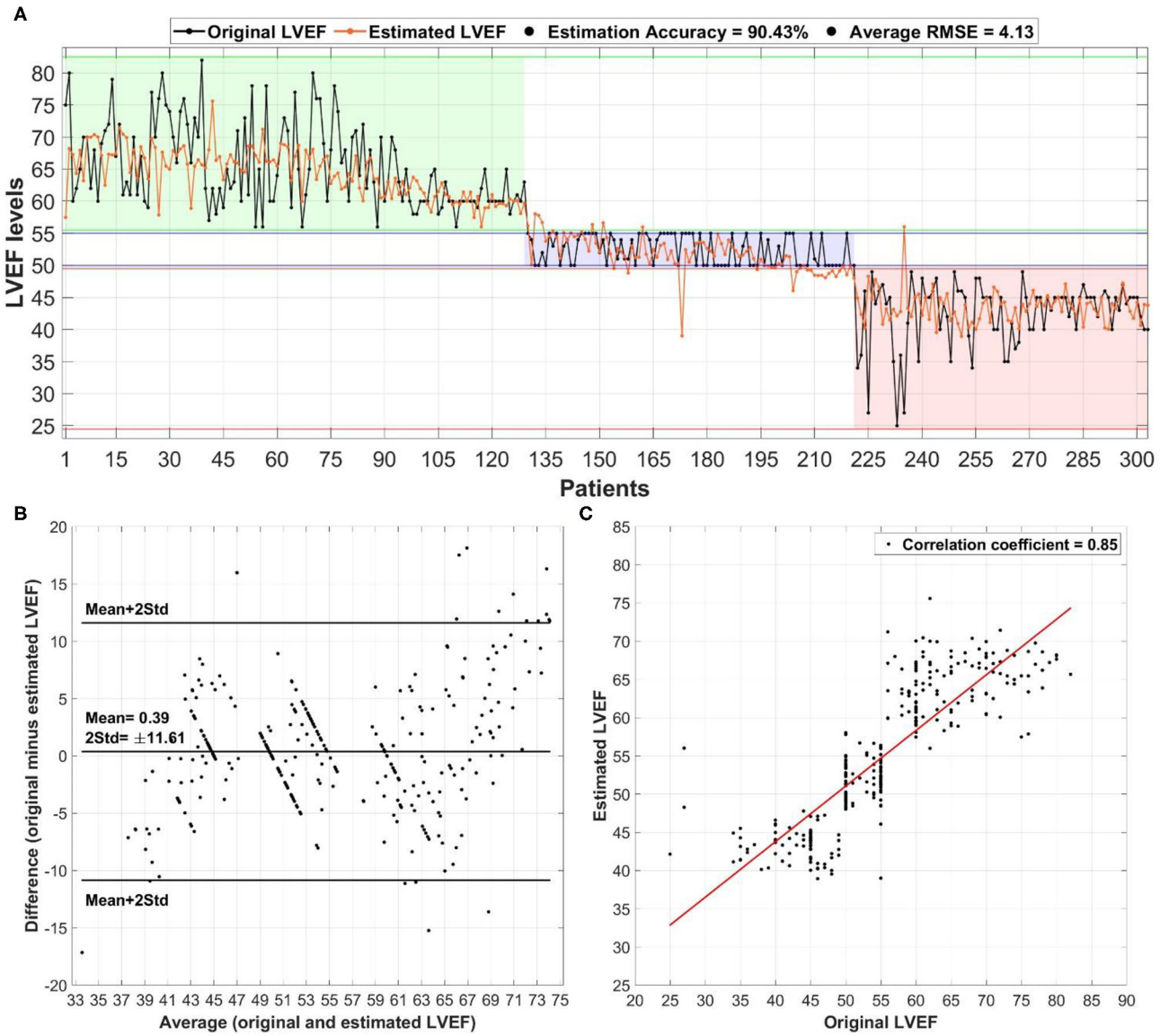


FIGURE 4 | The overall performance of the deep learning-based regression model in estimating left ventricular ejection fraction (LVEF) levels in heart failure patients: **(A)** estimation of LVEF relative to the original levels alongside the overall accuracy and average root mean square error (RMSE), **(B)** Bland-Altman plot for the average vs. difference between the estimated and original LVEF levels with the mean ± 2 std difference level, and **(C)** correlation plot between estimated and original LVEF levels with the corresponding correlation coefficient.

(sensitivity, 1-specificity, and precision) than deep learning in discriminating between HFpEF and HFmEF.

DISCUSSION

In this study, we demonstrated the significance of utilizing deep learning as a tool to estimate LVEF levels in HF patients as well as to categorize HF patients in accordance with their LVEF levels, offering an easily used and automated assistive tool for everyday clinical practice. The adopted narrower band for the HFmEF highlights that even slightly reduced values of LVEF can have an effect on heart rhythm and hence change in patient condition. Therefore, it was essential to employ versatile criteria for various cohorts in order to enable the adaptive analysis of the collected patient data. In addition, the ability to use automated deep learning-based trained models could save crucial time in clinical circumstances. In addition, these models may be able of aiding in the clinical decision making in HF assessment by going through available patient information with less dependence on medical experts. Only few studies have identified and discussed clinical information that may be capable of classifying HFpEF, HFmEF, and HFrEF patients statistically as well as from a machine learning-based perspective. To fill this gap, in the

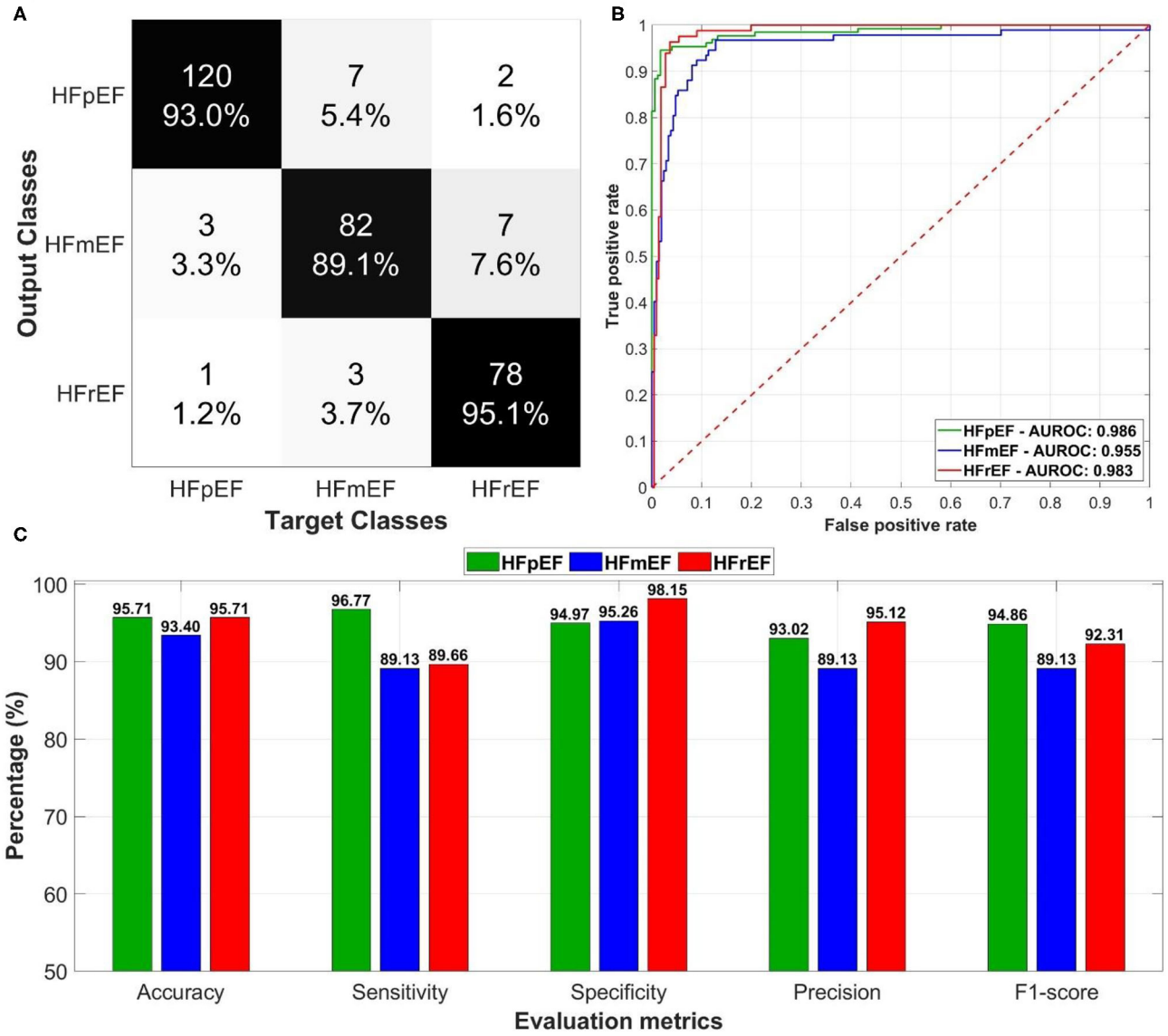


FIGURE 5 | The overall performance of the deep learning-based classification model in predicting left ventricular ejection fraction (LVEF) categories in heart failure patients: **(A)** confusion matrix of the predictions for the output and target classes, **(B)** receiver operating characteristics (ROC) curves for each LVEF category with the corresponding area under the ROC curves, and **(C)** evaluation metrics including accuracy, sensitivity, specificity, precision, and F1-score.

current study important clinical markers were first statistically identified and then projected into a novel arc radial visualization (ArcViz). Furthermore, a complete deep learning approach was developed that ensures higher levels of performance for automatic estimation of LVEF levels and differentiation between the three HF categories from clinical profiles only.

Clinical Markers Significance

Thirteen clinical markers often found in patient profiles were evaluated statistically as well as through a new dimensionality reduction approach (ArcViz). Among these markers, 7 were found to be important in classifying HF patients based on LVEF.

For cardiovascular risk factors, age was found to be an important marker in differentiating between HFpEF and HFmEF. However, HFpEF patients were more skewed toward the age region in ArcViz analysis in agreement with previous studies that have found that HFmEF patients were younger in age and closer to HFrEF in comparison to the HFpEF patients (24, 45, 46). Furthermore, BMI, although not significantly different, was better in differentiating HFrEF from the other two groups when applying ArcViz. Of interest and in agreement with our study, several previous studies (47, 48) suggested that higher BMI often associated with HFrEF patients (was beneficial to this patient group), as higher BMI may counteract catabolism inflammation and stress hormone activation in the HFrEF group. However, a

TABLE 2 | Performance comparison between deep learning support vector machines, and generalized linear model in estimating LVEF levels and predicting HF categories.

Trained models	Regression (LVEF estimations)				Classification (Categories prediction)											
	Accuracy %	RMSE	Bland-Altman Mean ± 2 Std	Correlation Coefficient	HFpEF			HFmEF			HFrEF			Overall		
					Sens. %	1-Spec. %	Prec. %	Sens. %	1-Spec. %	Prec. %	Sens. %	1-Spec. %	Prec. %	Acc. %
Deep learning	90.43	4.13	0.39 ± 11.61	0.85	96.77	5.03	93.02	89.13	4.74	89.13	89.66	1.85	95.12	90.10
Support vector machine	87.46	4.38	0.14 ± 11.89	0.82	94.78	10.64	84.50	81.00	5.42	88.04	88.64	1.76	94.12	88.45
Generalized linear Model	84.82	5.11	0.05 ± 14.04	0.75	88.14	13.51	80.62	76.09	10.43	76.09	84.95	1.43	96.34	84.14

LVEF, Left ventricular ejection fraction; HFpEF, Heart failure with preserved ejection fraction; HFmEF, Heart failure with mid-range ejection fraction; HFrEF, Heart failure with reduced ejection fraction; RMSE, Root mean square error; Sens, Sensitivity; Spec, Specificity; Prec, Precision; Acc, Accuracy.

high body weight in HFpEF patients is usually strongly associated with HF, causing this patient group to be at higher risk of developing further adverse cardiac events. Additionally, it was shown that BMI does not play a critical role in HF progression apart from the HFrEF category which has a higher 30-day mortality (49). In this study, the minimum-maximum range for BMI of HFpEF patients was 19.7–36.3 kg/m^2 with 21 obese patients (>30 kg/m^2). For HFmEF patients, the range was 18.0–37.7 kg/m^2 with 23 obese patients. For HFrEF patients, the range was 20.8–37.9 30 kg/m^2 with 17 obese patients. These ranges fit with the usual BMI range (20–40 kg/m^2) reflecting normal spread of BMI values across the included patients in the three LVEF categories. This could elaborate on the insignificance found using statistical analysis considering the narrow LVEF ranges in the three categories. Lastly, diabetes was found to be significantly different as well as the best in characterizing the three LVEF categories in ArcViz. Our findings show that the three LVEF categories can be better discriminated according to patients' diabetes diagnosis. All-cause mortality rates are reported to increase in diabetic HFrEF relative to HFpEF (50–52). However, patients within HFpEF and HFmEF groups showed higher burden of diabetes than HFrEF patients (53).

Important markers associated with cardiovascular medication included anti-arrhythmics and diuretics medication. Patients with HFpEF had a higher intake of both medications followed by HFrEF and lastly HFmEF. This shows a distinct medication procedure between the three LVEF categories using these two medications. The literature reports that HFpEF patients are more prone to atrial fibrillation, and thus, anti-arrhythmics medication is usually needed. Further, they were more likely to undergo repeated ablations compared to the HFrEF group (54, 55). This elaborates on the high number of patients taking anti-arrhythmics medication observed in this study for the HFpEF category with a better representation between HFpEF and HFrEF in the ArcViz analysis. In addition, use of diuretics medication was found to be highly discriminant between HFpEF and the other two LVEF categories. Previous studies reported that the prevalence of diuretics intake among HFmEF patients was found to be less than the prevalence in the HFpEF and HFrEF in agreement with the current study (45). Furthermore, diuretics are widely used in HFpEF and HFrEF patients to prevent symptoms of congestion in HF (56). This information supports the findings of this study by considering diuretics as a highly favored clinical marker in classifying patients into one of the three LVEF categories.

Lastly, in cardiovascular history, the best marker was the occurrence of prior MI. In a few previous studies (14, 57), a greater number of prior MI was observed in HFrEF compared to HFpEF. However, this could be due to the drop of LVEF levels in the HFrEF patients included in these studies, as a higher rate of prior MI is usually recorded if LVEF levels are <40% (58), which was found in very few cases in our study. In addition, VT was found to be the second most important cardiovascular history marker in characterizing the three LVEF categories using ArcViz analysis, especially the HFpEF, as well as being significant in discriminating HFrEF from the two other categories. This relates

to the higher burden of VT observed in patients with HFpEF over patients in the other LVEF categories (59, 60).

It is worth mentioning that sex (male/female) was not found to be significant nor effective in characterizing any LVEF category in ArcViz analysis. However, this could be due to the high number of male patients enrolled in this study and needs to be further investigated as sex has been shown to be a factor in the prevalence of HF (60–63).

Deep Learning as an Assistive Tool

Our study suggests deep learning as an assistive tool that could be capable of automatically reading and extracting characteristics from the clinical records of HF patients. In comparison with machine learning, our trained models allow for training on deeply extracted attributes between patients of each LVEF category. Thus, it was less biased than feature engineering techniques often used in conventional machine learning algorithms. Our novel deep learning models may assist clinicians based on the automated estimation of LVEF as well as the accurate classification into one of the three main HF categories (64). Furthermore, the models estimate and predict LVEF based on the cardiovascular risk factors, medication, and history. Additionally, the high levels of performance achieved in our deep learning models suggest the potential of relatively simple, yet effective, artificial intelligence algorithms in identifying certain clinical characteristics that differentiate between LVEF categories that may not be possible in conventional approaches. Although deep learning has outperformed other machine learning models in this work including SVM and GLM, further testing on external patient cohorts are still needed to elaborate further on the general validity of the achieved performance.

Limitations

Although our study shows that deep learning-based models have performed efficiently in LVEF predictions, it has a number of shortcomings. First, we have utilized 13 features (the 7 most important ones were selected later) that were available in the databases used in this study. However, additional markers need to be further investigated, especially echocardiographic attributes, i.e., left ventricular diastolic and systolic dimensions (LVDD and LVDS), to provide more information on their effects on LVEF predictions. Moreover, even though the dataset used in this study combined patients from American and Greek populations, the trained models should be tested further on wider sets of patients to ensure additional generality of the performance. Future studies should focus on using external validation sets from different patient cohorts to imply general validity of the trained models. In addition, the proposed models were trained and tested on a specific range for each LVEF category as recommended by the ASE/EACVI guidelines. Despite having a narrower border-line for the HFmEF ($50\% \leq EF \leq 55\%$), further studies of other LVEF guidelines and recommended LVEF category ranges may enhance the effectiveness of deep learning in LVEF predictions. Furthermore, validation on longitudinal data needs to be undertaken to identify efficacy of intervention over time based on the current models. Lastly, patients cohort in this study includes a much higher percentage of males compared to females. In addition, the median BMI of the included patients was 27.28 kg/m^2 indicating overweight but not obese subjects with a narrow BMI range between the three LVEF categories. Future studies with cohorts differing with reference to all demographic categories including BMI are needed in order to demonstrate the efficacy of the proposed methods in all possible populations/clinical scenarios with narrow or wide ranges for clinical information across LVEF categories.

Conclusions

Overall, our novel deep learning-based models showed high levels of performance in automatically estimating LVEF levels as well as classifying HF patients into one of the three LVEF categories, suggesting it as a promising assistive tool in clinical settings. The developed approach may lead to a better understanding, from a machine learning (or deep learning) perspective, of the clinical variables most suitable for discriminating HFpEF, HFmEF, and HFrEF. The proposed study is to extend the applicability of use of LVEF to communities where the required instruments are not available due to economic hardship or lack of clinical expertise. Future research can add additional demographic and clinical information to the deep learning models alongside clinical profiles for an even better performance and understanding of the differences between each LVEF category. Our outcomes may also facilitate the development of a model for the prediction of the HF phenotype or its changes during the followed therapy of HF, offering a versatile tool for the further exploration of disease pathophysiology or the objective assessment of the different therapeutic schemes in future patients with HF.

DATA AVAILABILITY STATEMENT

Patient data used in this study will be provided by contacting the corresponding author upon reasonable request. Deep learning networks, final trained models, and regression/classification main code (including data preparation) are available at: https://github.com/malkhodari/Alkhodari_frontiers. Requests to access these datasets should be directed to Mohanad.alkhodari@ku.ac.ae.

ETHICS STATEMENT

Patients enrollment protocol in the Intercity Digital Electrocardiography (ECG) Alliance (IDEAL) study was conducted according to Title 45, U.S. Code of Federal Regulations, Part 46, protection of human subjects (revised: November 13, 2001 – effective: December 13, 2001) and in accordance with the Declaration of Helsinki. Furthermore, the research subject review board of the University of Rochester approved the IDEAL protocol. All patients provided a consent form before participating in the study. Furthermore, the enrollment protocol of patients for the PRESERVE EF study was approved by the ethics committee at each of the seven selected cardiology departments at Greece and was endorsed by the Hellenic Society of Cardiology. All patients signed

a consent form prior to enrollment in the study at each cardiology department.

AUTHOR CONTRIBUTIONS

MA, HJ, LH, and AHK: designed research idea. MA: performed literature search, algorithm implementation, statistical analysis, and deep learning modeling. HJ and AK: advised on the categorization of clinical information in patient profiles. SS and AHK: clinical data acquisition and preparation. MA: wrote the initial draft of the manuscript. HJ, AK, SS, LH, and AHK: edited the final version of the manuscript. All authors reviewed and agreed on the manuscript and ensured that any questions on the work are appropriately resolved.

FUNDING

This work was supported by a grant (award number: 8474000132) from the Healthcare Engineering Innovation Center (HEIC) at Khalifa University, Abu Dhabi, UAE, and by grant (award number: 29934) from the Department of Education and Knowledge (ADEK), Abu Dhabi, UAE.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Sergey Sulima from Helmholtz Zentrum München, Institute of Biological and Medical Imaging, Neuherberg, Germany, for his English and scientific language editing of the final manuscript.

REFERENCES

1. Roger VL. Epidemiology of heart failure. *Cir Res*. (2013) 113:646–59. doi: 10.1161/CIRCRESAHA.113.300268
2. Velagaleti RS, Vasan RS. Heart failure in the twenty-first century: is it a coronary artery disease or hypertension problem? *Cardiol Clin*. (2007) 25:487–95. doi: 10.1016/j.ccl.2007.08.010
3. Mackay J, Mensah GA, Greenlund K. *The Atlas of Heart Disease and Stroke*. World Health Organization (2004). Available online at: (accessed June 16, 2012).
4. Ponikowski P, Anker SD, AlHabib KF, Cowie MR, Force TL, Hu S, et al. Heart failure: preventing disease and death worldwide. *ESC Heart Fail*. (2014) 1:4–25. doi: 10.1002/ehf2.12005
5. Curtis JP, Sokol SI, Wang Y, Rathore SS, Ko DT, Jadbabaie F, et al. The association of left ventricular ejection fraction, mortality, and cause of death in stable outpatients with heart failure. *J Am Coll Cardiol*. (2003) 42:736–42. doi: 10.1016/S0735-1097(03)00789-7
6. Mele D, Nardozza M, Ferrari R. Left ventricular ejection fraction and heart failure: an indissoluble marriage? *Eur J Heart Fail*. (2018) 20:427–30. doi: 10.1002/ejhf.1071
7. Lang RM, Bierig M, Devereux RB, Flachskampf FA, Foster E, Pellikka PA, et al. Recommendations for chamber quantification. *Eur J echocardiogr*. (2006) 7:79–108. doi: 10.1016/j.euje.2005.12.014
8. Fonarow GC, Hsu JJ. Left ventricular ejection fraction: what is "normal"? *JACC Heart Fail*. (2016) 4:511–3. doi: 10.1016/j.jchf.2016.03.021
9. Tsao CW, Lyass A, Larson MG, Cheng S, Lam CS, Aragam JR, et al. Prognosis of adults with borderline left ventricular ejection fraction. *JACC Heart Fail*. (2016) 4:502–10. doi: 10.1016/j.jchf.2016.03.003
10. McDonagh A, Metra M, Adamo M, Gardner S, Baumbach A, Böhm M, et al. 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*. (2021) 42:3599–726. doi: 10.1093/eurheartj/ehab368
11. Choi M, Park S, Youn C. Update on heart failure management and future directions. *Korean J Intern Med*. (2019) 34:11–43. doi: 10.3904/kjim.2018.428
12. Kalogirou F, Forsyth F, Kyriakou M, Mantle R, Deaton C. Heart failure disease management: a systematic review of effectiveness in heart failure with preserved ejection fraction. *ESC Heart Fail*. (2020) 7:195–213. doi: 10.1002/ehf2.12559
13. Hoekstra T, Lesman-Leegte I, van Veldhuisen DJ, Sanderman R, Jaarsma T. Quality of life is impaired similarly in heart failure patients with preserved and reduced ejection fraction. *Eur J Heart Fail*. (2011) 13:1013–8. doi: 10.1093/eurjhf/hfr072
14. Vedin O, Lam CS, Koh AS, Benson L, Teng THK, Tay WT, et al. Significance of ischemic heart disease in patients with heart failure and preserved, midrange, and reduced ejection fraction: a nationwide cohort study. *Circ Heart Fail*. (2017) 10:e003875. doi: 10.1161/CIRCHEARTFAILURE.117.003875
15. Rastogi A, Novak E, Platts AE, Mann DL. Epidemiology, pathophysiology and clinical outcomes for heart failure patients with a mid-range ejection fraction. *Eur J Heart Fail*. (2017) 19:1597–605. doi: 10.1002/ejhf.879
16. Zhou Q, Li P, Zhao H, Xu X, Li S, Zhao J, et al. Heart failure with mid-range ejection fraction: a distinctive subtype or a transitional stage? *Front Cardiovasc Med*. (2021) 8:678121. doi: 10.3389/fcvm.2021.678121
17. Tromp J, Khan MA, Mentz RJ, O'Connor CM, Metra M, Dittrich HC, et al. Biomarker profiles of acute heart failure patients with a mid-range ejection fraction. *JACC Heart Fail*. (2017) 5:507–17. doi: 10.1016/j.jchf.2017.04.007
18. Cheng RK, Cox M, Neely ML, Heidenreich PA, Bhatt DL, Eapen ZJ, et al. Outcomes in patients with heart failure with preserved, borderline, and reduced ejection fraction in the Medicare population. *Am Heart J*. (2014) 168:721–30. doi: 10.1016/j.ahj.2014.07.008
19. Chioncel O, Lainscak M, Seferovic PM, Anker SD, Crespo-Leiro MG, Harjola VP, et al. Epidemiology and one-year outcomes in patients with chronic heart failure and preserved, mid-range and reduced ejection fraction: an analysis of the ESC heart failure long-term registry. *Eur J Heart Fail*. (2017) 19:1574–85. doi: 10.1002/ejhf.813
20. Löfman I, Szummer K, Dahlström U, Jernberg T, Lund LH. Associations with and prognostic impact of chronic kidney disease in heart failure with preserved, mid-range, and reduced ejection fraction. *Eur J Heart Fail*. (2017) 19:1606–14. doi: 10.1002/ejhf.821
21. Zhao H, Zeng Q, Xu D. Similarities and differences between HFmrEF and HFpEF. *Front Cardiovasc Med*. (2021) 8:1–14. doi: 10.3389/fcvm.2021.678614
22. Zhu K, Ma T, Su Y, Pan X, Huang R, Zhang F, Xu D. Heart failure with mid-range ejection fraction: every coin has two sides. *Front Cardiovasc Med*. (2021) 8:683418. doi: 10.3389/fcvm.2021.683418
23. Palazzuoli A, Beltrami M. Are HFpEF and HFmrEF so different? The need to understand distinct phenotypes. *Front Cardiovasc Med*. (2021) 8:676658. doi: 10.3389/fcvm.2021.676658
24. Lopatin Y. Heart failure with mid-range ejection fraction and how to treat it. *Card Fail Rev*. (2018) 4:9–13. doi: 10.15420/cfr.2018:10:1
25. Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Med Inform Decis Mak*. (2020) 20:16. doi: 10.1186/s12911-020-1023-5
26. Bazoukis G, Stavrakis S, Zhou J, Bollepalli SC, Tse G, Zhang Q, et al. Machine learning versus conventional clinical methods in guiding management of heart failure patients—a systematic review. *Heart Fail Rev*. (2021) 26:23–34. doi: 10.1007/s10741-020-10007-3
27. 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
28. Fahmy AS, Rowin EJ, Manning WJ, Maron MS, Nezafat R. Machine learning for predicting heart failure progression in hypertrophic cardiomyopathy. *Front Cardiovasc Med*. (2021) 8:647857. doi: 10.3389/fcvm.2021.647857
29. Alkhodari M, Jelinek HF, Werghi N, Hadjileontiadis LJ, Khandoker AH. Estimating left ventricle ejection fraction levels using circadian heart rate variability features and support vector regression models. *IEEE J Biomed Health Inform*. (2020) 25:746–54. doi: 10.1109/JBHI.2020.3002336
30. Bayasi N, Tekeste T, Saleh H, Khandoker A, Mohammad B, Ismail M. Adaptive technique for P and T wave delineation in electrocardiogram signals. In; *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* Chicago, IL (2014).

31. 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
32. Segar MW, Patel KV, Ayers C, Basit M, Tang WW, Willett D, et al. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail*. (2020) 22:148–58. doi: 10.1002/ejhf.1621
33. Hedman ÅK, Hage C, Sharma A, Brosnan MJ, Buckbinder L, Gan LM, et al. Identification of novel pheno-groups in heart failure with preserved ejection fraction using machine learning. *Heart*. (2020) 106:342–9. doi: 10.1136/heartjnl-2019-315481
34. Woolley RJ, Ceelen D, Ouwerkerk W, Tromp J, Figarska SM, Anker SD, et al. Machine learning based on biomarker profiles identifies distinct subgroups of heart failure with preserved ejection fraction. *Eur J Heart Fail*. (2021) 23:983–91. doi: 10.1002/ejhf.2144
35. University of Rochester Medical Center. Telemetric and Holter ECG Warehouse (THEW). Available online at: http://thew-project.org/databases.htm (accessed December 24, 2019).
36. Burattini L, Burattini R. *Characterization of Repolarization Alternans in the Coronary Artery Disease. In Coronary Artery Diseases*. IntechOpen. (2012). doi: 10.5772/28766
37. Gatzoulis KA, Tsiachris D, Arsenos P, Antoniou CK, Dilaveris P, Sideris S, et al. Arrhythmic risk stratification in post-myocardial infarction patients with preserved ejection fraction: the PRESERVE EF study. *Eur Heart J*. (2019) 40:2940–9. doi: 10.1093/eurheartj/ehz260
38. Gatzoulis KA, Tsiachris D, Arsenos P, Dilaveris P, Sideris S, Simantirakis E, et al. Post myocardial infarction risk stratification for sudden cardiac death in patients with preserved ejection fraction: PRESERVE-EF study design. *Hellenic J Cardiol*. (2014) 55:361–8.
39. Chatterjee S, McLeish DL. Fitting linear regression models to censored data by least squares and maximum likelihood methods. *Commun Stat Theory Methods*. (1986) 15:3227–43. doi: 10.1080/03610928608829305
40. Greenwood PE, Nikulin MS. *A guide to chi-squared testing*. New York: John Wiley & Sons (1996).
41. Bertini E, Dell'Aquila L, Santucci G. Springview: cooperation of radviz and parallel coordinates for view optimization and clutter reduction. In: *Coordinated and Multiple Views in Exploratory Visualization (CMV'05)*. London: IEEE (2005). p. 22–9.
42. Rubio-Sánchez M, Raya L, Diaz F, Sanchez A. A comparative study between radviz and star coordinates. *IEEE Trans Vis Comput Graph*. (2015) 22:619–28. doi: 10.1109/TVCG.2015.2467324
43. Van Long T. Arcviz: an extended radial visualization for classes separation of high dimensional data. In: *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*. Ho Chi Minh: IEEE (2018). p. 158–62.
44. Giavarina D. Understanding bland altman analysis. *Biochem Med*. (2015) 25:141–51. doi: 10.11613/BM.2015.015
45. Nadruz W Jr, West E, Santos M, Skali H, Groarke JD, Forman DE, et al. Heart failure and midrange ejection fraction: implications of recovered ejection fraction for exercise tolerance and outcomes. *Circ Heart Fail*. (2016) 9:e002826. doi: 10.1161/CIRCHEARTFAILURE.115.002826
46. Delepaul B, Robin G, Delmas C, Moine T, Blanc A, Fournier P, et al. Who are patients classified within the new terminology of heart failure from the 2016 ESC guidelines? *ESC Heart Fail*. (2017) 4:99–104. doi: 10.1002/ehf2.12131
47. Haass M, Kitzman DW, Anand IS, Miller A, Zile MR, Massie BM, et al. Body mass index and adverse cardiovascular outcomes in heart failure patients with preserved ejection fraction: results from the Irbesartan in Heart Failure with Preserved Ejection Fraction (I-PRESERVE) trial. *Circ Heart Fail*. (2011) 4:324–31. doi: 10.1161/CIRCHEARTFAILURE.110.959890
48. Savji N, Meijers WC, Bartz TM, Bhambhani V, Cushman M, Nayor M, et al. The association of obesity and cardiometabolic traits with incident HFpEF and HFrEF. *JACC Heart Fail*. (2018) 6:701–9. doi: 10.1016/j.jchf.2018.05.018
49. Powell-Wiley TM, Ngwa J, Kebede S, Lu D, Schulte PJ, Bhatt DL, et al. Impact of body mass index on heart failure by race/ethnicity from the Get With The Guidelines–Heart Failure (GWTG–HF) registry. *JACC Heart Fail*. (2018) 6:233–42. doi: 10.1016/j.jchf.2017.11.011
50. Al-Jarallah M, Rajan R, Al-Zakwani I, Dashti R, Bulbanat B, Ridha M, et al. Mortality and morbidity in HFrEF, HFmrEF, and HFpEF patients with diabetes in the middle east. *Oman Med J*. (2020) 35:e99. doi: 10.5001/omj.2020.17
51. Bell DS, Goncalves E. Heart failure in the patient with diabetes: epidemiology, aetiology, prognosis, therapy and the effect of glucose-lowering medications. *Diabetes Obes Metab*. (2019) 21:1277–90. doi: 10.1111/dom.13652
52. Khandoker AH, Jelinek HF, Palaniswami M. Risk stratification of cardiac autonomic neuropathy based on multi-lag ToneŰEntropy. *Med Biol Eng Comput*. (2013) 51(5):537–46.
53. Albakri A. Heart failure with mid-range ejection fraction: a review of clinical status and meta-analysis of clinical management methods. *Trends Res*. (2018) 1:1–2. doi: 10.15761/TR.1000121
54. Masarone D, Limongelli G, Rubino M, Valente F, Vastarella R, Ammendola E, et al. Management of arrhythmias in heart failure. *J Cardiovasc Dev.Dis*. (2017) 4:3. doi: 10.3390/jcdd4010003
55. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC, et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American college of cardiology/American heart association task force on clinical practice guidelines and the heart rhythm society. *J Am Coll Cardiol*. (2019) 74:104–32. doi: 10.1016/j.jacc.2019.01.011
56. van Veldhuisen DJ, McMurray JJ. Pharmacological treatment of heart failure with preserved ejection fraction: a glimpse of light at the end of the tunnel? *Eur J Heart Fail*. (2013) 15:5–8. doi: 10.1093/eurjhf/hfs194
57. Lund LH, Claggett B, Liu J, Lam CS, Jhund PS, Rosano GM, et al. Heart failure with mid-range ejection fraction in CHARM: characteristics, outcomes and effect of candesartan across the entire ejection fraction spectrum. *Eur J Heart Fail*. (2018) 20:1230–9. doi: 10.1002/ejhf.1149
58. Lam CS, Gamble GD, Ling LH, Sim D, Leong KTG, Yeo PSD, et al. Mortality associated with heart failure with preserved vs. reduced ejection fraction in a prospective international multi-ethnic cohort study. *Eur Heart J*. (2018) 39:1770–80. doi: 10.1093/eurheartj/ehy005
59. Gutierrez A, Ash J, Akdemir B, Alexy T, Cogswell R, Chen J, et al. Nonsustained ventricular tachycardia in heart failure with preserved ejection fraction. *Pacing Clin Electrophysiol*. (2020) 43:1126–31. doi: 10.1111/pace.14043
60. Ash J, Akdemir B, Gutierrez A, Chen J, Adabag S. Ventricular tachycardia is a common arrhythmia among patients with heart failure with preserved ejection fraction. *Circulation*. (2019) 140 (Suppl. 1):A11658. doi: 10.1161/circ.140.suppl_1.11658
61. Regitz-Zagrosek V. Sex and gender differences in heart failure. *Int J Heart Fail*. (2020) 2:157–81. doi: 10.36628/ijhf.2020.0004
62. Lam CS, Arnott C, Beale AL, Chandramouli C, Hilfiker-Kleiner D, Kaye DM, et al. Sex differences in heart failure. *Eur Heart J*. (2019) 40:3859–68c. doi: 10.1093/eurheartj/ehz835
63. Tadic M, Cuspidi C, Plein S, Belyavskiy E, Heinzel F, Galderisi M. Sex and heart failure with preserved ejection fraction: from pathophysiology to clinical studies. *J Clin Med*. (2019) 8:792. doi: 10.3390/jcm8060792
64. Alkhodari M, Jelinek H, Saleem S, Hadjileontiadis L, Khandoker A. Revisiting left ventricular ejection fraction levels: a circadian heart rate variability-based approach. *IEEE Access*. (2021) 9:130111–26. doi: 10.1109/ACCESS.2021.3114029

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 23 November 2021
doi: 10.3389/fcvm.2021.768245

MOCOnet: Robust Motion Correction of Cardiovascular Magnetic Resonance T1 Mapping Using Convolutional Neural Networks

Ricardo A. Gonzales[1†], Qiang Zhang[1†], Bartłomiej W. Papież[2,3], Konrad Werys[1], Elena Lukaschuk[1], Iulia A. Popescu[1], Matthew K. Burrage[1], Mayooran Shanmuganathan[1], Vanessa M. Ferreira[1] and Stefan K. Piechnik[1]*

[1] Oxford Centre for Clinical Magnetic Resonance Research (OCMR), Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom, [2] Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom, [3] Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, United Kingdom

OPEN ACCESS

Edited by:
Shane Nanayakkara,
Alfred Hospital, Australia

Reviewed by:
Haikun Qi,
ShanghaiTech University, China
Ruud B. van Heeswijk,
Centre Hospitalier Universitaire Vaudois (CHUV), Switzerland
Filippo Cademartiri,
Gabriele Monasterio Tuscany Foundation (CNR), Italy

***Correspondence:**
Stefan K. Piechnik
stefan.piechnik@cardiov.ox.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: 31 August 2021
Accepted: 27 October 2021
Published: 23 November 2021

Citation:
Gonzales RA, Zhang Q, Papież BW, Werys K, Lukaschuk E, Popescu IA, Burrage MK, Shanmuganathan M, Ferreira VM and Piechnik SK (2021) MOCOnet: Robust Motion Correction of Cardiovascular Magnetic Resonance T1 Mapping Using Convolutional Neural Networks. Front. Cardiovasc. Med. 8:768245. doi: 10.3389/fcvm.2021.768245

Background: Quantitative cardiovascular magnetic resonance (CMR) T1 mapping has shown promise for advanced tissue characterisation in routine clinical practise. However, T1 mapping is prone to motion artefacts, which affects its robustness and clinical interpretation. Current methods for motion correction on T1 mapping are model-driven with no guarantee on generalisability, limiting its widespread use. In contrast, emerging data-driven deep learning approaches have shown good performance in general image registration tasks. We propose MOCOnet, a convolutional neural network solution, for generalisable motion artefact correction in T1 maps.

Methods: The network architecture employs U-Net for producing distance vector fields and utilises warping layers to apply deformation to the feature maps in a coarse-to-fine manner. Using the UK Biobank imaging dataset scanned at 1.5T, MOCOnet was trained on 1,536 mid-ventricular T1 maps (acquired using the ShMOLLI method) with motion artefacts, generated by a customised deformation procedure, and tested on a different set of 200 samples with a diverse range of motion. MOCOnet was compared to a well-validated baseline multi-modal image registration method. Motion reduction was visually assessed by 3 human experts, with motion scores ranging from 0% (strictly no motion) to 100% (very severe motion).

Results: MOCOnet achieved fast image registration (<1 second per T1 map) and successfully suppressed a wide range of motion artefacts. MOCOnet significantly reduced motion scores from 37.1±21.5 to 13.3±10.5 ($p < 0.001$), whereas the baseline method reduced it to 15.8±15.6 ($p < 0.001$). MOCOnet was significantly better than the baseline method in suppressing motion artefacts and more consistently ($p = 0.007$).

Conclusion: MOCOnet demonstrated significantly better motion correction performance compared to a traditional image registration approach. Salvaging data affected by motion with robustness and in a time-efficient manner may enable better image quality and reliable images for immediate clinical interpretation.

Keywords: cardiovascular magnetic resonance, deep learning, image registration, ShMOLLI, T1 mapping

1. INTRODUCTION

Quantitative T1 mapping is a novel approach in cardiovascular magnetic resonance (CMR) for myocardial tissue characterisation (1). Native and post-contrast T1 mapping offer quantitative, pixel-wise measures to detect tissue changes in myocardial composition (2) and have been used in the assessment of myocardial inflammation (3), oedema (4, 5), infiltration (6), diffuse fibrosis (7), and other pathologies (8). Stress T1 mapping has the potential to assess coronary artery disease without the need for gadolinium-based contrast agents (9–11).

T1 mapping is obtained from pixel-wise exponential recovery curve fitting of multiple T1-weighted images. With advances made from the original Look-Locker spectroscopic method (12), current mapping techniques employ intermittent image acquisition using electrocardiographic gating during multiple heartbeats (2). The shortened modified Look-Locker inversion recovery (ShMOLLI) (13) allowed shorter breath-holds with 9 heartbeats with high precision and reproducibility. Although acquiring multiple T1-weighted images at the same cardiac phase largely reduces the influence of cardiac motion, undesired respiratory motion still poses significant challenges (14). Uncorrected and unrecognised respiratory motion artefacts may cause errors in T1 estimation and incorrect diagnoses (13).

Retrospective motion correction (MOCO) on the multiple T1-weighted images can significantly improve the robustness and clinical utility of mapping techniques (15). Such correction is accomplished by aligning the T1-weighted images before reconstruction. The main challenge is the variation in image contrast and signal nulling of the multiple T1-weighted images acquired at different inversion times. Model-driven registration methods for MOCO were developed to circumvent this limitation with promising results (16–19). However, careful inspection for uncorrected residual motion or distortions from failures in motion correction is still needed (20). Although visual assessment in CMR is still the clinical standard for image interpretation (21), constant and long manual labour is prone to error due to inconsistency and operator fatigue, as well as slow clinical workflow if handling a large volume of images.

With the advent of deep learning, convolutional neural networks (CNN) have enabled unprecedented progress in image processing, shifting the paradigm from predefined, hand-crafted rules to automated learning procedures aided by large data. The rapid adaptation of deep learning approaches within CMR provides fast, consistent, and accurate pipelines primarily for image segmentation and analysis (22) significantly reducing physician labour hours. The field of clinical image registration with deep learning is also primed to replace iterative registration methods, with potential to improve accuracy, time efficiency and quality control (23), and applicability to cover the unmet need of MOCO in T1 maps. We hypothesised that a data-driven method for myocardial motion correction would suppress motion artefacts with more robustness and generalisability to serve large clinical datasets.

In this work, we present MOCOnet, a novel deep learning approach for myocardial motion correction developed using CMR T1 mapping from the UK Biobank (24). We adapted an encoder-decoder architecture with warping layers to aid the learning of such deformation in a coarse-to-fine manner. Given a set of T1-weighted images, MOCOnet can predict the deformation required to correct any present motion artefacts in a time-efficient manner. MOCOnet was tested for its motion correction performance against a well-validated multi-modal image registration method, using multiple blinded expert observers to validate the motion correction effectiveness.

2. MATERIALS AND METHODS

2.1. Cardiac T1 Mapping and Motion Artefact

Cardiac ShMOLLI T1 mapping is calculated by fitting exponential recovery curves to 7 inversion recovery-weighted (IRW) images with multiple inversion times (**Figure 1A**) and acquired within a short 9-heartbeat single breath-hold (13). The reconstructed T1 map (**Figure 1B**) enables pixel-wise quantification of T1 values. The associated map of coefficient of explained variance (R^2 map; **Figure 1C**) allows quality monitoring of the curve fitting in reference to a mono-exponential T1 relaxation recovery model. A closer proximity to the reference displays a uniform white appearance of relevant regions of interest in the R^2 map, whereas motion in the IRW images (**Figure 1D**, arrowed) decreases the T1 map interpretability (**Figure 1E**, arrowed), corresponding to the dark bands at the motion-affected areas in the R^2 map (**Figure 1F**, arrowed). Besides motion artefacts, the R^2 map is also sensitive to off-resonance, fat inclusion, mistriggering, and other artefacts (5, 25).

2.2. Non-rigid Registration Approach

Given that a T1 map with motion artefacts is composed of 7 unaligned IRW images, a motion-corrected T1 map can be achieved by aligning the IRW images. The motion artefact can be synthesised as a deformation of aligned IRW images with a displacement vector field (DVF). The non-rigid registration problem is then solved by estimating the inverse DVF of a given set of unaligned IRW images.

2.3. Multi-Scale Registration Neural Network

The proposed learning-based model corrects a T1 map by estimating the inverse DVF in each of its 7 IRW images to enable a non-rigid registration between them, before the T1 map reconstruction. The multi-scale registration CNN (**Figure 2**) adopts an encoder-decoder U-Net-like structure (26) and employs warping layers (27) between the contracting and expansive paths at each scale. The feature maps are down-sampled with a series of 3×3 convolutional layers followed by a batch normalisation layer, a leaky rectified linear unit and a max-pooling layer, and similarly up-sampled with a transposed convolutional layer. The warping layers speed up the training by imposing a loss function on a multi-scale manner and increase the registration accuracy by correcting motion starting from

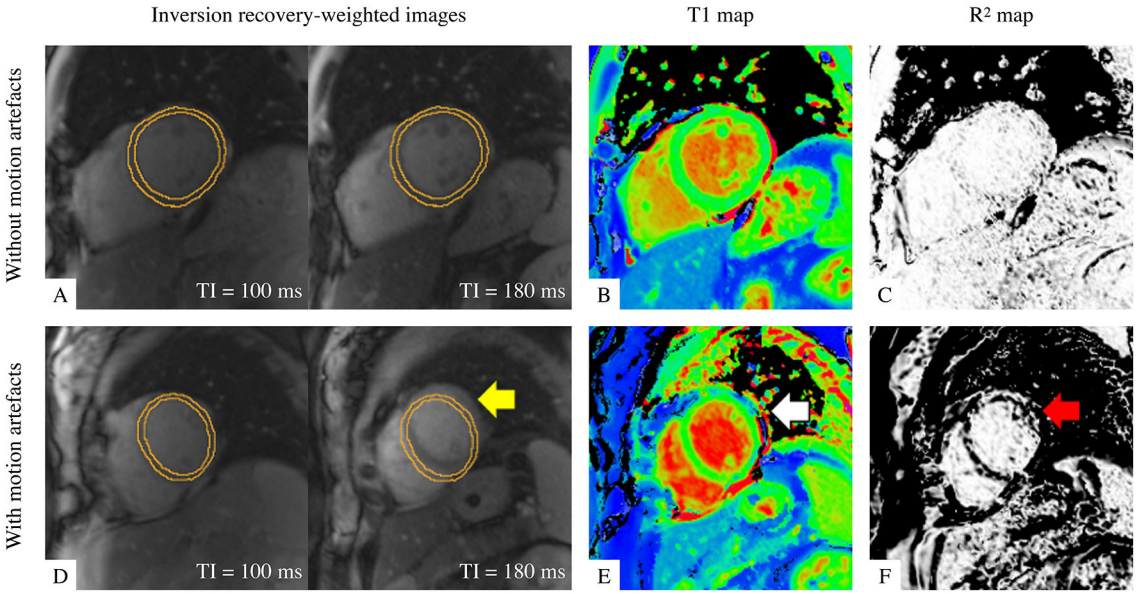


FIGURE 1 | Illustration of T1 maps with good quality (top row) and with motion artefact (bottom row). **(A,D)** Two examples out of seven of inversion-recovery weighted (IRW) images required for T1 map reconstruction are shown, time-stamped with their corresponding inversion times (TI) and overlaid by identical manual myocardial contours for identifying motion. **(B,E)** ShMOLLI T1 maps. **(C,F)** R^2 quality control maps. A good quality T1 map is indicated by **(A)** myocardium in same position and **(C)** "all white" in the left ventricular myocardium indicating high T1 fitting confidence. A T1 map with motion artefact is evident by misalignment in IRW images (yellow arrow), suspicious features in T1 map (white arrow) and dark bands in R^2 map in the left ventricular myocardium as evidence of poor T1 fitting (red arrow).

coarse levels and passing the residual motion to higher resolution layers for fine motion correction.

The IRW images are first fed in a sequence of convolution and downsampling operations to produce features at multiple scales on a per-channel basis. The features, from low to high resolutions, are then used as input of convolution modules to produce DVFs. Each convolution module takes as input the features from the previous step, the DVF at the previous scale, and the warped features from the downsampling stage. Applying warping at each of the 4 scales enables the use of residual motion information to be corrected and refined in the next scale. Hence, the neural network generates the DVFs in a coarse-to-fine manner and adds more details with higher resolution in each subsequent level, with a loss function defined at each scale to further supervise the learning manner.

2.4. Imaging Data and Inclusion Criteria

The imaging data comprised of over 5,000 CMR native T1 maps from the UK Biobank Imaging Component (24), acquired at the mid-ventricular short-axis view using the ShMOLLI T1 mapping sequence (13). For quality control, a trained human operator (EL), with over 10 years of experience in CMR image analysis, assessed the presence of any artefact in the left ventricular myocardium in the 7 IRW images for each T1 map. A total of 1,536 T1 maps were scored strictly as good quality with no artefact. The remaining data were marked to have either mild to severe motion or other imaging artefacts and were excluded from the training dataset. This strict quality control ensured that the neural network learnt to align the images accurately with no distraction from residual motion artefacts in the training data, i.e., with images that did not require any motion correction.

2.5. Training Procedure

The quality-controlled imaging data were used to generate a training dataset with 10% of the data preserved for validation. Artificial DVFs were generated as previously described (28) and applied to the IRW images without motion artefacts to synthesise random non-rigid motion without requiring segmentations (28). Specifically, 7 DVFs were generated with random parameters preserving anatomical topology. Mean displacement value at each pixel was calculated and removed from all 7 DVFs to focus on relative displacement between images. The generated DVFs were applied to each of the IRW images, respectively to produce deformed IRW images. The proposed model was trained to predict 7 inverse DVFs from 7 deformed IRW images with the synthetic, inverse DVFs as ground truth (**Figure 3A**).

2.6. Testing Procedure

Once trained, MOCOnet reads a given set of 7 IRW images with or without motion artefacts and estimates the deformation required to correct any present motion (**Figure 3B**), without ground truth. The T1 map is then reconstructed offline using motion-corrected images with an open source library for CMR parametric mapping (29).

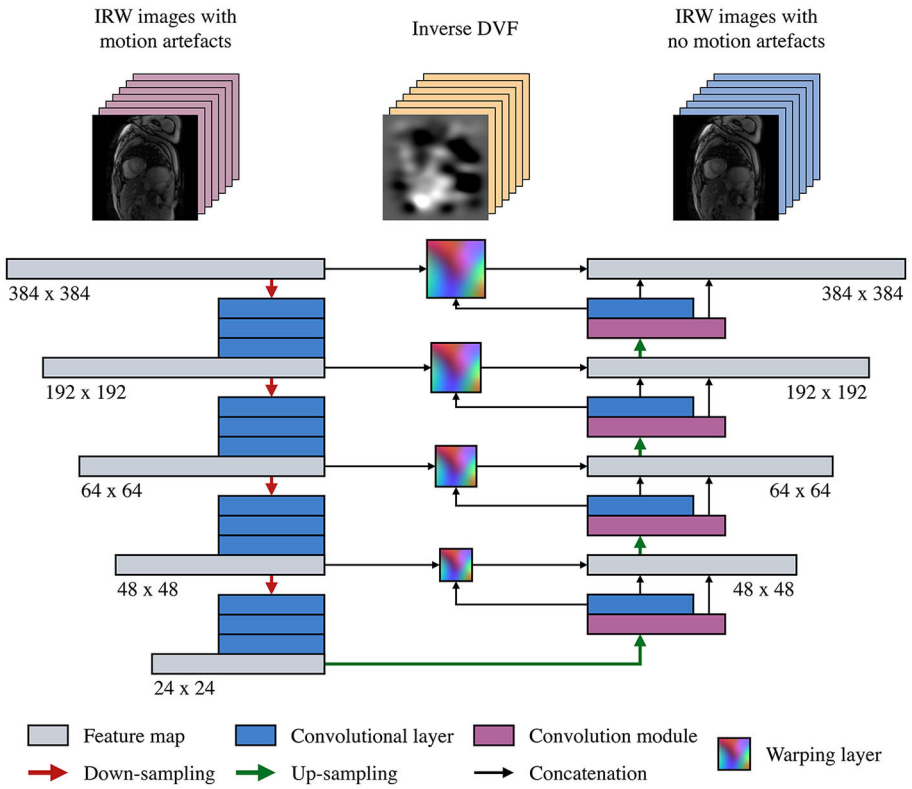


FIGURE 2 | Structure of the proposed motion correction convolutional neural network (MOCOnet). A stack of seven inversion recovery-weighted (IRW) images is input into the encoder-decoder structure on a per-channel basis. The warping layers estimate the optical flow from all the channels in a coarse-to-fine manner at each scale. The last warping layer generates the inverse distance vector field (DVF), i.e., the deformation required to correct the motion artefacts, in a groupwise manner.

2.7. Implementation Specification

All images were zero-padded to the same size of 384 × 384 pixels and image intensities were pre-processed with quantile normalisation to ensure generalisability (30). The multi-scale loss was calculated as the average mean square errors of the predicted DVFs at each scale and resolution. The neural network was optimised using the Adam method (31) with an initial learning rate of 0.001 and a learning rate scheduler to reduce the learning rate during the training, and mini-batch size of 4. Training was stopped once the validation loss did not decrease for 50 epochs. The network was trained for approximately 48 h until the training curve converged with low bias and variance using a NVIDIA TITAN XP GPU and implemented in TensorFlow (32). After the training, correcting motion for each set of 7 IRW images took less than 1 s on GPU or a modern CPU.

2.8. Validation

2.8.1. Baseline Deformable Image Registration Method

The proposed method's performance was compared against a well-validated multi-modal image registration algorithm (33) as the baseline method. The registration method alleviated the problem of artificial motion discontinuities by combining a bilateral filter with an additional deformation field-based filter and a diffusion regularisation algorithm, serving as an excellent registration approach without requiring a prior image segmentation task as conventional methods. The baseline method, implemented in C, used the first IRW image as a reference image for all subsequent pairwise registrations and took approximately 30 s per T1 map on a modern CPU.

2.8.2. Test on Respiratory Motion With Human Observer Scores

A multi-observer experiment was designed to evaluate the effectiveness and robustness of motion correction, and potential noise introduced to cases originally with no motion. From the UK Biobank, a test set of 200 real acquired T1 maps with various degree of motion artefacts was selected based on the existing quality scores by an experienced human observer. Specifically, 50 samples presented severe motion artefacts affecting all myocardial segments, 100 presented moderate motion affecting individual segments, and 50 presented mild to no motion.

The extent of motion on the test set was assessed in a 5-point categorical scale: 'no motion', 'mild motion', 'moderate motion', 'severe motion', and 'very severe motion', with a numerical scale between 0 to 100% behind the interface, to ensure both

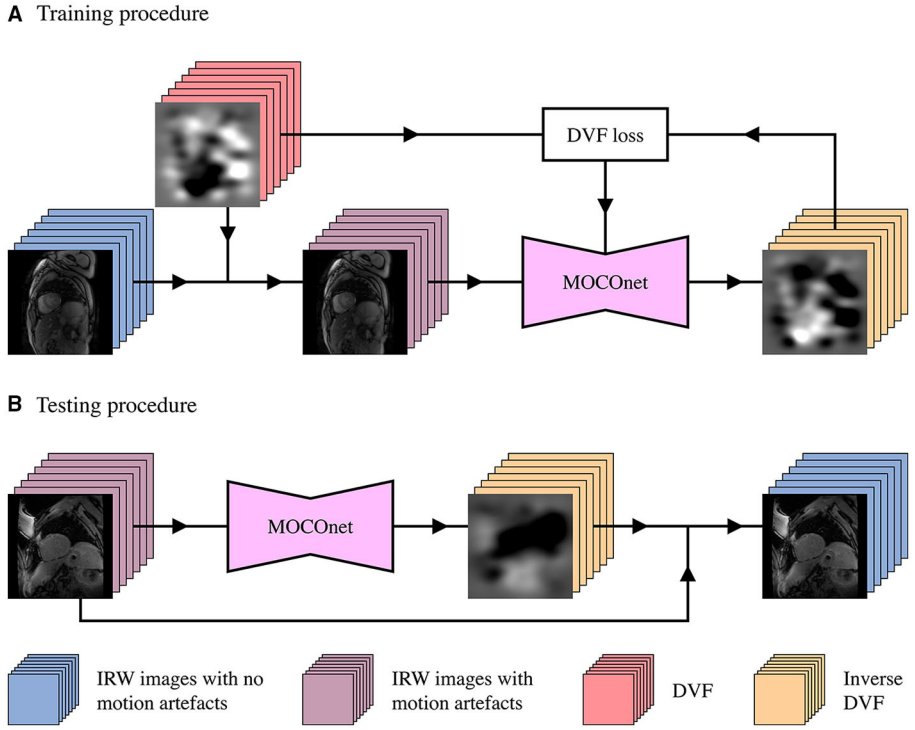


FIGURE 3 | Development workflow of the proposed motion correction convolutional neural network (MOCOnet) for myocardial ShMOLLI T1 mapping. **(A)** MOCOnet was trained on 1,536 sets of seven inversion recovery-weighted (IRW) images with no motion artefacts which were synthetically deformed with displacement vector fields (DVFs), to predict the inverse DVF required to correct the motion. **(B)** MOCOnet was tested on 200 T1 maps with a varied degree of motion artefacts. Each stack denotes a set of seven images; each junction denotes the DVFs application to the IRW images; the box with DVF loss represents the weight adjustment during training.

intuitiveness for human operators and practicality for statistical analyses. The baseline and proposed methods were applied to all samples unselectively, giving in total 400 motion-corrected samples. One hundred and twenty only samples (20%) were randomly chosen from the mixed 600 samples and duplicated to evaluate intra-observer variability. Three trained human observers (IP, MB and MS) were instructed to score the resultant 720 samples for the extent of motion. All observers were blinded to the original artefact scores and which motion correction method was applied. To reduce the variance of the human scores X_i, the weighted average score $\overline{X}$ of the three observers ($i = 3$) was calculated as $\overline{X} = \sum W_i X_i / \sum W_i$. The weights W_i were calculated by the inverse of intra-observer variance σ_i (34, 35) based on the duplicated 20% cases, i.e., $W_i = 1/\sigma_i^2$ for the i-th observer. The expected standard error of the weighted average scores was $SE(\overline{X}) = \sqrt{\sum W_i}^{-1}$.

2.8.3. Statistical Analysis

Quality scores were reported as mean ± standard deviation. Non-parametric Wilcoxon signed-rank test was used to assess the statistical difference between the data with and without motion correction by the baseline and proposed methods. Given the modest number of repeated comparisons within each group the statistical significance threshold was kept at standard $p < 0.05$ (36). Statistical analysis was performed using the Python programming language.

3. RESULTS

The results of human observer validation on the 200 cases from the UK Biobank are reported in **Table 1**. Intra-observer variabilities of the three observers on the 20% duplicated cases were 10.6, 17.3 and 21.9, respectively. Standard error of the final weighted-average scores that were used to compare the motion correction methods was 8.3 at a scale from 0 to 100%. Overall, both methods significantly reduced the motion artefacts, from an average motion score of 37.1 ± 21.5 to 15.8 ± 15.6 (baseline method) and 13.3 ± 10.5 (MOCOnet; both $p < 0.001$). MOCOnet was significantly more effective at reducing motion artefacts than the baseline method for the subgroups with severe motion ($N = 50$, $p = 0.006$) and moderate motion ($N = 100$, $p = 0.04$). For the subgroup with mild to no motion ($N = 50$), both methods significantly further reduced the motion artefacts (both $p < 0.001$), and neither added noise, nor was significantly different from each other ($p = 0.2$). Overall, MOCOnet suppressed motion artefacts to a higher extent and in a more consistent way compared to the baseline method, as evidenced by its lower maximum score and variability ($N = 200$, $p = 0.007$). The boxplot of motion scores (**Figure 4**) further

TABLE 1 | Human observer assessment of motion extent (%) on 200 T1 maps before motion correction, and after the baseline and proposed method (MOCOnet) for motion correction.

	All data ($N = 200$)	Group 1 ($N = 50$) Severe motion	Group 2 ($N = 100$) Moderate motion	Group 3 ($N = 50$) Mild to no motion
Before MOCO	37.1 ± 21.5	55.8 ± 18.7 (99.3)	35.5 ± 18.9 (80.5)	21.7 ± 13.8 (62.1)
Baseline method	15.8 ± 15.6	25.8 ± 19.8 (93.4)	14.7 ± 13.9 (65.7)	**8.1** ± 6.5 (34.2)
MOCOnet	**13.3 ± 10.5**	**18.6 ± 14.3 (86.9)**	**12.7 ± 9.2 (46.4)**	9.4 ± **6.4 (19.8)**

The quality scores are inverse variance-weighted scores of three human observers and reported in mean ± SD (maximum value). The best results are highlighted in bold.

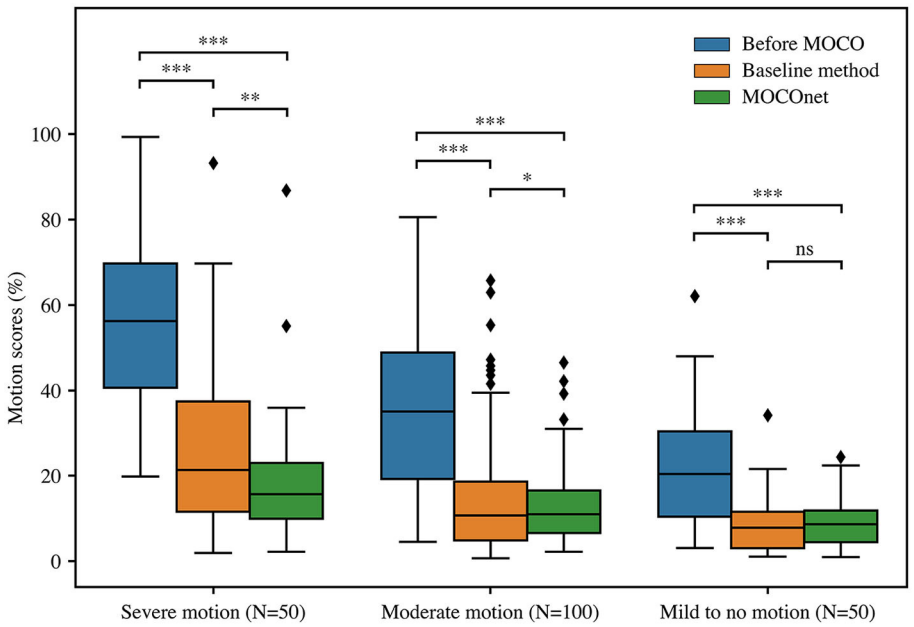


FIGURE 4 | Motion correction (MOCO) performance of the baseline and the proposed deep learning-based motion correction (MOCOnet) methods. Box and whisker plot of motion scores in non-parametric terms of three data groups, before (blue) and after motion correction by the baseline (orange) and proposed MOCOnet (green) methods. Reported values are inverse variance-weighted scores of three human observers. MOCOnet achieved the best results and significantly reduced the motion artefacts. $^{*}p = 0.04$; $^{**}p < 0.01$; $^{***}p < 0.001$; ns = not significant.

illustrates the above dependencies in non-parametric terms. This demonstrates that MOCOnet achieved a tighter span of perceived motion estimates, with better perceived robustness to outliers.

MOCOnet successfully learnt from synthetic random motion to predict the required DVFs to correct the motion of IRW images ensuring a motion-corrected T1 map in real acquired data. **Figure 5** exemplifies the robustness of the method. One training sample was falsely considered to have no motion artefacts, as evidenced by the overlaid contours of both myocardium and stomach but this did not overfit the learning or affect the final results. The data-driven process aided the learning of the general rule, as MOCOnet managed to correct the error in this training sample, instead of replicating it.

4. DISCUSSION

In this work, MOCOnet, a novel end-to-end motion correction neural network for CMR T1 maps, was developed using a large-scale dataset and validated with expert human analysts. MOCOnet was able to automatically predict the deformation required to correct real motion artefact cases. The proposed method has a fast-processing speed of <1 s per T1 map and does not require modification of image acquisition sequences, external hardware, or user intervention, enabling direct implementation to clinical practise.

Although the principle of estimating the required DVFs on a given set of images to correct their mutual alignment was tested on myocardial ShMOLLI T1 maps, the problem formulation and solution are not limited to this mapping method or region of interest. The deformation estimation is alleviated by considering the images 'as is' with a data-driven procedure (37) without heavily relying on the differences in contrast, the specific inversion recovery times or a prior user input. This principle can be directly applied to other T1 mapping methods that require multiple T1-weighted images to be aligned in a groupwise manner to ensure an accurate exponential recovery curve fitting (38), to other organs that are evaluated with parametric mapping, such as the brain (39) and liver (40), and to other imaging modalities with varied image contrast (41).

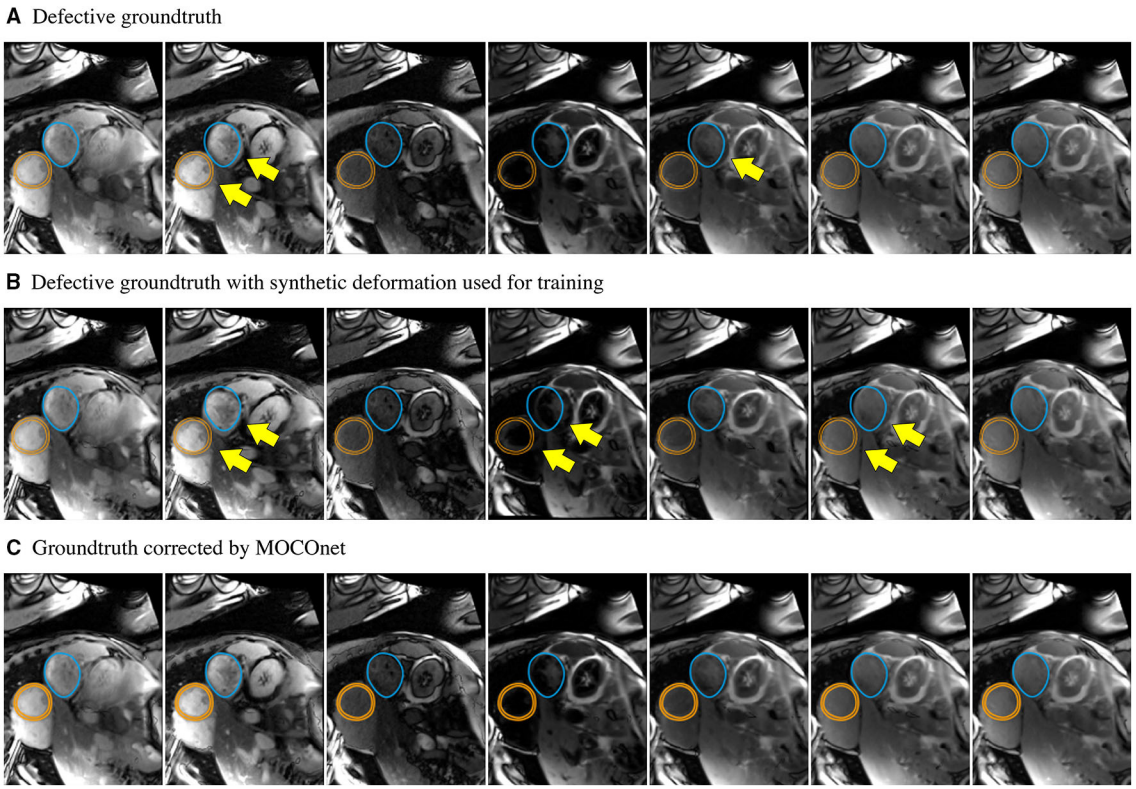


FIGURE 5 | Robustness of the proposed motion correction convolutional neural network (MOCOnet) for myocardial ShMOLLI T1 mapping from a noisy training sample. **(A)** Training sample falsely considered free of motion (1 in 1,536) as manually depicted with unaligned myocardium (orange) and stomach (blue) with yellow arrows throughout the inversion recovery-weighted images. **(B)** Applied deformation to the training sample used for training. **(C)** Sample corrected by MOCOnet after training demonstrating the successful learning of the general rule without replicating the data.

The potential clinical impact of the method is promising. A large portion of the UK Biobank T1 mapping data analysed in this study presented mild to severe motion, hampering the diagnostic utility of T1 mapping. Although recent progress on automated motion artefact detection methods (42) may alleviate the quality monitoring process, rescanning to ensure a free-of-motion T1 map would increase scan times and reduce patient throughput. The presented data-driven MOCOnet approach provides an attractive solution to retrospectively suppress the motion using most of the acquired data to enhance T1 map quality, which is expected to salvage data corrupted by motion, reduce the need for rescanning and improve diagnosis. MOCOnet also holds promise for stress T1 mapping applications (9–11, 38) which may be subject to greater motion artefact. With the rapidly evolving field of deep learning, further research can be done to assess potential benefits of incorporating a more diverse variety of learning-based registration methods (23, 43) into a quality-control driven pipeline (44–46) to verify the registration accuracy on-the-fly including the R^2 maps. With further work, MOCOnet together with T1 protocol quality assurance (47, 48) and automated myocardial segmentation (45) could ultimately lead to a comprehensive framework for robust T1 mapping for clinical use.

Despite a good performance in motion correction, as evidenced with the large improvement in the motion score, it is revealed by human observer experiments that MOCOnet could still fail in correcting images with severe motion. The challenge is not only due to difficulty in motion correction, but also through-plane motion, resulting in fitting of T1 values using signals at different tissue location. Breath holding remains crucial in acquiring good quality T1 maps. Future work will include validation on a multi-vendor, multi-centre population, expansion to other regions of interest, and direct implementation onto the scanner for robust inline motion artefact correction to generate good quality and reliable images for immediate clinical interpretation.

5. CONCLUSION

MOCOnet is an effective and robust convolutional neural network for correction of artefacts from myocardial motion. The technique can be readily deployed for post-processing of T1 mapping to restore T1 values in images affected by motion artefacts. This non-rigid registration solution can be further extended to other mapping methods, for generating good quality and reliable images for immediate clinical interpretation. MOCOnet can eventually enhance parametric mapping methods

paving the way towards more reliable quantitative CMR medical imaging.

DATA AVAILABILITY STATEMENT

The imaging data were provided by the UK Biobank under the technical access agreement. Data access must be obtained directly from the UK Biobank.

AUTHOR CONTRIBUTIONS

RG, QZ, BP, and KW contributed to the design of the study. RG drafted the manuscript. QZ curated the database and analysed the results. EL quality controlled the UK Biobank data. IP, MB, and MS scored the motion extent. VF and SP provided research guidance and conceived the study. All authors critically edited and revised the article.

FUNDING

RG acknowledges support for his D.Phil. studies from the Clarendon Fund, the Balliol College and the Radcliffe Department of Medicine, University of Oxford. QZ, VF, and SP acknowledge John Fell Oxford University Press Research Fund. QZ, MB, VF, and SP acknowledge support from the Oxford BHF Centre of Research Excellence (RE/18/3/34214). BP acknowledges Rutherford Fund at Health Data Research UK (MR/S004092/1). MB is supported by a British Heart Foundation (BHF) Clinical Research Training Fellowship (FS/19/65/34692). MS is supported by the Alison Brading Memorial Graduate Scholarship in Medical Science, Lady Margaret Hall, University of Oxford. IP, MB, VF, and SP acknowledge support from the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre at The Oxford University Hospitals NHS Foundation Trust.

REFERENCES

1. Karamitsos TD, Arvanitaki A, Karvounis H, Neubauer S, Ferreira VM. Myocardial tissue characterization and fibrosis by imaging. *JACC Cardiovasc Imaging.* (2020) 13:1221–34. doi: 10.1016/j.jcmg.2019.06.030
2. 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 (EACVI). *J Cardiovasc Magn Reson.* (2017) 19:75. doi: 10.1186/s12968-017-0389-8
3. Ferreira VM, Schulz-Menger J, Holmvang G, Kramer CM, Carbone I, Sechtem U, et al. Cardiovascular magnetic resonance in nonischemic myocardial inflammation. *J Am Coll Cardiol.* (2018) 72:3158–3176. doi: 10.1016/j.jacc.2018.09.072
4. Messroghli DR, Niendorf T, Schulz-Menger J, Dietz R, Friedrich MG. T1 mapping in patients with acute myocardial infarction. *J Cardiovasc Magn Reson.* (2003) 5:353–9. doi: 10.1081/JCMR-120019418
5. Ferreira VM, Piechnik SK, Dall'Armellina E, Karamitsos TD, Francis JM, Choudhury RP, et al. Non-contrast T1-mapping detects acute myocardial edema with high diagnostic accuracy: a comparison to T2-weighted cardiovascular magnetic resonance. *J Cardiovasc Magn Reson.* (2012) 14:42. doi: 10.1186/1532-429X-14-42
6. Karamitsos TD, Piechnik SK, Banypersad SM, Fontana M, Ntusi NB, Ferreira VM, et al. Noncontrast T1 Mapping for the diagnosis of cardiac amyloidosis. *JACC Cardiovasc Imaging.* (2013) 6:488–97. doi: 10.1016/j.jcmg.2012.11.013
7. Everett RJ, Stirrat CG, Semple SIR, Newby DE, Dweck MR, Mirsadraee S. Assessment of myocardial fibrosis with T1 mapping MRI. *Clin Radiol.* (2016) 71:768–78. doi: 10.1016/j.crad.2016.02.013
8. Liu JM, Liu A, Leal J, McMillan F, Francis J, Greiser A, et al. Measurement of myocardial native T1 in cardiovascular diseases and norm in 1291 subjects. *J Cardiovasc Magn Reson.* (2017) 19:74. doi: 10.1186/s12968-017-0386-y
9. Liu A, Wijesurendra RS, Francis JM, Robson MD, Neubauer S, Piechnik SK, et al. Adenosine stress and rest T1 mapping can differentiate between ischemic, infarcted, remote, and normal myocardium without the need for gadolinium contrast agents. *JACC Cardiovasc Imaging.* (2016) 9:27–36. doi: 10.1016/j.jcmg.2015.08.018
10. Burrage MK, Shanmuganathan M, Masi A, Hann E, Zhang Q, Popescu IA, et al. Cardiovascular magnetic resonance stress and rest T1-mapping using regadenoson for detection of ischemic heart disease compared to healthy controls. *Int J Cardiol.* (2021) 333:239–45. doi: 10.1016/j.ijcard.2021.03.010
11. Burrage MK, Shanmuganathan M, Zhang Q, Hann E, Popescu IA, Soundarajan R, et al. Cardiac stress T1-mapping response and extracellular volume stability of MOLLI-based T1-mapping methods. *Sci Rep.* (2021) 11:13568. doi: 10.1038/s41598-021-92923-4
12. Look DC, Locker DR. Time saving in measurement of NMR and EPR relaxation times. *Rev Sci Instruments.* (1970) 41:250–1. doi: 10.1063/1.1684482
13. Piechnik SK, Ferreira VM, Dall'Armellina E, Cochlin LE, Greiser A, Neubauer S, et al. Shortened Modified Look-Locker Inversion recovery (ShMOLLI) for clinical myocardial T1-mapping at 1.5 and 3 T within a 9 heartbeat breathhold. *J Cardiovasc Magn Reson.* (2010) 12:69. doi: 10.1186/1532-429X-12-69
14. Messroghli DR, Plein S, Higgins DM, Walters K, Jones TR, Ridgway JP, et al. Human myocardium: single-breath-hold MR T1 mapping with high spatial resolution—reproducibility study. *Radiology.* (2006) 238:1004–12. doi: 10.1148/radiol.2382041903
15. Kellman P, Wilson JR, Xue H, Ugander M, Arai AE. Extracellular volume fraction mapping in the myocardium, part 1: evaluation of an automated method. *J Cardiovasc Magn Reson.* (2012) 14:63. doi: 10.1186/1532-429X-14-63
16. Chefd'hotel C, Hermosillo G, Faugeras O. Flows of diffeomorphisms for multimodal image registration. In: *Proceedings IEEE International Symposium on Biomedical Imaging.* (2002). p. 753–6. doi: 10.1109/ISBI.2002.1029367
17. Xue H, Shah S, Greiser A, Guetter C, Littmann A, Jolly MP, et al. Motion correction for myocardial T1 mapping using image registration with synthetic image estimation. *Magn Reson Med.* (2012) 67:1644–55. doi: 10.1002/mrm.23153
18. Zhou R, Huang W, Yang Y, Chen X, Weller DS, Kramer CM, et al. Simple motion correction strategy reduces respiratory-induced motion artifacts for k-t accelerated and compressed-sensing cardiovascular magnetic resonance perfusion imaging. *J Cardiovasc Magn Reson.* (2018) 20:6. doi: 10.1186/s12968-018-0427-1
19. Becker KM, Blaszczyk E, Funk S, Nuesslein A, Schulz-Menger J, Schaeffter T, et al. Fast myocardial T1 mapping using cardiac motion correction. *Magn Reson Med.* (2020) 83:438–451. doi: 10.1002/mrm.27935
20. Robinson AA, Chow K, Salerno M. Myocardial T1 and ECV Measurement. *JACC Cardiovasc Imaging.* (2019) 12:2332–44. doi: 10.1016/j.jcmg.2019.06.031
21. Schulz-Menger J, Bluemke DA, Bremerich J, Flamm SD, Fogel MA, Friedrich MG, et al. Standardized image interpretation and post-processing in cardiovascular magnetic resonance - 2020 update. *J Cardiovasc Magn Reson.* (2020) 22:19. doi: 10.1186/s12968-020-00610-6
22. 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
23. Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. Deep learning in medical image registration: a review. *Phys Med Biol.* (2020) 65:20TR01. doi: 10.1088/1361-6560/ab843e

24. 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
25. Piechnik SK, Jerosch-Herold M. Myocardial T1 mapping and extracellular volume quantification: an overview of technical and biological confounders. *Int J Cardiovasc Imaging.* (2018) 34:3–14. doi: 10.1007/s10554-017-1235-7
26. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical Image Computing and Computer-Assisted Intervention-MICCAI.* (2015). Cham: Springer International Publishing (2015). p. 234–41. doi: 10.1007/978-3-319-24574-4_28
27. Sun D, Yang X, Liu M, Kautz J. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA: IEEE Computer Society (2018). p. 8934–43. doi: 10.1109/CVPR.2018.00931
28. Sokooti H, De Vos B, Berendsen F, Lelieveldt BP, Išgum I, Staring M. Nonrigid image registration using multi-scale 3D convolutional neural networks. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2017.* Cham: Springer International Publishing (2017). p. 232–9. doi: 10.1007/978-3-319-66182-7_27
29. Werys K, Dragonu I, Zhang Q, Popescu I, Hann E, Puchta H, et al. Total mapping toolbox (TOMATO): an open source library for cardiac magnetic resonance parametric mapping. *SoftwareX.* (2020) 11:100369. doi: 10.1016/j.softx.2019.100369
30. Onofrey JA, Casetti-Dinescu DI, Lauritzen AD, Sarkar S, Venkataraman R, Fan RE, et al. Generalizable multi-site training and testing of deep neural networks using image normalization. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019).* (2019). p. 348–51. doi: 10.1109/ISBI.2019.8759295
31. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv e-prints.* (2014).
32. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. *arXiv e-prints.* (2016).
33. Papież BW, Heinrich MP, Fehrenbach J, Risser L, Schnabel JA. An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration. *Med Image Anal.* (2014) 18:1299–311. doi: 10.1016/j.media.2014.05.005
34. Cochran WG. The combination of estimates from different experiments. *Biometrics.* (1954) 10:101–29. doi: 10.2307/3001666
35. Lee CH, Cook S, Lee JS, Han B. Comparison of two meta-analysis methods: inverse-variance-weighted average and weighted sum of z-scores. *Genomics Inform.* (2016) 14:173–180. doi: 10.5808/GI.2016.14.4.173
36. Armstrong RA. When to use the Bonferroni correction. *Ophthalmic Physiol Optics.* (2014) 34:502–8. doi: 10.1111/opo.12131
37. Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing medical imaging data for machine learning. *Radiology.* (2020) 295:4–15. doi: 10.1148/radiol.2020192224
38. Piechnik SK, Neubauer S, Ferreira VM. State-of-the-art review: stress T1 mapping–technical considerations, pitfalls and emerging clinical applications. *Magn Reson Mater Phys Biol Med.* (2018) 31:131–41. doi: 10.1007/s10334-017-0649-5
39. Bergamino M, Bonzano L, Levrero F, Mancardi GL, Roccatagliata L. A review of technical aspects of T1-weighted dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) in human brain tumors. *Phys Med.* (2014) 30:635–43. doi: 10.1016/j.ejmp.2014.04.005
40. Jaubert O, Arrieta C, Cruz G, Bustin A, Schneider T, Georgiopoulos G, et al. Multi-parametric liver tissue characterization using MR fingerprinting: Simultaneous T1, T2, T2*, and fat fraction mapping. *Magn Reson Med.* (2020) 84:2625–35. doi: 10.1002/mrm.28311
41. Zheng JQ, Lim NH, Papież BW. D-net: siamese based network for arbitrarily oriented volume alignment. In: Reuter M, Wachinger C, Lombaert H, Paniagua B, Goksel O, Rekik I, editors. *Shape in Medical Imaging.* Cham: Springer International Publishing (2020). p. 73–84. doi: 10.1007/978-3-030-61056-2_6
42. Zhang Q, Hann E, Werys K, Wu C, Popescu IA, Lukaschuk E, et al. Deep learning with attention supervision for automated motion artefact detection in quality control of cardiac T1-mapping. *Artif Intell Med.* (2020) 110:101955. doi: 10.1016/j.artmed.2020.101955
43. Lara Hernandez KA, Rienmüller T, Baumgartner D, Baumgartner C. Deep learning in spatiotemporal cardiac imaging: a review of methodologies and clinical usability. *Comput Biol Med.* (2021) 130:104200. doi: 10.1016/j.compbiomed.2020.104200
44. Hann E, Piechnik S, Popescu IA, Zhang Q, Werys K, Ferreira V. *Method and Apparatus for Quality Prediction.* WIPO: Patent WO 2020/161481 A1 (2020).
45. Hann E, Popescu IA, Zhang Q, Gonzales RA, Barutçu A, Neubauer S, et al. Deep neural network ensemble for on-the-fly quality control-driven segmentation of cardiac MRI T1 mapping. *Med Image Anal.* (2021) 71:102029. doi: 10.1016/j.media.2021.102029
46. Hann E, Gonzales RA, Popescu IA, Zhang Q, Ferreira VM, Piechnik SK. Ensemble of deep convolutional neural networks with monte carlo dropout sampling for automated image segmentation quality control and robust deep learning using small datasets. In: Papie z BW, Yaqub M, Jiao J, Namburete AIL, Noble JA, editors. *Medical Image Understanding and Analysis.* Cham: Springer International Publishing (2021). p. 280–93. doi: 10.1007/978-3-030-80432-9_22
47. Zhang Q, Piechnik S, Werys K, Popescu IA, Ferreira V. *Validation of Quantitative Magnetic Resonance Imaging Protocols.* WIPO: Patent WO 2020/234570 A1 (2020).
48. Zhang Q, Werys K, Popescu IA, Biasiolli L, Ntusi NAB, Desai M, et al. Quality assurance of quantitative cardiac T1-mapping in multicenter clinical trials — A T1 phantom program from the hypertrophic cardiomyopathy registry (HCMR) study. *Int J Cardiol.* (2021) 330:251–8. doi: 10.1016/j.ijcard.2021.01.026

Conflict of Interest: SP has patent authorship rights for U.S. patent 9285446 B2 systems and methods for Shortened Look-Locker Inversion Recovery ShMOLLI cardiac gated mapping of T1, granted March 15, 2016; licensed to Siemens Medical. KW is an employee of Circle Cardiovascular Imaging since 2019.

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 16 December 2021
doi: 10.3389/fcvm.2021.771246

Machine Learning for the Prediction of Complications in Patients After Mitral Valve Surgery

Haiye Jiang[1,2†], Leping Liu[3†], Yongjun Wang[4], Hongwen Ji[5], Xianjun Ma[6], Jingyi Wu[7], Yuanshuai Huang[8], Xinhua Wang[9], Rong Gui[3], Qinyu Zhao[10*] and Bingyu Chen[11*]*

[1] Clinical Laboratory, The Third Xiangya Hospital, Central South University, Changsha, China, [2] Hunan Engineering Technology Research Center of Optoelectronic Health Detection, Changsha, China, [3] Department of Transfusion, The Third Xiangya Hospital, Central South University, Changsha, China, [4] Department of Blood Transfusion, The Second Xiangya Hospital, Central South University, Changsha, China, [5] Department of Anesthesiology, Fuwai Hospital National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China, [6] Department of Blood Transfusion, Qilu Hospital of Shandong University, Jinan, China, [7] Department of Transfusion, Xiamen Cardiovascular Hospital Xiamen University, Xiamen, China, [8] Department of Transfusion, The Affiliated Hospital of Southwest Medical University, Luzhou, China, [9] Department of Transfusion, Beijing Aerospace General Hospital, Beijing, China, [10] College of Engineering & Computer Science, Australian National University, Canberra, ACT, Australia, [11] Department of Transfusion, Zhejiang Provincial People's Hospital, Hangzhou, China

OPEN ACCESS

Edited by:
Paul Leeson,
University of Oxford, United Kingdom

Reviewed by:
Antonino S. Rubino,
University of Campania Luigi Vanvitelli, Italy
Behshid Ghadrdoost,
Shaheed Rajaei Cardiovascular Medical and Research Center, Iran

***Correspondence:**
Rong Gui
aguirong@163.com
Bingyu Chen
1844035880@qq.com
Qinyu Zhao
qinyu.zhao@anu.edu.au

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Heart Surgery, a section of the journal Frontiers in Cardiovascular Medicine

Received: 06 September 2021
Accepted: 02 November 2021
Published: 16 December 2021

Citation:
Jiang H, Liu L, Wang Y, Ji H, Ma X, Wu J, Huang Y, Wang X, Gui R, Zhao Q and Chen B (2021) Machine Learning for the Prediction of Complications in Patients After Mitral Valve Surgery. Front. Cardiovasc. Med. 8:771246. doi: 10.3389/fcvm.2021.771246

Background: This study intended to use a machine learning model to identify critical preoperative and intraoperative variables and predict the risk of several severe complications (myocardial infarction, stroke, renal failure, and hospital mortality) after cardiac valvular surgery.

Study Design and Methods: A total of 1,488 patients undergoing cardiac valvular surgery in eight large tertiary hospitals in China were examined. Fifty-four perioperative variables, such as essential demographic characteristics, concomitant disease, preoperative laboratory indicators, operation type, and intraoperative information, were collected. Machine learning models were developed and validated by 10-fold cross-validation. In each fold, Recursive Feature Elimination was used to select key variables. Ten machine learning models and logistic regression were developed. The area under the receiver operating characteristic (AUROC), accuracy (ACC), Youden index, sensitivity, specificity, F1-score, positive predictive value (PPV), and negative predictive value (NPV) were used to compare the prediction performance of different models. The SHapley Additive ex Planations package was applied to interpret the best machine learning model. Finally, a model was trained on the whole dataset with the merged key variables, and a web tool was created for clinicians to use.

Results: In this study, 14 vital variables, namely, intraoperative total input, intraoperative blood loss, intraoperative colloid bolus, Classification of New York Heart Association (NYHA) heart function, preoperative hemoglobin (Hb), preoperative platelet (PLT), age, preoperative fibrinogen (FIB), intraoperative minimum red blood cell volume (Hct), body mass index (BMI), creatinine, preoperative Hct, intraoperative minimum Hb, and intraoperative autologous blood, were finally selected. The eXtreme Gradient Boosting algorithms (XGBOOST) algorithm model presented a significantly better predictive performance (AUROC: 0.90) than the other models (ACC: 81%, Youden index: 70%, sensitivity: 89%, specificity: 81%, F1-score:0.26, PPV: 15%, and NPV: 99%).

Conclusion: A model for predicting several severe complications after cardiac valvular surgery was successfully developed using a machine learning algorithm based on 14 perioperative variables, which could guide clinical physicians to take appropriate preventive measures and diminish the complications for patients at high risk.

Keywords: machine learning, cardiac valvular surgery, complications, predict, model

INTRODUCTION

The prevalence of unhealthy lifestyles, such as long-term high-fat diet and lack of exercise, has caused the higher and higher incidence of cardiac diseases. Patients with cardiac diseases will suffer serious morbidity and mortality without reasonable interventions, which increased the number of cardiac surgery significantly. It was discovered that more than 1 million patients with heart disease need to be treated with cardiac surgery every year worldwide (1). There has also been a sharp increase in the number of patients with valvular diseases, many of which are severe and must be treated with cardiac surgery to replace insufficient valves (2, 3). A large number of cardiac patients always along with various of complications after cardiac valvular surgery, these postoperative complications mainly including myocardial infarction, stroke, acute renal failure, death, and so on (4).

The high incidence of postoperative complications in cardiac surgery plays an important role in the exacerbation of hospital stay and hospitalization cost, reducing the quality of life and even elevating mortality after cardiac surgery (5). An eligible surgical treatment involves not only a smooth operation but also the early prediction of risks, provision of appropriate recommendations, and timely adoption of effective medical measures to avoid postoperative complications (6).

The most important process for a qualified medical treatment is the early prediction of postoperative complications (7). Clinicians generally give judgments whether patients have postoperative complications mainly based on the tests of clinical laboratory and examinations, or their clinical experiences, when patients have corresponding clinical indications after cardiac surgery (8). On the one hand, the tests or examinations for postoperative complications are time-sensitive; on the other hand, clinical experience is subjective, and many young clinicians do not have mature clinical experience. Based on the above situation, patients who underwent cardiac surgery always miss the optimal treatment window for postoperative complications. Thus, it is urgent to construct a risk predictive system that could implement the best outcome for patients.

Previous studies on predicting postoperative risks after cardiac surgery mainly on account of traditional stastics methods, such as linear models or logistic regression (9). However, these traditional methods usually focus on one or few clinical indicators. More and more studies have found that the preoperative and intraoperative indicators of a patient have an impact on the outcome of the patient (10). Meanwhile, many studies have proved that a prediction model based on machine learning has high accuracy in predicting clinical outcomes (11, 12). Therefore, we aim to construct a model based on machine learning to predict the postoperative outcomes of patients using various preoperative and intraoperative indicators, so as to provide theoretical guidance for clinical practice.

The purpose of this study was to determine the preoperative and intraoperative risk factors associated with postoperative complications in patients undergoing cardiac valvular surgery and to develop a machine learning model to predict postoperative complications.

MATERIALS AND METHODS

Study Subjects

Participants were patients aged more than 18 years but <75 who underwent cardiac valvular surgery (mitral valve replacement, mitral valvuloplasty, and tricuspid valvuloplasty) from January 2016 to December 2018 at one of the following eight tertiary hospitals: the Second Xiangya Hospital of Central South University, the Third Xiangya Hospital of Central South University, Beijing Aerospace General Hospital, Qilu Hospital of Shandong University, Fuwai Hospital National Center for Cardiovascular Diseases, Zhejiang Provincial People's Hospital, the Affiliated Hospital of Southwest Medical University, and Xiamen Cardiovascular Hospital Xiamen University. We collected 38 cases of biological valve replacement from the Third Xiangya Hospital from 2019 to 2020 for verification.

The types of surgery for cardiac valvular surgery in our study include the classical mitral valve replacement, mitral valvuloplasty, and tricuspid valvuloplasty. Since the four types of surgery account for the majority of the population, only these three procedures were included in this study.

Patients who underwent other types of surgery (coronary artery bypass grafting, CABG, atrial septal defect repair, etc.), re-cardiac surgery, or emergency surgery, and those whose missing rates of data were more than 80% were excluded.

Postoperative myocardial infarction, postoperative stroke, postoperative renal failure, and postoperative hospital mortality that occurred 48 h after the initial surgery were defined as relevant outcomes; Then we labeled patients who had at least one complication as "complication occurred" and patients who did not have any complication as "complication did not occur."

Approval was obtained from the institutional review board of the Third Xiangya Hospital of Central South University for this study (NCT03885570). The study was reported according to the recommendations of the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement. No written consent was required in view of the purely observational nature of the study.

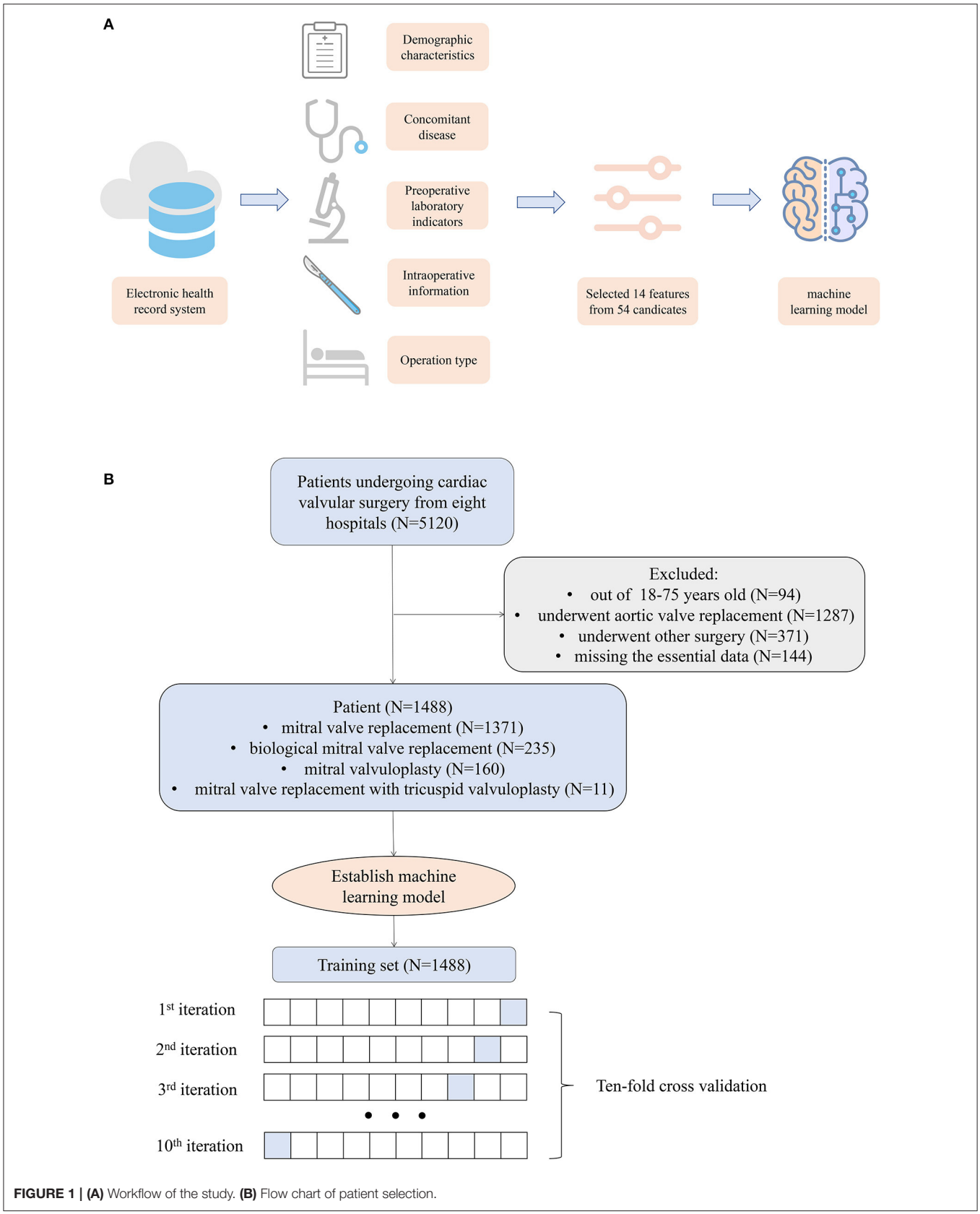


FIGURE 1 | (A) Workflow of the study. **(B)** Flow chart of patient selection.

TABLE 1 | Preoperation and intraoperative information.

Variable		All (n = 1,488)	None complication group (n = 1,433)	Complication group (n = 55)	p-value
n		1,488	1,433	55	
Gender, *n* (%)	Female	907 (60.95)	873 (60.92)	34 (61.82)	0.994
	Male	581 (39.05)	560 (39.08)	21 (38.18)	
Age, mean (SD)		52.69 (10.36)	52.44 (10.24)	59.00 (11.44)	<0.001
BMI, mean (SD)		22.84 (3.39)	22.85 (3.39)	22.55 (3.45)	0.541
Blood group, *n* (%)	A	494 (33.20)	474 (33.08)	20 (36.36)	0.436
	AB	125 (8.40)	123 (8.58)	2 (3.64)	
	B	350 (23.52)	334 (23.31)	16 (29.09)	
	O	519 (34.88)	502 (35.03)	17 (30.91)	
Atrial fibrillation, *n* (%)	1	765 (51.41)	734 (51.22)	31 (56.36)	0.541
LV dilatation, *n* (%)	1	653 (43.88)	633 (44.17)	20 (36.36)	0.314
Hypertension, *n* (%)	0	1,259 (84.61)	1,217 (84.93)	42 (76.36)	<0.001
	1	100 (6.72)	88 (6.14)	12 (21.82)	<0.001
	2	60 (4.03)	59 (4.12)	1 (1.82)	<0.001
	3	69 (4.64)	69 (4.82)	0 (0.00)	<0.001
Diabetes, *n* (%)	0	1,433 (96.30)	1,378 (96.16)	55 (100.00)	0.334
	I type	14 (0.94)	14 (0.98)	0 (0.00)	0.334
	II type	41 (2.76)	41 (2.86)	0 (0.00)	0.334
Anemia, *n* (%)	1	481 (32.33)	460 (32.10)	21 (38.18)	0.424
Drug for anemia, *n* (%)	1	5 (0.34)	4 (0.28)	1 (1.82)	0.172
Cerebrovascular disease, *n* (%)	1	1,485 (99.80)	1,430 (99.79)	55 (100.00)	1
Mechanical valve, *n* (%)	1	1,082 (72.72)	1,056 (73.69)	26 (47.27)	<0.001
Mitral valvuloplasty, *n* (%)	1	160 (10.75)	155 (10.82)	5 (9.09)	0.854
Biological valve, *n* (%)	1	235 (15.79)	211 (14.72)	24 (43.64)	<0.001
NYHA, *n* (%)	1.0	24 (1.70)	23 (1.66)	1 (3.23)	<0.001
	2.0	286 (20.21)	282 (20.38)	4 (12.90)	<0.001
	3.0	971 (68.62)	955 (69.00)	16 (51.61)	<0.001
	4.0	134 (9.47)	124 (8.96)	10 (32.26)	<0.001
ASA, *n* (%)	1	22 (1.48)	5 (0.35)	17 (30.91)	<0.001
	2	75 (5.04)	62 (4.33)	13 (23.64)	<0.001
	3	1,046 (70.30)	1,030 (71.88)	16 (29.09)	<0.001
	4	345 (23.19)	336 (23.45)	9 (16.36)	<0.001
Op time (min), median [Q1,Q3]		225.00 [190.00, 265.00]	221.00 [190.00, 263.00]	291.50 [240.00, 350.00]	<0.001
CPB time (min), median [Q1,Q3]		93.00 [74.00, 118.00]	93.00 [73.75, 117.00]	117.00 [89.50, 149.50]	<0.001
Aortic cross clamp time (min), median [Q1,Q3]		59.00 [43.25, 80.00]	58.00 [43.00, 79.00]	72.00 [59.00, 95.00]	<0.001
Cardiopulmonary bypass precharge (ml), median [Q1,Q3]		1600.00 [1505.00, 1762.50]	1600.00 [1505.00, 1800.00]	1600.00 [1600.00, 1600.00]	0.103
Blood loss op (ml), median [Q1,Q3]		600.00 [420.00, 600.00]	600.00 [450.00, 600.00]	400.00 [300.00, 600.00]	<0.001
Crystal infusion volume op (ml), median [Q1,Q3]		2100.00 [1025.00, 2650.00]	2165.00 [1100.00, 2660.00]	1500.00 [1000.00, 2000.00]	0.008
Colloid bolus op (ml), median [Q1,Q3]		300.00 [0.00, 850.00]	320.00 [0.00, 1000.00]	0.00 [0.00, 0.00]	<0.001
Urine output op (ml), median [Q1,Q3]		700.00 [400.00, 1000.00]	700.00 [420.00, 1000.00]	450.00 [300.00, 752.50]	0.001
Total output op (ml), median [Q1,Q3]		2555.00 [1100.00, 3400.00]	2600.00 [1200.00, 3420.00]	0.00 [0.00, 1500.00]	<0.001

(Continued)

TABLE 1 | Continued

Variable	All (n = 1,488)	None complication group (n = 1,433)	Complication group (n = 55)	p-value
Total input op (ml), median [Q1,Q3]	2916.68 [2400.00, 3650.00]	2950.00 [2410.00, 3700.00]	2000.00 [1500.00, 2570.00]	<0.001
Autologous blood op (ml), median [Q1,Q3]	0.00 [0.00, 250.00]	0.00 [0.00, 250.00]	0.00 [0.00, 0.00]	<0.001
Machine blood, median [Q1,Q3]	800.00 [500.00, 1000.00]	800.00 [500.00, 1000.00]	500.00 [400.00, 925.00]	0.017
SO2 min op (%), median [Q1,Q3]	97.70 [94.00, 99.70]	97.50 [94.00, 99.70]	98.15 [95.95, 99.18]	0.706
RBC (10^{12}/l), mean (SD)	4.50 (0.67)	4.51 (0.67)	4.32 (0.71)	0.054
WBC (10^9/l), mean (SD)	6.61 (3.38)	6.64 (3.42)	5.82 (1.79)	0.002
HB (g/l), mean (SD)	130.18 (20.85)	130.34 (20.69)	126.16 (24.62)	0.220
HCT (/l), mean (SD)	40.37 (5.60)	40.40 (5.56)	39.54 (6.56)	0.345
Hb min op, mean (SD)	84.58 (16.63)	84.45 (16.70)	87.91 (14.57)	0.092
HCT min op, mean (SD)	24.75 (4.97)	24.66 (4.98)	27.03 (4.23)	<0.001
PLT (10^9/l), median [Q1,Q3]	193.50 [155.00, 241.00]	194.00 [156.00, 241.25]	160.00 [116.50,234.00]	0.002
Creatinine (μmol/l), median [Q1,Q3]	71.80 [60.80, 85.00]	71.50 [60.60, 85.00]	76.90 [69.22, 92.67]	0.002
TP (g/l), median [Q1,Q3]	68.10 [63.80, 72.80]	68.10 [63.80, 72.72]	68.95 [65.82, 73.42]	0.275
Albumin (g/l), mean (SD)	39.88 (4.56)	39.92 (4.54)	38.86 (4.94)	0.126
Globulin (g/l), median [Q1,Q3]	28.00 [25.10, 31.50]	27.90 [25.00, 31.50]	29.85 [27.70, 33.35]	0.002
ALT (IU/l), median [Q1,Q3]	19.85 [13.00, 31.00]	19.90 [13.00, 31.22]	19.00 [14.00, 26.75]	0.508
AST (IU/l), median [Q1,Q3]	22.75 [18.00, 29.48]	22.70 [18.00, 29.33]	25.00 [20.00, 30.85]	0.095
PT (s), median [Q1,Q3]	13.10 [12.00,14.40]	13.20 [12.00,14.40]	11.75 [10.90,13.40]	<0.001
INR, median [Q1,Q3]	1.06 [1.00, 1.18]	1.06 [1.00, 1.18]	1.13 [1.06, 1.79]	<0.001
FIB (g/l), median [Q1,Q3]	2.90 [2.44, 3.49]	2.91 [2.44, 3.48]	2.86 [2.48, 3.71]	0.924
LVEF (%), median [Q1,Q3]	62.00 [57.00, 67.00]	62.00 [57.00, 67.00]	61.00 [56.00, 65.25]	0.152
Trans RBC before (u), median [Q1,Q3]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.048
Trans FFP before (ml), median [Q1,Q3]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00,0.00]	0.603
Trans PLT before, median [Q1,Q3]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.001
Trans cryoprecipitate before (U), median [Q1,Q3]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.845
Trans RBC op (U), median [Q1,Q3]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 1.75]	0.065
Trans FFP op (ml), median [Q1,Q3]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 290.00]	0.010
Trans PLT op, median [Q1,Q3]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.628
Trans cryoprecipitate op (U), median [Q1,Q3]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.842

SD, standard deviation; RBC, red blood cell; WBC, white blood cell; Hb, hemoglobin; Hct, red blood cell volume; PLT, platelet; TP, total protein, ALT, alanine aminotransferase; AST, aspartate aminotransferase; PT, prothrombin time; INR, international normalized ratio; FIB, fibrinogen; LVEF, left ventricular ejection fractions; FFP, fresh frozen plasma; CPB, cardiopulmonary bypass precharge; SaO_2, oxygen saturation; NYHA, New York Heart Association; ASA, The American Society of Anesthesiologists; op means the intraoperative variable.

Study Design and Data Collection

A total of 54 preoperative variables (within 24 h before the day of surgery), intraoperative variables, and postoperative variables (occurred 48 h after the initial surgery) were collected. For some preoperative variables with multiple measurements, the values closest to the start time of the surgery were assessed. The collected preoperative information included the demographic characteristics of the patients (gender, age, and body mass index, BMI), clinical characteristics (blood group, atrial fibrillation, LV dilatation), concomitant disease (hypertension, diabetes, anemia, cerebrovascular disease), history of drug use (drug for anemia), preoperative laboratory indicators (red blood cell, RBC, white blood cell WBC), hemoglobin (Hb), red blood cell volume (Hct), platelet (PLT), creatinine, total protein (TP), albumin, globulin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), prothrombin time (PT),

international normalized ratio (INR), fibrinogen (FIB), left ventricular ejection fractions (LVEF), preoperative transfusion of RBC, preoperative transfusion of fresh frozen plasma (FFP), preoperative transfusion of PLT, preoperative transfusion of cryoprecipitate, operation type (mitral valve replacement, mitral valvuloplasty, and tricuspid valvuloplasty), intraoperative information (operation time; cardiopulmonary bypass precharge CPB, time; aortic cross clamp time; cardiopulmonary bypass precharge; blood loss; crystal infusion volume; colloid bolus; urine output; total output; total input; autologous blood; machine blood; minimum oxygen saturation, SaO_2; minimum Hb; minimum Hct; intraoperative transfusion of RBC; intraoperative transfusion of FFP; intraoperative transfusion of PLT; intraoperative transfusion of cryoprecipitate), and others (Classification of New York Heart Association, NYHA, heart function; The American Society of Anesthesiologists, ASA, classification). All the variables were obtained from the electronic health record systems of the eight hospitals. Two authors (LL and HJ) had access to the systems and collected the data.

The data collected by different hospitals were converted and unified. For example, 1 mg/dl of creatinine is equal to 88.4 μmol/l. The three main types of operation were transformed into ordinal variables: mitral valve replacement, mitral valvuloplasty, and tricuspid valvuloplasty.

Statistical Analysis

Continuous variables between complication and non-complication groups were compared by either the Student *t*-test or the rank-sum test as appropriate. The chi-square test or Fisher's exact test was performed to compare the differences in the categorical variables.

Then, the recursive feature elimination (RFE) algorithm was used to identify crucial variables, and we developed a machine learning model named eXtreme Gradient Boosting (XGBOOST) (13–15). In brief, RFE is a feature selection way that recursively fits a model derived from smaller feature sets until a specified termination criterion is reached. In each loop, features are graded by their importance in the trained model. By recursively eliminating one feature with the lowest importance, RFE intends to eliminate dependencies and collinearity that maybe existing in the model. Lastly, the most important features were screened out, and the XGBOOST model was developed based on the feature set. Other features were not included, because they only brought a small increment in AUROC but significantly increased the difficulty of model applications. The proposed prediction model was built in the XGBoost package in Python language, and it was carried out using the 10-fold cross-validation method, and then the AUROC was calculated.

Besides, 10 other models, CatBoost, LightGBM, MLP, SVM, LR, Random Forest, Gradiant Boosting, KNN, AdaBoost, and Naive Bayes, were developed and compared with the proposed machine learning model. These models were also developed and validated by 10-fold cross-validation, and then the AUCs were calculated. The accuracy (ACC), Youden index, sensitivity, specificity F1 score, positive predictive value (PPV), and negative predictive value (NPV) were also analyzed.

Finally, the key variables identified by REF in each fold were merged, and the 15 most important variables were selected. The XGBOOST model was trained on the whole dataset using the merged variables. After the model was established, the SHapley Additive exPlanations (SHAP) package in Python was used to explain the model by analyzing two cases. The SHAP package interpreted the output of the machine learning model using a game-theoretic approach (16). For each prediction sample, the model connected optimal credit allocation with local explanations. Besides, a web tool was created for clinicians to use our model.

RESULTS

Study Population

As **Figure 1** demonstrates, 1,488 patients were finally included in this study, and the preoperative information of the cohort is described in **Table 1**. The average age of the patients was 52.59 years, men accounted for 39.05%, and the average BMI was 22.84. In the complication occurred cohort, 12.73% of the patients died in the hospital, 61.82% of the patients had a myocardial infarction after the operation, 30.91% of the patients had a stroke, and 74.55% of the patients had renal failure after the operation.

Key Variables

Fifteen variables, namely, intraoperative total input, intraoperative blood loss, intraoperative colloid bolus, NYHA, preoperative Hb, preoperative PLT, age, preoperative FIB, intraoperative minimum Hct, BMI, preoperative creatinine, preoperative Hct, intraoperative minimum Hb, intraoperative, and autologous blood were selected as crucial variables using the RFE algorithm. As expected, the patients had less intraoperative total output, hypertension, higher preoperative FIB, less intraoperative total input, higher preoperative creatinine, less intraoperative autologous blood, higher NYHA score, older age, higher intraoperative minimum HCT, lower preoperative Hb, lower preoperative PLT, lower intraoperative infusion volume, higher intraoperative minimum Hb, lower preoperative HCT, higher BMI, and lower intraoperative blood loss. After identifying the 15 variables, machine learning was used to predict several severe complications after cardiac valvular surgery. As shown in **Figure 2**, the AUC of the proposed model is 0.9. The proposed model significantly outperformed the conventional LR (AUC: 0.74) and seven other machine learning models. As described in **Table 2**, ACC, Youden index, sensitivity, specificity, F1-score, PPV, and NPV of the XGBoost model is 81, 70, 89, 81, 0.26, 15, and 99%, respectively. These indicators of LR were 67, 40, 69, 71, 0.15, 8, and 98%, respectively.

Application of the Model

The SHAP package analyzed the entire cohort, and showed the impact of each variable on predicting complications (**Figure 3**). The preoperative and intraoperation information of a patient was inputted into the model: age 61 years, BMI 23.44 kg/m^2, NYHA 2, intraoperative blood loss 360 ml, intraoperative colloid infusion 3,000 ml, intraoperative total input 4,350 ml,

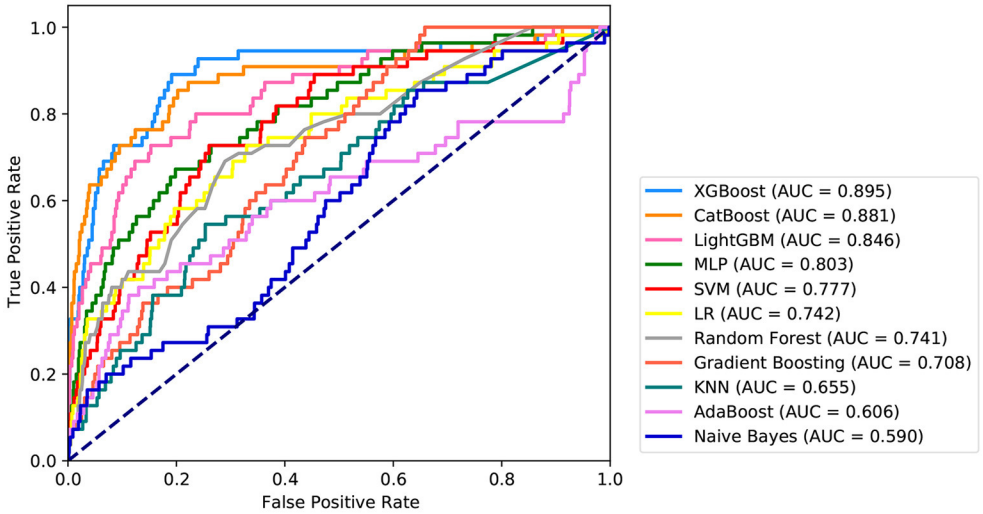


FIGURE 2 | Receiver operating characteristic curves for the machine learning model and logistic regression. XGBOOST, eXtremely Gradient Boosting; CatBoost, Categorical Boosting; LightGBM, Light Gradient Boosting; MLP, Multi-Layer Perceptron; SVM, Support Vector Machine; LR, Logistic Regression; KNN, K-Nearest Neighbor; AdaBoost, Adaptive boosting.

TABLE 2 | Performance of machine learning models.

Model	AUC	ACC (%)	Youden index (%)	Sensitivity (%)	Specificity (%)	F1 score	PPV (%)	NPV (%)
XGBoost	0.90	81	70	89	81	0.26	15	99
CatBoost	0.88	80	65	86	80	0.24	14	99
LightGBM	0.85	84	57	73	85	0.25	15	99
MLP	0.80	80	47	67	80	0.19	11	98
SVM	0.78	74	47	73	74	0.17	10	99
LR	0.74	67	40	73	67	0.14	8	98
Random forest	0.74	71	40	69	71	0.15	8	98
Gradient boosting	0.71	37	34	100	34	0.10	5	100
KNN	0.66	74	29	55	75	0.13	8	98
AdaBoost	0.61	85	27	40	87	0.16	10	97
Naive Bayes	0.59	38	21	86	36	0.09	5	98

XGBOOST, eXtremely Gradient Boosting; CatBoost, Categorical Boosting; LightGBM, Light Gradient Boosting; MLP, Multi-Layer Perceptron; SVM, Support Vector Machine; LR, Logistic Regression. KNN, K-Nearest Neighbor; AdaBoost, Adaptive boosting; ACC, accuracy, PPV, positive predictive value; NPV, negative predictive value.

intraoperative autologous blood collection 120 ml, preoperative Hb 143 g/l, intraoperative minimum Hb 57 g/l, preoperative Hct 43.1%, intraoperative minimum Hct 17%, preoperative PLT $85*10^9$/l, preoperative creatinine 80.21, and preoperative FIB 2.82 g/l. The model analyzed that the risk of adverse events in this patient was 92.4%, indicating that the probability of severe complications for the patients was high (**Figure 4A**, Example 1). The preoperative and intraoperation information of another patient was inputted into the model: age 42 years, BMI 22.89 kg/m^2, NYHA 4, intraoperative blood loss 800 ml, intraoperative colloid infusion 300 ml, intraoperative total input 2,400 ml, intraoperative autologous blood collection 0 ml, preoperative Hb 88 g/L, intraoperative minimum Hb 81 g/l, preoperative Hct 31%, intraoperative minimum Hct 81%, preoperative PLT $258*10^9$/l, preoperative creatinine 65.2 μmoI/l, and preoperative FIB 2.6 g/l. The predicted probability of adverse events in this patient was 5.3%, indicating that the patient had a good outcome (**Figure 4B**, Example 2). Furthermore, a website was established for clinicians to use the proposed model, http://www.aimedicallab.com/tool/aiml-valvecomp.html. As shown in **Supplementary Figure 1**, the predicted probabilities are significantly different between the positive and negative groups. If we use 50% as a cut off, our model will achieve a 100% accuracy.

DISCUSSION

As being mentioned above, the incidence of postoperative complications can be declined with an eligible medical treatment, including a smooth operation, the early prediction of postoperative risks, the provision of appropriate recommendations, and the timely adoption of effective medical

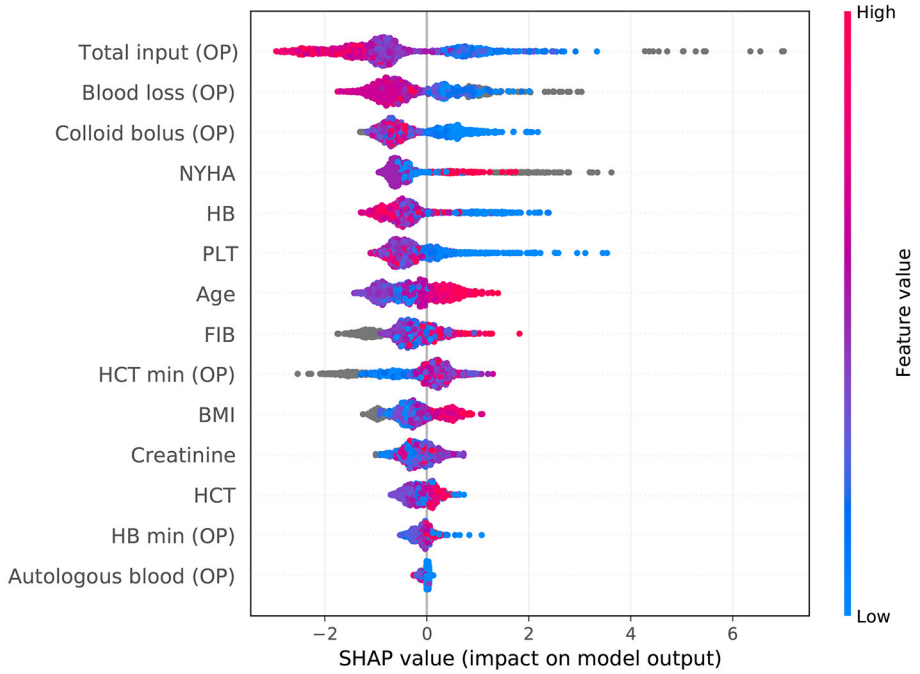


FIGURE 3 | SHAP analysis of the proposed model on the whole cohort. This figure described data from the whole cohort, with each point representing one patient. The color represents the value of the variable; blue represents the smaller value, and red represents the larger value; the horizontal coordinates represent a positive or negative correlation with severe complications risk, with a positive value indicating a good outcome and a negative value indicating a risk of severe complications. The absolute value of the horizontal coordinate indicates the contribution of variables; the greater the absolute value of the horizontal coordinate, the greater the contribution of the variables.

measures, which has been explored by many researches (6). In our study, the incidence of mortality in patients with postoperative complications was 7%, and the incidence of mortality in all cardiac surgery patients was 0.4%. The incidence of postoperative complications in all patients was 6.7%.

Based on previous studies, there is a general convergence of the incidence of postoperative complications among different hospitals. Some high-turnover institutions have low mortality rates and may be associated with higher complication rates (17). In other words, surgical patients in high-turnover facilities may experience one or more postoperative complications, but they have a low mortality rate, because these facilities have a higher rate of rescue success (18). Based on the above research studies, the ability to predict, identify, prepare, and implement the management of postoperative risks are vitally important to improve the outcomes of patients. In previous investigations, several kinds of statistic means have been discovered to figure out outcomes, analyze manifestation, and construct models for improving the outcomes of cardiac surgery (19, 20). Researchers who constructed these models were primarily interested in postoperative mortality and rarely predicted other alternative outcomes, such as postoperative complications, so it is really vital to develop an effective measuring system to predict postoperative outcomes. In this study, we introduced machine learning to build the prediction model.

One of the most significant aspect is increasing the area under the receiver operating characteristic (AUROC) curves of predictive models. The AUROC of traditional predictive models is no more than 0.8 or even lower (21, 22). In our research, the model based on machine learning exhibited a perfect performance. Different methods were used to prove that our prediction model has a good predictive effect on several different postoperative complications, all of the AUROCs were more than 0.8, some of which even reached 0.9. This proved that our model has a fantastic predictive effect on postoperative complications.

We also implemented two examples into our predictive model to confirm what variables were important to the predictive model, which can provide guidance for clinicians in making medical decisions, such as how to manage the cardiac surgery. In this study, we identified 14 key indicators that had a significant impact on clinical outcomes, suggesting that clinicians should take care changes in some important variables, such as NYHA, blood loss, and creatinine (23, 24). This research also found that clinicians should pay close attention to changes in blood clotting function and kidney function of cardiac surgery patients. Most importantly, it can indicate to clinicians how likely a patient is to develop complications after cardiac surgery. Based on the above model, we also built an online open website. We can easily obtain the incidence of postoperative complications for a particular patient by entering several important variables in the corresponding column of this website. It is proved that the accuracy of our

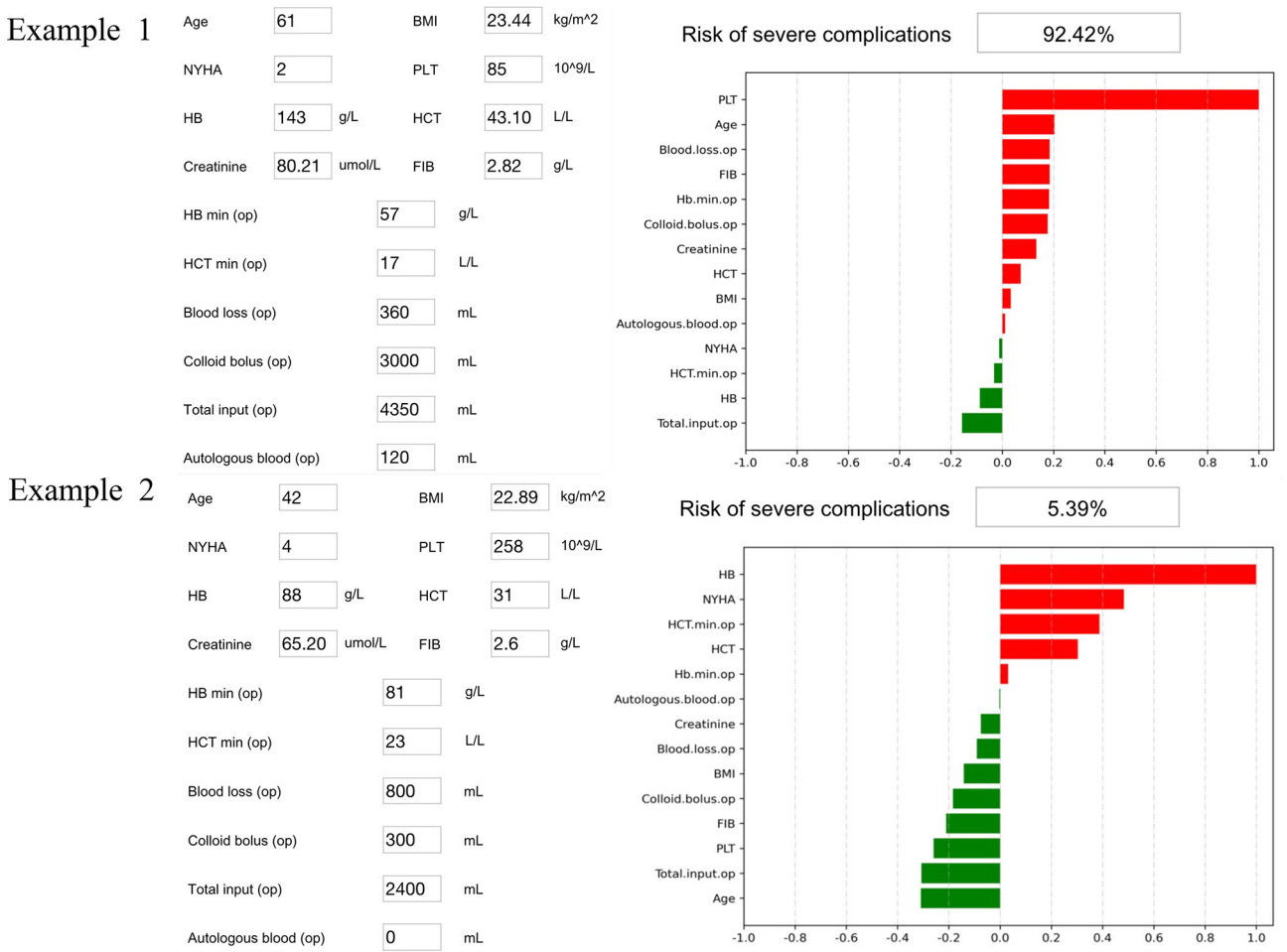


FIGURE 4 | Two examples of website tool usage. Enter the values of 14 key variables to predict the risk of severe complications and show the contribution of each value to the outcome. Example 1 has a higher risk of severe complications, and example 2 may have a better prognosis.

prediction model is very high, which can provide guidance for clinicians.

Inevitably, our research still leaves some to be desired. On the one hand, this study was a retrospective study with selection bias and confounding factors. We have enhanced the reliability of our results by incorporating multicenter data and performing robust cross-validation. At the same time, we will add prospective studies to our future studies to reduce these errors. A randomized controlled trial associated with this research should be conducted. However, the design of this type of RCT remains unclear. On the other hand, the entire process of machine learning to complete tasks operates in a black box, lacks interpretability, and is not as intuitive and clear as traditional linear models. Our results showed that the machine learning model had incomparable prediction efficiency compared with traditional linear model prediction. This study did not include patients undergoing minimally invasive mitral valve replacement, and this algorithm is not applicable to such patients. Because of the increasing use of this surgical procedure, we will include such patients in subsequent studies. Meanwhile, we have implemented a web page to promote clinical application, which is actually very meaningful and convenient.

In this study, a postoperative complication prediction model after cardiac surgery was exploited based on a machine learning algorithm, with a splendid prediction performance and convenient implementation. This model has the ability to recognize minimal risk of postoperative complications. Meanwhile, the best outcomes of patient prognosis can be achieved through an individualized assessment system. To reduce selection bias, a prospective management database for surgery patients should be built. Based on preoperative and intraoperative variables, machine learning models can be constructed and validated by the variables of surgery patients in the future. Last but not least, to measure the performance of machine learning models, a randomized controlled trial associated with this research should be conducted. It can provide suggestions for clinical work, and reduce the risk of patients and improve patient outcomes.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because. The Ethics Committee did not agree to disclose the data in a public database. Requests to access the datasets should be directed to Leping Liu, 1105380949@qq.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of the Third Xiangya Hospital of Central South University (NCT03885570). Written informed consent was not required for this study, in accordance with the local legislation and institutional requirements.

AUTHOR CONTRIBUTIONS

HJia, YW, HJi, XM, JW, YH, and XW: clinical data collection. HJia and BC: data analysis. RG, HJia, and LL: writing of the article. RG: designing of the study. QZ: construction of the model. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (Nos. 81573091 and 81802668), the Natural Science Foundation of Hunan Province (Nos. 2018JJ3776 and 2017JJ3467), and the Fundamental Research Funds for the Central Universities of Central South University under Grant No. 2020zzts892.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2021.771246/full#supplementary-material

REFERENCES

1. Pelosi P, Ball L, Schultz MJ. How to optimize critical care resources in surgical patients: intensive care without physical borders. *Curr Opin Crit Care.* (2018) 24:581–7. doi: 10.1097/mcc.0000000000000557
2. Dixon B, Santamaria JD, Reid D, Collins M, Rechnitzer T, Newcomb AE, et al. The association of blood transfusion with mortality after cardiac surgery: cause or confounding? (CME). *Transfusion.* (2013) 53:19–27. doi: 10.1111/j.1537-2995.2012.03697.x
3. Ghaferi AA, Birkmeyer JD, Dimick JB. Complications, failure to rescue, and mortality with major inpatient surgery in medicare patients. *Ann Surg.* (2009) 250:1029–34. doi: 10.1097/sla.0b013e3181bef697
4. Wang TKM, Akyuz K, Crane AD, Xu S, Xu B, Gillinov A, et al. Comparison of risk scores at predicting mortality and morbidities after isolated tricuspid valve surgery. *Circulation.* (2020) 37:126–34. doi: 10.1161/circ.142.suppl_3.12871
5. Berg KS, Stenseth R, Wahba A, Pleym H, Videm V. How can we best predict acute kidney injury following cardiac surgery? A prospective observational study. *Eur J Anaesthesiol.* (2013) 30:704–12. doi: 10.1097/EJA.0b013e328365ae64
6. Al-Lawati A, Cheung A. Transcatheter mitral valve replacement. *Interv Cardiol Clin.* (2016) 5:109–15. doi: 10.1016/j.iccl.2015.08.010
7. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit Care.* (2019) 23:112. doi: 10.1186/s13054-019-2411-z
8. Jiang W, Teng J, Xu J, Shen B, Wang Y, Fang Y, et al. Dynamic predictive scores for cardiac surgery-associated acute kidney injury. *J Am Heart Assoc.* (2016) 5:e003754. doi: 10.1161/jaha.116.003754
9. Ridgway ZA, Howell SJ. Cardiopulmonary exercise testing: a review of methods and applications in surgical patients. *Eur J Anaesthesiol.* (2010) 27:858–65. doi: 10.1097/EJA.0b013e32833c5b05
10. Dhippayom T, Dilokthornsakul P, Laophokhin V, Kitikannakorn N, Chaiyakunapruk N. Clinical burden associated with postsurgical complications in major cardiac surgeries in Asia-Oceania countries: a systematic review and meta-analysis. *J Cardiac Surg.* (2020) 35:2618–26. doi: 10.1111/jocs.14855
11. Remenyi B, ElGuindy A, Smith SC Jr, Yacoub M, Holmes DR Jr. Valvular aspects of rheumatic heart disease. *Lancet.* (2016) 387:1335–46. doi: 10.1016/s0140-6736(16)00547-x
12. Westphal S, Stoppe C, Gruenewald M, Bein B, Renner J, Cremer J, et al. Genome-wide association study of myocardial infarction, atrial fibrillation, acute stroke, acute kidney injury and delirium after cardiac surgery - a sub-analysis of the RIPHeart-Study. *Bmc Cardiovasc Disord.* (2019) 19:26. doi: 10.1186/s12872-019-1002-x
13. Yamamoto M, Kawaguchi Y, Ichida A, Matsumura M, Sakamoto Y, Arita J, et al. Evaluation of preoperative nutritional variables to predict postoperative complications after pancreaticoduodenectomy. *Nutrition.* (2019) 67–68S:100006. doi: 10.1016/j.nutx.2020.100006 PubMed
14. Jiang W, Xu J, Shen B, Wang C, Teng J, Ding X. Validation of four prediction scores for cardiac surgery-associated acute kidney injury in Chinese patients. *Braz J Cardiovasc Surg.* (2017) 32:481–6. doi: 10.21470/1678-9741-2017-0116
15. Pillutla V, Goodwin TJ, Tutungi E, Gao H, Almeida A. The use of a preoperative mitral valve model to guide mitral valve repair. *Heart Lung Circ.* (2020) 29:1704–12. doi: 10.1016/j.hlc.2020.01.018
16. Bodenhofer U, Haslinger-Eisterer B, Minichmayer A, Hermanutz G, Meier J. Machine learning-based risk profile classification of patients undergoing elective heart valve surgery. *Eur J Cardio Thorac Surg.* (2021) 60:1378–85. doi: 10.1093/ejcts/ezab219
17. Zhou C, Wang R, Jiang W, Zhu J, Liu Y, Zheng J, et al. Machine learning for the prediction of acute kidney injury and paraplegia after thoracoabdominal aortic aneurysm repair. *J Cardiac Surg.* (2020) 35:89–99. doi: 10.1111/jocs.14317
18. Fan Y, Li Y, Bao X, Zhu H, Lu L, Yao Y, et al. Development of machine learning models for predicting postoperative delayed remission in patients with cushing's disease. *J Clin Endocrinol Metab.* (2021) 106:e217–31. doi: 10.1210/clinem/dgaa698
19. Zhang Z, Zhao Y, Canes A, Steinberg D, Lyashevska O. Predictive analytics with gradient boosting in clinical medicine. *Ann Transl Med.* (2019) 7:152. doi: 10.21037/atm.2019.03.29
20. Mace JE, Xie R, Deng L, Asban A, Kim W, Pereira SJ, et al. Cardiac surgery and postoperative renal allograft failure. *Ann Thorac Surg.* (2020) 110:1904–8. doi: 10.1016/j.athoracsur.2020.03.066
21. Bohnen JD, Mavros MN, Ramly EP, Chang Y, Yeh DD, Lee J, et al. Intraoperative adverse events in abdominal surgery what happens in the operating room does not stay in the operating room. *Ann Surg.* (2017) 265:1119–25. doi: 10.1097/sla.0000000000001906
22. Benedetto U, Dimagli A, Sinha S, Cocomello L, Gibbison B, Caputo M, et al. Machine learning improves mortality risk prediction after cardiac surgery: systematic review and meta-analysis. *J Thorac Cardiovasc Surg.* (2020). doi: 10.1016/j.jtcvs.2020.07.105. [Epub ahead of print].

23. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell.* (2020) 2:56–67. doi: 10.1038/s42256-019-0138-9
24. Pasquali SK, He X, Jacobs JP, Jacobs ML, O'Brien SM, Gaynor JW. Evaluation of failure to rescue as a quality metric in pediatric heart surgery: an analysis of the STS Congenital Heart Surgery Database. *Ann Thorac Surg.* (2012) 94:573–9; discussion 9–80. doi: 10.1016/j.athoracsur.2012.03.065

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cardiovascular Medicine

CORRECTION
published: 08 February 2022
doi: 10.3389/fcvm.2022.854588

Corrigendum: Machine Learning for the Prediction of Complications in Patients After Mitral Valve Surgery

Haiye Jiang[1†], Leping Liu[2†], Yongjun Wang[3], Hongwen Ji[4], Xianjun Ma[5], Jingyi Wu[6], Yuanshuai Huang[7], Xinhua Wang[8], Rong Gui[2], Qinyu Zhao[2,9*] and Bingyu Chen[10*]*

[1] Clinical Laboratory, The Third Xiangya Hospital, Central South University, Changsha, China, [2] Department of Transfusion, The Third Xiangya Hospital, Central South University, Changsha, China, [3] Department of Blood Transfusion, The Second Xiangya Hospital, Central South University, Changsha, China, [4] Department of Anesthesiology, Fuwai Hospital National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China, [5] Department of Blood Transfusion, Qilu Hospital of Shandong University, Jinan, China, [6] Department of Transfusion, Xiamen Cardiovascular Hospital Xiamen University, Xiamen, China, [7] Department of Transfusion, The Affiliated Hospital of Southwest Medical University, Luzhou, China, [8] Department of Transfusion, Beijing Aerospace General Hospital, Beijing, China, [9] College of Engineering & Computer Science, Australian National University, Canberra, ACT, Australia, [10] Department of Transfusion, Zhejiang Provincial People's Hospital, Hangzhou, China

Keywords: machine learning, cardiac valvular surgery, complications, predict, model

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Rong Gui
aguirong@163.com
Bingyu Chen
1844035880@qq.com
Qinyu Zhao
qinyu.zhao@anu.edu.au

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Heart Surgery, a section of the journal Frontiers in Cardiovascular Medicine

Received: *14 January 2022*
Accepted: *17 January 2022*
Published: *08 February 2022*

Citation:
Jiang H, Liu L, Wang Y, Ji H, Ma X, Wu J, Huang Y, Wang X, Gui R, Zhao Q and Chen B (2022) Corrigendum: Machine Learning for the Prediction of Complications in Patients After Mitral Valve Surgery. Front. Cardiovasc. Med. 9:854588. doi: 10.3389/fcvm.2022.854588

A Corrigendum on

Machine Learning for the Prediction of Complications in Patients After Mitral Valve Surgery
by Jiang, H., Liu, L., Wang, Y., Ji, H., Ma, X., Wu, J., Huang, Y., Wang, X., Gui, R., Zhao, Q., and Chen, B. (2021). Front. Cardiovasc. Med. 8:771246. doi: 10.3389/fcvm.2021.771246

In the published article, there was an error in affiliation for authors Xinhua Wang and Rong Gui. Instead of "Xinhua Wang[7], Rong Gui[8*]", it should be "Xinhua Wang[8], Rong Gui[2*]".

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

frontiers in Cardiovascular Medicine

CORRECTION
published: 24 March 2022
doi: 10.3389/fcvm.2022.890752

Corrigendum: Machine Learning for the Prediction of Complications in Patients After Mitral Valve Surgery

Haiye Jiang[1,2†], Leping Liu[3†], Yongjun Wang[4], Hongwen Ji[5], Xianjun Ma[6], Jingyi Wu[7], Yuanshuai Huang[8], Xinhua Wang[9], Rong Gui[3], Qinyu Zhao[10*] and Bingyu Chen[11*]*

[1] Clinical Laboratory, The Third Xiangya Hospital, Central South University, Changsha, China, [2] Hunan Engineering Technology Research Center of Optoelectronic Health Detection, Changsha, China, [3] Department of Transfusion, The Third Xiangya Hospital, Central South University, Changsha, China, [4] Department of Blood Transfusion, The Second Xiangya Hospital, Central South University, Changsha, China, [5] Department of Anesthesiology, Fuwai Hospital National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China, [6] Department of Blood Transfusion, Qilu Hospital of Shandong University, Jinan, China, [7] Department of Transfusion, Xiamen Cardiovascular Hospital Xiamen University, Xiamen, China, [8] Department of Transfusion, The Affiliated Hospital of Southwest Medical University, Luzhou, China, [9] Department of Transfusion, Beijing Aerospace General Hospital, Beijing, China, [10] College of Engineering & Computer Science, Australian National University, Canberra, ACT, Australia, [11] Department of Transfusion, Zhejiang Provincial People's Hospital, Hangzhou, China

Keywords: machine learning, cardiac valvular surgery, complications, predict, model

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Rong Gui
aguirong@163.com
Bingyu Chen
1844035880@qq.com
Qinyu Zhao
qinyu.zhao@anu.edu.au

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Heart Surgery, a section of the journal Frontiers in Cardiovascular Medicine

Received: *06 March 2022*
Accepted: *07 March 2022*
Published: *24 March 2022*

Citation:
Jiang H, Liu L, Wang Y, Ji H, Ma X, Wu J, Huang Y, Wang X, Gui R, Zhao Q and Chen B (2022) Corrigendum: Machine Learning for the Prediction of Complications in Patients After Mitral Valve Surgery. Front. Cardiovasc. Med. 9:890752. doi: 10.3389/fcvm.2022.890752

A Corrigendum on

Machine Learning for the Prediction of Complications in Patients After Mitral Valve Surgery
by Jiang, H., Liu, L., Wang, Y., Ji, H., Ma, X., Wu, J., Huang, Y., Wang, X., Gui, R., Zhao, Q., and Chen, B., (2021). Front. Cardiovasc. Med. 8:771246. doi: 10.3389/fcvm.2021.771246

In the published article, there were two errors in the affiliations. The first error is that "Haiye Jiang[1†]" should have a new affiliation 2 "Hunan Engineering Technology Research Center of Optoelectronic Health Detection, Changsha, Hunan, China," and change it to "Haiye Jiang[1,2†]". The second error is "Qinyu Zhao[2,9*]" should have affiliation 2 removed "Department of Transfusion, The Third Xiangya Hospital, Central South University, Changsha, Hunan, China," and change it to " Qinyu Zhao[10*]".

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

frontiers in Cardiovascular Medicine

REVIEW
published: 21 December 2021
doi: 10.3389/fcvm.2021.787614

The Role of AI in Characterizing the DCM Phenotype

Clint Asher[1,2*†], *Esther Puyol-Antón*[1†], *Maleeha Rizvi*[1,2], *Bram Ruijsink*[1,2,3], *Amedeo Chiribiri*[1,2], *Reza Razavi*[1,2‡] *and Gerry Carr-White*[1,2‡]

[1] *Department of Cardiovascular Imaging, School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom,* [2] *Department of Cardiology, Guys and St Thomas' NHS Trust, London, United Kingdom,* [3] *Division of Heart and Lungs, Department of Cardiology, University Medical Center Utrecht, Utrecht, Netherlands*

Dilated Cardiomyopathy is conventionally defined by left ventricular dilatation and dysfunction in the absence of coronary disease. Emerging evidence suggests many patients remain vulnerable to major adverse outcomes despite clear therapeutic success of modern evidence-based heart failure therapy. In this era of personalized medical care, the conventional assessment of left ventricular ejection fraction falls short in fully predicting evolution and risk of outcomes in this heterogenous group of heart muscle disease, as such, a more refined means of phenotyping this disease appears essential. Cardiac MRI (CMR) is well-placed in this respect, not only for its diagnostic utility, but the wealth of information captured in global and regional function assessment with the addition of unique tissue characterization across different disease states and patient cohorts. Advanced tools are needed to leverage these sensitive metrics and integrate with clinical, genetic and biochemical information for personalized, and more clinically useful characterization of the dilated cardiomyopathy phenotype. Recent advances in artificial intelligence offers the unique opportunity to impact clinical decision making through enhanced precision image-analysis tasks, multi-source extraction of relevant features and seamless integration to enhance understanding, improve diagnosis, and subsequently clinical outcomes. Focusing particularly on deep learning, a subfield of artificial intelligence, that has garnered significant interest in the imaging community, this paper reviews the main developments that could offer more robust disease characterization and risk stratification in the Dilated Cardiomyopathy phenotype. Given its promising utility in the non-invasive assessment of cardiac diseases, we firstly highlight the key applications in CMR, set to enable comprehensive quantitative measures of function beyond the standard of care assessment. Concurrently, we revisit the added value of tissue characterization techniques for risk stratification, showcasing the deep learning platforms that overcome limitations in current clinical workflows and discuss how they could be utilized to better differentiate at-risk subgroups of this phenotype. The final section of this paper is dedicated to the allied clinical applications to imaging, that incorporate artificial intelligence and have harnessed the comprehensive abundance of data from genetics and relevant clinical variables to facilitate better classification and enable enhanced risk prediction for relevant outcomes.

Keywords: dilated cardiomyopathy, cardiac magnetic resonance, late gadolinium enhancement, artificial intelligence, deep learning

OPEN ACCESS

Edited by:
Shane Nanayakkara, Alfred Hospital, Australia

Reviewed by:
Ian Paterson, University of Alberta, Canada
Filippo Cademartiri, Gabriele Monasterio Tuscany Foundation (CNR), Italy
Paul Leeson, University of Oxford, United Kingdom

****Correspondence:***
Clint Asher
clint.1.asher@kcl.ac.uk

†These authors share first authorship

‡These authors share senior authorship

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

Received: *01 October 2021*
Accepted: *02 December 2021*
Published: *21 December 2021*

Citation:
Asher C, Puyol-Antón E, Rizvi M, Ruijsink B, Chiribiri A, Razavi R and Carr-White G (2021) The Role of AI in Characterizing the DCM Phenotype. Front. Cardiovasc. Med. 8:787614. doi: 10.3389/fcvm.2021.787614

INTRODUCTION

Dilated Cardiomyopathy (DCM) merely describes a dilated and dysfunctional left ventricle (LV) in the absence of significant coronary disease, valvular dysfunction, or poorly controlled hypertension. Documentation of LV size and ejection fraction (EF) are the established measurements by echocardiography or CMR that define phenotype and determine risk stratification. CMR is considered the gold standard, as it provides accurate volume assessment, morphology, function, and tissue characterization all within a single assessment to better describe underlying cardiac pathology.

It is increasingly appreciated that DCM is not simply the single disease entity of "non-ischemic" heart failure, but rather, represents a unique family of heart muscle diseases with complex interactions between genetic predisposition and environmental precipitants (see **Figure 1**) (1–3).

As such a clinical spectrum of DCM exists, with variable expression of arrhythmic and functional changes over time (4). Genetic testing clearly provides a fundamental insight into discriminating part of these diverse DCM subtypes; however, the complex interplay of genetics and environmental influences dictates for a deeper characterization of the DCM phenotype through advanced imaging techniques. This would also be warranted in the matter of risk stratification, which to date, remains particularly challenging for this cohort and appears to be inadequate when focused on the single parameter of LVEF (5). Evidently, a non-negligible proportion of DCM patients suffering from sudden cardiac death have much milder reductions in LVEF that do not meet consensus criteria for primary prevention implantable cardioverter-defibrillator (ICD) (6). Furthermore, at least a third of adverse events can occur later in the course of the disease, negating some of the reliability of static, solitary measures of systolic function in predicting long-term outcomes in DCM cohorts (7). There remains a relative lack of robust markers for stratifying patients with the DCM phenotype, and this is highlighted in the DANISH study, suggesting a limited benefit of primary prevention ICD on overall mortality in patients with non-ischaemic heart failure (8). By extracting a multitude of information generated from images and clinical datasets, Artificial Intelligence (AI) potentially holds the essential link to uncovering some of the complex associations between clusters of DCM patients in a fully automated manner. By shifting toward better characterization, it may be ultimately possible to integrate these disease characteristics and multiple novel markers, thereby advancing the refined risk stratification needed in DCM cohorts. This capability does not replace, but rather should augment the clinical decision process in a more efficient, user-friendly way, that hopefully translates into improved patient care.

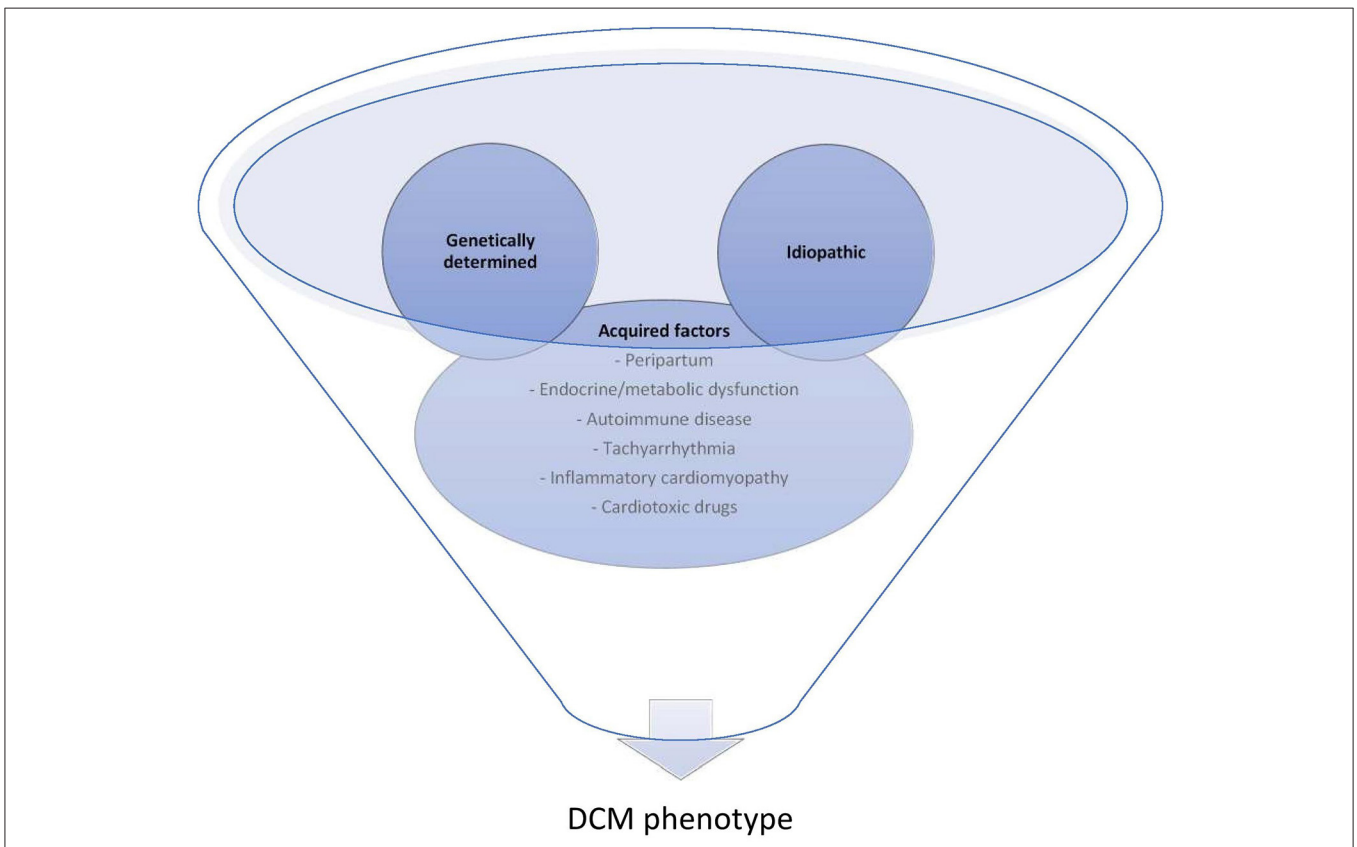


FIGURE 1 | Complex interplay of environment with genetic factors contribute to the DCM phenotype. Commonly overlooked acquired factors that are either reversible factors for those with "idiopathic" DCM or can contribute to the clinical expression or progression of those with underlying genetic predisposition.

The rest of the article is organized as follows, firstly, we provide a summary of current methods for diagnosis and characterization of DCM utilizing CMR techniques, followed by recent and key applications of AI within this scope. Subsequently, we highlight the use of AI for risk prediction in DCM and methods that combine imaging and genetic information in DCM characterization. Finally, we provide relevant discussions on current research efforts and future work towards more comprehensive and personalized imaging stratification of this heterogenous phenotype.

CURRENT DIAGNOSIS AND CHARACTERIZATION OF DCM

DCM is a heterogenous myocardial disease characterized by several degrees of reduced LVEF. Whilst the majority with this phenotype benefit from outcomes that improve year on year with up to 90% alive and well at 10 years from diagnosis, the natural history remains variable, with often unidentified initiating triggers, and some individuals unfortunate enough to succumb to unheralded life-threatening arrhythmias and sudden cardiac death at the onset of their clinical presentation (9–13).

Understanding the characteristics, evolution and long-term prognosis are key challenges to enabling proper etiological classification, customized surveillance and initiation of appropriate, effective treatment in a timely fashion. Although evaluation in practice rarely deviates from the protocol-driven investigation of heart failure, the heterogenous nature of the disease that directly results in variable clinical and phenotypic expression, dictates for a comprehensive, DCM-focused investigation strategy. Furthermore, risk stratification based on the simplistic evaluation of LV dimensions and LVEF is clearly inadequate across the phenotypic spectrum and our current grasp of suitable predictors of outcomes is still limited.

The Role of Imaging in DCM

Following detailed history, clinical examination, electrocardiogram (ECG) and laboratory tests that may elucidate features of a specific underlying etiology or secondary organ dysfunction, imaging techniques play a crucial role in confirming the diagnosis, ruling out other competing causes for LV dysfunction, further evaluation of the etiology and in guiding treatment strategies.

Whilst two-dimensional echocardiography is often first line in the diagnostic imaging pathway and has an additional role in both early and follow up function assessment in DCM patients, its role in defining an underlying etiology is limited, particularly with the compromise that occurs in light of inadequate acoustic windows and poor endocardial border definition. Furthermore, due to the inherent geometric assumptions that perform well in healthy individuals with normal sized hearts, volume assessment in those with distorted ventricular size and shape is less reliable, with significant intra- and interobserver variability.

CMR is well-placed in this respect, with unrestricted field of view and high spatial resolution to capture global and regional changes in structure and function irrespective of ventricular geometry or patient habitus (14). As there is less operator dependence for endocardial delineation, the interobserver reproducibility variability for volume and EF quantification is less for CMR than it is in echocardiography (14). This is ideal for both the initial evaluation, where decisions on initiation of medical therapy are based on LVEF thresholds, but also to carefully monitor progression of the disease and the appropriate selection of those who require device implantation. The integration of perfusion and whole heart angiography enables the exclusion of significant coronary disease with a high accuracy, thereby reducing the need for separate ischaemia assessment by computed tomography (CT) or invasive coronary angiography in the initial work up of DCM (15). Thus far, routine use of CMR for diagnosis alone has not been shown to significantly improve the clinical identification of non-ischaemic heart failure causes (16). However, complementary information is offered with tissue characterization and parametric mapping sequences that enable assessment of changes to intrinsic myocardial properties correlating with altered biological pathways. These additional features offer the potential to aid the differentiation of the underlying etiology, enable prognostic assessments and guide treatment options. Although there remains a lack of data from large randomized controlled trials asserting the role of contemporary CMR on impacting patient outcomes, the evolving landscape of techniques and applications for in-depth phenotyping paired with advanced analytics pave an important path toward CMR-guided precision care in the DCM population.

CMR for Dynamic Cardiac Assessment

Standard CMR provides the gold standard for biventricular volume assessment, further allowing for accurate documentation of systolic function, which is imperative for the investigation of all comers with heart failure. Even though current clinical practice focusses on these static measures obtained from only two end time points of the cardiac cycle, due to real time acquisition over multiple phases, cine-CMR possesses additional information on dynamic volumetric changes. Consequently, it is feasible to generate volume/time profile curves that allow evaluation of continuous ventricular volume changes and extraction of more sensitive parameters of cardiac function such as peak filling rates (see **Figure 2**) (17). From this, additional indices of filling and ejection are possible to obtain simultaneously, with the potential for more detailed analysis of both systolic and diastolic function (18). Differing LV filling patterns have already been suggested to exist amongst DCM patients with direct implications on the classification of functional status and predicting adverse outcomes (19, 20). However, most studies that assessed these parameters in the DCM phenotype were significantly limited in the diversity of structural and functional heterogeneity seen in most contemporary DCM cohorts, thus hindering full exploration into the evolutions of filling and ejection patterns in different subgroups and their varied clinical outcomes (18, 21). Such studies are warranted but the current tools for obtaining these parameters are limited by the extent of user-interface involved in semiautomatic processing, whereby contours are determined not only at each slice level but also at

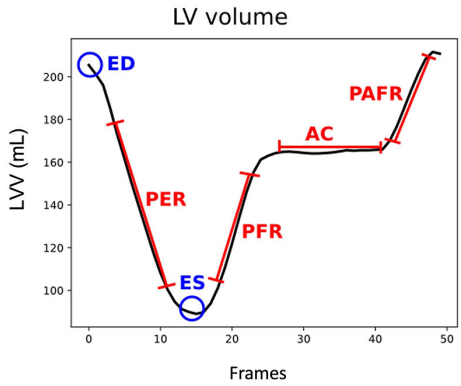


FIGURE 2 | LV volume (LVV) curve for a cardiac cycle, in blue end diastole (ED) and end systole (ES) frames, in red peak ejection rate (PER), peak filling rate (PFR), atrial contribution (AC), and peak atrial filling rate (PAFR) parameters.

each time point or phase prior to the computation of volumes needed to produce the curves for each patient.

CMR Tissue Characterization

Unique tissue characterization sequences add a further dimension to the investigative prowess of CMR in the evaluation of the DCM phenotype. The ability to non-invasively assess and quantitate myocardial tissue properties makes CMR well-suited to unravel the onset and extent of pathogenic processes occurring within the myocardium, that could only previously be determined through high-risk invasive cardiac biopsy.

Late Gadolinium Enhancement- CMR

Tissue characterization using the late gadolinium enhancement (LGE)-CMR technique enables the identification and quantification of regional areas of replacement fibrosis; this refers histologically to a process of reparative microscopic scarring occurring in response to myocyte necrosis (21). It has been found to be a clinically useful tool for distinguishing DCM from other important differentials of LV dysfunction such as coronary disease or sarcoidosis, subtyping the etiology of DCM, as well as for predicting the trajectory of the disease (**Figure 3**) (22, 23). Up to 45% of DCM patients are affected, usually in a mid-wall distribution, with <15% showing an ischaemic pattern that crucially, would not be sufficient to explain the degree of ventricular dysfunction (22, 24).

There is substantial clinical evidence that the presence of fibrosis and its detection *via* LGE-CMR heralds a strong and independent predictor of adverse outcomes in patients with non-ischaemic cardiomyopathy even in the absence of heart failure symptoms (24–32). This is a powerful parameter in the era of personalized risk stratification, especially when current criteria for prophylactic ICD implantation on the basis of significant LV dysfunction has low sensitivity for identifying some high risk patients whose clinical outcomes are not consistently related to LVEF (5).

The identification and extent of LGE at an early stage of the cardiomyopathic disease provides additional information beyond LVEF, thus enabling earlier prognostic characterization and drawing attention to those who might benefit from closer surveillance or earlier consideration of advanced therapies (25, 30). In the study by Gulati et al. (25), mid-wall fibrosis detected by LGE-CMR imaging in a longitudinal study of 472 patients with DCM, was incrementally associated with all-cause mortality and cardiovascular death or transplantation across the entire range of LVEF. In another study of 150 patients, up to 30% with the mutation *PLN* (phospholamban) p.Arg14del had LGE on CMR with a normal LVEF, suggesting this to be an early feature and higher risk of arrhythmias in carriers of this mutation, but also attesting to the phenotypic insights CMR offers for those with underlying genetic substrate (33).

The identification of LGE in clinical practice and certainly demonstrated in the majority of studies evaluating LGE-CMR in DCM, occurs mostly by visual analysis which is clearly subject to inter-observer variability (24). Elucidating the extent of LGE is apparently complementary to detecting its presence in terms of the additional risk stratification beyond conventional criteria. Neilan et al. (31) assessed the extent of LGE using quantitative methods in 162 patients with non-ischaemic cardiomyopathy and assessed for the annual major adverse cardiac events (MACE), including cardiovascular death and appropriate ICD therapy. Over a follow up period for a mean of 29 ± 18 months, quantified LGE extent demonstrated the strongest predictor of MACE over age, sex and LVEF in multivariate analyses with an adjusted HR 7.61, $p < 0.0001$.

Although quantitative methods might provide more consistent validation for the presence of LGE and a measure of the extent of fibrosis, there are also a number of practical limitations; these include the lack of universal access to quantitative software packages, variable extent of fibrosis quantified by different methods and dependence on supplementary, time-consuming contouring of LGE areas (34, 35). Moreover, LGE which relies on differences in signal intensity between healthy myocardium and focal fibrotic areas, appears to be limited in its ability to assess and quantitate diffuse (non-focal) myocardial injury and interstitial fibrosis (36, 37). From a technical perspective, LGE is also affected by inconsistencies in acquisition parameters, such as choice inversion time (TI), and in post-processing when signal intensity thresholds may be arbitrarily applied to distinguish normal myocardium from fibrotic tissue. Finally, despite the strong prognostic value in identifying high risk patients, randomized controlled trials evaluating LGE-based risk stratification are still warranted prior to any guideline recommendation on its use in managing non-ischaemic heart failure cohorts.

T1 Mapping

Refined methods in quantitative assessment of tissue characteristics enable routine measurement of diffuse fibrosis, without the reliance on regional differences in tissue contrast intensity (38). Novel techniques comprising of native (non-contrast) and contrast-enhanced T1 mapping represent advances in CMR that enable detection of pathological changes occurring within myocytes and the interstitium in a number of disease states (38). Native T1 is additionally helpful in those unable to

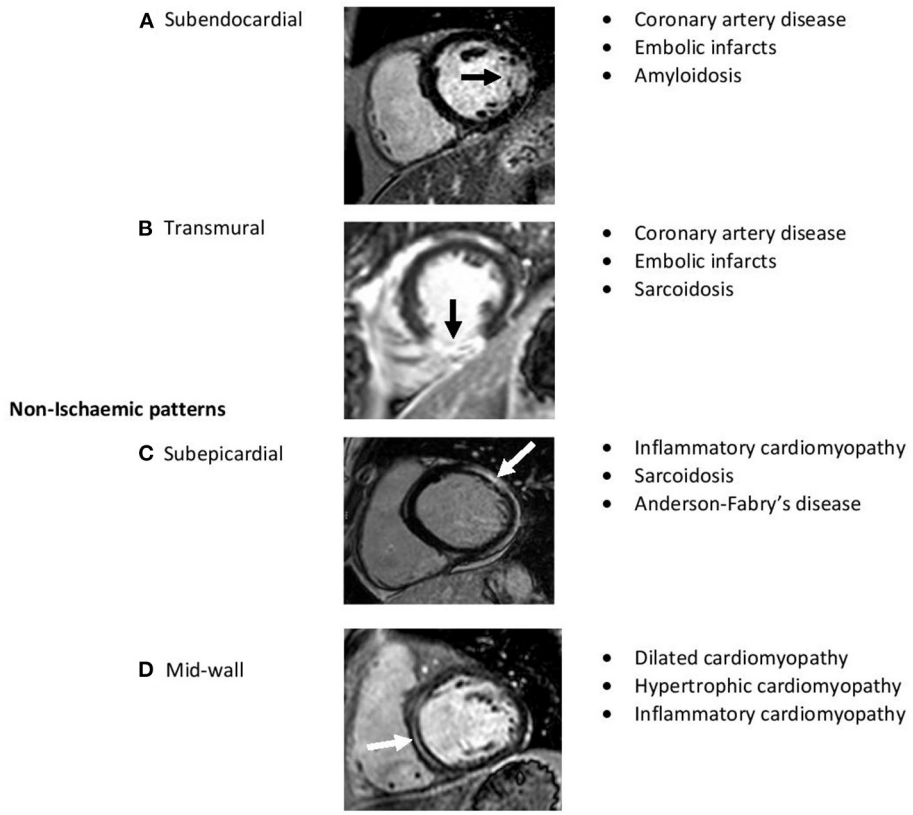


FIGURE 3 | Short-axis late-gadolinium-enhanced CMR images demonstrating hyperenhancement (arrows) indicative of scar. The differing patterns help characterize various myocardial diseases. **(A,B)** Represent typical ischaemic scar pattens involving subendocardium. **(C,D)** Represent non-ischaemic scar patterns which typically involve epicardium to mid wall.

have contrast due to contraindications such as pregnancy or severe renal failure. The acquisition of relaxation times during the same cardiac phase enables T1 values to be displayed as a pixelwise map, which can be used to directly quantify myocardial T1 values globally and at specific regions. As this process is not reliant on tissue contrast differences, T1 mapping overcomes the limitations of LGE imaging in detecting diffusely diseased myocardium but has the potential to detect and track myocardial structural alterations throughout the clinical course of disease expression (39, 40).

In DCM, the feasibility of T1 mapping as a surrogate of diffuse fibrosis has been demonstrated at different stages of the clinical phenotype, suggesting a potential biomarker role for non-hazardous follow up in the progression of different DCM cohorts (36, 41–43). This notion is further upheld in the multicenter study of over 600 DCM patients, where T1 indices both regionally and globally showed significant predictive associations with all-cause mortality and likelihood of heart failure-related mortality or hospitalization over a median follow up of 22 months, $p < 0.001$ (44). In a recent study of DCM patients affected by complex ventricular arrhythmias, events thought to be attributable to pathologic remodeling and the inter-related process of diffuse fibrosis, global native T1 time was found to be independently associated with ventricular arrhythmias even after adjustment for LVEF and scar on LGE-imaging (odds ratio 1.14, 95% confidence interval 1.03–1.25; $p = 0.008$) (45). Whilst these studies demonstrate the incremental value T1 mapping may provide in the evaluation of DCM, substantial overlap in T1 values is apparent between those with adverse outcomes and those without (44–46). Accounting for this precise continuum of T1 values with pixel-to-pixel mapping may more reliably differentiate higher and lower risk groups of patients but would be technically difficult and laborious with current manual techniques.

Pre- and post-contrast T1 mapping can also be adjusted for haematocrit, i.e., correcting for the blood volume of distribution, and this introduces an additional technique known as the extracellular volume fraction (ECV), for more focused examination of alterations occurring specifically within the extracellular interstitial compartments (40). ECV appears to have direct relationship with the extent of diffuse fibrosis with good correlation to histopathological quantification and therefore offers a non-invasive, quantifiable assessment of interstitial disease that shows significant promise in prediction of heart

failure related outcomes in DCM patients (36, 44, 47). Currently, ECV still has a limited role in differentiating DCM from other causes of non-ischaemic heart failure, due to the significant overlap of values seen across various myocardial diseases (48). However, its particular advantage appears to lie in its reduced sensitivity to variation in scanner field strength, which lends itself applicable toward multi-center and vendor evaluations to assess the extent of its usefulness over LGE-CMR in future DCM studies (47).

T2 and T2* Mapping

T2 weighted sequences exploit the biological parameter of T2 relaxation times associated with tissue water content. As such, T2 images and subsequently quantitative T2 mapping can be used for the assessment of myocardial oedema, adding to the aetiological evaluation of active myocardial inflammation as occurs in acute myocarditis (49). The clinical application of T2 mapping to provide additional diagnostic information in distinguishing DCM from healthy myocardium, with the former showing larger and more progressive myocardial water content, was recently supported in a meta-analysis (standardized mean difference 1.90, $p < 0.01$) (50, 51). This could have a pivotal role in the evaluation and differentiation of those who have the functional appearance of DCM due to athletic training from those with pathological myocardial disease (52). However, differentiation of DCM from other forms of non-ischaemic cardiomyopathy is limited in this respect, due to similar changes in T2 values, and due to differences in the sequence acquisition these values may vary from center to center (50, 53). Further research is needed in regards to standardization, verification of its usefulness and timing in the diagnostic pathway, and to better understand the pathophysiological basis for an increase in T2 values in DCM without preceding myocarditis.

T2*(star) relaxation mapping is a parameter that shortens due to the local magnetic field homogeneity that occurs with progressive iron deposition. This is useful for the assessment and quantification of iron loading within the myocardium, which can occasionally be associated with a DCM-like phenotype (54). It is a clinically validated tool, with better predictive capability than serum iron biochemistry and can detect the effects of myocardial iron loading earlier than conventional cardiac function assessments (54, 55). As a result, rapid hematological diagnostic pathways can be primed without delay and the response to treatment serially monitored non-invasively using this tool (55).

CMR for Prognostication in DCM

CMR can confirm and reproduce the assessment of LV mass, volumes, and LVEF, all of which are important indicators for a worse prognosis in severe DCM and other causes of heart failure; the latter two markers being key targets for reverse remodeling and myocardial recovery (56–59). The main limitation of these measures for predictive outcomes is that they are often assessed at initial evaluation, failing to account for the dynamic nature of the disease with favorable response to therapy for a significant proportion of patients; concurrently, they are less sensitive for those with mild-moderate dysfunction who are still prone to significant risk of sudden cardiac death (25, 26, 60).

Risk stratification in this setting is difficult and the current focus of this has shifted toward a multiparametric, dynamic approach, which attempts to incorporate potential biomarkers from biochemistry, ECG signals and imaging (12).

There is increasing evidence for applications within CMR to guide prognostication and subsequent clinical management in DCM. Whilst the majority of these applications for risk prediction are captured through routine assessment, the additional tools, and longitudinal follow-up capability is still regarded as an investigational field of interest within the setting of CMR (14). These current and potential clinical CMR applications in the risk assessment of DCM are outlined in **Table 1**.

Much of the current CMR tools for characterization and predicting outcomes in DCM rely on multiple dedicated imaging sequences, followed by significant time devoted to qualitative post-processing in the evaluation of structure, function and tissue characterization. Despite their feasibility and utility, they are often not fully exploited in clinical practice due to these time constraints on clinical workflow. Even if employed, this often occurs *ad-hoc* and limited to one or two additional parameters evaluated in uniform manner, rather than assimilated in multi-parametric fashion for personalized characterization and risk stratification. Fully integrated analysis of all these features and metrics could aid better selection of patients who might benefit from earlier medical intervention, need closer surveillance regardless of LVEF, and those who we can more confidently discharge or halt medical therapies following improvement in their cardiac status (5, 13, 66).

The Role of Genetics in DCM Characterization

It is increasingly appreciated that DCM has a genetic basis, with disease causing variants identified in up to 40% of families of DCM and 25% of presumed sporadic cases (67). Some of these genetic mutations can predispose carriers toward significant brady- or tachy-arrhythmias, or in the presence of environmental factors such as alcohol, can be the driver for a more severe phenotype, and there is a suggestion that a genetic basis could explain the higher prevalence of DCM seen in particular ethnic groups (67–69). There has been an expansion in the reported breadth of genes associated with the DCM phenotype, particularly in recent times with the arrival of next generation sequencing methodology (70, 71). However, robust genotype-phenotype correlations are not always feasible as the genes implicated encode proteins with a variety of different functional properties, making it challenging to harmonize the extent of genetic influence on the spectrum of structural and functional changes in those with DCM (71). Furthermore, it is challenging to clinically define and manage the large number of variants of uncertain significance (VUS), inadvertently arising as a result of the high throughput of current genetic testing.

Being able to discern the full scope of genetic influence in those with DCM will further help tease underlying drivers of disease manifestation and offer the opportunity to establish a

TABLE 1 | The current and potential clinical CMR applications for predictive outcomes in DCM.

CMR Biomarker	Current use	Studies supporting biomarker for prognostication in DCM cohorts	No. of patients studied	Median F/U	HR/OR for primary end point (95% CI) $p<0.05$
LV volume and LVEF	Clinical use	Masci et al. (30)	125	1.2 years	Primary endpoint = CV death and HF hospitalization. LVEDVi HR 1.02 (1.00–1.03), LVEF HR 0.94 (0.90–0.99).
		Gulati et al. (25)	472	5.3 years	Primary endpoint = ACM, cardiac transplantation. LVEF per 1% HR 0.95 (0.93–0.96). LV-EDV index per 10 ml/m^2 HR 1.09 (1.05–1.13), LVMi per 10 g/m^2 1.12 (1.04–1.19).
		Masci et al. (26)	228	1.9 years	Primary endpoint = CV death, congestive heart failure, aborted SCD. LVEDVi HR 1.008(1.000–1.016), LVEF HR 0.962 (0.934–0.990), LVMi HR 1.018 (1.006–1.030).
		Buss et al. (60)	210	5.3 years	Primary endpoint = aborted SCD, CV death, cardiac transplantation. LVEDi HR 1.02 (1.01–1.03), LVEF HR 0.91 (0.88–0.94), LVMi HR 1.11 (1.04–1.18).
RV volume and RVEF	Clinical use	Alpendurada et al. (61)	60	2.2 years	Primary endpoint = ACM, CV hospitalization. RVEF HR 0.96 (0.94–0.99) TAPSE HR 0.88 (0.80–0.96).
		Gulati et al. (62)	250	6.8 years	Primary endpoint = ACM, cardiac transplantation. RVEDVi per 10 ml/m^2 HR 1.14 (1.05–1.25), RVEF HR 0.95 (0.93–0.97).
		Becker et al. (63)	168	2.2 years	Primary endpoint = ACM, cardiac transplantation, sustained ventricular arrhythmia, appropriate ICD therapy. RVEF per 10% HR 0.74 (0.57–0.95).
LA volume and dimension	Clinical use	Gulati et al. (64)	483	5.3 years	Primary endpoint = ACM or cardiac transplantation. LAVi per 10 ml/m^2 HR 1.08 (1.01–1.15).
LGE	Clinical use	Assomull et al. (28)	101	1.8 years	Primary endpoint = ACM, hospitalisations for CV event. LGE HR 3.4 (1.4–8.7).
		Cho et al. (32)	79	1.6 years	Primary endpoint = rehospitalisation, cardiac transplantation or death. LGE HR 8.06 (1.03–63.41).
		Masci et al. (30)	125	1.2 years	Primary endpoint = CV death and HF hospitalization. LGE HR 3.96 (1.53–10.3).
		Leyva et al. (27)	97	2.8 years	Primary endpoint = CV death and transplantation. LGE HR 22.0 (4.73–102).
		Neilan et al. (31)	162	2.4 years	Primary endpoint = MACE, which included composite of cardiovascular death and a ventricular arrhythmia, terminated by the ICD. LGE presence HR 14.5 (6.06–32.61).
		Gulati et al. (25)	472	5.3 years	Primary endpoint = ACM, cardiac transplantation. LGE per 1% increment 1.11 (1.06–1.17).
		Masci et al. (26)	228	1.9 years	Primary endpoint = CV death, congestive heart failure, aborted SCD. LGE extent HR 5.104 (2.783–9.361).
		Perazzolo Marra et al. (29)	137	3 years	Primary endpoint = SCD, sustained ventricular arrhythmia, appropriate ICD intervention. LGE presence HR 4.17 (1.56–11.2).
		Puntmann et al. (44)	637	1.8 years	Primary endpoint = ACM. LGE presence HR 2.9 (1.4–6.3).
T1 Mapping	Research tool	Barison et al. (43)	89	2 years	Primary endpoint = composite of cardiovascular death, hospitalization for heart failure, and appropriate defibrillator intervention. ECV HR 8.59 × 107 (1,503–4.80 × 1,012).
		Puntmann et al. (44)	637	1.8 years	Primary endpoint = ACM. Native T1 HR 1.1 (1.06–1.15), ECV per % change HR 1.1(1.05–1.14).
		Nakamori et al. (45)	107	Retrospective events	Primary endpoint = ventricular arrhythmia. Native T1 each 10-ms increment OR 1.14 (1.03–1.25).
FT-CMR: LV strain	Research tool	Buss et al. (60)	210	5.3 years	Primary endpoint = combination of CV death, heart transplantation, and aborted SCD. GLS HR 1.33 (1.21–1.47), GCS HR 1.23 (1.13–1.34), GRS HR 0.89 (0.84–0.95).
		Romano et al. (65)	507	4.4 years	Primary endpoint = all-cause death. GLS HR 1.402 (1.299–1.513).

LV, left ventricular; LVEF, left ventricular ejection fraction; HF, heart failure; LVEDVi, indexed left ventricular end diastolic volume; ACM, all-cause mortality; LV-EDV, left ventricular end diastolic volume; LVMi, indexed left ventricular mass; SCD, sudden cardiac death; CV, cardiovascular; RV, right ventricular; RVEF, right ventricular ejection fraction; TAPSE, tricuspid annular plane systolic excursion; RVEDVi, indexed right ventricular end diastolic volume; ICD, implantable cardioverter defibrillator; LA, left atrial; LAVi, indexed left atrial volume; LGE, late gadolinium enhancement; MACE, major adverse cardiac events; ECV, extracellular volume; FT-CMR, feature tracking-cardiac magnetic resonance imaging; GLS, global longitudinal strain; GCS, global circumferential strain; GRS, global radial strain.

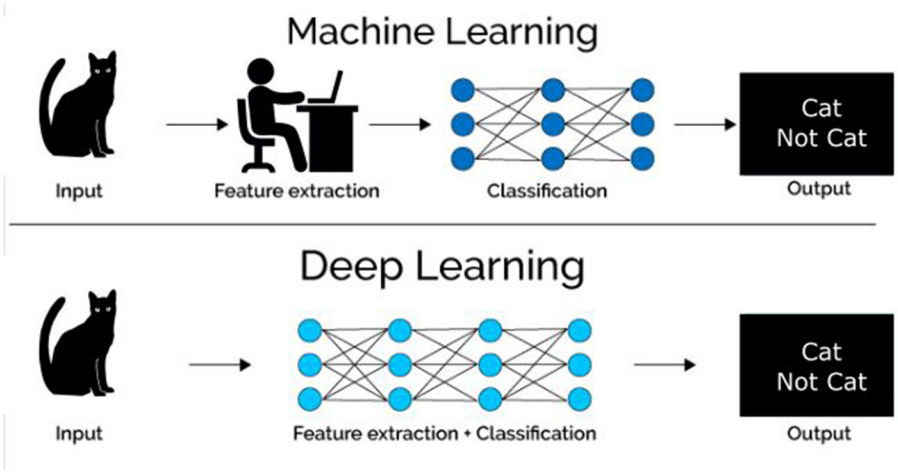


FIGURE 4 | Visual example of the differences between machine learning (ML) and deep learning (DL) methods.

deeper characterization of the phenotype. Coupled with the challenge to examine genetic influence on the spectrum of structural and functional changes, the addition of mutation status to clinical and imaging parameters may improve risk stratification and potential treatment strategies beyond the consensus management for heart failure with reduced ejection fraction (HFrEF) (72).

AI APPLICATIONS IN THE CMR CHARACTERIZATION OF DCM

AI is the division of computer science that deals with the ability of computer systems to use algorithms in order to interpret and learn from data, and successfully perform tasks that would normally require human intellect and input. Over time, we have seen AI gaining popularity in medicine, having applications within medical record mining, predictive clinical application systems, virtual patient care and, its widest application, medical imaging (73, 74). In short, AI has the potential to perform routine tasks more efficiently or provide new insights into disease processes, that were previously not achievable by manual review and analysis due to time and labor constraints (75).

AI, machine learning (ML), and deep learning (DL) are three terms often used interchangeably but are essentially hierarchical. AI is the overarching concept aiming to develop computers with human intelligence. ML is the subfield of AI that gives computers the ability to learn without being explicitly programmed (76). DL is a subset of ML algorithms called neural networks. Neural networks are algorithms that mimic the human brain's behavior in decision-making and try to find the most optimal path to a solution. Traditionally, ML methods contain a feature engineering phase, where experts propose a set of hand-crafted features to facilitate the learning from examples. This phase is very important and affects the overall performance of the learning system. In a DL pipeline, feature extraction is embedded in the learning algorithm where features are extracted in a fully automated way and without any intervention of a human expert (see **Figure 4** for visual example of the ML and DL method). A number of fundamental neural network architectures lie at the basis of DL models, and we provide a basic introduction to their concepts. However, for a more comprehensive overview of these architectures and DL algorithms for cardiac image segmentation, we refer the interested reader to Chen et al. (77), Convolutional neural networks (CNN) are the most popular class of DL network, widely applied in CMR, utilizing a patch-based image extraction approach (see **Figure 5A** for an example of a CNN network). As opposed to this conventional neural network, a fully convolutional neural network (FCNN) performs more efficient and accurate pixelwise segmentation by leveraging upsampling layers to concatenate multi-scale features obtained through a series of convolutions applied to the entire image (see **Figure 5B** for an example of a FCNN) (77). Finally, another emerging class of DL algorithms are the Generative adversarial networks (GAN). These consist of a pair of neural networks, contesting one against another ("adversarial"), in order to generate new, synthetic instances of data that can pass for real data (see **Figure 5C** for an example of GAN architecture).

There are three important types of AI algorithms: (1) Supervised learning algorithms try to model relationships and dependencies between the target prediction output and the input features or observations such that we can predict the output values for new data based on the learnt relationships. This is the most partial and widely adopted form of AI, but it requires a large amount of labeled training datasets; (2) Unsupervised learning is where there are no corresponding output variables, and the goal is to discover relationships between the input features or reveal the latent variables behind the observations; and (3) Reinforcement learning aims to learn a mapping from situations to actions so as to maximize a scalar reward or reinforcement signal. A key difference with supervised learning is that the reinforcement learning agent is never told the optimal action,

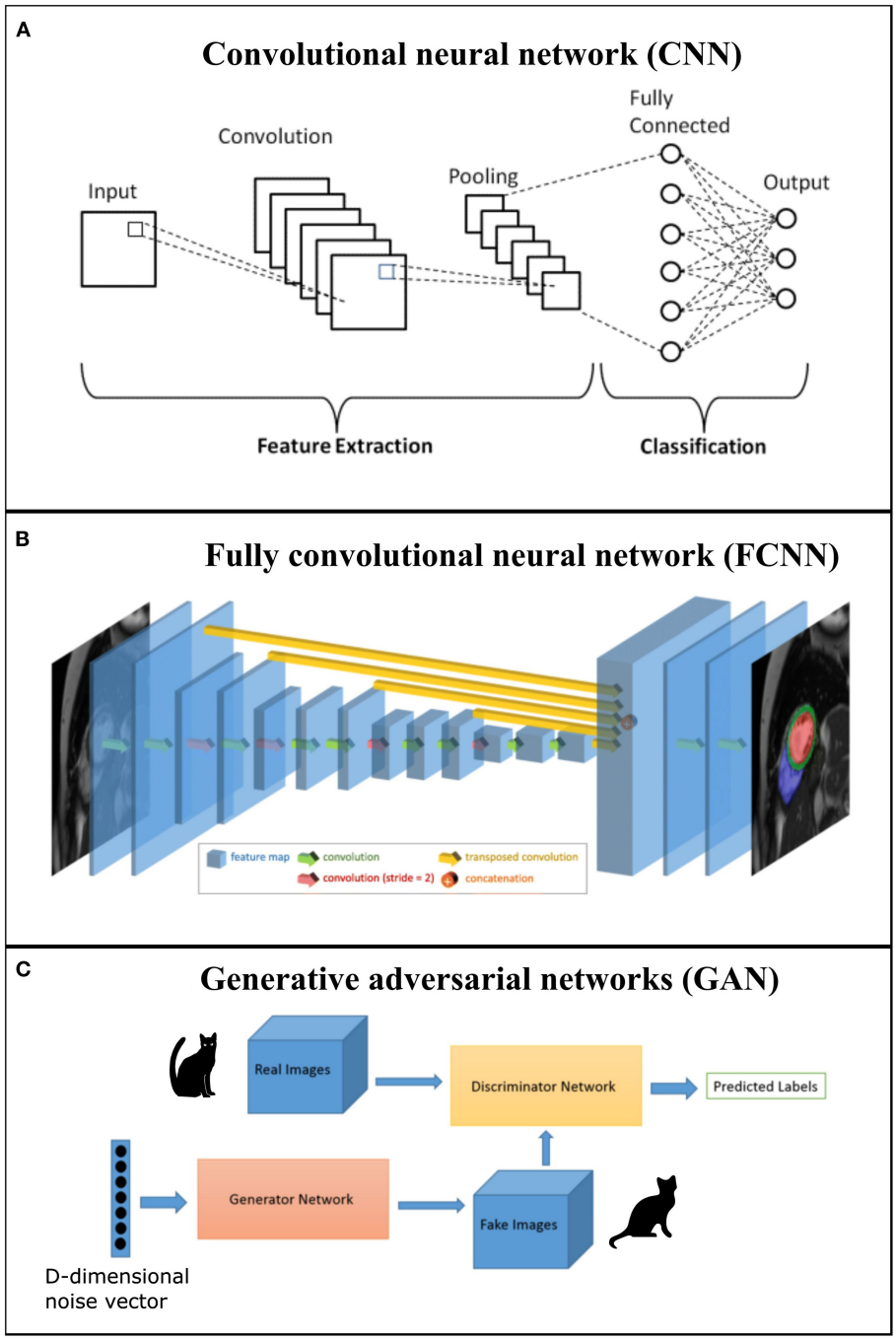


FIGURE 5 | (A) Example of a convolution neural network (CNN) where first section corresponds to the feature extraction and second section to classification; **(B)** Example of a generic fully convolutional neural network (FCNN) with feature map volumes that are color-coded by size. Figure adapted from Bai et al. (78); **(C)** Example of a generative adversarial networks (GAN) that comprises two networks (generator and discriminator).

instead it receives an evaluation signal indicating the goodness of fit for the selected action.

Medical image analysis involves the use of images generated in clinical practice, that can be interpreted to improve our ability to solve clinical problems and make treatment decisions more effective (79). As the increasing wealth of digital data becomes more accessible, clinicians need to be able to find more efficient ways of meaningfully combining this data to boost precision-based healthcare.

Due to the spatial and temporal pathologic heterogeneity of particular clinical phenotypes, such as DCM, the ability to accurately identify and extract relevant imaging biomarkers in

routine clinical practice is prone to subjective errors and has low reproducibility when carried out by hand. Over the last decade, AI has made significant progress in the field of medical imaging, improving techniques involved in acquisition, analysis, and interpretation with gradually less human oversight involved in the process (74). These efforts reduce the variability associated with subjective image interpretation, and moreover, enable feature extraction around regions of interest. It has the potential to provide quantifiable features that relate more objectively and in more detail with relevant clinical information (80).

Overcoming the Barriers to Fully Automated Image Segmentation

The opening act to CMR characterization involves image acquisition and segmentation, prior to feature extraction. The quality of this step is essential to the outcome of further downstream analysis and provides baseline cardiac parameters as well as the untapped features that might better describe cardiac function in specific cohorts. Assessment of the anatomical features following cine acquisition includes assessment of myocardium, pericardium, all 4 cardiac chambers, valves and vascular connections. Typically, a visual quality assessment is needed first to ensure the signal-to-noise ratio is enhanced by adequate positioning and breath-holding technique, limited blurring by cardiac gating, and appropriate planning for each subsequent image plane acquisition. Following the anatomical review, long and short axis cines are acquired that enable dynamic views of the global heart function. Segmentation by manual planimetry, or by semi-automated methods with clinician oversight enables the reproducible 3-dimensional (3D) assessment of atrial and ventricular volumes, LV mass and EF quantification.

Studies have shown that DL methods can outperform conventional ML, and in some cases, even better in both detection and segmentation tasks analyzed by human expertise (80, 81). CNNs are the technique of choice and the most successful type of models for image analysis (82, 83). Efficacy of the DL models is often assessed in the form of pixel classification accuracy. Although different methods for assessing this exist, the preferred evaluation metric for DL-based segmentation approaches is the Dice metric, which evaluates the overlap between automated segmentation and the ground truth segmentation. The Dice metric has values between 0 and 100%, where 0 denotes no overlap and 100% denotes perfect agreement.

One of the main challenges to implementation of CNNs in medical imaging is the lack of high-quality expert annotated data, available for training the DL network. Furthermore, these datasets often suffer from class imbalances due to certain conditions being encountered less frequently, thereby making it more difficult for a CNN to generalize and limiting large scale CMR evaluations. As highlighted in **Table 2**, whilst the segmentation performance of state-of-the-art DL methods is commended, it is evident that the number of DCM cases encountered in these evaluations has significantly low representation. Given the heterogeneity of this condition with many individuals at presentation subject to highly remodeled ventricles and rotated cardiac axes, these current automated segmentation methods may not yet be robust enough for the deployment and evaluation of this phenotype.

To ensure any of these methods can translate into clinically useful tools in the evaluation of DCM, it is essential they are complemented by high quality datasets, that help improve the accuracy of segmentation and classification tasks, whilst providing large variability in terms of the clinical phenotype and image acquisition modules, thus enabling generalizability (87, 91). However, manual annotations of large datasets that are able to encompass this scale of heterogeneity is no easy feat, being costly and requiring extensive expert time for good quality annotation. This could be partly overcome with data sharing initiatives and collaborations between CMR centers to obtain large repositories of images with associated clinical information. This is not inevitably a seamless solution, as there are often ethical and legal requirements to satisfy within all participating sites, with limits set on how and where specific data can be utilized during the development and deployment of the pipeline.

Encouragingly, over the years open technical challenges and several publicly available datasets have been made available, helping to unravel this generalizability issue (93, 94). The UK Biobank (UKBB), although limited to a single CMR vendor, provides one of the largest imaging datasets facilitating the exploration of DL capabilities in a large general population whilst solving issues relating to ethics and clinical data aggregation. This was harnessed recently in a genome-wide association study of CMR-derived LV measurements in ∼36,000 participants from the UKBB to study the relationship between genetic variants associated with LV structure and function, and risk of incident DCM (95).

Another way of improving the generalizability during training and take advantage of the limited amount of high-quality labeled data is the strategy of data augmentation. It is possible to artificially increase the variation of examples encountered by applying random transformations such as image rotation by certain degrees, image scaling to increase variations in organ size, changing image orientation with random horizontal or vertical flips and even inclusion of random "noise" to images (94, 96, 97). Whilst this option effectively enables the acquisition of more labeled data, the diversity in practice may still be limited in terms of reflecting the full spectrum of the DCM phenotype and the pixel-level differences of images obtained from different CMR vendors. The breakthrough in improving this generalization of networks to reliably segment heterogenous phenotypes acquired from different CMR vendors and clinical sites was demonstrated recently by Chen et al. (94). Unlike the efforts to solely re-train or fine tune networks to improve the performance on a specific dataset, they explored the pre-processing step of data normalization enabling their network to deal with the distribution changes amongst input features from multi-source images. This overcomes the small differences in features arising from images obtained from different scanners and the overfitting to distribution changes that occurs with network development from a single source. Along with data augmentation strategies, their approach achieved encouraging

TABLE 2 | State of the art DL architecture on CMR datasets and number of DCM cases encountered in test datasets.

Selected Work, References	DL Architecture and type of images	Structures Segmented	No. subjects in total used for training/validation/ testing	Dice metric between AS and MS: LV cavity	Dice metric between AS and MS: LV myocardium	Dice metric between AS and MS: RV cavity	No. of DCM test cases
Bai et al. (78)	2D FCNN, SAX images	Biventricular and atria	Training 3,975 Validation 300 Testing 600	Mean 0.94 (SD 0.04)	Mean 0.88 (SD 0.03)	Mean 0.90 (SD 0.05)	142
Tran (84)	2D FCNN with transfer training, SAX images	Biventricular	Training 131 Validation 100 Testing 115 (LV), 32 (RV)	Mean 0.92 (SD 0.03)	Mean 0.96 (SD 0.01)	Mean 0.84 (SD 0.21)	Unspecified; mix of cardiac conditions
Isensee et al. (85)	Ensemble FCNN (2D and 3D U-net), SAX images over full cardiac cycle	Biventricular	Training 100 Testing 50	Mean 0.945	Mean 0.905	Mean 0.908	10
Tao et al. (86)	2D FCNN, SAX images from multivendor dataset	LV/Myocardium	Training 400 Testing 196	Mean 0.92 (SD 0.06)	Mean 0.94 (SD 0.05)		46
Khened et al. (87)	2D Densenet (FCNN), SAX images	Biventricular	Training 700 Validation 300 Testing 490	Mean 0.93 (SD 0.05)	Mean 0.89 (SD 0.03)	Mean 0.91 (SD 0.05)	10
Jang et al. (88)	2D M-net (FCNN), weighted cross entropy loss, SAX images	Biventricular	Training 80 Testing 20	Mean 0.938 (SD 0.05)	Mean 0.879 (SD 0.04)	Mean 0.890 (SD 0.07)	10
Fahmy et al. (89)	2D FCNN with alignment and T1 estimation, SAX images	LV/Myocardium	Training 63 Testing 147		Mean 0.85 (SD 0.07)		Unspecified; mix of cardiac conditions
Avendi et al. (90)	2D CNN for localizing LV, stacked autoencoders for shape inference. Deformable model for segmentation, SAX images	LV	Training 45 Validation 30 Testing 30	Mean 0.94 (SD 0.02)			Unspecified; mix of cardiac conditions
Avendi et al. (91)	2D CNN for localizing RV, stacked autoencoder for automatic initialization. Deformable model for segmentation.	RV	Training 16 Testing 16			Mean 0.83 (SD 0.14)	Unspecified; mix of cardiac conditions from dataset of 48 patients
Oktay et al. (92)	2D FCNN with anatomical shape priors, SAX images	LV/Myocardium	Training 900 Validation 100 Testing 200	Mean 0.939 (SD 0.02)	Mean 0.81 (SD 0.03)		0

DL, deep learning; AS, automated segmentation; MS, manual segmentation; LV, left ventricle; RV, right ventricle; FCNN, fully convolutional neural network; CNN, convolutional neural network; SAX, short axis.

results in terms of reliable segmentation accuracy across test images from multi-scanner and site domains (mean Dice metric of 0.91 for the left ventricle, 0.81 for the myocardium, and 0.82 for the right ventricle from a single site dataset; and 0.89 for the left ventricle, 0.83 for the myocardium from a multi-site dataset).

DL-Based Global Assessment of Function

Recent DL techniques have enabled rapid expansion in the CMR domain to achieve robust contour identification and accurate classification performance, whilst significantly minimizing the extent of post-processing involved in volumetric data calculations (98). Emerging approaches that have recently become commercially available, have further demonstrated the feasibility and precision of fully automated dynamic measurement of LV volumes (99, 100). Utilizing anatomical localization methods to determine relevant boundaries between structures, contours are created in consecutive frames of the cardiac cycle with LV volume/time curves derived at no extra time expense and with high correlation to the manual reference technique (99). Whilst such applications show promise for the application to evaluating DCM cohorts, the development and training of this technique has been based on very limited representation from such patients, with less accuracy seen in those with significantly dilated ventricles and whose impaired breath-holding technique can lead to significantly more artifacts, reducing image quality. Furthermore, the details of the DL pipeline are not disclosed by the manufacturer and this lack of transparency will make it difficult to optimize the current algorithm in order to generalize to other scanners and more diverse patient cohorts.

Motivated by these limitations, Ruijsink et al. (17) developed a robust, accurate and fully automated framework for CMR cardiac function analysis which included comprehensive quality control detection using a CNN to limit erroneous output. Segmentation of both ventricles in all frames was then executed utilizing a 17-layer 2D-FCNN, prior to an iterative alignment process to correct for any differences in breath-holding and motion. After validating their framework that presented with high correlation to manual analysis, biventricular volume curves were generated for over 2,000 healthy individuals to obtain a more detailed description of cardiac function, inclusive of diastolic parameters such as peak early filling rate, atrial contribution, and peak atrial filling rate. These parameters stratified healthy patients by age categories, with lower filling rates correlating with older age–a relationship consistent with the known increase in ventricular stiffness with age (101). Considering that these LV filling patterns also appear capable of distinguishing the different categories of diastolic dysfunction characterized on echocardiography, it is anticipated that this method could enable within DCM subgroups detection of those with persistent diastolic impairment despite LV systolic recovery on medical therapy, and identify patients with subclinical disease who will require closer surveillance (102). These parameters are feasible with no additional imaging outside routine care and can occur at no extra time-cost whilst the routine clinical analysis is ongoing. In terms of the potential clinical application to evaluation of the DCM phenotype, it is significant that the method employed by Ruijsink et al. (17) performed similarly well in unseen patients with cardiomyopathy as well as those without cardiac disease. It has been reported from studies on emerging 4D flow CMR, that DCM patients have altered and heterogenous diastolic flow patterns that occur due to abnormal filling kinetics and varying degrees of pathological geometrical configuration of the LV (103). This highlights the potential role offered by fully automated LV filling assessment in differentiating those with persistent altered filling patterns and abnormal diastolic flow, thereby remaining at risk of relapse compared to those who have truly achieved recovery and remission. Based on this promising AI tool, current work by this group is also exploring the innovative use of GANs to generate realistic CMR images from any domain in order to advance the generalization of the network and robustly deal with clinical CMR data from multiple centers, vendors, and field strengths (104). Given the feasibility to evaluate both ventricular filling profiles, and the suggested prognostic role of serial revaluation of RV function in the follow up of DCM, the characteristics and clinical utility of RV filling patterns over time will be another area of application in the DCM population (105).

DL-Based Tissue Characterization

State-of-the-art algorithms utilized in scar segmentation are commonly semi-automated, fixed-model approaches where the pixel intensities of scar regions are exploited through a process of thresholding (106). This requires a user-selected area of interest and knowledge of the nearby intensities of healthy nulled myocardium, prior to operating a region growing process to segment the scar region. These methods are currently popular for segmenting contiguous regions of scar, and are highlighted for their reproducibility, with encouraging performance against consensus segmentation by experts (106). Whilst simple to implement, they remain heavily user dependent for pre-processing with respect to definition of the myocardial borders, activating the boundaries of interest, initialization for region growing in each slice, and the subjective baseline selection of remote healthy myocardium as well as the perceived extent of scar. More automated approaches have been developed to help minimize the degree of user interaction whilst maintaining reproducible performance (106). These methods mostly utilize clustering techniques to fit data of different tissue signal intensities in order to characterize the voxels belonging to scar regions (106). Whilst they show good correlation with the fixed-model approaches in accurately identifying LGE, they unfortunately are not robust enough for clinical translation due to failure to accurately segment scar where CMR-LGE images are affected by noise or share homogenous signal intensity distributions within myocardial boundaries and other nearby tissues (106–109). This limitation is particularly important as most of these traditional methods have been validated on CMR-LGE images obtained from patients with coronary disease, where the pattern of scar is subendocardial as opposed to that seen, if present, in DCM and other non-ischaemic cohorts, occurring in the mid to epicardial wall segments, where tissue intensity homogeneities are more likely to be encountered.

As the attention of CMR segmentation transitions toward more DL-based approaches, it is hoped that these innovative techniques will also facilitate a more practical and reliable means of achieving standardized quantification of LGE. This is highly desirable, given the suggestion that even after adjustment of LVEF, the proportion of LGE assists the clinical stratification of DCM patients who are prone to a higher risk of death and

hospitalization (28). Not to mention, the ability to efficiently characterize border zone tissue–areas of variable transition between scar and normal tissue that have arrhythmogenic potential, thereby helping to identify those at risk of malignant arrhythmias and more likely to benefit from ICDs (110).

A successful FCNN architecture, the ENet was recently harnessed to deal with the task of scar segmentation (111). Several variants of this popular architecture have been adapted to enhance the accuracy of different cardiac segmentation tasks (see **Table 2**). However, the pursuit of scar segmentation is a relatively new concept. In this work by Moccia et al. (111), the ENet was adapted and evaluated to see if pre-identified LV regions could enable more accurate scar segmentation than current methods, and furthermore whether a fully automated output of scar segmentation was feasible and maintained a similar or improved accuracy. As a proof of concept, this method showed both protocols were able to identify scar on the CMR-LGE images without the need for pre-processing extraction steps. However, it was the semi-automated method with a priori knowledge of the restricted myocardial boundary in which to search for scar, that outperformed state of the art CMR-LGE segmentation algorithms and was closest to expert annotation (with a sensitivity of 0.88 and Dice coefficient of 0.71). This is still an important breakthrough, holding advantages over current efforts to quantify scar by minimizing subjective evaluation, user interaction and any parameter tuning prior to implementation. Expanding the training datasets to incorporate the variability of scar seen in those with a sole DCM phenotype and those with accompanying embolic sub-endocardial scar, could be an encouraging start to help encode the high variability of scar dimensions seen in this population. This study provides an important step forward in the clinical practice of scar quantification, and by enhancing the pixel classification through training, the ENet would not only acquire improved segmentation performance, but would be more generalizable to the DCM population.

In conjunction with acquiring diverse DCM datasets, image data augmentation is another common method to artificially boost training datasets in order to improve performance and generalizability of a deep learning model. This may be particularly relevant with regards to the DCM population, where a disconnect exists between high demand for sufficient training images and the variability of scar presence across the spectrum of patients, in essence, limiting the real-world availability of training examples (24). This was recently explored in the simulation of scar tissue on the LGE-CMR images of healthy patients (112). Lau et al. (112) utilized their GAN framework which could additionally incorporate domain-specific knowledge, to simulate various scar tissue shapes in different positions. These images were highly realistic as demonstrated by the improved segmentation prediction of scar tissue pixels correctly identified during testing from 75.9 to 80.5% and the qualitative assessment that imaging experts were unable to reliably distinguish between simulated and authentic scar.

A stream of work has focused on extending the use of a trained FCNN to assist analysis of myocardial tissue characterization by means of automated native T1 mapping (89). With good agreement to manual calculations, this showed promise for an automated pipeline to minimize the workflow involved in quantifying global T1 characteristics. This was validated on a single scanner, with further study needed to see if this method can be applied to other mapping sequences such as the modified Look-Locker inversion recovery (MOLLI) or the contemporary shortened modified Look-Locker inversion (shMOLLI) method, that is more acceptable and compatible with typical limits for end-expiration breath-holding in patients (113). Puyol-Antón et al. (114) evaluated an automated framework for tissue characterization using the shMOLLI method at 1.5 Tesla using a Probabilistic Hierarchical Segmentation (PHiSeg) network. This method models the probability distribution of pixel-wise segmentation samples from the input image and generates an uncertainty map to quantify the degree of error in segmentation, so that erroneous representations are not utilized for T1 mapping. A morphological operation was then applied to detect the LV-RV intersection and delineate LV free wall from the interventricular septum. T1 ranges were obtained from the uncovered myocardial regions of interest with correction for T1 from the ventricular blood pools to improve discrimination between healthy subjects and those with cardiovascular disease (115). Using this proposed method, they characterized global and regional T1 values from over 10,000 subjects from the UKBB dataset which included a significant proportion of non-ischaemic cardiomyopathies. In line with present comprehension, they demonstrated that for those conditions in which diffuse fibrosis is more prevalent such as DCM, hypertrophic cardiomyopathy (HCM), and cardiac sarcoidosis, they found significantly higher T1 values (all $p < 0.05$). The quality control process is an important feature for clinical scalability of this tool and would enable this supplementary prognostic information to be added to each DCM case in a uniform manner with no added time-expense. Furthermore, it would enable large scale application to assess the role that native T1 analysis may have in deriving enhanced prediction of adverse outcomes for particular subgroups of DCM, particularly those who remain at risk despite having only mild or moderately impaired LV function. In order to be more generalizable, the proposed model requires validation on datasets acquired from the various different vendors available currently in clinical practice.

CLINICAL RISK PREDICTION AND THE ROLE FOR AI

Whilst clinical risk prediction models for prognostic assessment exists for heart failure populations in general, these tend to be below par when utilized in DCM patients (116). For the most part, these models are derived mostly from heart failure due to ischaemic etiology, which on average is associated with a higher mortality risk, and tends to affect older individuals who have other associated cardiovascular risk factors (116). DCM tends to affect younger patients with the vast majority having mild dysfunction remaining stable for many years. Alternatively, they can also be characterized by incidences of sudden progressive dysfunction, or by those without severe LV

dysfunction who remain susceptible to ventricular arrhythmias and sudden cardiac death; both of which would not be accounted for by conventional risk prediction models (1, 117, 118).

Emerging techniques in AI pertaining to the exploration of informative clinical biomarkers potentially offers a better appreciation of the phenotypic heterogeneity underlying DCM, with refined clinical implications in risk stratification, earlier detection and personalized treatment strategies (116). In this section, we provide an overview of these AI-based clinical applications that are primed to advance the field of clinical risk prediction in DCM.

Chen et al. (119) recently evaluated their ML model based on 32 features obtained from baseline patient characteristics, bloods, ECG, echocardiography and CMR, and assessed its performance in predicting cardiovascular events in a group of severe DCM patients. Feature selection occurred with Information Gain, an attribute selection technique that enables rapid classification of the most relevant features to the cardiovascular events. Although a number of ML models performed well in terms of accuracy and ability to discriminate between an event and non-event for each feature, a naïve Bayes classifier was selected as the model of choice due to the additional transparency offered with the generation of conditional probabilities associated with each outcome. This was the most meaningful in terms of clinical translation, as the relevant significant features could form part of the clinician's probabilistic reasoning in the decision aid for guiding a patient's treatment. This would need further exploration in subsequent iterations of the model and prospective clinical trials in order to evaluate the capability to assign risk to particular patients. Nonetheless, by handling most of these features that are often used variably in practice for risk prediction such as LGE extent, degree of mitral regurgitation, and QRS duration, this model outperformed current scoring systems and LVEF alone for the prediction of cardiovascular events in each patient [AUC, 0.887 (95% confidence interval, 0.813–0.961)].

By integrating longitudinal clinical, biochemical and echocardiography imaging data from over 4,000 patients with cancer, Zhou et al. (120) built predictive supervised ML models for applicable cardiovascular outcomes such as heart failure and *de novo* cancer therapy-related cardiac dysfunction (CTRCD). Based on a number of model iterations from five different classification methods, logistic regression provided the optimal classification performance, with an area under the receiver operating characteristic curve of 0.882 (95% CI, 0.878–0.887) for heart failure and 0.802 (95% CI, 0.797–0.807) for *de novo* CTRCD. They identified a combination of 9 clinically relevant variables that were strong predictors for these outcomes ($p < 0.05$) and maintained this high performance even when tested on data from separate time points to the training dataset. As one of the potentially reversible causes of DCM, this generalizability and high performance in predicting CTRCD over time makes ML models such as this a potentially promising tool for real-world cardiac risk assessment in cancer patients throughout their treatment journey. As these models are evaluated in larger cohorts with fine-tuning and model-specific variable selection to enhance performance, this group are also collaborating with clinicians to develop integrated risk calculators with outcomes in order to test the prospective potential of ML-derived biomarkers in cardio-oncology practice.

Treatment of DCM is predominantly as part of the management of heart failure with reduced ejection fraction. This is directed at reversal of adverse LV adaptive mechanisms that occur in progressive LV dysfunction, so called LV reverse remodeling (LVRR), and is a key determinant of prognosis in DCM (1). Up to 40% of patients are reported to experience this within two years, due to removal of the precipitating factor or induced through medical therapies and/or cardiac resynchronization therapy (CRT) in those who have left bundle branch block and subsequent dyssynchronous ventricular activation (1, 121). Beyond medical therapy, CRT in this setting has clear efficacy in terms of improving symptoms and reducing mortality (122). However, determining those who will "respond" to this therapy moving toward and maintaining remission in the long-term, as opposed to those who may be non-responders, is still a current challenge in the clinical setting (1, 123). Whilst multiple clinical, imaging, and even device implant factors are associated with the likelihood of positive response to CRT, gaps of knowledge still remain regarding timing of this evaluation and how to leverage this information to identify evidence of early remodeling (121, 123).

A number of ML algorithms have explored the combined assessment of different clinical variables in predicting response to CRT and recovery of myocardial function. Multiple kernel learning (MKL) has been used by different groups as it offers the possibility of combining data from different sources as different kernel matrices, and it learns the importance of each kernel. For example, a framework was developed by Peressutti et al. (124), which captured LV motion information from spatio-temporal atlases deployed in CMR imaging from a mixed cardiomyopathy cohort. They then applied a supervised MKL to combine and evaluate the relationship between the rich motion descriptors and selected clinical information derived from clinical reports, ECG and data from echocardiography. Although applied to a limited cohort of 34 patients selected for CRT, this coupling of electro-mechanical LV data to clinical metrics achieved an accuracy of 94% in predicting super-responders and 91% for non-responders, at 6 months post CRT implant. Future work incorporating anatomical descriptors into the atlases could potentially inform of mechanistic differences between responders and non-responders.

Cikes et al. (125) utilized an unsupervised MKL algorithm in a heart failure cohort of over 100 patients recruited from the Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy (MADIT-CRT trial). This trial had previously demonstrated the added benefits CRT added to ICD in terms of decreased risk of heart failure events in those with a low LVEF and wide QRS duration on ECG (126). In order to provide meaningful classification of this phenotypically heterogenous cohort, this algorithm was used to cluster patients by clinical characteristics, biochemical biomarkers, ECG, and echocardiography-derived patterns. They observed specific phenogroups with characteristics predictive of best volume reduction, CRT response, and overall better treatment effect on heart failure outcomes [hazard ratio (HR)

0.35; 95% confidence interval (CI) 0.19–0.64; $P = 0.0005$ and HR 0.36; 95% CI 0.19–0.68; $P = 0.001$].

Although most of the examples above come from more diverse heart failure cohorts, ML within these settings clearly has the potential for novel integration of the readily available and extensive clinical, biochemical, and imaging parameters to phenotype heterogenous diseases, such as DCM. They offer the added advantage of exploiting this information as biomarker data to unearth and compare the similarities between subgroups, and importantly provide a degree of interpretability for the associations identified (125). Understanding the value and accuracy of this output is not only relevant to understanding how to improve the ML algorithm's operation, but is fundamental for bridging the gap to advances in the clinical application of these tools. This has been a need particularly for DL algorithms, which have made impressive leaps in performance and accuracy in some image classification tasks, but are often depicted as "black boxes," offering little understanding to the prediction of their results.

Puyol-Antón et al. (127) offered in the first of its kind, an interpretable approach to a DL model for the prediction of CRT response. This framework was based around a DL-based generative model known as a variational autoencoder (VAE) which encodes the segmented biventricular data into a low dimensional latent space, followed by a primary task classifier of predicting those who would respond to CRT utilizing pre-treatment CMR images. A secondary classifier which follows a similar structure to the first, and then incorporates clinical domain knowledge to provide an explanatory concept within the encoded space. By example, they utilized the concept of septal flash–an identified pattern of early septal contraction and a marker of interventricular dyssynchrony (128). The classifiers enabled the separation of CRT responders and non-responders in the image domain with visualization of where the learned features of CRT responders corresponded to the clinical domain knowledge. This has important implications beyond predicting CRT response in DCM, with the potential ability of DL models to explore multiple validated clinical parameters involved in arrhythmia prediction and reverse remodeling as explanatory concepts, thus granting a better understanding of the disease process pathways and the varying responses of different subgroups (129).

AI TOOLS FOR INTEGRATED IMAGING-GENETICS IN DCM

Supervised ML approaches improved the prediction of DCM patients mostly likely to experience LV reverse remodeling following the novel therapeutic approach of immunoadsorption and immunoglobulin substitution (130). The integration of overlapping myocardial gene expression patterns, using a support vector machine and random forest analysis, enabled the development of a robust classifier that helped distinguish responders to therapy, and enhanced predictions beyond clinical parameters and antibody response levels alone [sensitivity of 100% (95% CI 85.8–100%); specificity up to 100% (95% CI 79.4–100%); cut-off value: −0.28].

Similarly, Schmitz et al. (131) demonstrated that ML algorithms could be applied to identify predictive combinations of clinical and genetic markers that could enhance the classification of heart failure patients likely to respond to CRT treatment. This work proposed the concept of underlying genetic substrates that may exclusively or through interaction with other factors contribute to the remodeling phenotype of certain heart failure cohorts. This additional predictive information may provide some understanding of the variable responses to CRT therapy and help improve outcomes.

High fidelity ML models incorporating genetic sequencing, 2D and 3D CMR, explored the complexities surrounding the molecular mechanisms of DCM pathogenesis, mediated by titin-truncating variants (TTN) (132). These variants frequently associate with the DCM phenotype in sporadic and familial forms, and are also reported to occur in just under 1% of the general population, where their clinical significance is less clear (133, 134). However, following mass univariate analyses in healthy individuals, integrating multiple cardiac parameters obtained through CMR imaging, anthropometric variables and their relationship to detailed sequencing of the TTN genotype, Schafer et al. (132) demonstrated association between TTN status and higher LV volume due to eccentric remodeling. In leveraging this high-resolution phenotyping, this study highlighted the feasibility and benefits of ML in estimating the effect size of candidate pathogenic mutations on multiple metrics of cardiac morphology and function that are applicable to a deeper characterization of the DCM phenotypic spectrum. Furthermore, such studies are needed to help define the clinical indicators of an inherited DCM and the mechanistic interactions between genetic variants and other conditions that share some clinical features such as peripartum cardiomyopathy (135).

The potential of AI for rapid, purposeful extraction of high-quality imaging-derived phenotypes assimilated with genetics is also a promising arena for DL methods. Following rapid LV analysis in ∼17,000 individuals by a FCNN highly optimized to automatic segmentation, the largest genome wide association study (GWAS) of image-derived phenotypes identified 14 significant loci for different LV traits that related to cardiac morphogenesis and risk of heart failure development (136). Furthermore, there were distinct loci that associated with LV remodeling, and others that were causal genes for multiple LV traits such as BAG 3 and TTN; two genes that also share implications in the pathogenesis of DCM. These findings emphasize a potential genetic basis underlining many of the structural and functional LV imaging traits routinely acquired through CMR imaging. With the unparalleled performance of fully automated imaging analysis by DL, it may be feasible to integrate this information and enrich our understanding of the pathogenic evolution of heart failure syndromes occurring in some DCM subtypes.

These promising applications highlight the unrivaled capability of AI to integrate complex structural, functional and genetic characteristics of DCM to better understand and characterize the phenotype. However, in order to universally

translate to the clinical setting, they warrant validation and replication across the spectrum of LV dysfunction in order to tease the different pathways that are involved in the evolution of DCM.

FUTURE PERSPECTIVES AND CONCLUSIONS

Advances in the applications of AI based medical innovations are rapidly increasing with a particular surge of interest within the specialties of Radiology and Cardiology (137, 138). Even now, across Europe and America, a number of innovations have already received Conformité Européenne (CE) marked or Food and Drug Administration (FDA) approval for the introduction of AI based solutions to simplify detection of cardiovascular risk and enable efficient, personalized disease prediction across a range of imaging modalities and clinical platforms (137, 138).

Despite this rapidly evolving landscape for AI opportunities in cardiac healthcare, there are still some limitations that need to be addressed before such applications can be successfully deployed into clinical practice. Firstly, the generalization of the methods, as most are only validated with high-quality data from standardized research environments which don't necessarily generalize well to external databases. To overcome this limitation, we think that AI models need to be validated in external databases that reflect real-world, heterogenous populations, and tested using decentralized techniques such as federated learning in order for them to be relevant and personalized to specific cohorts. A pioneer example of such initiative is the partnership between the British Heart Foundation and the Health Data Research UK (HDRUK), enabling access to the UK's large-scale and diverse cardiovascular data resource, where population-wide data analysis can be utilized to extract valuable information from unstructured data and investigate novel insights into cardiac disease pathways.

Another well-known pitfall of AI models is that they are "black boxes," being difficult to gauge how they reach their output decisions and predictions. Explainable AI is a new branch of AI that aims to add interpretability to the models. From our point of view, this is likely to facilitate faster adoption of AI systems into the clinical healthcare setting and will help foster vital transparency and trust with their users.

For the DCM population, this further research from AI tools is welcomed and needed to find meaningful insights that are able to enhance the rapid, reliable automation of all relevant imaging indices for characterizing the phenotype. If these could help define the relationships between imaging phenotypes, genomic features and the impact of specific precipitant factors, then it may be possible to generate biomarker profiles to discover clusters of DCM patients that have similar outcomes, to better understand their similarities and furthermore, understand the influence of different treatment strategies. These biomarker indicators would be important in redefining risk stratification beyond LVEF, enabling a multi-parametric approach that can feasibly assess dynamic changes in cardiac status and help tailor treatments to the needs of a specific subtype and more specifically, the individual.

AUTHOR CONTRIBUTIONS

CA, EP-A, BR, RR, AC, and GC-W conceived and designed the work. CA and EP-A searched and read the CMR and AI literature, and drafted the manuscript with support from MR and BR. AC, RR, and GC-W provided the critical revision with valuable feedback to improve the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This research was funded in whole, or in part by the Wellcome Trust/EPSRC Center for Medical Engineering at Kings College London (WT 203148/Z/16/Z). EP-A was supported by the EPSRC (EP/R005516/1, EP/P001009/1) and by core funding from the Wellcome/EPSRC Center for Medical Engineering (WT203148/Z/16/Z). BR was supported by the NIHR Cardiovascular MedTech Co-operative award to the Guy's and St Thomas' NHS Foundation Trust. CA and MR are supported by the UKRI London Medical Imaging and Artificial Intelligence Center for Value Based Healthcare (RE15376).

ACKNOWLEDGMENTS

We would like to thank our colleague Dr. Rosita Zakeri who assisted with the conception of **Figure 3**.

REFERENCES

1. Merlo M, Cannatà A, Gobbo M, Stolfo D, Elliott PM, Sinagra G. Evolving concepts in dilated cardiomyopathy. *Eur J Heart Fail.* (2018) 20:228–39. doi: 10.1002/ejhf.1103
2. Japp AG, Gulati A, Cook SA, Cowie MR, Prasad SK. The diagnosis and evaluation of dilated cardiomyopathy. *J Am Coll Cardiol.* (2016) 67:2996–3010. doi: 10.1016/j.jacc.2016.03.590
3. Rapezzi C, Arbustini E, Caforio ALP, Charron P, Gimeno-Blanes J, Helio T, et al. Diagnostic work-up in cardiomyopathies: bridging the gap between clinical phenotypes and final diagnosis. A position statement from the ESC Working Group on Myocardial and Pericardial Diseases. *Eur Heart J.* (2013) 34:1448–58. doi: 10.1093/eurheartj/ehs397
4. Pinto YM, Elliott PM, Arbustini E, Adler Y, Anastasakis A, Böhm M, et al. Proposal for a revised definition of dilated cardiomyopathy, hypokinetic non-dilated cardiomyopathy, and its implications for clinical practice: a position statement of the ESC working group on myocardial and pericardial diseases. *Eur Heart J.* (2016) 37:1850–8. doi: 10.1093/eurheartj/ehv727
5. Cannatà A, De Angelis G, Boscutti A, Normand C, Artico J, Gentile P, et al. Arrhythmic risk stratification in non-ischaemic dilated cardiomyopathy beyond ejection fraction. *Heart.* (2020) 106:656–64. doi: 10.1136/heartjnl-2019-315942
6. Halliday BP, Cleland JGF, Goldberger JJ, Prasad SK. Personalizing risk stratification for sudden death in dilated cardiomyopathy: the past, present, and future. *Circulation.* (2017) 136:215–31. doi: 10.1161/CIRCULATIONAHA.116.027134

7. Stolfo D, Ceschia N, Zecchin M, De Luca A, Gobbo M, Barbati G, et al. Arrhythmic risk stratification in patients with idiopathic dilated cardiomyopathy. *Am J Cardiol.* (2018) 121:1601–9. doi: 10.1016/j.amjcard.2018.02.055
8. Køber L, Thune JJ, Nielsen JC, Haarbo J, Videbk L, Korup E, et al. Defibrillator implantation in patients with nonischemic systolic heart failure. *N Engl J Med.* (2016) 375:1221–30. doi: 10.1056/NEJMoa1608029
9. Stecker EC, Vickers C, Waltz J, Socoteanu C, John BT, Mariani R, et al. Population-based analysis of sudden cardiac death with and without left ventricular systolic dysfunction. *J Am Coll Cardiol.* (2006) 47:1161–6. doi: 10.1016/j.jacc.2005.11.045
10. Merlo M, Sinagra G, Carniel E, Slavov D, Zhu X, Barbati G, et al. Poor prognosis of rare sarcomeric gene variants in patients with dilated cardiomyopathy. *Clin Transl Sci.* (2013) 6:424–8. doi: 10.1111/cts.12116
11. Kayvanpour E, Sedaghat-Hamedani F, Amr A, Lai A, Haas J, Holzer DB, et al. Genotype-phenotype associations in dilated cardiomyopathy: meta-analysis on more than 8000 individuals. *Clin Res Cardiol.* (2017) 106:127–39. doi: 10.1007/s00392-016-1033-6
12. Akhtar M, Elliott PM. Risk stratification for sudden cardiac death in non-ischaemic dilated cardiomyopathy. *Curr Cardiol Rep.* (2019) 21:155. doi: 10.1007/s11886-019-1236-3
13. Verdonschot JAJ, Hazebroek MR, Ware JS, Prasad SK, Heymans SRB. Role of targeted therapy in dilated cardiomyopathy: the challenging road toward a personalized approach. *J Am Heart Assoc.* (2019) 8:1–18. doi: 10.1161/JAHA.119.012514
14. Francone M. Role of cardiac magnetic resonance in the evaluation of dilated cardiomyopathy: diagnostic contribution and prognostic significance. *ISRN Radiol.* (2014) 2014:1–16. doi: 10.1155/2014/365404
15. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J.* (2020) 41:407–77. doi: 10.1093/eurheartj/ehz425
16. Paterson DI, Wells G, Erthal F, Mielniczuk L, O'Meara E, White J, et al. OUTSMART HF. *Circulation.* (2020) 141:818–27. doi: 10.1161/CIRCULATIONAHA.119.043964
17. Ruijsink B, Puyol-Antón E, Oksuz I, Sinclair M, Bai W, Schnabel JA, et al. Fully automated, quality-controlled cardiac analysis from CMR. *JACC Cardiovasc Imaging.* (2020) 13:684–95. doi: 10.1016/j.jcmg.2019.05.030
18. Soldo SJ, Norris SL, Gober JR, Haywood LJ, Colletti PM, Terk M. MRI-derived ventricular volume curves for the assessment of left ventricular function. *Magn Reson Imaging.* (1994) 12:711–7. doi: 10.1016/0730-725X(94)92195-4
19. Pinamonti B, Zecchin M, Di Lenarda A, Gregori D, Sinagra G, Camerini F. Persistence of restrictive left ventricular filling pattern in dilated cardiomyopathy: an ominous prognostic sign. *J Am Coll Cardiol.* (1997) 29:604–12. doi: 10.1016/S0735-1097(96)00539-6
20. Vanoverschelde J-LJ, Raphael DA, Robert AR, Cosyns JR. Left ventricular filling in dilated cardiomyopathy: Relation to functional class and memodynamics. *J Am Coll Cardiol.* (1990) 15:1288–95. doi: 10.1016/S0735-1097(10)80016-6
21. Suzuki J, Caputo GR, Masui T, Chang J-M, O'Sullivan M, Higgins CB. Assessment of right ventricular diastolic and systolic function in patients with dilated cardiomyopathy using cine magnetic resonance imaging. *Am Heart J.* (1991) 122:1035–40. doi: 10.1016/0002-8703(91)90469-X
22. McCrohon JA, Moon JCC, Prasad SK, McKenna WJ, Lorenz CH, Coats AJS, et al. Differentiation of heart failure related to dilated cardiomyopathy and coronary artery disease using gadolinium-enhanced cardiovascular magnetic resonance. *Circulation.* (2003) 108:54–9. doi: 10.1161/01.CIR.0000078641.19365.4C
23. Mitropoulou P, Georgiopoulos G, Figliozzi S, Klettas D, Nicoli F, Masci PG. Multi-modality imaging in dilated cardiomyopathy: with a focus on the role of cardiac magnetic resonance. *Front Cardiovasc Med.* (2020) 7:97. doi: 10.3389/fcvm.2020.00097
24. Di Marco A, Anguera I, Schmitt M, Klem I, Neilan TG, White JA, et al. Late gadolinium enhancement and the risk for ventricular arrhythmias or sudden death in dilated cardiomyopathy. *JACC Hear Fail.* (2017) 5:28–38. doi: 10.1016/j.jchf.2016.09.017
25. Gulati A, Jabbour A, Ismail TF, Guha K, Khwaja J, Raza S, et al. Association of fibrosis with mortality and sudden cardiac death in patients with nonischemic dilated cardiomyopathy. *JAMA.* (2013) 309:896. doi: 10.1001/jama.2013.1363
26. Masci PG, Doulaptsis C, Bertella E, Del Torto A, Symons R, Pontone G, et al. Incremental prognostic value of myocardial fibrosis in patients with non–ischemic cardiomyopathy without congestive heart failure. *Circ Hear Fail.* (2014) 7:448–56. doi: 10.1161/CIRCHEARTFAILURE.113.000996
27. Leyva F, Taylor RJ, Foley PWX, Umar F, Mulligan LJ, Patel K, et al. Left ventricular midwall fibrosis as a predictor of mortality and morbidity after cardiac resynchronization therapy in patients with nonischemic cardiomyopathy. *J Am Coll Cardiol.* (2012) 60:1659–67. doi: 10.1016/j.jacc.2012.05.054
28. Assomull RG, Prasad SK, Lyne J, Smith G, Burman ED, Khan M, et al. Cardiovascular magnetic resonance, fibrosis, and prognosis in dilated cardiomyopathy. *J Am Coll Cardiol.* (2006) 48:1977–85. doi: 10.1016/j.jacc.2006.07.049
29. Perazzolo Marra M, De Lazzari M, Zorzi A, Migliore F, Zilio F, Calore C, et al. Impact of the presence and amount of myocardial fibrosis by cardiac magnetic resonance on arrhythmic outcome and sudden cardiac death in nonischemic dilated cardiomyopathy. *Hear Rhythm.* (2014) 11:856–63. doi: 10.1016/j.hrthm.2014.01.014
30. Masci PG, Barison A, Aquaro GD, Pingitore A, Mariotti R, Balbarini A, et al. Myocardial delayed enhancement in paucisymptomatic nonischemic dilated cardiomyopathy. *Int J Cardiol.* (2012) 157:43–7. doi: 10.1016/j.ijcard.2010.11.005
31. Neilan TG, Coelho-Filho OR, Danik SB, Shah RV, Dodson JA, Verdini DJ, et al. CMR quantification of myocardial scar provides additive prognostic information in nonischemic cardiomyopathy. *JACC Cardiovasc Imaging.* (2013) 6:944–54. doi: 10.1016/j.jcmg.2013.05.013
32. Cho JR, Park S, Choi BW, Kang S-M, Ha J-W, Chung N, et al. Delayed enhancement magnetic resonance imaging is a significant prognostic factor in patients with non-ischemic cardiomyopathy. *Circ J.* (2010) 74:476–83. doi: 10.1253/circj.CJ-09-0446
33. te Rijdt WP, ten Sande JN, Gorter TM, van der Zwaag PA, van Rijsingen IA, Boekholdt SM, et al. Myocardial fibrosis as an early feature in phospholamban p.Arg14del mutation carriers: phenotypic insights from cardiovascular magnetic resonance imaging. *Eur Hear J.* (2019) 20:92–100. doi: 10.1093/ehjci/jey047
34. Gräni C, Eichhorn C, Bière L, Kaneko K, Murthy VL, Agarwal V, et al. Comparison of myocardial fibrosis quantification methods by cardiovascular magnetic resonance imaging for risk stratification of patients with suspected myocarditis. *J Cardiovasc Magn Reson.* (2019) 21:14. doi: 10.1186/s12968-019-0520-0
35. Park YJ, Park S-J, Kim E-K, Park KM, Lee S-C, On YK, et al. Semi-quantitative versus quantitative assessments of late gadolinium enhancement extent for predicting spontaneous ventricular tachyarrhythmia events in patients with hypertrophic cardiomyopathy. *Sci Rep.* (2020) 10:2920. doi: 10.1038/s41598-020-59804-8
36. aus dem Siepen F, Buss SJ, Messroghli D, Andre F, Lossnitzer D, Seitz S, et al. T1 mapping in dilated cardiomyopathy with cardiac magnetic resonance: quantification of diffuse myocardial fibrosis and comparison with endomyocardial biopsy. *Eur Hear J.* (2015) 16:210–6. doi: 10.1093/ehjci/jeu183
37. Mewton N, Liu CY, Croisille P, Bluemke D, Lima JAC. Assessment of myocardial fibrosis with cardiovascular magnetic resonance. *J Am Coll Cardiol.* (2011) 57:891–903. doi: 10.1016/j.jacc.2010.11.013
38. Moon JC, Messroghli DR, Kellman P, Piechnik SK, Robson MD, Ugander M, et al. Myocardial T1 mapping and extracellular volume quantification: a Society for Cardiovascular Magnetic Resonance (SCMR) and CMR Working Group of the European Society of Cardiology consensus statement. *J Cardiovasc Magn Reson.* (2013) 15:92. doi: 10.1186/1532-429X-15-92
39. Puntmann VO, Voigt T, Chen Z, Mayr M, Karim R, Rhode K, et al. Native T1 mapping in differentiation of normal myocardium from diffuse disease in hypertrophic and dilated cardiomyopathy. *JACC Cardiovasc Imaging.* (2013) 6:475–84. doi: 10.1016/j.jcmg.2012.08.019
40. Haaf P, Garg P, Messroghli DR, Broadbent DA, Greenwood JP, Plein S. Cardiac T1 Mapping and Extracellular Volume (ECV) in clinical

practice: a comprehensive review. *J Cardiovasc Magn Reson.* (2017) 18:89. doi: 10.1186/s12968-016-0308-4
41. Nakamori S, Dohi K, Ishida M, Goto Y, Imanaka-Yoshida K, Omori T, et al. Native T1 mapping and extracellular volume mapping for the assessment of diffuse myocardial fibrosis in dilated cardiomyopathy. *JACC Cardiovasc Imaging.* (2018) 11:48–59. doi: 10.1016/j.jcmg.2017.04.006
42. Mordi I, Carrick D, Bezerra H, Tzemos N. T1 and T2 mapping for early diagnosis of dilated non-ischaemic cardiomyopathy in middle-aged patients and differentiation from normal physiological adaptation. *Eur Hear J.* (2016) 17:797–803. doi: 10.1093/ehjci/jev216
43. Barison A, Del Torto A, Chiappino S, Aquaro GD, Todiere G, Vergaro G, et al. Prognostic significance of myocardial extracellular volume fraction in nonischaemic dilated cardiomyopathy. *J Cardiovasc Med.* (2015) 16:681. doi: 10.2459/JCM.0000000000000275
44. Puntmann VO, Carr-White G, Jabbour A, Yu C-Y, Gebker R, Kelle S, et al. T1-mapping and outcome in nonischemic cardiomyopathy. *JACC Cardiovasc Imaging.* (2016) 9:40–50. doi: 10.1016/j.jcmg.2015.12.001
45. Nakamori S, Bui AH, Jang J, El-Rewaidy HA, Kato S, Ngo LH, et al. Increased myocardial native T 1 relaxation time in patients with nonischemic dilated cardiomyopathy with complex ventricular arrhythmia. *J Magn Reson Imaging.* (2018) 47:779–86. doi: 10.1002/jmri.25811
46. Piers SRD, Everaerts K, van der Geest RJ, Hazebroek MR, Siebelink H-M, Pison LAFG, et al. Myocardial scar predicts monomorphic ventricular tachycardia but not polymorphic ventricular tachycardia or ventricular fibrillation in nonischemic dilated cardiomyopathy. *Hear Rhythm.* (2015) 12:2106–14. doi: 10.1016/j.hrthm.2015.05.026
47. Vita T, Gräni C, Abbasi SA, Neilan TG, Rowin E, Kaneko K, et al. Comparing CMR mapping methods and myocardial patterns toward heart failure outcomes in nonischemic dilated cardiomyopathy. *JACC Cardiovasc Imaging.* (2019) 12:1659–69. doi: 10.1016/j.jcmg.2018.08.021
48. Jellis CL, Kwon DH. Myocardial T1 mapping: modalities and clinical applications. *Cardiovasc Diagn Ther.* (2014) 4:126–37. doi: 10.3978/j.issn.2223-3652.2013.09.03
49. Lurz P, Eitel I, Adam J, Steiner J, Grothoff M, Desch S, et al. Diagnostic performance of CMR imaging compared with EMB in patients with suspected myocarditis. *JACC Cardiovasc Imaging.* (2012) 5:513–24. doi: 10.1016/j.jcmg.2011.11.022
50. Snel GJH, van den Boomen M, Hernandez LM, Nguyen CT, Sosnovik DE, Velthuis BK, et al. Cardiovascular magnetic resonance native T2 and T2* quantitative values for cardiomyopathies and heart transplantations: a systematic review and meta-analysis. *J Cardiovasc Magn Reson.* (2020) 22:34. doi: 10.1186/s12968-020-00627-x
51. Nishii T, Kono AK, Shigeru M. Cardiovascular magnetic resonance T2 mapping can detect myocardial edema in idiopathic dilated cardiomyopathy. *Int J Cardiovasc Imaging.* (2014) 30:65–72. doi: 10.1007/s10554-014-0414-z
52. Fogante M, Agliata G, Basile MC, Compagnucci P, Volpato G, Falanga U, et al. Cardiac imaging in athlete's heart: the role of the radiologist. *Medicina.* (2021) 57:455. doi: 10.3390/medicina57050455
53. Haberkorn SM, Spieker M, Jacoby C, Flögel U, Kelm M, Bönner F. State of the art in cardiovascular T2 mapping : on the way to a cardiac biomarker? *Curr Cardiovasc Imaging Rep*. (2018) 11:15. doi: 10.1007/s12410-018-9455-3
54. Anderson LJ, Holden S, Davis B, Prescott E, Charrier CC, Bunce NH, et al. Cardiovascular T2-star (T2 *) magnetic resonance for the early diagnosis of myocardial iron overload. *Eur Heart J.* (2001) 22:2171–9. doi: 10.1053/euhj.2001.2822
55. Lota AS, Gatehouse PD, Mohiaddin RH. T2 mapping and T2 * imaging in heart failure. *Heart Fail Rev*. (2017) 22:431–40. doi: 10.1007/s10741-017-9616-5
56. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* (2016) 37:2129–200. doi: 10.1093/eurheartj/ehw128
57. Rihal CS, Nishimura RA, Hatle LK, Bailey KR, Tajik AJ. Systolic and diastolic dysfunction in patients with clinical diagnosis of dilated cardiomyopathy. Relation to symptoms and prognosis. *Circulation.* (1994) 90:2772–9. doi: 10.1161/01.CIR.90.6.2772
58. Diaz RA, Obasohan A, Oakley CM. Prediction of outcome in dilated cardiomyopathy. *Heart.* (1987) 58:393–9. doi: 10.1136/hrt.58.4.393
59. McNamara DM, Starling RC, Cooper LT, Boehmer JP, Mather PJ, Janosko KM, et al. Clinical and demographic predictors of outcomes in recent onset dilated cardiomyopathy. *J Am Coll Cardiol.* (2011) 58:1112–8. doi: 10.1016/j.jacc.2011.05.033
60. Buss SJ, Breuninger K, Lehrke S, Voss A, Galuschky C, Lossnitzer D, et al. Assessment of myocardial deformation with cardiac magnetic resonance strain imaging improves risk stratification in patients with dilated cardiomyopathy. *Eur Hear J.* (2015) 16:307–15. doi: 10.1093/ehjci/jeu181
61. Alpendurada F, Guha K, Sharma R, Ismail TF, Clifford A, Banya W, et al. Right ventricular dysfunction is a predictor of non-response and clinical outcome following cardiac resynchronization therapy. *J Cardiovasc Magn Reson.* (2011) 13:68. doi: 10.1186/1532-429X-13-68
62. Gulati A, Ismail TF, Jabbour A, Alpendurada F, Guha K, Ismail NA, et al. The prevalence and prognostic significance of right ventricular systolic dysfunction in nonischemic dilated cardiomyopathy. *Circulation.* (2013) 128:1623–33. doi: 10.1161/CIRCULATIONAHA.113.002518
63. Becker MAJ, Lingen ACJ, Wubben M, Ven PM, Rossum AC, Cornel JH, et al. Characteristics and prognostic value of right ventricular (dys)function in patients with non-ischaemic dilated cardiomyopathy assessed with cardiac magnetic resonance imaging. *ESC Hear Fail.* (2021) 8:1055–63. doi: 10.1002/ehf2.13072
64. Gulati A, Ismail TF, Jabbour A, Ismail NA, Morarji K, Ali A, et al. Clinical utility and prognostic value of left atrial volume assessment by cardiovascular magnetic resonance in non-ischaemic dilated cardiomyopathy. *Eur J Heart Fail.* (2013) 15:660–70. doi: 10.1093/eurjhf/hft019
65. Romano S, Judd RM, Kim RJ, Kim HW, Klem I, Heitner JF, et al. Feature-tracking global longitudinal strain predicts death in a multicenter population of patients with ischemic and nonischemic dilated cardiomyopathy incremental to ejection fraction and late gadolinium enhancement. *JACC Cardiovasc Imaging.* (2018) 11:1419–29. doi: 10.1016/j.jcmg.2017.10.024
66. Halliday BP, Wassall R, Lota AS, Khalique Z, Gregson J, Newsome S, et al. Withdrawal of pharmacological treatment for heart failure in patients with recovered dilated cardiomyopathy (TRED-HF): an open-label, pilot, randomised trial. *Lancet.* (2019) 393:61–73. doi: 10.1016/S0140-6736(18)32484-X
67. Peters S, Kumar S, Elliott P, Kalman JM, Fatkin D. Arrhythmic genotypes in familial dilated cardiomyopathy: implications for genetic testing and clinical management. *Hear Lung Circ.* (2019) 28:31–8. doi: 10.1016/j.hlc.2018.09.010
68. Ware JS, Amor-Salamanca A, Tayal U, Govind R, Serrano I, Salazar-Mendiguchía J, et al. Genetic etiology for alcohol-induced cardiac toxicity. *J Am Coll Cardiol.* (2018) 71:2293–302. doi: 10.1016/j.jacc.2018.03.462
69. Xu H, Dorn II G, Shetty A, Parihar A, Dave T, Robinson S, et al. A genome-wide association study of idiopathic dilated cardiomyopathy in African Americans. *J Pers Med.* (2018) 8:11. doi: 10.3390/jpm8010011
70. Hershberger RE, Morales A, Siegfried JD. Clinical and genetic issues in dilated cardiomyopathy: a review for genetics professionals. *Genet Med.* (2010) 12:655–67. doi: 10.1097/GIM.0b013e3181f2481f
71. Muir AR, Menown IB. Genetic biomarkers in cardiovascular disease. *Biomark Med.* (2013) 7:497–9. doi: 10.2217/bmm.13.82
72. Verdonschot JAJ, Hazebroek MR, Wang P. Sanders-van Wijk S, Merken JJ, Adriaansen YA, et al. Clinical phenotype and genotype associations with improvement in left ventricular function in dilated cardiomyopathy. *Circ Hear Fail.* (2018) 11:e005220. doi: 10.1161/CIRCHEARTFAILURE.118.005220
73. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. *J Med Internet Res.* (2020) 22:e15154. doi: 10.2196/15154
74. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. *Minim Invasive Ther Allied Technol.* (2019) 28:73–81. doi: 10.1080/13645706.2019.1575882
75. Colling R, Pitman H, Oien K, Rajpoot N, Macklin P, Bachtiar V, et al. Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *J Pathol.* (2019) 249:143–50. doi: 10.1002/path.5310
76. Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. II—recent progress. *IBM J Res Dev.* (1967) 11:601–17. doi: 10.1147/rd.116.0601

77. Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med.* (2020) 7:25. doi: 10.3389/fcvm.2020.00025
78. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson.* (2018) 20:1–12. doi: 10.1186/s12968-018-0471-x
79. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical image analysis using convolutional neural networks: a review. *J Med Syst.* (2018) 42:226. doi: 10.1007/s10916-018-1088-1
80. Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O. Deep learning with convolutional neural network in radiology. *Jpn J Radiol.* (2018) 36:257–72. doi: 10.1007/s11604-018-0726-3
81. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer.* (2018) 18:500–10. doi: 10.1038/s41568-018-0016-5
82. 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
83. 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
84. Tran PV. A fully convolutional neural network for cardiac segmentation in short-axis MRI. *arxiv.* (2016) arXiv:1604.00494
85. 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: *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. doi: 10.1007/978-3-319-75541-0_13
86. 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) 290:81–8. doi: 10.1148/radiol.2018180513
87. 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
88. Jang Y, Hong Y, Ha S, Kim S, Chang H-J. Automatic segmentation of LV and RV in cardiac MRI. In: Pop M, Sermesant M, Jodoin P-M, Lalande A, Zhuang X, Yang G, editors. *Statistical Atlases and Computational Models of the Heart ACDC and MMWHS Challenges*. Cham: Springer (2018). p. 161–9. doi: 10.1007/978-3-319-75541-0_17
89. Fahmy AS, El-Rewaidy 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) 21:7. doi: 10.1186/s12968-018-0516-1
90. 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
91. Avendi MR, Kheradvar A, Jafarkhani 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
92. 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
93. van Assen M, Cornelissen LJ. Artificial intelligence: from scientific curiosity to clinical precocity? *JACC Cardiovasc Imaging.* (2020) 13:1172–4. doi: 10.1016/j.jcmg.2019.09.008
94. Chen C, Bai W, Davies RH, Bhuva AN, Manisty C, Moon JC, et al. Improving the generalizability of convolutional neural network-based segmentation on CMR images. *Front Cardiovasc Med.* (2020) 7:105. doi: 10.3389/fcvm.2020.00105
95. Pirruccello JP, Bick A, Wang M, Chaffin M, Friedman S, Yao J, et al. Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nat Commun.* (2020) 11:2254. doi: 10.1038/s41467-020-15823-7
96. Rizwan I, Haque I, Neubert J. Deep learning approaches to biomedical image segmentation. *Informatics Med Unlocked.* (2020) 18:100297. doi: 10.1016/j.imu.2020.100297
97. Gibson E, Li W, Sudre C, Fidon L, Shakir DI, Wang G, et al. NiftyNet: a deep-learning platform for medical imaging. *Comput Methods Programs Biomed.* (2018) 158:113–22. doi: 10.1016/j.cmpb.2018.01.025
98. 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
99. Goyal N, Mor-Avi V, Volpato V, Narang A, Wang S, Salerno M, et al. Machine learning based quantification of ejection and filling parameters by fully automated dynamic measurement of left ventricular volumes from cardiac magnetic resonance images. *Magn Reson Imaging.* (2020) 67:28–32. doi: 10.1016/j.mri.2019.12.004
100. Backhaus SJ, Staab W, Steinmetz M, Ritter CO, Lotz J, Hasenfuß G, et al. Fully automated quantification of biventricular volumes and function in cardiovascular magnetic resonance: applicability to clinical routine settings. *J Cardiovasc Magn Reson.* (2019) 21:24. doi: 10.1186/s12968-019-0532-9
101. Ruijsink JB, Puyol-Anton E, Sinclair M, Baji W, King A, Razavi R. 4382Fully automated assessment of filling and ejection rates of the ventricle. Reference values for healthy volunteers from the UK-biobank cohort. *Eur Heart J.* (2018) 39(suppl_1):5–6. doi: 10.1093/eurheartj/ehy563.4382
102. Mendoza D, Codella N, Wang Y, Prince M, Sethi S, Manoushagian S, et al. Impact of diastolic dysfunction severity on global left ventricular volumetric filling - Assessment by automated segmentation of routine cine cardiovascular magnetic resonance. *J Cardiovasc Magn Reson.* (2010) 12:1–11. doi: 10.1186/1532-429X-12-46
103. Eriksson J, Bolger AF, Ebbers T, Carlhäll C-J. Assessment of left ventricular hemodynamic forces in healthy subjects and patients with dilated cardiomyopathy using 4D flow MRI. *Physiol Rep.* (2016) 4:e12685. doi: 10.14814/phy2.12685
104. Mariscal Harana J, Vergani V, Asher C, Razavi R, King A, Ruijsink B, et al. Large-scale, multi-vendor, multi-protocol, quality-controlled analysis of clinical cine CMR using artificial intelligence. *Eur Hear J.* (2021) 22(Supplement_2):2021. doi: 10.1093/ehjci/jeab090.046
105. Merlo M, Gobbo M, Stolfo D, Losurdo P, Ramani F, Barbati G, et al. The prognostic impact of the evolution of RV function in idiopathic DCM. *JACC Cardiovasc Imaging.* (2016) 9:1034–42. doi: 10.1016/j.jcmg.2016.01.027
106. Karim R, Bhagirath P, Claus P, James Housden R, Chen Z, Karimaghaloo 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
107. Hennemuth A, Friman O, Huellebrand M, Peitgen H-O. Mixture-model-based segmentation of myocardial delayed enhancement MRI. In: Camara O, Mansi T, Pop M, Rhode K, Sermesant M, Young A, editors. *Statistical Atlases and Computational Models of the Heart Imaging and Modelling Challenges STACOM 2012 Lecture Notes in Computer Science*. Berlin: Springer Berlin Heidelberg (2013). p. 87–96. doi: 10.1007/978-3-642-36961-2_11
108. Detsky JS, Paul G, Dick AJ, Wright GA. Reproducible classification of infarct heterogeneity using fuzzy clustering on multicontrast delayed enhancement magnetic resonance images. *IEEE Trans Med Imaging.* (2009) 28:1606–14. doi: 10.1109/TMI.2009.2023515
109. 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
110. Jablonowski R, Chaudhry U, van der Pals J, Engblom H, Arheden H, Heiberg E, et al. Cardiovascular magnetic resonance to predict appropriate implantable cardioverter defibrillator therapy in ischemic and nonischemic cardiomyopathy patients using late gadolinium enhancement border zone. *Circ Cardiovasc Imaging.* (2017) 10:1–9. doi: 10.1161/CIRCIMAGING.116.006105
111. 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 Physics, Biol Med.* (2019) 32:187–95. doi: 10.1007/s10334-018-0718-4

112. Lau F, Hendriks T, Lieman-Sifry J, Sall S, Golden D. ScarGAN: Chained Generative Adversarial Networks to Simulate Pathological Tissue on Cardiovascular MR Scans. In: Stoyanov D, Taylor Z, Carneiro G, Mahmood TS, Hein LM,Tavaras JMRS, et al. editors. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2018, ML-CDS 2018*. Lecture Notes in Computer Science, Vol. 11045. Cham: Springer (2018). doi: 10.1007/978-3-030-00889-5_39
113. Piechnik SK, Ferreira VM, Dall'Armellina E, Cochlin LE, Greiser A, Neubauer S, et al. Shortened Modified Look-Locker Inversion recovery (ShMOLLI) for clinical myocardial T1- mapping at 15 and 3 T within a 9 heartbeat breathhold. *J Cardiovasc Magn Reson.* (2010) 12:1–11. doi: 10.1186/1532-429X-12-69
114. Puyol-Antón E, Ruijsink B, Baumgartner CF, Masci PG, Sinclair M, Konukoglu E, et al. Automated quantification of myocardial tissue characteristics from native T1mapping using neural networks with uncertainty-based quality-control. *J Cardiovasc Magn Reson.* (2020) 22:1–15. doi: 10.1186/s12968-020-00650-y
115. Nickander J, Lundin M, Abdula G, Sörensson P, Rosmini S, Moon JC, et al. Blood correction reduces variability and gender differences in native myocardial T1 values at 1. 5 T cardiovascular magnetic resonance - a derivation/validation approach. *J Cardiovasc Magn Reson.* (2017) 19:1–11. doi: 10.1186/s12968-017-0353-7
116. Sammani A, Baas AF, Asselbergs FW, te Riele ASJM. Diagnosis and risk prediction of dilated cardiomyopathy in the era of big data and genomics. *J Clin Med.* (2021) 10:921. doi: 10.3390/jcm10050921
117. Glöcklhofer CR, Steinfurt J, Franke G, Hoppmann A, Glantschnig T, Perez-Feliz S, et al. A novel LMNA nonsense mutation causes two distinct phenotypes of cardiomyopathy with high risk of sudden cardiac death in a large five-generation family. *EP Eur.* (2018) 20:2003–13. doi: 10.1093/europace/euy127
118. Peña-Peña ML, Monserrat L. Risk stratification in patients with nonisquemic dilated cardiomyopathy. The role of genetic testing. *Rev Española Cardiol.* (2019) 72:333–40. doi: 10.1016/j.rec.2018.10.017
119. 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
120. Zhou Y, Hou Y, Hussain M, Brown S, Budd T, Tang WHW, et al. Machine learning–based risk assessment for cancer therapy–related cardiac dysfunction in 4300 longitudinal oncology patients. *J Am Heart Assoc.* (2020) 9:19628. doi: 10.1161/JAHA.120.019628
121. Merlo M, Caiffa T, Gobbo M, Adamo L, Sinagra G. Reverse remodeling in Dilated Cardiomyopathy: insights and future perspectives. *IJC Hear Vasc.* (2018) 18:52–7. doi: 10.1016/j.ijcha.2018.02.005
122. Wells G, Parkash R, Healey JS, Talajic M, Arnold JM, Sullivan S, et al. Cardiac resynchronization therapy: a meta-analysis of randomized controlled trials. *Can Med Assoc J.* (2011) 183:421–9. doi: 10.1503/cmaj.101685
123. Sieniewicz BJ, Gould J, Porter B, Sidhu BS, Teall T, Webb J, et al. Understanding non-response to cardiac resynchronisation therapy: common problems and potential solutions. *Heart Fail Rev.* (2019) 24:41–54. doi: 10.1007/s10741-018-9734-8
124. 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
125. 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
126. Moss AJ, Hall WJ, Cannom DS, Klein H, Brown MW, Daubert JP, et al. Cardiac-resynchronization therapy for the prevention of heart-failure events. *N Engl J Med.* (2009) 361:1329–38. doi: 10.1056/NEJMoa0906431
127. Puyol-Antón E, Chen C, Clough JR, Ruijsink B, Sidhu BS, Gould J, et al. Interpretable deep models for cardiac resynchronisation therapy response prediction. In: Anton EP, Pop M, Sermesant M, Campello V, Lalande A, Lekadir K, Suinesiaputra A, Camara O, Young A, editors. *Statistical Atlases and Computational Models of the Heart M&Ms and EMIDEC Challenges.* Cham: Springer (2020). p. 284–93. doi: 10.1007/978-3-030-59710-8_28
128. Stătescu C, Ureche C, Enachi Ştefana, Radu R, Sascău RA. Cardiac resynchronization therapy in non-ischemic cardiomyopathy: role of multimodality. *Imaging Diagnostics.* (2021) 11:625. doi: 10.3390/diagnostics11040625
129. Puyol-Antón E, Ruijsink B, Clough JR, Oksuz I, Rueckert D, Razavi R, et al. Assessing the impact of blood pressure on cardiac function using interpretable biomarkers and variational autoencoders. In: Camara O, Mansi T, Pop M, Rhode K, Sermesant M, Young M, editors. *Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2020). p. 22–30. doi: 10.1007/978-3-030-39074-7_3
130. 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
131. 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. *Circ Cardiovasc Genet.* (2014) 7:760–70. doi: 10.1161/CIRCGENETICS.113.000384
132. 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
133. Akinrinade O, Koskenvuo JW, Alastalo T-P. Prevalence of titin truncating variants in general population. *PLoS ONE.* (2015) 10:e0145284. doi: 10.1371/journal.pone.0145284
134. Herman DS, Lam L, Taylor MRG, 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
135. Ware JS Li J, Mazaika E, Yasso CM, DeSouza T, Cappola TP, et al. Shared genetic predisposition in peripartum and dilated cardiomyopathies. *N Engl J Med.* (2016) 374:233–41. doi: 10.1056/NEJMoa1505517
136. 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
137. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med.* (2020) 3:118. doi: 10.1038/s41746-020-00324-0
138. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Heal.* (2021) 3:e195–203. doi: 10.1016/S2589-7500(20)30292-2

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.

frontiers in Cardiovascular Medicine

REVIEW
published: 04 January 2022
doi: 10.3389/fcvm.2021.765693

Machine Learning for Clinical Decision-Making: Challenges and Opportunities in Cardiovascular Imaging

Sergio Sanchez-Martinez[1], *Oscar Camara*[2], *Gemma Piella*[2], *Maja Cikes*[3], *Miguel Ángel González-Ballester*[2,4], *Marius Miron*[5], *Alfredo Vellido*[6], *Emilia Gómez*[2,5]*, *Alan G. Fraser*[7] and *Bart Bijnens*[1,4,8]

[1] August Pi i Sunyer Biomedical Research Institute (IDIBAPS), Barcelona, Spain, [2] Department of Information and Communication Technologies, University Pompeu Fabra, Barcelona, Spain, [3] Department of Cardiovascular Diseases, University of Zagreb School of Medicine, University Hospital Centre Zagreb, Zagreb, Croatia, [4] ICREA, Barcelona, Spain, [5] Joint Research Centre, European Commission, Seville, Spain, [6] Computer Science Department, Intelligent Data Science and Artificial Intelligence (IDEAI-UPC) Research Center, Universitat Politècnica de Catalunya, Barcelona, Spain, [7] School of Medicine, Cardiff University, Cardiff, United Kingdom, [8] Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium

OPEN ACCESS

Edited by:
Pablo Lamata,
King's College London,
United Kingdom

Reviewed by:
Jennifer Mancio,
University of Porto, Portugal
Karthik Seetharam,
West Virginia State University,
United States

***Correspondence:**
Emilia Gómez
emilia.gomez-gutierrez@ec.europa.eu

Specialty section:
This article was submitted to Cardiovascular Biologics and Regenerative Medicine, a section of the journal Frontiers in Cardiovascular Medicine

Received: *27 August 2021*
Accepted: *07 December 2021*
Published: *04 January 2022*

Citation:
Sanchez-Martinez S, Camara O, Piella G, Cikes M, González-Ballester MÁ, Miron M, Vellido A, Gómez E, Fraser AG and Bijnens B (2022) Machine Learning for Clinical Decision-Making: Challenges and Opportunities in Cardiovascular Imaging. Front. Cardiovasc. Med. 8:765693. doi: 10.3389/fcvm.2021.765693

The use of machine learning (ML) approaches to target clinical problems is called to revolutionize clinical decision-making in cardiology. The success of these tools is dependent on the understanding of the intrinsic processes being used during the conventional pathway by which clinicians make decisions. In a parallelism with this pathway, ML can have an impact at four levels: for data acquisition, predominantly by extracting standardized, high-quality information with the smallest possible learning curve; for feature extraction, by discharging healthcare practitioners from performing tedious measurements on raw data; for interpretation, by digesting complex, heterogeneous data in order to augment the understanding of the patient status; and for decision support, by leveraging the previous steps to predict clinical outcomes, response to treatment or to recommend a specific intervention. This paper discusses the state-of-the-art, as well as the current clinical status and challenges associated with the two later tasks of interpretation and decision support, together with the challenges related to the learning process, the auditability/traceability, the system infrastructure and the integration within clinical processes in cardiovascular imaging.

Keywords: artificial intelligence, machine learning, deep learning, clinical decision making, cardiovascular imaging, diagnosis, prediction

INTRODUCTION

Artificial intelligence (AI) systems are programmed to achieve complex tasks by perceiving their environment through data acquisition, interpreting the collected data and deciding the best action(s) to take to achieve a given goal. As a broad scientific discipline, AI includes several approaches and techniques, such as machine learning, machine reasoning, and robotics (1). Machine learning (ML) is the subfield of AI that focuses on the development of algorithms that allow computers to automatically discover patterns in the data and improve with experience, without being given a set of explicit instructions. Among ML techniques, Deep Learning (DL) is

the subfield concerned with algorithms inspired by the structure and function of the brain called artificial neural networks. Unlike other ML techniques, DL bypasses the need of using hand-crafted features as input, automatically figuring out the data features that are important for solving complex problems. This is the main reason why DL stands out as the current state-of-the-art in virtually all medical imaging related tasks.

In the particular case of clinical decision-making in cardiology, ML methods would perceive an individual by collecting and interpreting his/her clinical data and would reason on them to suggest actions to maintain or improve that individual's cardiovascular health. This mimics the clinician's approach when examining and treating a sick patient, or when suggesting preventive actions to avoid illness. Therefore, in order to assess the challenges and opportunities of ML systems for clinical decision-making in cardiology, an in-depth understanding of this process, when performed by cardiologists, is paramount.

Figure 1 summarises a typical paradigm for clinical decision-making. It starts by data acquisition, including the clinical history of the patient, demographics, physiological measurements, electrocardiogram, imaging and laboratory tests, and the relevant indices collected from these data. Next, clinicians construct and interpret the state of the patient by comparison with population-based information learned during their education or daily practice, or information derived from guidelines. This interpretation is based on reasoning on the data using the human innate capability of contextualizing information through pattern recognition. Furthermore, clinicians assess the uncertainty associated with measurements and the completeness of the available information to estimate how much they can rely on the data. Finally, they consider the knowledge from the (natural as well as treated) expected evolution of populations related to the patient's status to make decisions. The resulting actions can be to either collect more data to minimize the uncertainty associated with the decision, to make an intervention (drug/device therapy, surgery, etc.) to improve the patient's outcomes, or to send the patient home (whether or not with planned observation follow-up).

In the era of evidence-based, personalised medicine (2), millions of individuals are carefully examined, which results in a deluge of complex, heterogeneous data. The use of algorithmic approaches to digest these data and augment clinical decision-making is now feasible due to the ever-increasing computing power, and the latest advances in the ML field (3). Indeed, big data leveraged by ML can provide well-curated information to clinicians so they can make better informed diagnoses and treatment recommendations, while also estimating probabilities and costs for the possible outcomes. ML-augmented decisions made by clinicians have the potential to improve outcomes, lower costs of care, and increase patient and family satisfaction.

ML analyses have, to date, demonstrated human-like performance in low-level tasks where pattern recognition or perception play a fundamental role. Some examples are **data acquisition**, standardization and classification (4, 5), and **feature extraction** (6, 7). For higher-level tasks involving reasoning, such as patient's **status interpretation** and **decision support**, ML allows for the integration of complex, heterogenous data in the decision-making process, but these are still immature and need substantial validation (8). In parallel to **Figure 1**, which illustrated the process of making clinical decisions, **Figure 2** describes the tasks involved in this process according to how ML could contribute, and highlights how the risks to a patient from erroneous conclusions increase with each step.

There exist other review papers that cover the topic of AI in cardiovascular imaging from a broader perspective (9), or that highlight the synergy between machine learning and mechanistic models that enable the creation of a "digital twin" in pursuit of precision cardiology (10). Complementarily, this paper focuses on ML as a subfield of AI and on clinical decision-making as an essential part of cardiovascular medicine. Although cardiovascular imaging only constitutes a limited portion of the data spectrum in cardiology, we emphasize the imaging field in our literature review given that it is one of the areas to which ML has contributed the most (11). In the following, we discuss the higher-level tasks related to clinical decision-making that involve reasoning on clinical data, namely interpretation, and decision support. For each of these tasks, we review the ML state-of-the-art (indicating whether the implementation was based on DL or other ML technique), comment on the current penetration of ML tools into clinical practice (see *Clinical status* subsections), and elaborate on the current challenges that limit their implementation in clinical practice. Finally, we discuss the general challenges that may appear when tackling any clinical problem with ML approaches.

This paper addresses potential questions arising from data scientists, industrial partners and funding institutions, helping them understand clinical decision-making in cardiology and identify potential niches for their solutions to be helpful. At the same time, the paper aims at informing cardiologists about which ML tools could target their problems and what are their current limitations.

STATUS INTERPRETATION—COMPARISON TO POPULATION

Let us assume that the clinical data of a patient have been properly acquired and relevant features are readily available. The next stage in the decision process consists in interpreting his/her status by comparison to populations. This comparison requires data normalization. When complex data are involved, such as cardiac images, the traditional approach to normalization is to build a statistical atlas–a reference model that captures the variability associated to a population (12). To build the atlas, the training data must be transformed into a common spatio(-temporal) framework, which can be achieved by registration. In this sense, registration appears as a crucial step for status interpretation toward diagnosis, and deep learning has emerged as a suitable tool to register 3D cardiac volumes (13), 3D pre-operative cardiac models to 2D intraoperative x-ray fluoroscopy to facilitate image-guided interventions (14), or cardiac MRI sequences (15).

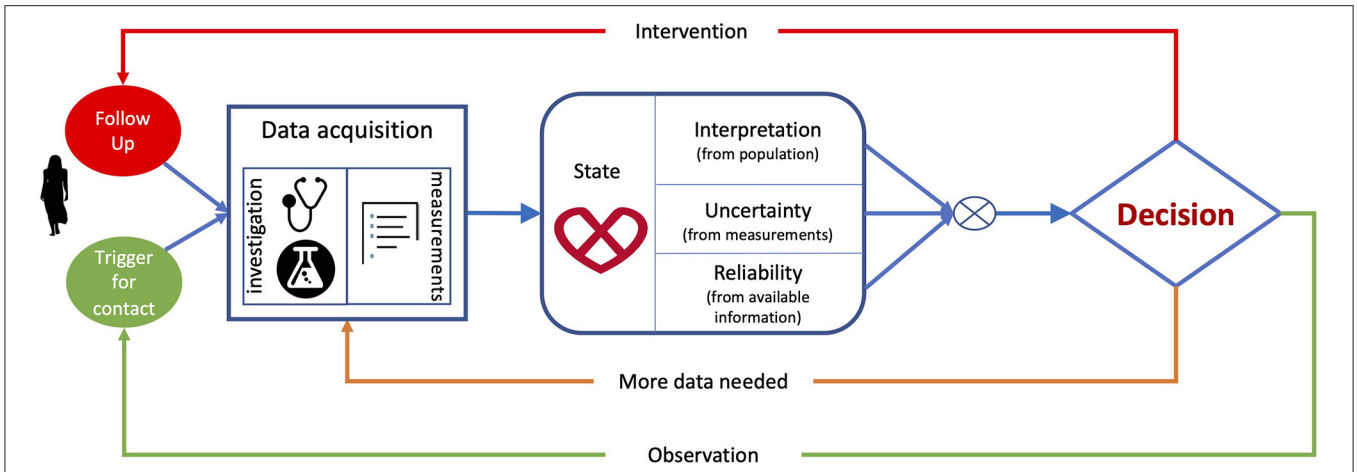


FIGURE 1 | Clinical decision-making flowchart, from data acquisition and extraction, to patient's status interpretation and associated decision.

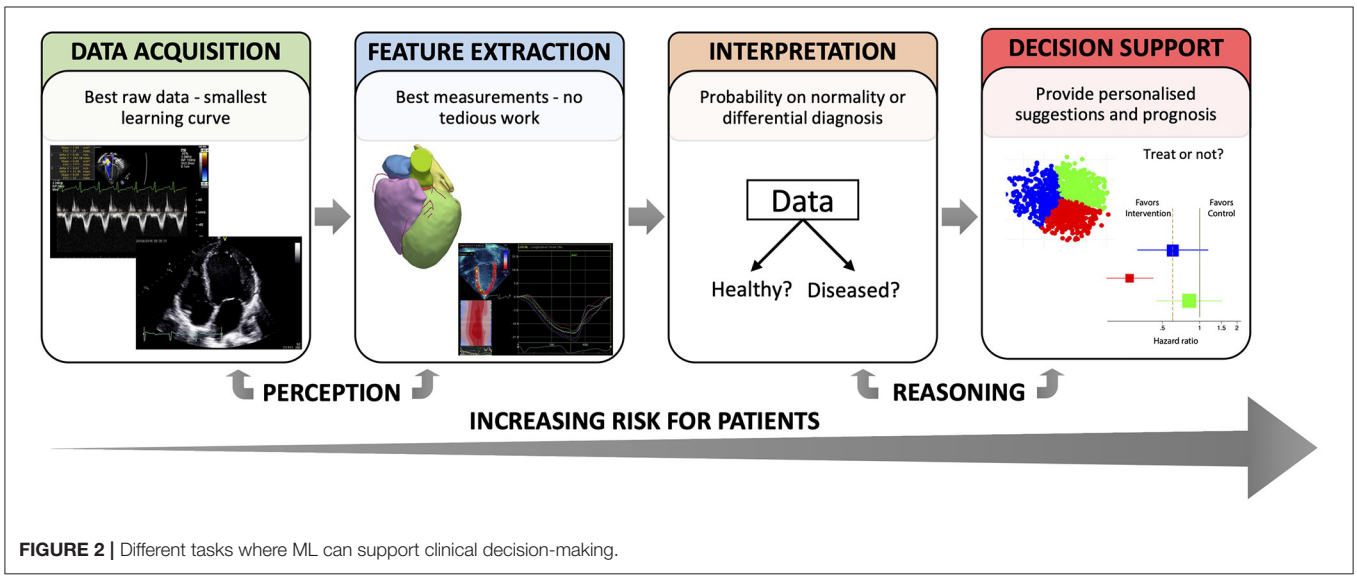


FIGURE 2 | Different tasks where ML can support clinical decision-making.

The ML interpretation of the state of the patient can augment the diagnosis made by clinicians. Indeed, a recently published meta-analysis highlighted the promising potential of ML and DL models to predict conditions such as coronary artery disease, heart failure, stroke, and cardiac arrhythmias using data derived from routinely used imaging techniques and ECG (16). Based on imaging, ensemble ML models, which group the prediction of different weak learners, have demonstrated higher accuracy than expert readers for the diagnosis of obstructive coronary artery disease (17); a DL model automated the diagnosis of acute ischemic infarction using CT studies (18); and another DL model achieved 92.3% accuracy for left ventricular hypertrophy classification analysing echocardiographic images (19). Different ML models have also operated on electronic health records (EHR) for triaging of low-risk vs. high-risk cardiovascular patients, grading findings as requiring non-urgent, urgent or critical attention, as a strategy to improve efficiency and allocation of the finite resources available in the emergency department (20). Lastly, a ML ensemble model combined clinical data, quantitative stenosis, and plaque metrics from CT angiography to effectively detect lesion-specific ischemia (21).

Another data-driven example of status interpretation is unsupervised machine learning for dimensionality reduction; a label-agnostic approach that orders individuals according to their similarity, i.e., those with a similar clinical presentation are grouped together, whereas those showing distinct pathophysiological features are positioned far apart (22). This allows identifying different levels of abnormality, or assessing the effect of therapies and interventions, as these are aimed to restore an individual toward increased "normality." An implementation of unsupervised dimensionality reduction provided useful insight into treatment response in large patient populations (23), and quantified patient changes after an intervention using temporally dynamic data (24).

Clinical Status

ML approaches for interpreting a patient's status enhance discovery in massive databases by offering the possibility to identify similar cases, build normality statistics, and spot outliers. Whether these approaches are intended for diagnosis or risk assessment, they could contribute to deliver better healthcare.

Unfortunately, many current ML applications for interpreting clinical data present a technically sound contribution, but do not address real clinical needs, and they often focus on binary classification of normal vs. abnormal (19), which strongly limits their use in routine clinical practice. Furthermore, studies showing impact on hard clinical endpoints rather than on surrogate measures are still needed. The way forward is through further integration of technical and clinical contributions, and through the elaboration of consensus recommendations on how to tackle a clinical necessity using ML.

DECISION-MAKING (PREDICTION)

Based on the interpretation of the patient's status, clinicians should decide on whether: (1) observe the patient and wait until an event triggers the need for a decision; (2) collect more data to improve the odds of making the right decision; or (3) perform an intervention and monitor the outcome (see **Figure 1**). Machine learning methods can help clinicians to decide which pathway to follow (25), in a way that is cost-effective (26).

Several studies have assessed the predictive power of ML techniques based on imaging. An echocardiography-based DL model was shown to improve the prediction of in-hospital mortality among coronary heart disease and heart failure patients as compared to traditionally used prediction models (27). An ensemble ML approach interrogating SPECT myocardial perfusion studies demonstrated superior performance at predicting early revascularization in patients with suspected coronary artery disease as compared to an experienced reader (28), or in combination with clinical and ECG data outperformed the reading physicians at predicting the occurrence of major adverse cardiovascular events (29). Lastly, a DL implementation fed with CT scans from asymptomatic as well as stable and acute chest pain cohorts demonstrated the added clinical value of automated systems to predict cardiovascular events (30). Leaving imaging aside, deep learning based on clinical, laboratory and demographic data, ECG parameters, and cardiopulmonary exercise testing estimated prognosis and guided therapy in a large population of adults with congenital heart disease (31).

The interplay between different a priori non-related imaging tests has recently been discovered by DL through the identification of previously unnoticed associations. For example, breast arterial calcifications and the likelihood of patients at a high cardiovascular risk was sorted out using a DL model that operated with mammograms (32). Similarly, the power of ML in combination with the non-invasiveness of retinal scanning has been used to predict abnormalities in the macrovasculature based on the microvascular features of the eye. One such example is the DL model that predicted cardiovascular risk factors using retinal fundus photographs, thus allowing for an easier and cost-effective cardiovascular risk stratification (33), or the DL implementation that inferred coronary artery calcium (CAC) scores from retinal photographs, which turned out to be as accurate as CT scan-measured CAC in predicting cardiovascular events (34).

Clinical Status

The few examples of FDA-cleared cutting-edge ML applications to cardiovascular imaging that are thus suitable for routine use, focus on the low-level tasks of data acquisition and feature extraction, both in cardiac MRI (35) and echocardiography (36, 37), although the latter contribution did actually prove useful to predict a poor prognosis in acute COVID-19 patients based on DL-enabled automated quantification of echocardiographic images. However, the use of these ML applications for prediction and decision-making is still in its early days, as most models are still incapable of making predictions at the individual level (8, 38). More effort is needed toward integration in a clinical environment, interpretability, and validation if we want to see these models embedded in routine patient care.

CHALLENGES COMMON TO STATUS INTERPRETATION & DECISION-MAKING

Applications concerning patient's status interpretation and decision-making, which entails learning what is the risk associated with each possible clinical decision, imply a much higher risk as compared to the low-level tasks of data acquisition and feature extraction, since decisions derived from them could harm patients. Accordingly, ML outcomes need to be intuitively interpretable by the cardiologist and validated in a much more exhaustive way (as required by medical device regulators; e.g., class IIa or IIb routes to commercialization), ultimately with the launch of randomized prospective trials.

One of the main challenges for ML approaches to status interpretation lies in the extraction of meaningful concepts from raw data. This challenge entails many others, related to the data themselves. The first concerns the reliability and representativeness of training and outcome data. If representative, ML models need to find a reliable metric to compare heterogeneous data, which is not trivial. Furthermore, for a successful interpretation, data collection protocols should be designed to cover gender-, ethnicity- and age-related changes, and capture the rare outliers (39). On top of this, ML systems should be designed to consider longitudinal data, as to assess a patient over time, e.g., during a stress protocol (40) or disease progression (41). Finally, ML models are trained on three different kinds of data; ranging from higher to lower quality and completeness: (1) randomized clinical trials, (2) cohorts, and (3) clinical routine real-world data. The exchange of knowledge throughout these collections of data is challenging, since what was learned from highly curated data (e.g., randomized clinical trials) may not generalize to routinely collected data.

Another crucial problem associated to currently available data is bias, i.e., when the training sample is not representative

of the population of interest (see section "General challenges" for more details). Accordingly, caution is needed when testing a trained model in new clinical centres. As ML users can attest, there will always be a trade-off between improving the system performance locally and having systems that generalize well (42). Automation bias, defined as the human tendency to accept a computer-generated solution without searching for contradictory information (43), may also affect clinical interpretation and decision-making. As shown by Goddard et al. (44), when the ML solution is reliable it augments human performance, but when the solution is incorrect human errors increase. Thus, who is to blame if a diagnostic algorithm fails? The further we move along the clinical decision-making flowchart (**Figure 2**), the more ethical and legal barriers the ML practitioner/company faces. To mitigate some of these issues, the training data should be accessible, and the learning systems equipped with tools that allow reconstructing the reasoning behind the decision taken.

Table 1 organizes the ideas discussed for status interpretation and decision-making in the form of a SWOT analysis.

GENERAL CHALLENGES

We have previously described the specific challenges that may arise when ML models are given the tasks of interpreting the patient's status or making predictions to guide the clinical decision. In the following, we discuss the general challenges that may appear when tackling any clinical problem with ML approaches. These are divided into different sections, depending on whether they relate to the learning itself, the auditability/traceability aspects, the system/infrastructure, or the integration within clinical processes.

Learning

(Non-standardized) Data

Medical data are normally kept in many separate systems, which hampers accessibility and makes comparisons at a population level nearly impossible. Electronic health records mostly contain unstructured data, and so they are underutilized by care providers and clinical researchers. Machine learning systems can help organize and standardize information, or can be designed to directly integrate unstructured complex data for high-throughput phenotyping to identify patient cohorts (45).

Bias and Confounding

As discussed in the previous section, bias is another risk that arises with the use of ML. Indeed, a recent review of cardiovascular risk prediction models revealed potential problems in the generalizability of multicentre studies that often show a wide variation in reporting, and thus these models may be biased toward the methods of care routinely used in the interrogated centres (46). For example, in the case of cardiac MRI, protocols are not standardised, varying by institution and machine vendor (47). This bias may amplify the gap in health outcomes between the dominant social group, whose data are used to train algorithms, and the minorities (8). Luckily enough, there are studies that make sure that all minorities are represented in the training data (12, 39), but this should be the rule, not the exception. Another challenge when applying ML models that are designed to recognise patterns lies in the tendency of these to overfit the dataset because they fail at distinguishing a true contributing factor to the clinical outcome from noise (48).

ML solutions can similarly inherit human-like biases (49), such as the model whose recommendation of care after a heart attack was unequal among sex groups (50), as a consequence of a biased training dataset. This bias may also appear in ML-powered echocardiography, where studies are dependent on both the operator performing the study and the interpreter analysing it (47). This bias occurs because we ask ML solutions to predict which decisions the humans profiled in the training data would have made, thus we should not expect the ML method to be fair. The effect of human cognitive biases in ML algorithms have already been addressed and different "debiasing" techniques have been proposed (51), but this is a topic that certainly needs more attention.

Similar to bias, the learning process can be undermined by confounding, i.e., the finding of a spurious association between the input data and the outcome under study. Such is the case of the deep learning model that attempted at predicting ischaemia by looking at ECG records, but rather learnt at detecting the background electrical activity noise present in the ischaemic ECG examples, which was not present in the control cases (52). Unsupervised learning can be beneficial to avoid confounding, as it does not force the output of the model to match a given label but rather finds natural associations within the data (53). Similarly, randomization of experiments is highly recommended to avoid confounding (54).

Validation and Continuous Improvement

Even if an algorithm proves to outperform humans in prediction tasks, systematic debugging, audit and extensive validation should be mandatory. For ML algorithms to be deployed in hospitals, they must improve patient as well as financial outcomes (8). Validation should be through multi-centre randomized prospective trials, to assess whether models trained at one site can be applied elsewhere. Examples of prospective ML trials assessed in a "real world" clinical environment are scarce–only 6% of 516 surveyed studies performed external validation, according to (55). Among these rare examples, finding prospective validation studies to prove the suitability of ML-enabled applications in cardiovascular imaging is even rarer. In (56), a prospective study concluded that a ML model that integrates clinical and quantitative imaging-based features outperforms the prediction of myocardial infarction and death as compared to standard clinical risk assessment. Attia et al. (57) conducted a prospective study to validate a DL algorithm that detected left ventricular systolic dysfunction. Another pivotal prospective multicentre trial was launched to demonstrate the feasibility of a ML-powered image guided acquisition software that helps novices to perform transthoracic echocardiography studies (58). Lastly, a validation study was performed to prove the feasibility of using DL to automatically segment and quantify the ventricular volumes in cardiac MRI (35).

TABLE 1 | SWOT analysis—status interpretation and decision-making.

Strengths	Weaknesses
• Allow objective and thorough comparison to populations • Allow the integration of complex, heterogenous features • May enhance the prediction of clinical outcomes, or the prediction of response to a given treatment or intervention.	• Need well-curated, representative databases for training • Affected by data reliability, representativeness, and bias • Need to extract meaningful, interpretable concepts • Need thorough validation–prospective trials • Need to integrate longitudinal data • Ensure transference of knowledge across populations • Need to prove clinical benefit • Need to be integrated within clinical systems • Need to prove cost-effectiveness
Opportunities	**Threats**
• Stimulate the man/machine collaboration • Reach diagnosis in a shorter time • Separate ambiguous cases that deserve more attention from clear cases–triaging • Help in the organization of healthcare—diagnosis, risk assessment and urgency assessment • Lower cost of healthcare by suggesting cost-effective decisions	• Harm patients if wrong decisions are taken—high-risk • Disappoint users, especially after all the striking news on ML failures • Affect human decisions in a negative way—automation bias • Make decisions for the average patient, not at the individual level

One of the greatest benefits of ML models resides in their ability to improve their performance as more data become available. However, this might be challenging particularly for neural networks, which are prone to "catastrophic forgetting"—to abruptly forget previously learned information upon learning from new data. Furthermore, re-training on the whole database is time and resource consuming. To solve these problems, federated learning, a novel de-centralized computational architecture where machines run models locally to improve them with a single user's data (59), could be helpful. Given the evolving nature of ML models, medical device providers are obliged to periodically monitor the performance of their programs, using a continuous validation paradigm (60).

Auditability/Traceability

Interpretability vs. Explainability

Interpretability is understood as the ability to explain or to present in understandable terms to a human (61). In a strictly-regulated field such as cardiovascular medicine, the lack of interpretability of ML models is one of the main limitations hindering adoption (62). Indeed, from the example of predicting ischaemia by looking at ECG records discussed above (52), it is evident that the non-intelligible use of ML outputs can lead to controversial results and therefore translation to clinical practice should be done cautiously. Unfortunately, many ML implementations available do not comply with the European General Data Protection Regulation (GDPR), which compels ML providers to reveal the information and logic involved in each decision (63).

When reasoning on the data to make decisions, the human brain can follow two approaches (64): the *fast/intuitive* (Type 1) vs. the *slow/reasoned* (Type 2) one. Type 1 is almost instantaneous and based on the human ability to apply heuristics to identify patterns from raw information. However, it is prone to error and bias, as it can lead to an incomplete *perception* of the patient due to low quality or lack of relevant data (65, 66). In contrast, Type 2 is deductive, deliberate, and demands a greater intellectual, time and cost investment, but often turns out to be more accurate.

The above is highly relevant for both "traditional" and ML-based clinical decision-making, as ML systems ultimately mimic different aspects of human reasoning and can lead to the same errors. For the sake of explaining ML decisions, researchers provide attention maps (67), reveal which data the model "looked at" for each individual decision (68), provide estimates of feature importance (69), or explain the local behaviour of complex models by changing the input and evaluating the impact on the prediction (70). However, caution is needed with this entire research trend, as explainability is not a synonym of interpretability (71). Explainable models tend to reach conclusions by fast/intuitive black-box reasoning (Type 1, see also causal vs. predictive learning in the following subsection), while interpretable models demand a slow/reasoned (Type 2) approach throughout the entire learning path. In this sense, explainable models that follow a fast/intuitive reasoning might incur more diagnostic errors than interpretable models, which follow an analytical reasoning (64). The research field that focuses on enhancing models' interpretability is still in its infancy. Generative synthesis (69), which uses ML to generate a simplified version of a neural network, or mathematical attempts to explain the inner working of neural networks (72) could provide key insight into why and how a network behaves the way it does, thus unravelling the black-box enigma (73).

For ML models to be applied in clinical decision-making, they cannot merely be interpretable, but they also must be credible. A credible model is an interpretable model that: (1) provides arguments for its predictions that are, at least in part, in-line with domain knowledge; (2) is at least as good as previous standards in terms of predictive performance (74). For ML models to achieve credibility, the medical expert must be included in the interpretation loop (75).

Together with models, we should also develop strategies to objectively evaluate their interpretability. This can be done by assessing fairness; privacy of data; generalizability; causality, to

prevent spurious correlations; or trust, to make sure the model is right for the right reasons (61). Depending on the application, the interpretability needs might be different. For cardiology applications in particular, ML for status interpretation and decision-making should be equipped with the most-advanced interpretability tools.

Causal ML Rather Than Predictive ML

Predictive ML based on correlations of input data and outcomes may not be enough to truly impact the healthcare system. Indeed, this form of learning can be misleading if important causal variables are not analysed. For auditability reasons, we should probably shift toward finding the root causes of *why* that decision was made, and interpreting the process followed by the algorithm to reach that (diagnostic) decision, i.e., *how* the diagnosis was made. These two questions are addressed by causal ML, a powerful type of analysis aimed at inferring the mechanisms of the system producing the output data. In practice, causal models provide detailed maps of interaction between variables, so the users can simulate cause and effect of future actions (76).

System-Related

Security

Machine learning raises a handful of data security and privacy issues, as DL models require enormous datasets for training purposes. The most secure way to transfer data between healthcare organizations is still unclear, and stakeholders no longer underestimate the hazards of a high-profile data leakage. Hacking is even more harmful, as hackers could manipulate a decision-making model to damage people at a large scale.

The European GDPR compels to adopt security measures against hacking and data breaches. A potential solution that has been largely discussed is Blockchain, a technology that enables data exchange systems that are cryptographically secured and irrevocable, by providing a public and immutable log of transactions and "smart contracts" to regulate data access. The downsides of Blockchain's technology are that it is slow, costly to maintain, and hard to scale (77). As an alternative, federated learning could guarantee the security of patient data (see "Validation and continuous improvement subsection"), as this model-training paradigm allows updating a learning model locally without sharing individual information with a central system (78).

Regulatory

The use of ML for clinical decision-making unavoidably brings legal challenges regarding medical negligence derived from learning failures. When such negligence arises, the legal system needs to provide guidance on what entity holds liability, for which recommendations have been developed (79). Furthermore, the evolving nature of ML models poses a unique challenge to regulatory agencies, and the best way to evaluate updates remains unclear (60). Policymakers should generate specific criteria for demonstrating non-inferiority of algorithms compared to existing standards, specially emphasizing the quality of the training data and the validation process (80). Regulatory bodies must also ensure that algorithms are used properly and for people's welfare. In summary, for ML technology to be adopted by cardiology departments within the next years many legal aspects still need to be addressed, and decision- and policy-makers should join efforts toward this end.

Integration (Man/Machine Coexistence)

The scenario of ML tools replacing humans in clinical medicine is highly unlikely (81). Besides the formidable challenges for ML solutions discussed above (8), cardiologists will still be needed to interact with the patients and perform physical examinations, navigate diagnostic procedures, integrate and adapt ML solutions according to the changing stages of disease or patient's preferences, inform the patient's family about therapy options, or console them if the disease stage is very advanced.

Accordingly, instead of a human-machine competition, we should rather think of a cooperation paradigm, where ML is used to augment human intelligence–targeting repetitive sub-tasks to assist physicians to reach a more informed decision, more efficiently. Indeed, ML and humans possess complementary skills: ML stands out at pattern recognition on massive amounts of data, whereas people are far better at understanding the context, abstracting knowledge from their experience, and transferring it across domains. Human-in-the-loop approaches facilitate cooperation by enabling users to interact with ML models without requiring in-depth technical knowledge. However, understanding where ML models can be used and at which level is crucial to avoid preventable errors attributed to automation bias (43). Examples of human-machine collaboration already exist. Indeed, a ML algorithm cleared by the FDA improved the diagnosis of wrist fractures when clinicians used it, as compared to clinicians alone (82). In diabetic retinopathy diagnosis (83), model assistance increased the accuracy of retina specialists above that of the unassisted reader or model alone. In cardiovascular imaging, most examples of human-machine collaboration thus far focus on segmentation, and detection-classification of imaging planes (84).

In light of this, the current clinical workflow could be rethought: the ML system would propose a diagnosis, the human operator then revises the data on which the conclusions are drawn, informing the system of potential measurement errors or confounders, and finally accepts or rejects the diagnosis. Thus, the human operator preserves the overall control, while machines perform measurements and integrate and compare data at request (75). Ultimately, this human-machine symbiosis will be beneficial to release physicians from low-level tasks such as cardiac measurements, data preparation, and standardization, to give them more time on higher-level tasks such as patient care and clinical decision-making (85).

ML Applied to Real Clinical Data

In human decision-making, a clinician would explore all available data and compare them to patients they have seen before or were trained to recognize. Once an individual is put into context with regards to expected normality and typical cases, previous knowledge on treatment effect is used to manage this individual patient. This 'eminence-based' approach is only within reach of very experienced clinicians. For standardization, many

professional organizations provide diagnostic guidelines based on data from large cohorts or clinical trials (86–88). Although guidelines have significantly contributed to improve medical care, they do not consider the full data available. In this sense, the use of ML seems amply justified.

Most ML models are trained with data collected following strict input criteria and well-defined protocols used in randomized clinical trials (89, 90), but routinely collected data is often much noisier, heterogeneous and incomplete. ML techniques need to deal with incompleteness, either by performing imputation or by adopting formulations that explicitly consider that the data can be incomplete. Furthermore, patients often lie outside the narrow selection criteria of cardiology trials (including co-morbidities, ethnicity, gender, age, lifestyle, etc.), may have been differently treated before the investigation, may present at a different stage of disease, and most importantly, may undergo different decision pathways during the study. On top of this, obtaining a hard outcome to train an algorithm is often difficult, e.g., to register death, the study would need to be conducted until everybody dies, which is unfeasible both for time and economic constraints. Even if registered, often the outcome is scarce, and when appearing, the reason for experiencing it may be different among patients (91).

All these aspects make it extremely challenging to associate input descriptors to outcomes using supervised predictive ML/DL techniques, which may fail to understand the context from which data have been drawn, and thus yield unwanted results that might harm patients. A more promising approach could be based on unsupervised dimensionality reduction, a label-agnostic approach where input descriptors are used to position individuals according to their similarity and combined with previous knowledge this similarity can be used to infer diagnosis or to predict treatment response (23).

FUTURE PERSPECTIVES

The foreseeable application of ML in the short to mid-term is to perform specific and well-defined tasks relating to data acquisition, predominantly by extracting standardized, high-quality information with the smallest possible learning curve. In this sense, DL solutions already help extracting information with minimal or even without the need of human intervention (8, 92), or aid selecting the images that are good enough for subsequent clinical interrogation (93). Another evident application of ML that will soon be ubiquitous in clinical practice is that of image analysis, which will discharge cardiologists from monotonous activities related to feature extraction from images (94), thus freeing them up to dedicate more time to higher-level tasks involving interpretation, patient care, and decision-making.

For the topics covered in this paper, i.e., the higher-level tasks involving reasoning, such as patient's status interpretation and decision support, ML applications are still immature and need substantial validation. A modest number of ongoing clinical trials have been conceived to tackle these drawbacks. One such example is the current investigation aiming at validating a DL model that diagnoses different arrhythmias (AF, supraventricular tachycardia, AV-block, asystole, ventricular tachycardia and ventricular fibrillation) on 12-lead ECGs and single-lead Holter monitoring registered in 25,458 participants (95). Another example is the clinical study that will interrogate stress echocardiography scans with ML models to discriminate normal hearts from those at risk of a heart attack in a prospective cohort of 1,250 participants (96). Considering the above, we do not expect to see a vast penetration of ML-enabled applications for patient's status interpretation and decision support in clinical practice in the foreseeable future.

Whatever the application, the penetration of ML models into routine practice will be subject to their seamless integration into the clinical decision pathway used by cardiologists. Furthermore, we consider that the upcoming policies for ML research in healthcare should address the challenges described in the previous section, which can only be achieved by multidisciplinary teams. On the algorithmic side, more attention should be dedicated to dealing with longitudinal data, and how to relate the ML conclusions with pathophysiological knowledge. Data integration and what is the best approach for dealing with incomplete data and outliers should be also surveyed. On the validation side, generalization performance should be systematically reported, and uncertainty quantification methods should be developed to establish trust in the (predictive) models. Finally, the practical considerations that will affect adoption of the ML technology, such as how ML software should be integrated with the archiving and communication system of the hospital or how it would be paid for by facilities, should be explored. For these, a clear demonstration of the cost-effectiveness of ML technology in healthcare systems and its positive impact on patients' outcomes is needed.

CONCLUSION

ML algorithms allow computers to automatically discover patterns in the data and improve with experience. Together with the enormous computational capacities of modern servers and the overwhelming amount of data resulting from the digitalization of healthcare systems, these algorithms open the door for a paradigm shift in clinical decision-making in cardiology. However, their seamless integration is dependent on the understanding of the intrinsic processes being used during the conventional pathway by which clinicians make decisions, which in turn helps identifying the areas where certain types of ML models can be most beneficial. If the obstacles and pitfalls that have been covered in this paper can be addressed satisfactorily, then ML might indeed revolutionize many aspects of healthcare, including cardiovascular medicine. For the promise to be fulfilled, engineers and clinicians will need to engage jointly in intensive development and validation of specific ML-enabled clinical applications.

AUTHOR CONTRIBUTIONS

SS-M participated in the conception and design of the review, drafted the work, approved the final version and agreed on the accuracy and integrity of the work. OC, GP, MC, MG-B, MM, AV, EG, and AF helped drafting the review, revised it critically

for important intellectual content, approved the final version, and agreed on the accuracy and integrity of the work. BB participated in the conception and design of the review, drafted the work, approved the final version, and agreed on the accuracy and integrity of the work. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Spanish Ministry of Economy and Competitiveness (María de Maeztu Programme for R&D [MDM-2015-0502], Madrid, Spain) and by the Fundació La Marató de TV3 (n°20154031, Barcelona, Spain). The work of SS-M was supported by IDIBAPS and by the HUMAINT project of the Joint Research Centre of the European Commission. AV's contribution is funded by Spanish research project PID2019-104551RB-I00.

ACKNOWLEDGMENTS

The author thanks to Nicolas Duchateau, who carefully read this manuscript and provided valuable feedback.

REFERENCES

1. Smuha N, AI, HLEG. High-level expert group on artificial intelligence. In: *Ethics Guidelines for Trustworthy AI*. (2019).
2. 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.
3. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. (2016) 316:2402. doi: 10.1001/jama.2016.17216
4. Nie D, Cao X, Gao Y, Wang L, Shen D. "Estimating CT image from MRI data using 3d fully convolutional networks," In: *International Workshop on Deep Learning in Medical Image Analysis*. (Cham: Springer), 170–178.
5. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *npj Digit Med.* (2018) 1:6. doi: 10.1038/s41746-017-0013-1
6. Desai AS, Jhund PS, Yancy CW, Lopatin M, Stevenson LW, Marco T De, et al. After TOPCAT: what to do now in heart failure with preserved ejection fraction. *Eur Heart J.* (2016) 47:1510–8. doi: 10.1093/eurheartj/ehw114
7. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal.* (2017) 36:61–78. doi: 10.1016/j.media.2016.10.004
8. 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
9. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP, et al. Artificial intelligence in cardiovascular imaging: jacc state-of-the-art review. *J Am Coll Cardiol.* (2019) 73:1317–35. doi: 10.1016/j.jacc.2018.12.054
10. 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 state of the art review. *Eur Heart J.* (2020) 41:4556–64. doi: 10.1093/eurheartj/ehaa159
11. 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. *The Lancet Digit Health*. (2019) 1:e271–97. doi: 10.1016/S2589-7500(19)30123-2
12. Fonseca CG, Backhaus M, Bluemke DA, Britten RD, Chung J, 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
13. Rohé M-M, Datar M, Heimann T, Sermesant M, Pennec X. SVF-Net: learning deformable image registration using shape matching. *Lecture Notes Comput Sci.* (2017) 10433:266–74. doi: 10.1007/978-3-319-66182-7_31
14. Toth D, Miao S, Kurzendorfer T, Rinaldi CA, Liao R, Mansi T, et al. 3D/2D model-to-image registration by imitation learning for cardiac procedures. *Int J Comput Assist Radiol Surg.* (2018) 13:1141. doi: 10.1007/s11548-018-1774-y
15. Krebs J, Delingette H, Mailhe 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
16. Krittanawong C, Virk HUH, Bangalore S, Wang Z, Johnson KW, Pinotti R, et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. *Scientific Rep.* (2020) 10:1–11. doi: 10.1038/s41598-020-72685-1
17. 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
18. Beecy AN, Chang Q, Anchouche K, Baskaran L, Elmore K, Kolli K, et al. A novel deep learning approach for automated diagnosis of acute ischemic infarction on computed tomography. *JACC: Cardiovascular Imaging.* (2018) 11:1723–5. doi: 10.1016/j.jcmg.2018.03.012
19. Madani A, Ong JR, Tibrewal A, Mofrad MRK. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *npj Digital Med.* (2018) 1:59. doi: 10.1038/s41746-018-0065-x
20. Jiang H, Mao H, Lu H, Lin P, Garry W, Lu H, et al. Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease. *Int J Med Inform.* (2021) 145:104326. doi: 10.1016/j.ijmedinf.2020.104326
21. Dey D, Gaur S, Ovrehus KA, Slomka PJ, Betancur J, Goeller M, et al. Integrated prediction of lesion-specific ischaemia from quantitative coronary CT angiography using machine learning: a multicentre study. *Euro Radiol.* (2018) 28:2655–2664. doi: 10.1007/s00330-017-5223-z
22. Loncaric F, Marti Castellote P-M, Sanchez-Martinez S, Fabijanovic D, Nunno L, Mimbrero M, et al. Automated pattern recognition in whole–cardiac cycle echocardiographic data: capturing functional phenotypes with machine learning. *J Am Soc Echocardiograph*. (2021) 6:14. doi: 10.1016/j.echo.2021.06.014
23. 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
24. Nogueira M, De Craene M, Sanchez-Martinez S, Chowdhury D, Bijnens B, Piella G. Analysis of nonstandardized stress echocardiography sequences using multiview dimensionality reduction. *Med Image Anal.* (2020) 60:101594. doi: 10.1016/j.media.2019.101594
25. Funkner AA, Yakovlev AN, Kovalchuk S V. Data-driven modeling of clinical pathways using electronic health records. *Procedia Comput Sci.* (2017) 121:835–42. doi: 10.1016/j.procs.2017.11.108
26. Morid MA, Kawamoto K, Ault T, Dorius J, Abdelrahman S. Supervised learning methods for predicting healthcare costs: systematic literature review and empirical evaluation. *AMIA Sympo Proceed.* (2017) 2017:1312–21.
27. Kwon J, Kim K-H, Jeon K-H, Park J. Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography. *Echocardiography*. (2019) 36:213–8. doi: 10.1111/echo.14220
28. Arsanjani R, Dey D, Khachatryan T, Shalev A, Hayes SW, Fish M, et al. Prediction of revascularization after myocardial perfusion spect by machine learning in a large population. *J Nuclear Cardiol Offic Publicat Am Soc Nucl Cardiol.* (2015) 22:877. doi: 10.1007/s12350-014-0027-x
29. Betancur J, Otaki Y, Motwani M, Fish MB, Lemley M, Dey D, et al. Prognostic value of combined clinical and myocardial perfusion imaging

data using machine learning. *JACC: Cardiovascul Imaging.* (2018) 11:1000–9. doi: 10.1016/j.jcmg.2017.07.024
30. Zeleznik R, Foldyna B, Eslami P, Weiss J, Alexander I, Taron J, et al. Deep convolutional neural networks to predict cardiovascular risk from computed tomography. *Nature Commun.* (2021) 12:1–9. doi: 10.1038/s41467-021-20966-2
31. Diller G-P, Kempny A, Babu-Narayan S V, Henrichs M, Brida M, Uebing A, et al. Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients. *Eur Heart J.* (2019) 40:1069–77. doi: 10.1093/eurheartj/ehy915
32. Wang J, Ding H, Bidgoli FA, Zhou B, Iribarren C, Molloi S, et al. Detecting cardiovascular disease from mammograms with deep learning. *IEEE Trans Med Imaging.* (2017) 36:1172–81. doi: 10.1109/TMI.2017.2655486
33. Poplin R, Varadarajan A V., Blumer K, Liu Y, McConnell M V, 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
34. Rim TH, Lee CJ, Tham Y-C, Cheung N, Yu M, Lee G, et al. Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. *Lancet Digit Health.* (2021) 3:e306–16. doi: 10.1016/S2589-7500(21)00043-1
35. Retson TA, Masutani EM, Golden D, Hsiao A. Clinical performance and role of expert supervision of deep learning for cardiac ventricular volumetry: a validation study. *Radiol Artifi Intell.* (2020) 2:e190064. doi: 10.1148/ryai.2020190064
36. Karagodin I, Singulane CC, Woodward GM, Xie M, Tucay ES, Rodrigues ACT, et al. Echocardiographic correlates of in-hospital death in patients with acute Covid-19 infection: the world alliance societies of echocardiography (WASE-COVID) study. *J Am Soc Echocardiograph.* (2021) 34:819–30. doi: 10.1016/j.echo.2021.05.010
37. Asch FM, Mor-Avi V, Rubenson D, Goldstein S, Saric M, Mikati I, et al. Deep learning–based automated echocardiographic quantification of left ventricular ejection fraction: a point-of-care solution. *Circul Cardiovascul Imaging.* (2021) 21:528–537. doi: 10.1161/CIRCIMAGING.120.012293
38. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ.* (2018) 363:4245. doi: 10.1136/bmj.k4245
39. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux A V, Folsom AR, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol.* (2002) 156:871–81. doi: 10.1093/aje/kwf113
40. Nogueira M, Piella G, Sanchez-Martinez S, Langet H, Saloux E, Bijnens B, et al. "Characterizing patterns of response during mild stress-testing in continuous echocardiography recordings using a multiview dimensionality reduction technique," In" *Functional Imaging and Modelling of the Heart—Conference Proceedings.*
41. Lee G, Kang B, Nho K, Sohn K-A, Kim D. MildInt: deep learning-based multimodal longitudinal data integration framework. *Front Genet.* (2019) 10:617. doi: 10.3389/fgene.2019.00617
42. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. *Lancet Digit Health.* (2020) 2:e489–92. doi: 10.1016/S2589-7500(20)30186-2
43. Cummings ML. "Automation bias in intelligent time critical decision support systems," In: *AIAA 3rd Intelligent Systems conference*. (2004), p. 6313.
44. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc : JAMIA.* (2012) 19:121–7. doi: 10.1136/amiajnl-2011-000089
45. Pathak J, Bailey KR, Beebe CE, Bethard S, Carrell DS, Chen PJ, et al. Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPn consortium. *J Am Med Inform Assoc.* (2013) 20:e341–8. doi: 10.1136/amiajnl-2013-001939
46. Wynants L, Kent DM, Timmerman D, Lundquist CM, Calster B. Untapped potential of multicenter studies: a review of cardiovascular risk prediction models revealed inappropriate analyses and wide variation in reporting. *Diagnostic Prognostic Res.* (2019) 3:9. doi: 10.1186/s41512-019-0046-9
47. Tat E, Bhatt DL, Rabbat MG. Addressing bias: artificial intelligence in cardiovascular medicine. *The Lancet Digital Health.* (2020) 2:e635–6. doi: 10.1016/S2589-7500(20)30249-1
48. Kagiyama N, Shrestha S, Farjo PD, Sengupta PP. Artificial intelligence: practical primer for clinical research in cardiovascular disease. *J Am Heart Assoc.* (2019) 8:12788. doi: 10.1161/JAHA.119.012788
49. Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science.* (2017) 356:183–6. doi: 10.1126/science.aal4230
50. Wilkinson C, Bebb O, Dondo TB, Munyombwe T, Casadei B, Clarke S, et al. Sex differences in quality indicator attainment for myocardial infarction: a nationwide cohort study. *Heart.* (2019) 105:516–23. doi: 10.1136/heartjnl-2018-313959
51. Kliegr T, Stěpán Bahník, Fürnkranz J. A review of possible effects of cognitive biases on interpretation of rule-based machine learning models. *AI.* (2021) 103458. doi: 10.1016/j.artint.2021.103458
52. Brisk R, Bond R, Finlay D, McLaughlin J, Piadlo A, Leslie SJ, et al. The effect of confounding data features on a deep learning algorithm to predict complete coronary occlusion in a retrospective observational setting. *Euro Heart J Digit Health.* (2021) 2:127–34. doi: 10.1093/ehjdh/ztab002
53. Hastie T, Tibshirani R, Friedman J. *The elements of Statistical Learning (Vol.1).* Springer, Berlin: Springer series in statistics (2001).
54. Pourhoseingholi MA, Baghestani AR, Vahedi M. How to control confounding effects by statistical analysis. *Gastroenterol Hepatol bed to bench.* (2012) 5:79–83.
55. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. *Korean J Radiol.* (2019) 20:405. doi: 10.3348/kjr.2019.0025
56. Commandeur F, Slomka PJ, Goeller M, Chen X, Cadet S, Razipour A, et al. Machine learning to predict the long-term risk of myocardial infarction and cardiac death based on clinical risk, coronary calcium, and epicardial adipose tissue: a prospective study. *Cardiovasc Res.* (2020) 116:2216–25. doi: 10.1093/cvr/cvz321
57. Attia ZI, Kapa S, Yao X, Lopez-Jimenez F, Mohan TL, Pellikka PA, et al. Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. *J Cardiovasc Electrophysiol.* (2019) 30:668–74. doi: 10.1111/jce.13889
58. Surette S, Narang A, Bae R, Hong H, Thomas Y, Cadieu C, et al. Artificial intelligence-guided image acquisition on patients with implanted electrophysiological devices: results from a pivotal prospective multi-center clinical trial. *Euro Heart J.* (2020) 41:6. doi: 10.1093/ehjci/ehaa946.0006
59. Konečný J, McMahan HB, Yu FX, Richtarik P, Suresh AT, Bacon D. Federated learning: strategies for improving communication efficiency. *arXiv [preprint].* (2016)
60. FDA. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD)-discussion paper and request for feedback. (2019).
61. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. *arXiv [preprint].* (2017).
62. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA.* (2017) 318:517. doi: 10.1001/jama.2017.7797
63. Goodman B, Flaxman S. European Union regulations on algorithmic decision-making and a "right to explanation." *AI Magazine.* (2017) 38:2741. doi: 10.1609/aimag.v38i3.2741
64. Croskerry P, A. universal model of diagnostic reasoning. *Acad Med.* (2009) 84:1022–8. doi: 10.1097/ACM.0b013e3181ace703
65. Benavidez OJ, Gauvreau K, Geva T. Diagnostic errors in congenital echocardiography: importance of study conditions. *J Am Soc Echocardiograph.* (2014) 27:616–23. doi: 10.1016/j.echo.2014.03.001
66. Balogh EP, Miller BT, Ball JR. *Improving diagnosis in health care and M National Academies of Sciences, Engineering.* Washington, DC: National Academies Press. (2015).
67. Lee H, Yune S, Mansouri M, Kim M, Tajmir SH, Guerrier CE, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nat Biomed Eng.* (2018) 1:9. doi: 10.1038/s41551-018-0324-9
68. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *npj Digital Med.* (2018) 1:18. doi: 10.1038/s41746-018-0029-1

69. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* (2018) 2:749–60. doi: 10.1038/s41551-018-0304-0
70. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. (2016). p. 1135–44. doi: 10.1145/2939672.2939778
71. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* (2019) 1:206–15. doi: 10.1038/s42256-019-0048-x
72. Mallat S. Understanding deep convolutional networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.* (2016) 374:20150203. doi: 10.1098/rsta.2015.0203
73. Kumar D, Taylor GW, Wong A. Discovery radiomics with CLEAR-DR: interpretable computer aided diagnosis of diabetic retinopathy. *IEEE Access.* (2017) 7:25891–6. doi: 10.1109/ACCESS.2019.2893635
74. Wang J, Oh J, Wang H, Wiens J. Learning credible models. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. (2018) 2417–26. doi: 10.1145/3219819.3220070
75. D'hooge J, Fraser AG. Learning about machine learning to create a self-driving echocardiographic laboratory. *Circulation.* (2018) 138:1636–8. doi: 10.1161/CIRCULATIONAHA.118.037094
76. Pearl J, Mackenzie D. *The Book of Why : The New Science of Cause And Effect*. London: Basic Books, Inc. (2018).
77. Song J. Why blockchain is hard. *Medium*. (2018)
78. Silva S, Gutman B, Romero E, Thompson PM, Altmann A, Lorenzi M. Federated learning in distributed medical databases: meta-analysis of large-scale subcortical brain data. In: *2019 IEEE 16th International Symposium on Biomedical Imaging*. IEEE (2018). p. 270–4. doi: 10.1109/ISBI.2019.8759317
79. Expert Group on Liability and New Technologies. *Report on Liability for Artificial Intelligence and Other Emerging Digital Technologies - European Commision*. (2019).
80. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng.* (2018) 2:719–31. doi: 10.1038/s41551-018-0305-z
81. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *NE J Med.* (2016) 375:1216–9. doi: 10.1056/NEJMp1606181
82. Voelker R. Diagnosing fractures with AI. *JAMA*. (2018) 320:23. doi: 10.1001/jama.2018.8565
83. Sayres R, Taly A, Rahimy E, Blumer K, Coz D, Hammel N, et al. Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy. *Ophthalmology*. (2018) 11:16. doi: 10.1016/j.ophtha.2018.11.016
84. Soares De Siqueira V, De Castro Rodrigues D, Nunes Dourado C, Marcos Borges M, Gomes Furtado R, Pereira Delfino H, et al. "Machine learning applied to support medical decision in transthoracic echocardiogram exams: a systematic review," In: *Proceedings - 2020 IEEE 44th Annual Computers, Software, and Applications Conference, COMPSAC 2020.* (2020), p. 400–407.
85. Sengupta PP, Adjeroh DA. Will artificial intelligence replace the human echocardiographer? *Circulation.* (2018) 138:1639–42. doi: 10.1161/CIRCULATIONAHA.118.037095
86. Ponikowski P, Voors AA, Anker SD, Bueno H, González-Juanatey JR, Harjola V-P, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* (2016) 37:2129–200. doi: 10.1093/eurheartj/ehw128
87. Brignole M, Auricchio A, Baron-Esquivias G, Bordachar P, Boriani G, Breithardt OA, et al. ESC Guidelines on cardiac pacing and cardiac resynchronization therapy. *Eur Heart J.* (2013) 15:1070–118. doi: 10.15829/1560-4071-2014-4-5-63
88. Epstein AE, Dimarco JP, Ellenbogen KA, Estes NAM, Freedman RA, Gettes LS, et al. ACC/AHA/HRS 2008 Guidelines of cardiac rhythm abnormalities. A report of the American College of Cardiology/American Heart Association task force on practice guidelines (writing committee to revise the ACC/AHA/NASPE 2002 Guideline update for implantation. *Circulation.* (2008) 117:350–408.
89. Kalscheur MM, Kipp RT, Tattersall MC, Mei C, Buhr KA, DeMets DL, et al. Machine Learning Algorithm Predicts Cardiac Resynchronization Therapy Outcomes. *Circulation: Arrhythm Electrophysiol.* (2018) 11:499. doi: 10.1161/CIRCEP.117.005499
90. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Gregory Hundley W, McClelland R, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circ Res.* (2017) 121:1092–101. doi: 10.1161/CIRCRESAHA.117.311312
91. Oladapo OT, Souza JP, Bohren MA, Tunçalp Ö, Vogel JP, Fawole B, et al. Better Outcomes in Labour Difficulty (BOLD) project: innovating to improve quality of care around the time of childbirth. *Reprod Health.* (2015) 12:48. doi: 10.1186/s12978-015-0027-6
92. Housden J, Wang S, Noh Y, Singh D, Singh A, Back J, et al. "Control strategy for a new extra-corporeal robotic ultrasound system," In: *MPEC* (2017).
93. Wu L, Cheng J-Z, Li S, Lei B, Wang T, Ni D, et al. fetal ultrasound image quality assessment with deep convolutional networks. *IEEE Trans Cybern.* (2017) 47:1336–49. doi: 10.1109/TCYB.2017.2671898
94. 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
95. NCT03662802. *Development of a Novel Convolution Neural Network for Arrhythmia Classification. ClinicalTrials.gov*. Available online at: https://clinicaltrials.gov/ct2/show/NCT03662802?term=Machine$+$Learning%2C$+$AI&draw=3&rank=29
96. NCT04193475. *Machine Learning in Quantitative Stress Echocardiography. ClinicalTrials.gov*. Available online at: https://clinicaltrials.gov/ct2/show/NCT04193475?term=machine$+$learning%2C$+$echocardiography&draw=2&rank=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.

frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 24 January 2022
doi: 10.3389/fcvm.2021.769927

WSSNet: Aortic Wall Shear Stress Estimation Using Deep Learning on 4D Flow MRI

Edward Ferdian[1]*, *David J. Dubowitz*[1], *Charlene A. Mauger*[1], *Alan Wang*[1,2] and *Alistair A. Young*[1,3]*

[1] Department of Anatomy and Medical Imaging, University of Auckland, Auckland, New Zealand, [2] Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand, [3] Department of Biomedical Engineering, King's College London, London, United Kingdom

Wall shear stress (WSS) is an important contributor to vessel wall remodeling and atherosclerosis. However, image-based WSS estimation from 4D Flow MRI underestimates true WSS values, and the accuracy is dependent on spatial resolution, which is limited in 4D Flow MRI. To address this, we present a deep learning algorithm (WSSNet) to estimate WSS trained on aortic computational fluid dynamics (CFD) simulations. The 3D CFD velocity and coordinate point clouds were resampled into a 2D template of 48 × 93 points at two inward distances (randomly varied from 0.3 to 2.0 mm) from the vessel surface ("velocity sheets"). The algorithm was trained on 37 patient-specific geometries and velocity sheets. Results from 6 validation and test cases showed high accuracy against CFD WSS (mean absolute error 0.55 ± 0.60 Pa, relative error 4.34 ± 4.14%, 0.92 ± 0.05 Pearson correlation) and noisy synthetic 4D Flow MRI at 2.4 mm resolution (mean absolute error 0.99 ± 0.91 Pa, relative error 7.13 ± 6.27%, and 0.79 ± 0.10 Pearson correlation). Furthermore, the method was applied on *in vivo* 4D Flow MRI cases, effectively estimating WSS from standard clinical images. Compared with the existing parabolic fitting method, WSSNet estimates showed 2–3× higher values, closer to CFD, and a Pearson correlation of 0.68 ± 0.12. This approach, considering both geometric and velocity information from the image, is capable of estimating spatiotemporal WSS with varying image resolution, and is more accurate than existing methods while still preserving the correct WSS pattern distribution.

Keywords: 4D Flow MRI, computational fluid dynamics, deep learning, wall shear stress (WSS), aorta

OPEN ACCESS

Edited by:
Paul Leeson,
University of Oxford, United Kingdom

Reviewed by:
Jan Brüning,
Charité University Medicine Berlin, Germany
Giuseppe Muscogiuri,
Istituto Auxologico Italiano (IRCCS), Italy

***Correspondence:**
Edward Ferdian
e.ferdian@auckland.ac.nz
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: *02 September 2021*
Accepted: *30 November 2021*
Published: *24 January 2022*

Citation:
Ferdian E, Dubowitz DJ, Mauger CA, Wang A and Young AA (2022) WSSNet: Aortic Wall Shear Stress Estimation Using Deep Learning on 4D Flow MRI. Front. Cardiovasc. Med. 8:769927. doi: 10.3389/fcvm.2021.769927

INTRODUCTION

Wall shear stress (WSS) is an important contributor to vessel wall remodeling and atherosclerosis (1–3). WSS is defined as the shear force produced by tangential blood flow on the vessel wall as a result of blood viscosity and is related to the gradient of velocity in the surface normal direction. Previous studies suggest that wall shear stress is an important biomarker for atherosclerosis formation (4, 5). Both low and high time-averaged WSS (TAWSS) have been suggested to be associated with pathology. Moreover, recent studies also found that a high oscillatory shear index (OSI) plays an important role in causing wall thickening (6). Early detection of these biomarkers may provide useful information for clinical practice. However, despite recent findings on the importance of WSS and related measures, there is yet no practical method to accurately measure WSS from clinical data.

Magnetic resonance imaging (MRI) phase contrast imaging methods enable non-invasive quantification of the three-dimensional time varying velocity field (4D Flow MRI) (7, 8). However, the spatiotemporal resolution of 4D Flow MRI is limited. Several existing methods have employed curve fitting to estimate WSS from velocity derivatives near the vessel wall (9–11). Stalder et al. (9) introduced a velocity-based method using B-spline interpolation. A later study investigated several approaches based on velocity mapping, Fourier velocity encoding, and intravoxel velocity SD mapping (12). Overall, all these methods are dependent on spatial resolution, segmentation accuracy, velocity encoding (VENC), and voxel position relative to the wall, with each method being more sensitive to different parameters. These methods show consistent reproducibility when the comparison between methods was performed relative to each other. However, the WSS obtained using these methods were consistently lower compared to the values obtained from computational fluid dynamics (CFD) (13, 14). This underestimation is likely due to the limited resolution of 4D Flow MRI, as hemodynamic parameters may be biased due to partial volume effects and temporal blurring. The need to have a higher resolution MRI is constrained by the limited amount of examination time.

Computational modeling enables physics-based simulation of clinical data at high resolution, constrained only by computation resources. While it is possible to achieve accurate estimates for hemodynamic variables, CFD simulations require patient-specific parameters. These boundary conditions are not always obtainable and often rely on assumptions such as vessel rigidity, incompressible fluid, and pressure estimations. Moreover, the amount of computation required to solve the numerical problem is often not feasible in a clinical setting.

Recent developments in medical imaging and deep learning have enabled the use of physics-based simulations as surrogates to train a deep learning model (15–18). These approaches offer high accuracy compared to the CFD ground truth by learning spatial representation of geometric features. Liang et al. (18) used a shape decoding technique to train a network to estimate aortic stress distributions based on the input mesh. Similarly, Acebes et al. (16) presented a CNN-based network to estimate endothelial cell activation potential (ECAP) using an unwrapped model of the left atria. Gharleghi et al. (15) also presented a deep learning method to estimate TAWSS in left main coronary bifurcations by using geometric information as the input. Conversely, conventional curve fitting methods have used velocity information at constant spatial locations (equidistant inward normal). By combining the use of CFD simulations, variable geometric features, velocity information, and deep learning, a fast and accurate method that can be applied to *in vivo* data can be developed.

This study proposes a deep learning approach to estimate WSS based on patient-specific aortic vessel geometries and velocity information. To achieve this, CFD simulations were generated for patient-specific geometries in order to extract a uniform grid sampling of spatial and velocity information at two inward distances from the aortic vessel wall. The locations of the sampled velocity sheets were encoded as coordinate flatmaps and varied over a range of values, enabling the network to learn the relationships between geometry, sampling distance, velocity, and WSS. WSS vectors were output as a uniform-grid flatmap, predicted at any given time frame, enabling the calculation of other WSS measures, such as TAWSS and OSI. The method was validated on synthetic 4D Flow MRI data derived from the CFD simulations. Additionally, the method was applied to clinical *in vivo* cases in comparison with the parabolic fitting method.

METHODS

The following section provides a detailed description of the methodology used in this study. First, the data generation process is described, including the geometry extraction from 4D Flow MRI, CFD simulation setup, and input data preparation step. Second, the description of WSSNet is presented, including the network architecture, loss function definition, and training hyperparameters setup. Finally, the performance of the network is evaluated with respect to the estimated WSS magnitudes, WSS distribution, time-averaged WSS (TAWSS), and OSI, with quantifications performed in the CFD dataset, synthetic MRI from CFD, and actual *in vivo* cases.

Geometry Extraction

Clinical cardiac 4D Flow MRI were obtained for a total of 59 volunteers and patients using a prototype sequence. Data were acquired using a 1.5T scanner (MAGNETOM Avanto fit, Siemens Healthcare, Erlangen, Germany). 4D Flow images were acquired with retrospective gating, encoding velocities of 150 cm/s (VENC) at 2.375 mm grid spacing, 2.75 mm slice thickness, covering the entire heart and great vessels. Other parameters included repetition and echo time (TR/TE) of 38.3 and 2.3 ms, respectively, and flip angle 7°, with 38–58 ms temporal resolution and ∼20 reconstructed frames.

Sixteen cases were excluded due to low image quality, leaving 43 cases for this study (34 healthy, 9 left ventricular hypertrophy). Patient-specific aortic geometries were extracted from the 43 *in vivo* cases. Phase contrast magnetic resonance angiography (PC-MRA) images (temporal mean) were constructed to define the anatomical structure. For each case, two segmentations were performed, one with aortic branches, and another one without. We refer to the segmentation without branches as the "*aorta-only-segmentation.*" For consistency, the branch segmentations contained 3 aortic branches: brachiocephalic artery (BCA), left common carotid artery (LCCA), and left subclavian artery (LSA). For the aorta-only-segmentation, these branches were simply excluded. Segmentations were performed semi-automatically using the ITK-SNAP (19) active contour method. The resulting segmentations were exported as a surface mesh.

The 3D aortic segmentations (ones with the aortic branches) were truncated at the ascending aorta distal to the aortic root and at the distal end of the thoracic aorta to obtain flat inlet and outlet surfaces to be used for CFD simulations. The aortic branches were also truncated ∼2 cm from their bases. An additional smoothing operation (10–12 iterations of vertices' distance averaging) was applied to smooth out the rough surface obtained from the segmentation. All these steps

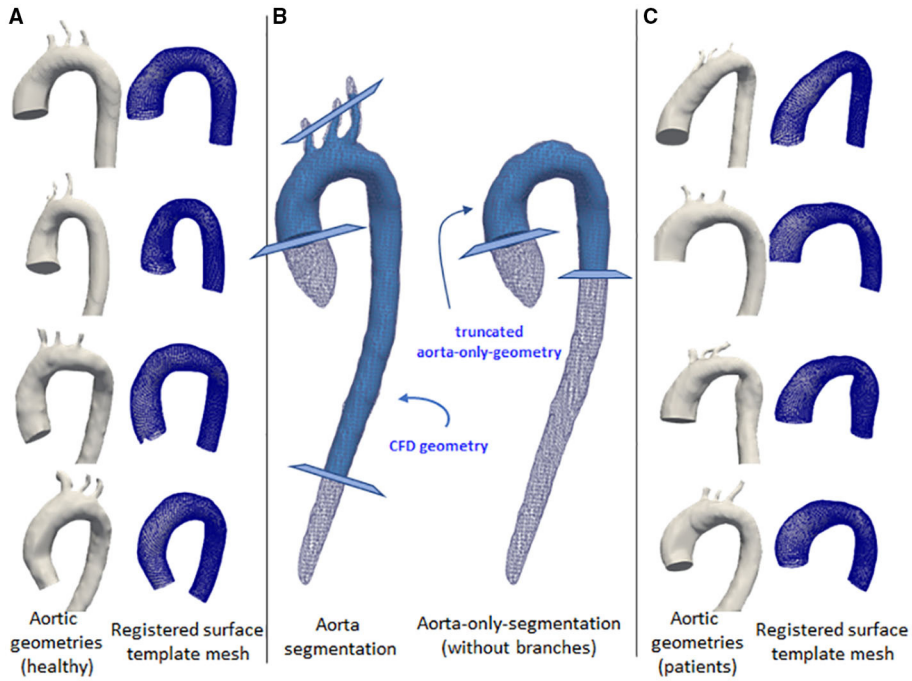


FIGURE 1 | Some examples of the aortic geometry in the dataset. The first and third columns show the corresponding aorta (left) and the registered surface template mesh (right). The first column **(A)** shows aorta geometries for normal volunteers. Third column **(C)** show aorta geometries from left ventricular hypertrophy cases. The middle column **(B)** shows the aortic segmentation with branches (left) and aorta-only-segmentation (right), shown with white wireframes. Light blue geometries show the processed segmentation for the use of computational fluid dynamics (CFD) simulation and mesh registration. The cross-sectional planes show the truncation lines for the geometry marking the inlet and outlets.

were performed using Blender 2.8 (20). Finally, Instant Meshes (21) was utilized to retopologize the complex meshes to more structured quad surfaces. This set of geometries was used to perform the CFD simulations.

The same steps were also applied for the aorta-only-segmentations (ones without the branches), except that they were truncated at around the mid-thoracic level of the descending aorta. This second set of geometries was used for the registration step using a surface template mesh, as explained below. **Figure 1** shows the two geometries and the locations of the truncation-lines.

CFD Simulations

The branched aortic geometries were imported to Ansys 19.2 (Canonsburg, PA, USA). A mesh independence study was performed on one of the geometries from the training set (case 12), over four grid resolutions (1.5, 1.0, 0.75-, and 0.5-mm tetrahedral elements) using a steady-state simulation. For all the four grid resolutions, the same boundary layer setup was used (1 mm total thickness, 10 prism layers with increasing thickness, the growth rate of 1.2). Inlet and outlets were refined with 0.3 mm tetrahedral elements. The selected meshing strategy (1 mm elements, 589k nodes, 1.5M elements) compared to the finer mesh (0.75 mm elements, 746k nodes, 1.9M elements) resulted in differences of <4% for average WSS, <3% for average velocity, and <2% for flow rate at three cross-sectional planes measured at the ascending aorta, aortic arch, and descending aorta. Accordingly, this meshing strategy was selected considering the computation cost, file export size, and computational resource availability.

All 43 geometries were discretized using the selected meshing strategy, with additional local refinements applied when necessary. This results in a mesh containing between 500k and 800k nodes for each geometry.

We imposed rigid and no-slip boundary conditions at the wall. Blood was modeled as a Newtonian fluid with a density of 1,060 kg/m^3 and a viscosity of 4×10^{-3} Pa.s. A plug flow profile was prescribed at the inlet. Two different variations of outlet boundary conditions were prescribed: (1) constant pressure (0 Pa) was set at all outlets, (2) flow percentage ratio, with 70% flow going to the descending aorta, and 15/5/10% going to BCA, LCCA, and LSA, respectively. Due to time and resource constraints, the different outlet boundary conditions (1) and (2) were applied separately for 25 and 18 geometries, respectively.

Time-varying patient-specific inflow velocity was extracted from one case (case 1) over a cardiac cycle (710 ms). All simulations were performed for two cardiac cycles, using the same time-periodic velocity profile. The simulations were run with a time step of 1 ms. Velocity and wall shear stress vectors were extracted from the last cardiac cycle of the simulation to avoid transient initialization effects. The data were obtained for

every 10th time step (dt = 10 ms), resulting in 72 time frames (71 from the last cycle and 1 from the last time frame of the previous cycle).

It is important to note in this study that while patient-specific velocity profiles could be acquired, the rationale of running these CFD simulations was to generate a dataset with sufficient flow variations to enable the network to learn the local relationships between velocity and WSS, mainly through the different geometries and temporal variation. Hence, the same boundary conditions were applied for all geometries. Consequently, the resulting CFD simulations were not compared against the actual measurements from 4D Flow MRI. As a result, the WSS obtained from CFD simulations were different from the *in vivo* cases, which was expected. Therefore, the network applicability in predicting unseen data can be tested.

Essentially, WSS can be formulated as

$$\tau_w = \mu \frac{\partial u}{\partial y}$$

which is a product of the dynamic viscosity of the fluid (u) and the velocity gradient near the wall (wall shear rate). While in general, blood flow in the aorta is laminar and during peak systole, the flow can become turbulent, specifically at the ascending aorta (22). In turbulent flow, the velocity gradients near the wall become very steep and, hence, also the wall shear stress, as the velocity follows a logarithmic profile. The use of the laminar model for aortic flow is known to underestimate WSS (23). For this reason, the use of a turbulence model helps to improve WSS calculations through the use of turbulence (eddy) viscosity, which is a part of turbulence computations.

The realizable k-ε turbulence model (24) was chosen to account for possible turbulence effects during the peak systolic phase, where the Reynolds number reached >5,000. The incompressible Navier-Stokes equations were solved iteratively in ANSYS Fluent 19.2, with convergence criteria of scaled residual value to be less than 10^{-5} for mass and momentum. Each simulation took between 40 and 50 core hours to solve on a high-performance parallel computing environment (1.5 GB/core).

Data Preparation

Due to the complex relationship between flow and velocity gradients, it is important to incorporate both the velocity and spatial information as inputs to the network. Node coordinates, velocity, and WSS vectors from CFD were processed to create pairs of input-output data for the network.

To have a standardized data structure, we utilized the surface template mesh representation from Liang et al. (18, 25), which was modified into a 48 × 93 quadrilateral mesh. The quadrilateral mesh was then unwrapped into a UV map, and a 2D flatmap representation with 48 and 93 corresponding to the size of the circumference (U) and longitudinal (V) directions, respectively. The template mesh extends from the ascending aorta, aortic arch, and proximal section of the descending aorta. Note that the template mesh did not model the branching vessels. The template mesh was unwrapped using Blender, with the shortest distance from ascending to descending aorta selected as the cut-line. Subsequently, the UV map was aligned to form rectangular elements. To speed up the mesh registration step, a coarse version of the template (12 × 24) was also constructed and paired with subdivision matrices to convert it back to its actual size (48 × 93) using subdivision surface (26). These two template meshes were used for registration. **Figure 1** shows the variations of geometry used to build the training dataset and each geometry is shown in pairs: the aortic geometry with branches used for the CFD simulations, and the registered mesh on the aorta-only-geometry used for the WSSNet. Comprehensive visualization of the template mesh and registration steps are shown in **Figure 2**.

Registrations were performed first on the coarse template to all the 43 aorta-only-geometries using Coherent Point Drift (CPD) (27, 28) using rigid, affine, and deformable transformations ($\alpha = 3$, $\beta = 15$). After the initial registration of the coarse mesh, the mesh was subdivided using the subdivision matrices. A second deformable transformation ($\alpha = 3$, $\beta = 7$) was performed to ensure the small details in the geometry were aligned properly and to correct the deflation effect of the subdivided surface. The two parameters α and β represent the trade-off between goodness of maximum likelihood fit and regularization, and the width of smoothing Gaussian filter, respectively (27).

Note that the coarse template is optional and was used to speed up the mesh registration process. Without a coarse template, the registration process would be performed directly using the normal template mesh with all 3 transformations (rigid, affine, and deformable) and no subdivision surface is necessary. However, different parameters for the deformable transformation may be required. The two step registrations were performed in this study to tune the parameters quickly on the coarse template mesh while ensuring they have sufficient registration accuracy for all geometries.

Finally, the registered surfaces were used to extract the wall shear stress vectors and magnitude from the CFD simulations. The spatial coordinates (x, y, z) of each mesh node were stored as a "flatmap" with 3 channels (1 for each axis), with the Cartesian coordinate system. The KDTree algorithm was used to obtain WSS vectors for every point on the registered surface by searching for the closest point in the CFD surface mesh, with a search radius of 5 mm from each surface node. Template nodes corresponding to the aortic branches were masked as "invalid" by applying a distance threshold of >1.2 mm radius. Manual inspection and corrections were performed subsequently to ensure other aortic surface regions were included, and only the aortic branches regions were invalid. Despite how the CFD simulations included branching vessels, the registered surfaces did not. On the base of the branching vessels on the registered surfaces, there were no actual WSS values, which renders these regions invalid. These invalid regions were not optimized during the loss calculation.

Additionally, velocity vectors were extracted in varying inward distances (0.3, 0.5, 0.6, 0.8, 1.0, and 2.0 mm) normal to the surface points. Each velocity vector corresponded to each point with a predefined distance from the registered surface, forming a layer of velocity values, which we call a "velocity sheet." Alongside this, the spatial coordinates of the internal surface were also stored as flatmaps. Due to the no-slip-wall

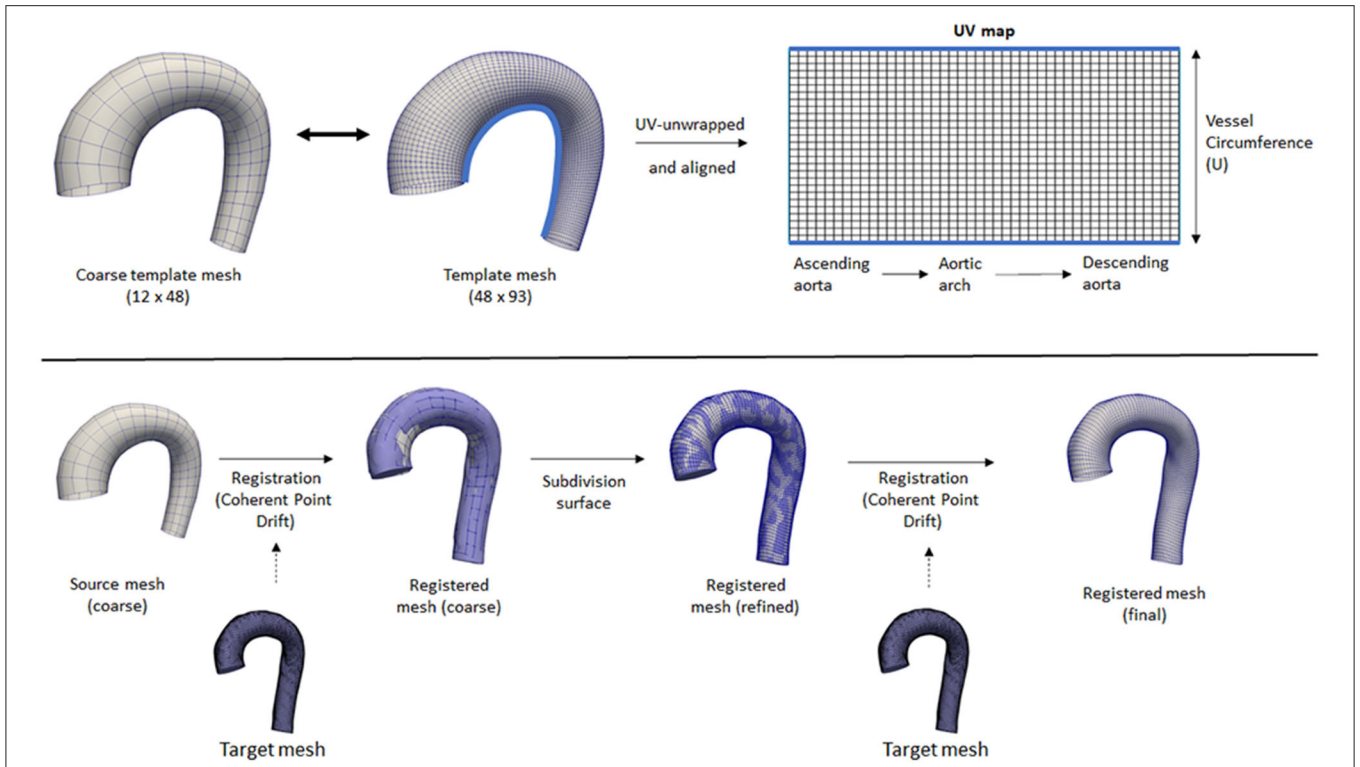


FIGURE 2 | Top: Coarse and fine template meshes used for registration. UV unwrapping was performed on the fine template mesh, with a light blue line showing the cut line. Bottom: An overview of the registration process using Coherent Point Drift. Registration was performed on the coarse template mesh, followed by a subdivision surface operation, followed by another registration step on the refined mesh.

boundary condition, the velocity sheet at the vessel surface was assumed to be 0 and, thus, was not extracted nor included as part of input data. The input data consisted of the registered surface mesh coordinates and the internal coordinates (points with variable inward distances normal from the surface) with their corresponding velocity vectors, while the ground truth label consisted of the 3D wall shear stress vectors at the registered surface mesh coordinates. An overview of the extracted information is shown in **Figure 3**.

Network and Training

With the input and output data effectively represented as 2D images, we could leverage the convolutional neural network (CNN) capability in learning spatial relationships. The input of the network was a 15-channel tensor, consisting of the Cartesian wall coordinates (x_0, y_0, z_0), two internal surface coordinates (x_1, y_1, z_1, and x_2, y_2, z_2), and two velocity sheets (v_{x1}, v_{y1}, v_{z1}, and v_{x2}, v_{y2}, v_{z2}). The output of the network is a 3-channel tensor, depicting the wall shear stress vectors (wss_x, wss_y, wss_z). A U-Net-like structure was used for the network architecture. The network consisted of 3 encoder and decoder blocks, with each block consisting of 2 convolutional layers with Rectified Linear Unit (ReLU) activation function, followed by batch normalization at the end of the block. Max pooling was applied on each of the encoder blocks, while bilinear interpolation was utilized to upsample each of the decoder blocks. The network architecture is shown in **Figure 4**.

The network was trained using a patch-based approach, with a 48 × 48 patch, which matched the size of the template mesh's circumference. The patches were selected randomly through the length of the vessel, acting as a sliding window. To ensure the network learned about the circular nature of the patch, we introduced periodic/circular padding. This is done by padding the top-most row within the patch with the bottom-row and *vice-versa*, and by duplicating the value of the left or right most column in the longitudinal direction. Periodic padding was applied before the first two convolutional layers.

Several augmentation strategies were applied to the dataset:

1. Distance to wall: to ensure that the network learns the spatial features, v_1 and v_2 are a combination of the available sheets, with the v_1 sheet closer to the wall than v_2. The combination of different velocity sheets was random within a pre-defined range.
2. Translation: to simulate translation to the training dataset, we selected a random node within the wall coordinate patch, and subtract that node position from all the coordinates, effectively setting it as the origin.
3. Rotation: random 3D rotations on a randomly selected plane were applied to the coordinate flatmaps and velocity sheets.

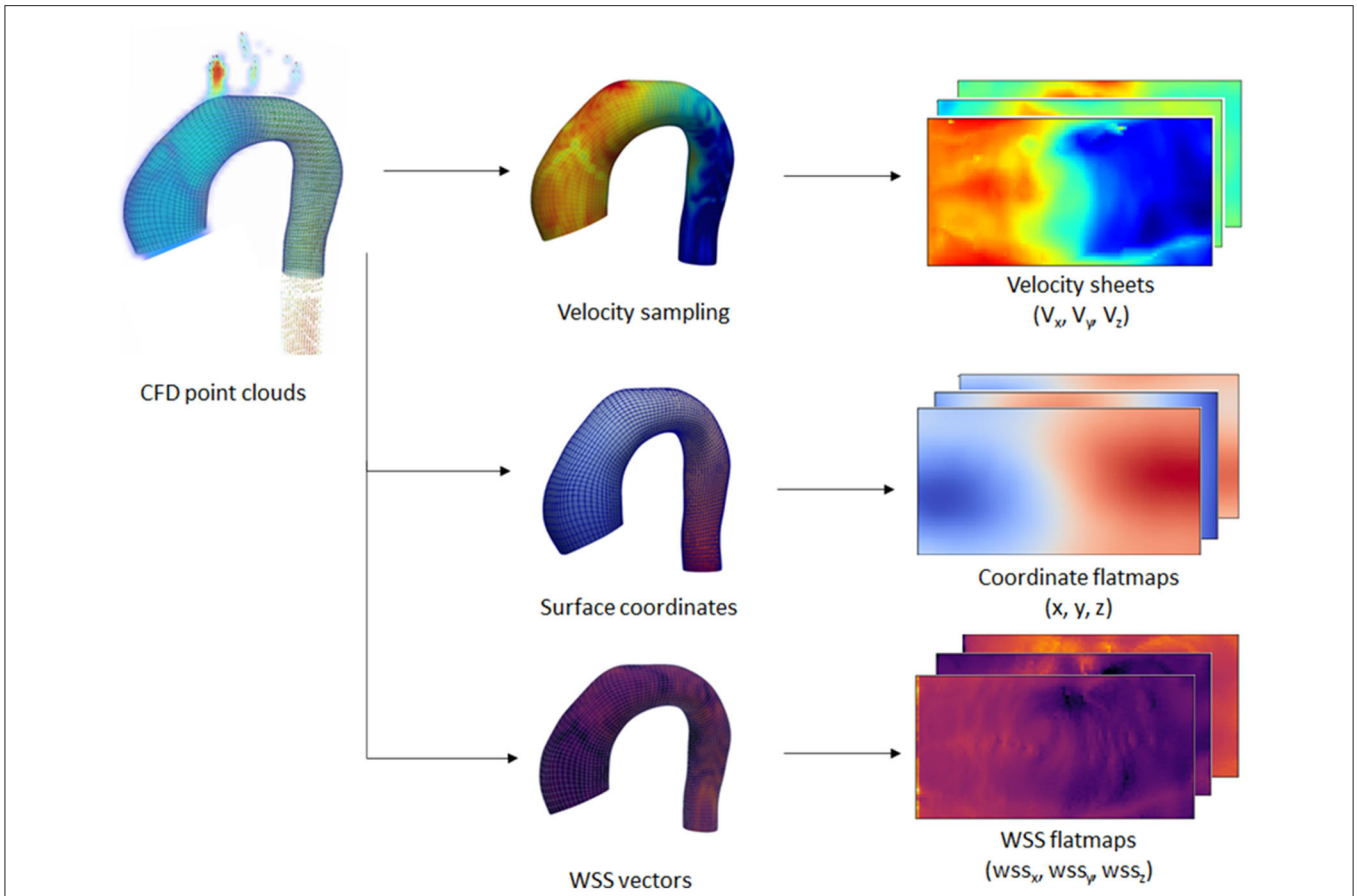


FIGURE 3 | Overview of the extraction process performed on the CFD point clouds dataset. Extraction was performed on the wall coordinates and several inner surface coordinates. Wall shear stress (WSS) vectors and velocity vectors were extracted at the wall and inner coordinates, respectively. The extracted information was transformed into 2D flatmaps based on the template mesh.

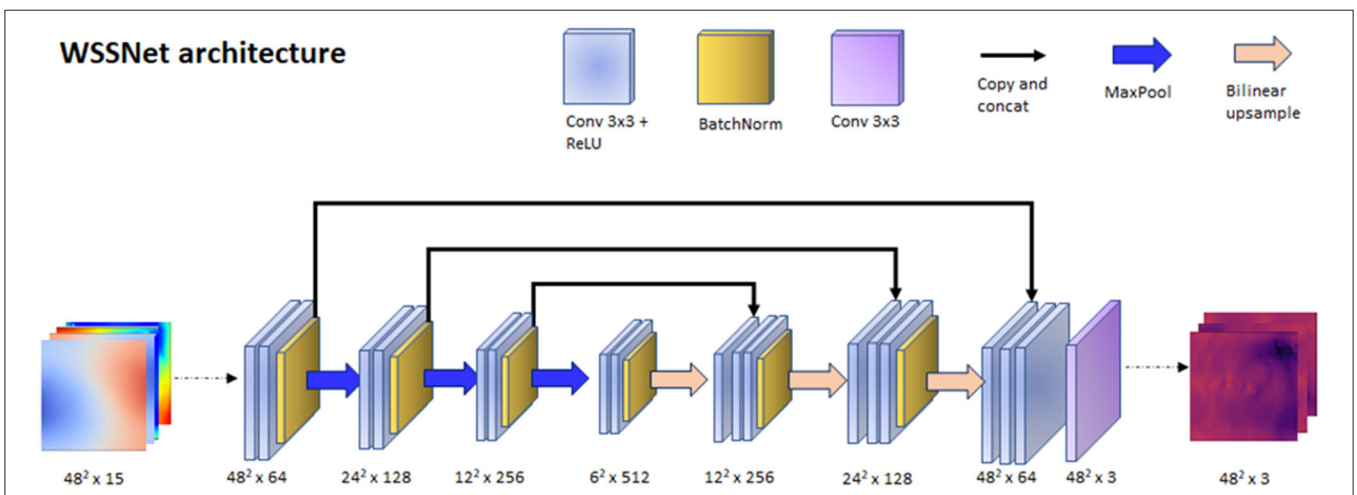


FIGURE 4 | WSSNet architecture. The network is based on U-Net architecture, which receives an input of 15 channels of 48×48 patches, consisting of coordinate flatmaps and velocity sheets, and outputs Cartesian wall shear stress vector patches.

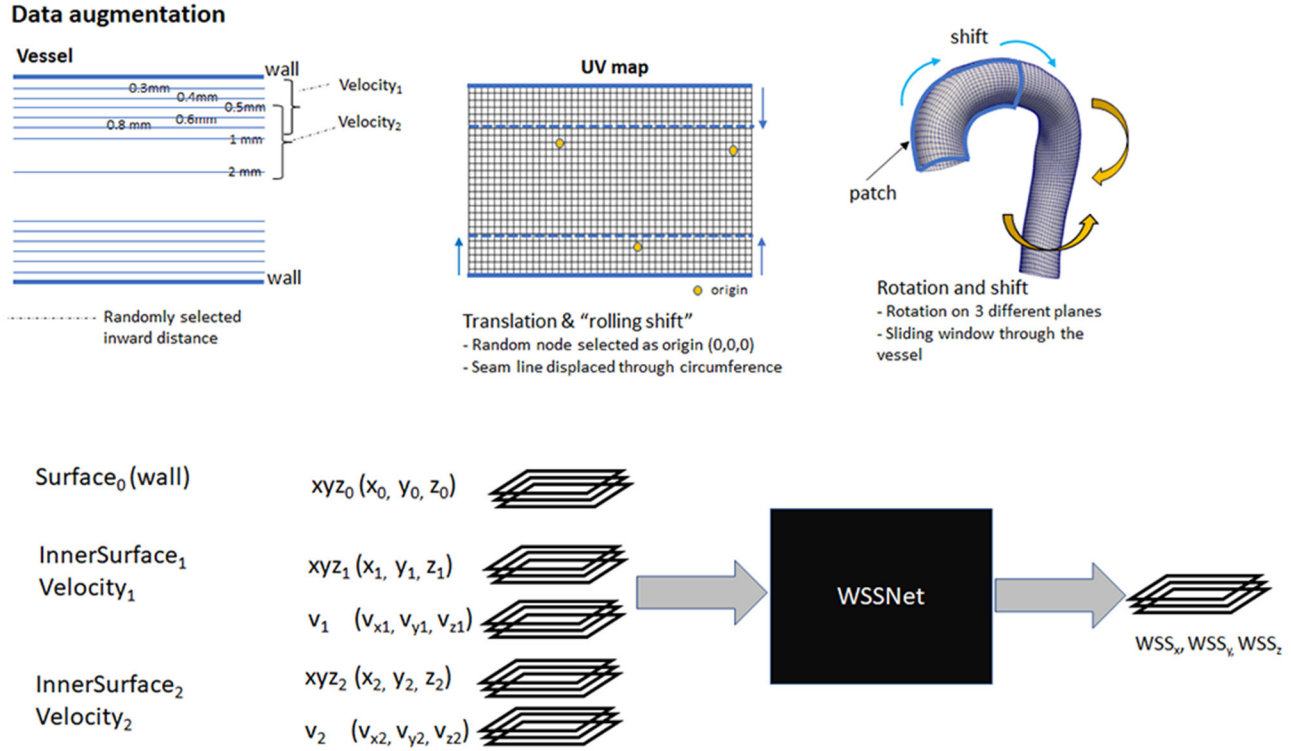


FIGURE 5 | Top: Overview of augmentation strategies. Velocity sheets were extracted at various distances from the surface (0.3–2.0 mm). During training, two velocity sheets were chosen randomly, with the first one closer to the vessel than the other. Bottom: a global overview of the input and output of WSSNet. Input consists of 15 channels, consisting of 3 coordinate flatmaps and 2 velocity sheets. The output consists of 3 channels, correspond to wall shear stress vectors in Cartesian coordinates.

4. Shift (sliding window): with the patch based approach, we shifted the patch in the longitudinal direction allowing the network to learn geometric and flow features on different regions of the vessel. This acts similar to the random selection of the patch.
5. Rolling-shift: with the cut-line of the template mesh predefined at the inner side of the aortic curve, the network might be fixated on the same geometric features (i.e., center rows having aortic branches). To introduce variation during the training, we perform a periodic-shift in the circumferential direction (U) by a maximum of 5 pixels.
6. Random noise: A 50% chance of adding Gaussian-smoothed Gaussian noise with an SD between 1 to 4% venc was added (venc = 1.5 m/s). The normally distributed noise was added to simulate the noise characteristics in the fluid domain. The Gaussian smoothing operation was added to simulate the interpolation that occurs when resampling CFD point clouds to a uniform grid.

The first 3 pixels from the inlet were excluded during training to avoid overestimated WSS caused by CFD boundary values. Nodes outside the mask (the base of the aortic branches) were also excluded because the WSS obtained are not true WSS and are basically obtained from the aortic branches. **Figure 5** summarizes the augmentation strategies, alongside a global overview of the network input and output.

The network was trained using Adam optimizer for 100 epochs, with a batch size (m) of 16. Cosine annealing learning rate was used on a repeating cycle for every 10 epochs, with a learning rate set between 10^{-4} and 10^{-7}. Tensorflow 2.0 (29) was used as the backend of the training. The network was trained on a Titan X GPU with 16GB memory.

From the 43 CFD simulations, 37 simulations were used for training, 3 for validation, and 3 simulations reserved for testing. The datasets were split randomly, with 8 left ventricular hypertrophy cases ending up in the training set, 1 case in the test set. It is worth noting that the data generated using CFD simulations do not represent the actual *in vivo* measurements.

The training set consisted of flatmaps extracted directly from the CFD point clouds. To ensure that the network can generalize well to 4D Flow MRI data, the validation and test sets consisted of flatmaps extracted from the following data representations: (1) CFD point clouds, (2) downsampled 2.4 mm^3 uniform grid (to mimic the MRI resolution), and (3) 2.4 mm^3 grid with noise (normal distribution, SD of 2% venc to simulate 4DFlow data). More detailed explanations on the sampling process

from CFD point clouds to simulate MRI are explained in the next section.

The training set comprised 46,676 unique flatmap combinations (mainly due to the combinations of velocity sheets with different distances), and the validation set consists of 3,996 unique flatmap combinations. Additionally, the sliding window strategy ensures the network "sees" different parts of the flatmap during the training process.

Loss Function Definition

A combination of loss functions was utilized, to ensure minimum difference of the WSS vectors and distributions (pattern similarity) between the predicted and reference values. First, we minimized the mean absolute error (MAE) between each of the predicted wall shear stress vector components and the reference values. Additionally, as we modeled the WSS flatmap as an image, we could optimize the pattern similarity between the predicted WSS magnitude and the ground truth WSS. Finally, an L2 regularization term was added to the network that can generalize to the new dataset, which was controlled by regularization weight (λ) scaled to the batch size (m). The complete loss function is given as

$$loss = l_{MAE} + \omega\, l_{SSIM} + \frac{\lambda}{2m}\sum_{i=1}^{N} w_i^2$$

with $\omega = 1.5$ and $\lambda = 10^{-2}$ to balance each of the loss terms to the same scale.

The Structural Similarity (SSIM) index, commonly used to measure the similarity of two images x and y, was added as a loss term to ensure WSS pattern similarity. SSIM is calculated based on three components: luminance (l), contrast (c), and structure (s). The luminance can be measured from the local average (μ) image values, while contrast is measured from the local SD (σ), and the structure index is measured using Pearson correlation (r).

Luminance comparison function $l(x, y)$ can be defined as

$$l(x,y) = \frac{2u_x u_y + C_1}{u_x^2 + u_y^2 + C_1}$$

while contrast comparison function $c(x, y)$ is defined by

$$c(x,y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

and structural comparison function $s(x, y)$ is used to measure the linear correlation between the two images:

$$s(x,y) = r = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3}$$

with σ_{xy} being the covariance of the two images, denoted as

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - u_x)(y_i - u_y)$$

C_1, C_2, and C_3 are constants added for numerical stability. $C_1 = (K_1\, L)^2$, $C_2 = (K_2\, L)^2$, and $C_3 = C_2/2$ are defined with $K_1 = 0.01$ and $K_2 = 0.03$ as in the original article (30), with L being the maximum true WSS within a patch.

Overall, SSIM is a combination of all the terms above:

$$SSIM(x,y) = l(x,y)^{\alpha} \;.\; c(x,y)^{\beta} \;.\; s(x,y)^{\gamma}$$

where α, β, and γ are positive numbers, denoting the relevance of each term, with $\alpha = \beta = \gamma = 1$. With that definition, SSIM loss is described as

$$l_{SSIM} = 1 - SSIM(x,y)$$

A built-in SSIM implementation from Tensorflow was used for the training process, with the default local region of 11 × 11 pixels.

Evaluation

Overall, evaluation of the network was performed in three different stages:

1. Evaluation on CFD simulation data (point cloud data) The network was validated on 6 CFD simulations with each of 72-time frames ($n = 432$). The input flatmaps were extracted directly from CFD point clouds.
2. Evaluation of CFD simulation data (synthetic MRI grid) The network was validated on 6 CFD simulations with each of 20 time frames ($n = 120$). The CFD point clouds were first interpolated into 3D grid representation (synthetic MRI) before extracting the flatmaps. The network was validated on two different grid resolutions (2.4 and 1.2 mm isotropic) and with/without noise, resulting in four sets of validations.
3. Evaluation on *in vivo* 4D Flow MRI The network was validated on all 43 *in vivo* cases, with each of 20 time frames ($n = 860$). These 43 cases were utilized previously for the aortic geometry extraction only. The flow information obtained from these *in vivo* cases does not resemble the generated CFD simulations.

Evaluation Metrics

For quantitative assessment, performance was evaluated with respect to the difference in point-to-point WSS magnitude per time frame, reported as mean absolute error (MAE) and relative error. Relative error was calculated as the ratio of absolute WSS difference and peak reference WSS value at the specified time frame. In addition, Pearson correlation (r) was also reported to evaluate the pattern similarity or WSS distribution for every time frame.

For quantitative and qualitative assessment, TAWSS and OSI were also calculated at each point in the template. TAWSS represents the average WSS over a cardiac cycle, while OSI represents the oscillation of the WSS direction over a cardiac cycle, computed as follows:

$$TAWSS = \frac{1}{T}\int_0^T \left|\overrightarrow{wss}\right| dt$$

$$OSI = 0.5 \times \left(1 - \frac{\left|\int_0^T \overrightarrow{wss}\, dt\right|}{\int_0^T \left|\overrightarrow{wss}\right| dt}\right)$$

where T is the span of time for a cardiac cycle. OSI ranges from 0 to 0.5, with 0 describing no change in WSS during a cardiac cycle, and 0.5 where there is a change of direction of 180° during a cardiac cycle. Additionally, a single-measure intraclass correlation coefficient (ICC) based on a two-way mixed-effects model was independently calculated for each measured TAWSS value to assess the degree of absolute agreement between WSSNet and ground truth CFD.

Surface Extraction Error

The registration step was performed to align and deform the template mesh to the aorta-only geometry, which is not without error. This imprecision of the vessel wall causes the inward points where velocity sheets were extracted to not be at the exact distance from the wall. While this imprecision introduces error to the training data, it also simulates actual inaccuracy in segmentation, where it is hampered by spatial resolution and partial volume effects. To measure the inaccuracy of the mesh registration step, we evaluated the surface distance for every node in the template mesh.

Validation on CFD Point Clouds

To evaluate the performance of WSSNet to reproduce CFD WSS estimations, the evaluation metrics were computed using the velocity data obtained directly from the CFD point clouds, on the 3 validation and 3 test cases, with each case consisting of 72 time frames (dt = 10 ms). For this validation, we evaluate the results by measuring MAE, relative error, Pearson correlation, TAWSS, and OSI compared to the ground truth WSS from CFD.

Finally, linear regression analysis was performed against reference values derived from CFD, assessing TAWSS and OSI, separately. Bland-Altman plots of the same data were also extracted to assess potential network bias. ICC was evaluated to assess the absolute agreement between the TAWSS values from the network against the ground truth CFD.

Validation on Synthetic MRI From CFD Point Clouds

To evaluate the capability of the network in predicting WSS from MR image resolution data, we first sampled the point cloud CFD dataset into a voxelized uniform image grid. To match the commonly acquired MRI spatial resolution, data were generated with isotropic spatial sampling of dx = 2.4. Furthermore, spatial sampling of dx = 1.2 mm was also performed to see how the network performs in different resolutions. Both resolutions were sampled at temporal resolution (dt) of 40 ms, within the range of common MRI acquisition. Additionally, noise (normal distribution, the SD set to 2% of the venc; venc = 1.5 m/s) was added to evaluate the performance of the network in the presence of noise. Thus, the evaluation set consisted of noise-free and noisy data at both resolutions.

Afterward, we performed the same procedures to extract the coordinate flatmaps and interpolate the velocity sheets from the synthetic MRI grid. Velocity sheets were extracted using linear interpolation at 1.0 and 2.0 mm at the inward direction normal to the surface. Subsequently, these coordinate flatmaps and velocity sheets were then used as input to the network.

As a comparison, the parabolic fitting method (9, 11) was selected as it is commonly employed and requires similar input. The method requires 3 velocity points, where the velocity at the wall is assumed as 0, due to the no-slip boundary condition. With the given resolution (dx = 2.4 and 1.2 mm), the parabolic fitting method is expected to underestimate the WSS values. For this validation, we performed the same quantification with the CFD point clouds dataset (MAE, relative error, Pearson correlation, TAWSS, and OSI) because the ground truth WSS was known.

Linear regression analysis was performed for both methods against reference values derived from CFD, assessing TAWSS and OSI at different resolutions: noise-free 2.4 mm, noisy 2.4 mm, noise-free 1.2 mm, and noisy 1.2 mm. Bland-Altman plots were extracted to assess potential network bias for the noisy 2.4 mm dataset. Additionally, Bland-Altman plots were extracted to assess the agreement between the parabolic fitting method and WSSNet predictions for the noisy 2.4 mm dataset. For each of the resolutions, ICC was evaluated for both methods to assess the absolute agreement between the TAWSS values from the network against the ground truth CFD.

Validation on *in vivo* Cases

To assess WSS on *in vivo* data, the method was applied to the *in vivo* 4D Flow data from the same 43 cases used in the patient specific CFD simulations. The same registered surface meshes were used to extract the coordinate flatmaps and velocity sheets. Velocity sheets were extracted with inward distances of 1.0 and 2.0 mm due to the inherent MRI resolution at approximately dx = 2.4 mm. **Figure 6** shows an overview of the analysis pipeline for 4D Flow MRI to WSS flatmap.

For these *in vivo* cases, the WSS reference values are not available. Due to the expected difference in WSS magnitudes (between WSSNet and parabolic fitting), only Pearson correlation was evaluated between the two methods. In addition, TAWSS and OSI were also visualized.

Linear regression analysis was performed against WSSNet predictions as reference values, assessing TAWSS and OSI, separately. Bland-Altman plots of the same data were also extracted to assess bias between the two methods.

RESULTS

Surface Extraction Error

The WSS ground truth was obtained from the CFD point closest to the registered template mesh nodes. Surface distance errors for the aortic vessel wall were 0.32 ± 0.14 mm, rising to 2.38 ± 1.08 mm at the base of the aortic branches where there was no true wall. Nevertheless, the error was small relative to the current MRI resolution (2.4 mm).

Validation on CFD Point Cloud Dataset

A complete evaluation was performed on the 3 validation and 3 test cases, with each case consisting of 72 time frames (dt = 10 ms). Overall, WSS estimates were accurate (MAE 0.55 ± 0.60 Pa, relative error 4.34 ± 4.14%) and showed excellent Pearson correlation with CFD WSS

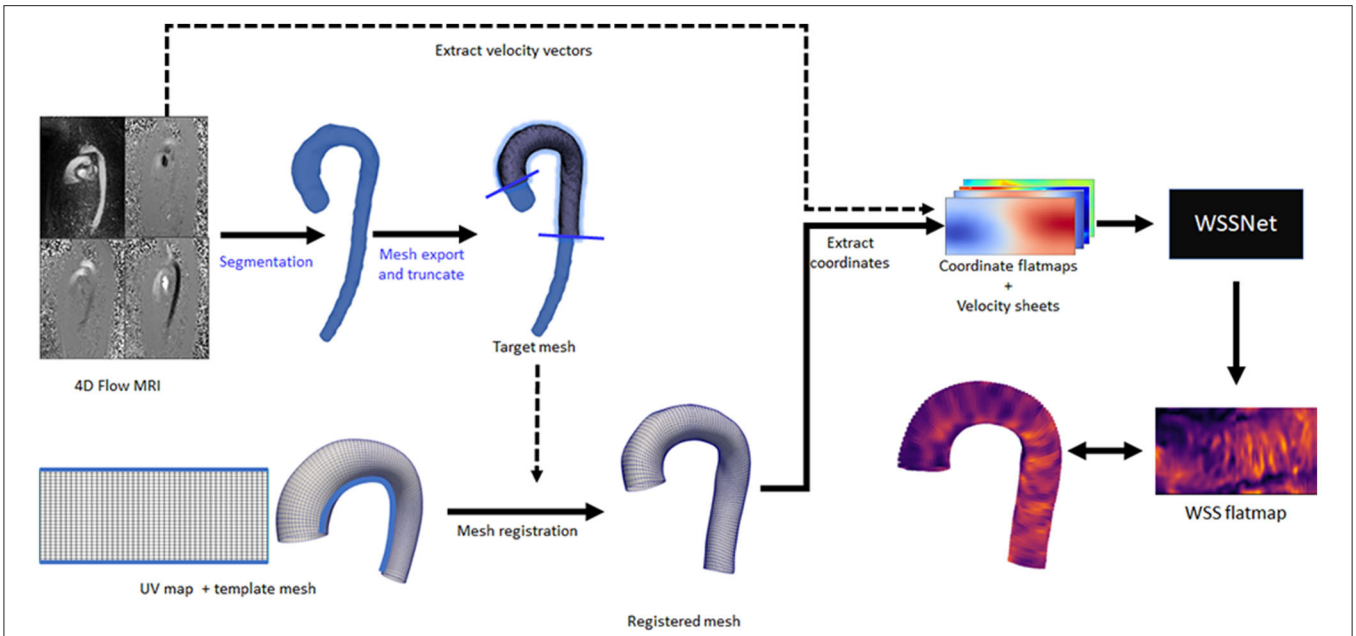


FIGURE 6 | Complete overview of the inference workflow for 4D Flow MRI. All the steps are fully automated, except for the segmentation and mesh truncation steps (marked with blue text).

TABLE 1 | Wall shear stress (WSS) magnitude and pattern similarity measurements for the validation and test cases from the CFD simulations.

Case (characteristics)	MAE (Pa)	Rel error (%)	Pearson correlation
Val #1 (normal)	0.44 ± 0.41	5.97 ± 4.82	0.92 ± 0.06
Val #2 (normal)	0.44 ± 0.50	3.70 ± 4.03	0.93 ± 0.03
Val #3 (normal)	0.49 ± 0.43	4.67 ± 4.22	0.92 ± 0.02
Test #1 (normal)	0.57 ± 0.75	2.87 ± 3.43	0.91 ± 0.04
Test #2 (LVH)	0.93 ± 1.04	5.76 ± 5.24	0.88 ± 0.09
Test #3 (normal)	0.44 ± 0.48	3.09 ± 3.12	0.94 ± 0.03
Overall	**0.55 ± 0.60**	**4.34 ± 4.14**	**0.92 ± 0.05**

For each case, results were measured from all time frames (n = 72, dt = 10 ms). MAE, mean absolute error; LVH, left ventricular hypertrophy. Average values are shown in bold.

(0.92 ± 0.05). More detailed quantitative measures per case are given in **Table 1**. **Figure 7** shows the qualitative results for each of the cases, represented by the TAWSS and OSI.

Figure 8 shows linear regression plots and Bland-Altman representations for the calculated TAWSS and OSI. In general, very high correlations are observed between TAWSS and OSI estimated by WSSNet and ground truth CFD, with linear regression slopes and coefficients of determination of k = 0.88 and R^2 = 0.90, and k = 0.88 and R^2 = 0.91 for TAWSS and OSI, respectively. Bland-Altman analysis shows a minimal bias (0.08 Pa) and limits of agreement of ± 1.29 Pa for TAWSS, and no bias for OSI with limits of agreement of ± 0.04. ICC shows excellent agreement (0.95) between TAWSS calculated from WSSNet and ground truth CFD.

Validation on CFD Synthetic MRI Dataset

Velocity sheets were extracted from the synthetic MRI dataset on the same 6 validation/test cases. Compared with the prediction on the sheets extracted from the CFD point clouds, a decrease in performance was observed when inference was performed on dx = 2.4 mm (MAE 0.94 ± 0.87 Pa), while Pearson correlation was still highly maintained (r = 0.82 ± 0.08). The addition of noise slightly decreased the performance further (MAE 0.99 ± 0.91 Pa, r = 0.79 ± 0.10). ICC showed good agreement with similar values, 0.86 and 0.85, for noise-free and noisy data, respectively.

At twice the resolution (dx = 1.2 mm), WSSNet showed better performance in predicting WSS (MAE = 0.65 ± 0.67 Pa, r = 0.89 ± 0.06) compared to the base resolution (dx = 2.4 mm). The addition of noise at 1.2 mm resolution impacted the performance slightly (MAE = 0.71 ± 0.71 Pa, r = 0.85 ± 0.10), showing reduced error and Pearson correlation. ICC went back up to 0.92 for both noise-free and noisy data at 1.2 mm resolution, getting closer to the CFD validation counterpart (0.95).

On the other hand, the parabolic fitting method showed much larger differences (MAE 2.89 ± 1.85 Pa and 2.33 ± 1.67 Pa at dx = 2.4 mm and 1.2 mm, respectively). The values were underestimated, mostly at regions of peak WSS. In terms of Pearson correlation, the parabolic fitting results showed moderate correlation with the CFD ground truth (r = 0.65 ± 0.12 and 0.69 ± 0.11, at dx = 2.4 and 1.2 mm, respectively).

Qualitative assessments are shown in **Figure 9**, represented as TAWSS and OSI. WSSNet predictions show good pattern similarity at both resolutions, with less detail recovered at dx = 2.4 mm. The parabolic fitting method showed WSS magnitude roughly 3 times lower than the CFD magnitude. For both algorithms, the OSI pattern appeared similar at both resolutions.

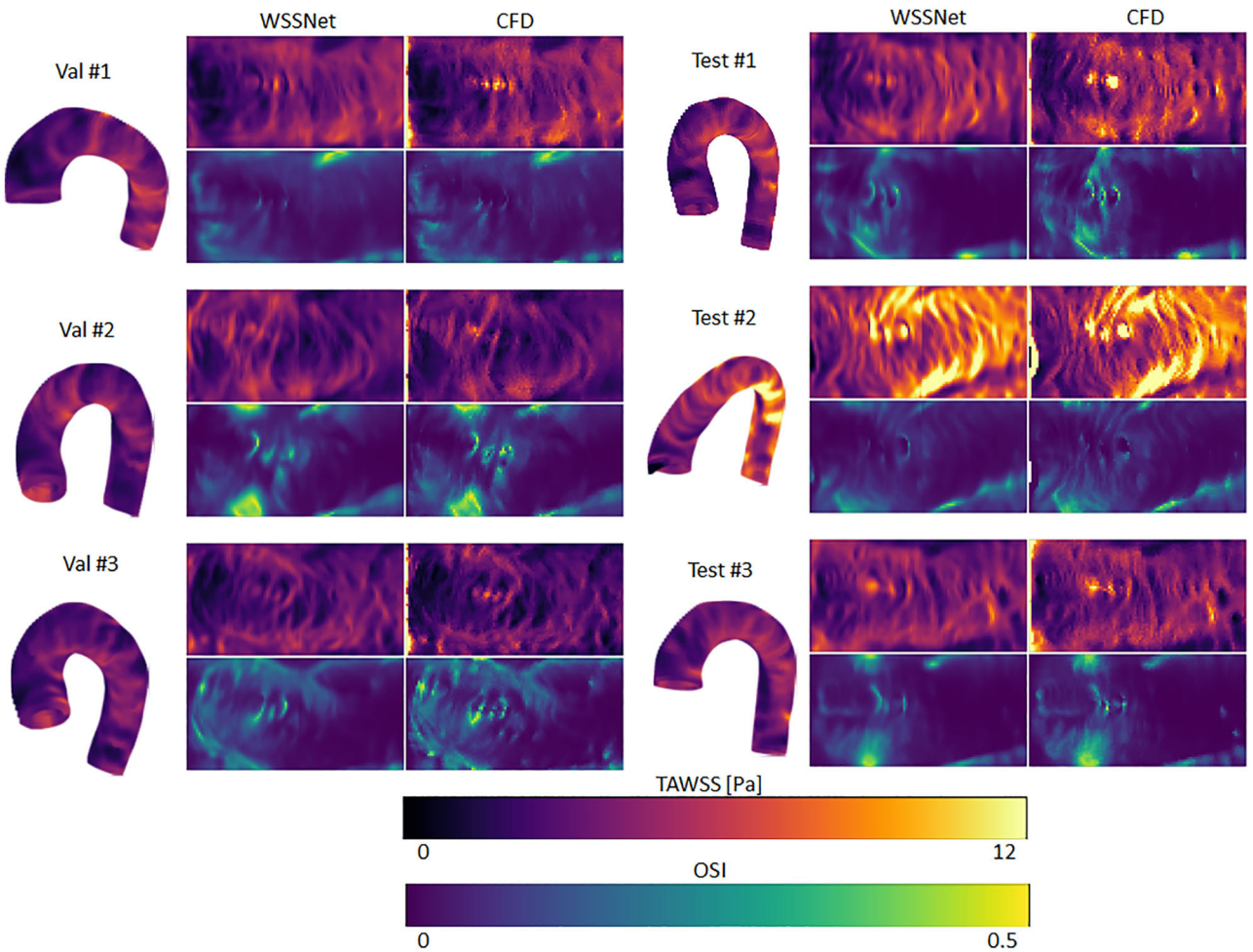


FIGURE 7 | Time-averaged WSS and oscillatory shear index (OSI) comparison between WSSNet and ground truth CFD. Time-averaged WSS (TAWSS) and OSI were calculated from all time frames (n = 72). 3D representations of the TAWSS from WSSNet are shown on the left side of each flatmap.

Table 2 shows a complete overview of the metrics for both methods on different resolutions.

Figure 10 shows linear regression plots and Bland-Altman representations for both methods. In general, good correlations are observed for TAWSS between WSSNet and CFD ground truth, with linear regression slopes and coefficients of determination of $k = 0.79$ and $R^2 = 0.75$ reported for the noise-free 2.4 mm resolution, and $k = 0.78$ and $R^2 = 0.74$ for the noisy 2.4 mm resolution. Slightly higher values are seen for WSSNet at 1.2 mm ($k = 0.82$ and $R^2 = 0.87$ for noise-free; $k = 0.81$ and $R^2 = 0.85$ for noisy). As a comparison, the parabolic fitting method shows poor correlations with CFD ground truth ($k = 0.14$ and $R^2 = 0.43$ for noise-free and noisy 2.4 mm; $k = 0.18$ and $R^2 = 0.56$ for noise-free and noisy 1.2 mm).

In terms of OSI, WSSNet shows moderate correlation at 2.4 mm ($k = 0.67$ and $R^2 = 0.73$ for noise-free; $k = 0.64$ and $R^2 = 0.70$ for noisy) and good correlation at 1.2 mm ($k = 0.83$ and $R^2 = 0.86$ for noise-free; $k = 0.80$ and $R^2 = 0.82$ for noisy). Similarly, the parabolic fitting method shows moderate correlation at 2.4 mm ($k = 0.65$, $R^2 = 0.66$ for noise-free; $k = 0.66$ and $R^2 = 0.64$ for noisy) and slightly better correlation at 1.2 mm ($k = 0.74$, $R^2 = 0.71$ for noise-free; $k = 0.74$, $R^2 = 0.68$ for noisy).

Bland-Altman plots were assessed at noisy 2.4 mm resolution to show the quality of results at common MRI resolution without any preprocessing. Bland-Altman plot indicated minimal TAWSS bias (0.32 Pa) with limits of agreement of 2.19 Pa between WSSNet and reference CFD. For OSI, the Bland Altman plot indicated minimal bias (−0.02) with limits of agreement of 0.08.

To compare the agreement between the parabolic fitting method and WSSNet, the Bland-Altman plot was also assessed at noisy 2.4 mm resolution. For TAWSS, Bland-Altman indicated a bias of −3.34 Pa, with higher TAWSS values showing larger differences (underestimation) by parabolic fitting method than WSSNet. Conversely, the Bland-Altman plot for OSI shows only a minimal bias (0.01) with narrow limits of agreement (0.05).

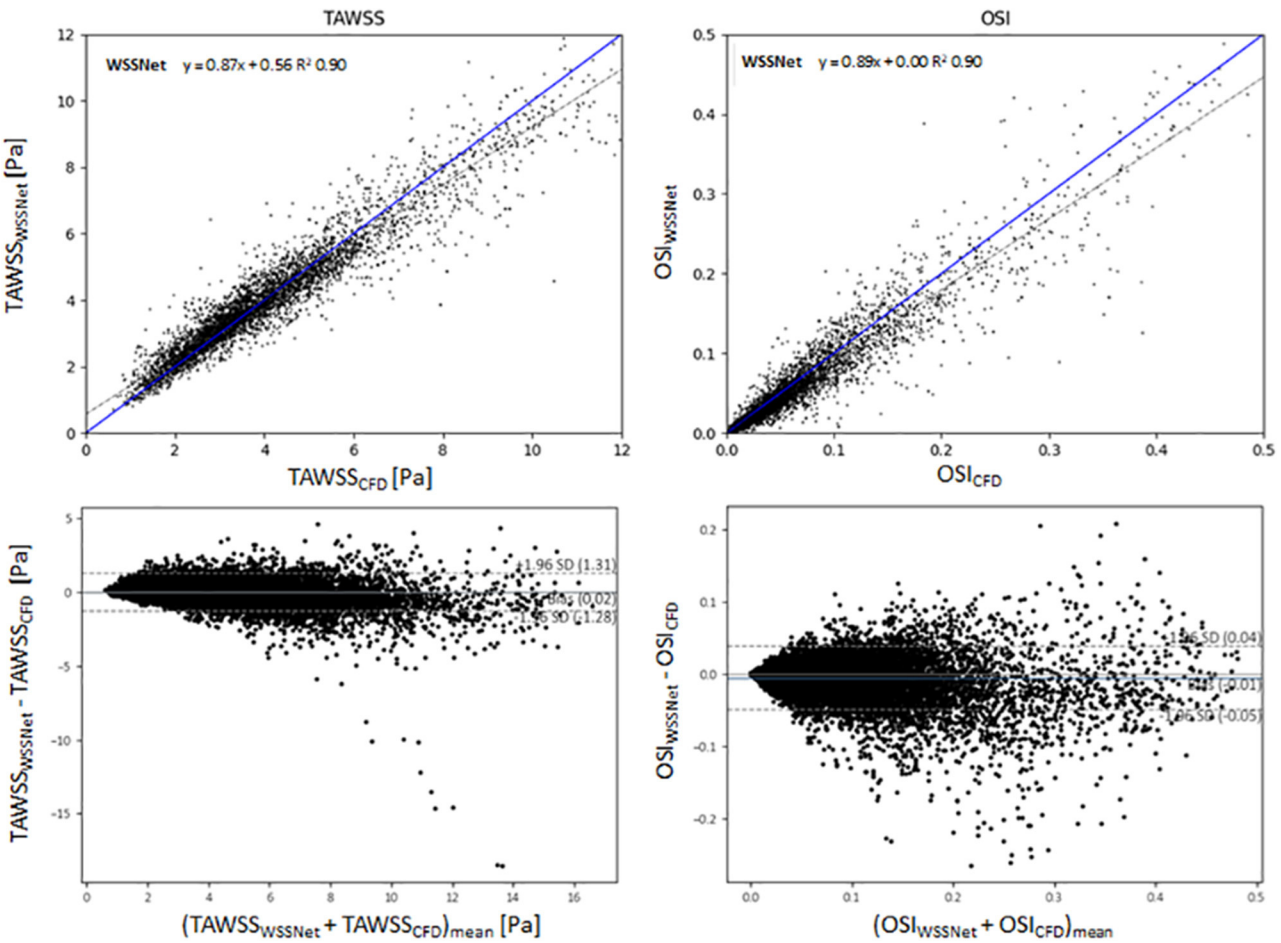


FIGURE 8 | Top: Regression plot for TAWSS and OSI between estimated values from WSSNet and ground truth CFD. TAWSS and OSI have computed over 6 cases (3 validation and 3 test) averaged over 72 time frames (dt = 10 ms). Bottom: Bland-Altman plots for TAWSS and OSI. The plots show a point-wise comparison within the flatmap. The plots show 20% of the data points, randomly selected.

Validation on *in vivo* 4D Flow MRI Cases

For the *in vivo* dataset, validation was performed on all 43 cases at the base resolution (2.375 mm x 2.375 × 2.75 mm) as is without any denoising. WSS was computed for all the cases with both the parabolic fitting method and WSSNet at every time frame. The resulting TAWSS and OSI were then compared between both methods.

Figure 11 shows linear regression and Bland-Altman plots for TAWSS and OSI comparing both methods, using WSSNet predictions as reference values. Relative to WSSNet, the parabolic fitting method shows poor correlation (k = 0.20 and $R^2 = 0.65$) for TAWSS but excellent OSI correlation (k = 0.91 and $R^2 = 0.71$). Bland-Altman plot of TAWSS shows a bias of −2.05 Pa with a similar downward trend observed in the synthetic MRI, with higher TAWSS showing more underestimations. Conversely, the Bland-Altman plot of OSI shows minimal bias (−0.02) with limits of agreement < 0.08.

Figure 12 shows visual comparisons of some of the cases using both methods. Visual inspection confirms the similarity between the computed TAWSS pattern, even though a clear difference in magnitude can be seen from the visualization. The spatiotemporal average WSS was 2.95 ± 1.57 Pa and 0.95 ± 0.46 Pa, for WSSNet and parabolic fitting method, respectively. Visual pattern similarity in OSI between both methods can be seen, which was observed in the synthetic MRI cases before. To quantify the similarity between WSS patterns, Pearson correlation was computed at every time frame, resulting in 0.68 ± 0.12.

DISCUSSION

This study demonstrates the feasibility of estimating WSS from low-resolution 4D Flow MR images using a deep learning method trained on a synthetic dataset acquired from CFD. Inference speed was 9 frames per second (26 cases per min) on a CPU for a typical 4D Flow MR image. The preprocessing step, which consisted of mesh registration and velocity sheet extraction, took ∼10 min. Currently, aortic curve segmentation remains a manual process, which can be improved in the

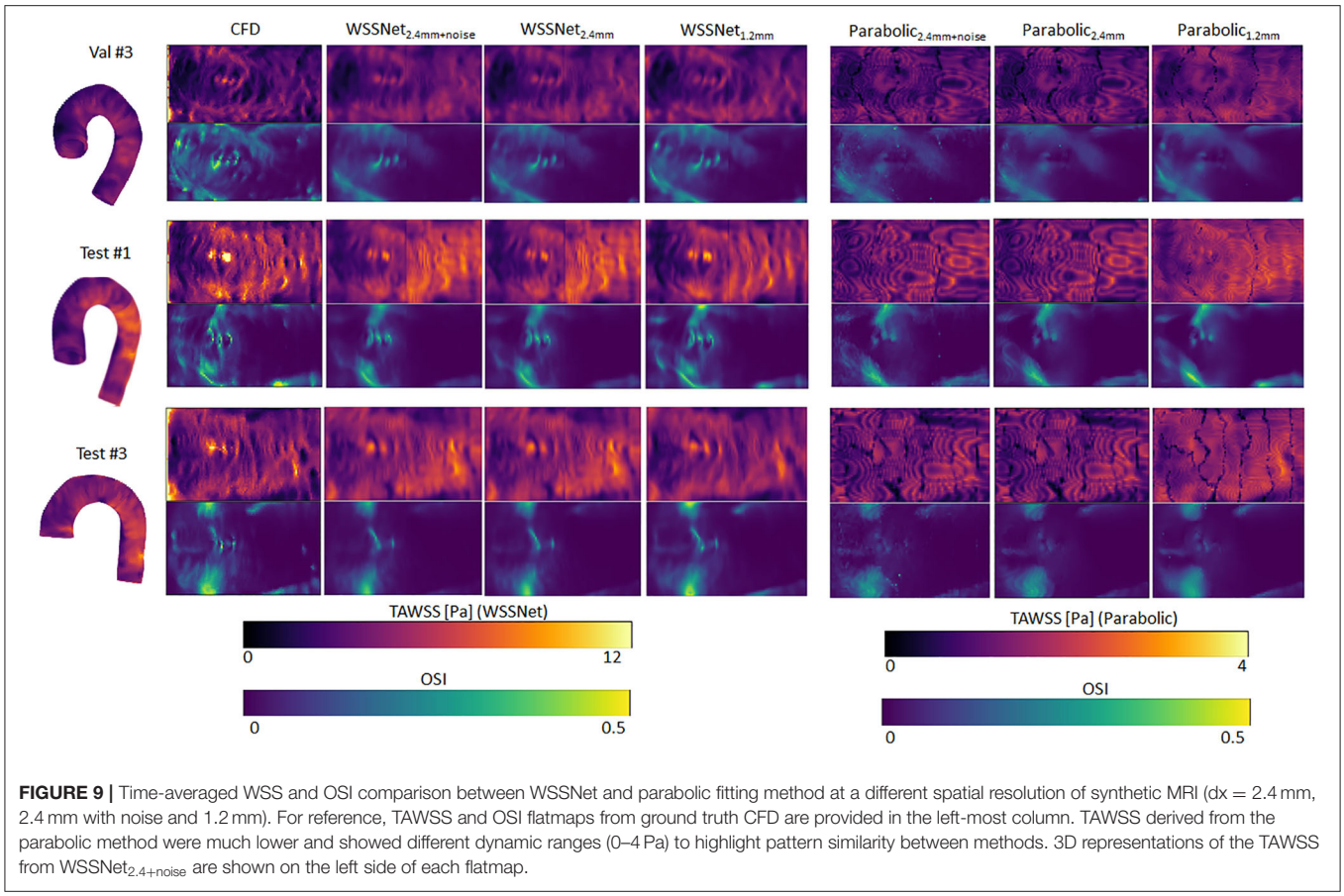


FIGURE 9 | Time-averaged WSS and OSI comparison between WSSNet and parabolic fitting method at a different spatial resolution of synthetic MRI (dx = 2.4 mm, 2.4 mm with noise and 1.2 mm). For reference, TAWSS and OSI flatmaps from ground truth CFD are provided in the left-most column. TAWSS derived from the parabolic method were much lower and showed different dynamic ranges (0–4 Pa) to highlight pattern similarity between methods. 3D representations of the TAWSS from $WSSNet_{2.4+noise}$ are shown on the left side of each flatmap.

TABLE 2 | WSS magnitude and pattern similarity measurements for the validation and test cases of the CFD simulations.

Method (resolution)	MAE (Pa)	Rel error (%)	Pearson correlation	ICC(A,1)
$WSSNet_{CFD}$	**0.55 ± 0.60**	**4.34 ± 4.14**	**0.92 ± 0.05**	**0.95**
$WSSNet_{2.4mm}$	0.94 ± 0.87	6.35 ± 5.67	0.82 ± 0.08	0.86
$WSSNet_{2.4mm+noise}$	0.99 ± 0.91	7.13 ± 6.27	0.79 ± 0.10	0.85
$WSSNet_{1.2mm}$	0.65 ± 0.67	4.80 ± 4.47	0.89 ± 0.06	0.92
$WSSNet_{1.2mm+noise}$	0.71 ± 0.71	5.66 ± 5.13	0.85 ± 0.10	0.92
$Parabolic_{2.4mm}$	2.89 ± 1.85	14.63 ± 6.83	0.65 ± 0.12	0.09
$Parabolic_{2.4mm+noise}$	2.89 ± 1.85	14.63 ± 10.69	0.59 ± 0.11	0.09
$Parabolic_{1.2mm}$	2.33 ± 1.67	11.23 ± 5.82	0.69 ± 0.11	0.13
$Parabolic_{1.2mm+noise}$	2.65 ± 1.75	13.07 ± 9.76	0.68 ± 0.11	0.13

The table shows the comparison of WSSNet and parabolic fitting on different spatial resolutions (dx = 2.4 and 1.2 mm isotropic) with and without noise (n = 120 each). The result from the CFD point clouds ($WSSNet_{CFD}$) is included for comparison (n = 432). Intra-class correlation (ICC) is computed from time-averaged WSS. Baseline values (WSSNet on CFD data) are shown in bold.

future. This workflow is several orders of magnitude faster than computational simulations while still offering good accuracy, which is not attainable using conventional methods at standard MRI resolution.

Spatial and Velocity Informed Neural Network

Previous studies have shown neural network's capability to estimate hemodynamic variables from geometric features (15–18). While these methods were able to estimate hemodynamic variables with sufficient accuracy, the estimated values produced were time-averaged or specific to a static boundary condition, as no other quantities (i.e., pressure, velocity) were provided as input values. Consequently, these methods learned strictly from spatial features and were agnostic to flow. In practice, clinical data contains temporal information, for instance, flow velocity, which can be used to derive other hemodynamic variables, such as pressure and WSS.

The aortic template mesh utilized in this study is based on Liang et al. (18) which consisted of 5,000 nodes, which when unwrapped into a UV map, becomes a 50 × 100 quadrilateral mesh. We adapted the template into a 48 × 93 grid to accommodate the circumferential direction in fitting the U-Net input size (48 × 48) and the longitudinal direction to enable compatibility with applying subdivision matrices on the coarse template (12 × 24). The subdivision matrices are equivalent to applying subdivision surfaces two times, which increased the number of nodes to $2n$ and $2n$-1 in a single pass, for circumferential and longitudinal direction, respectively. While the size of the circumference is fixed (48), different templates

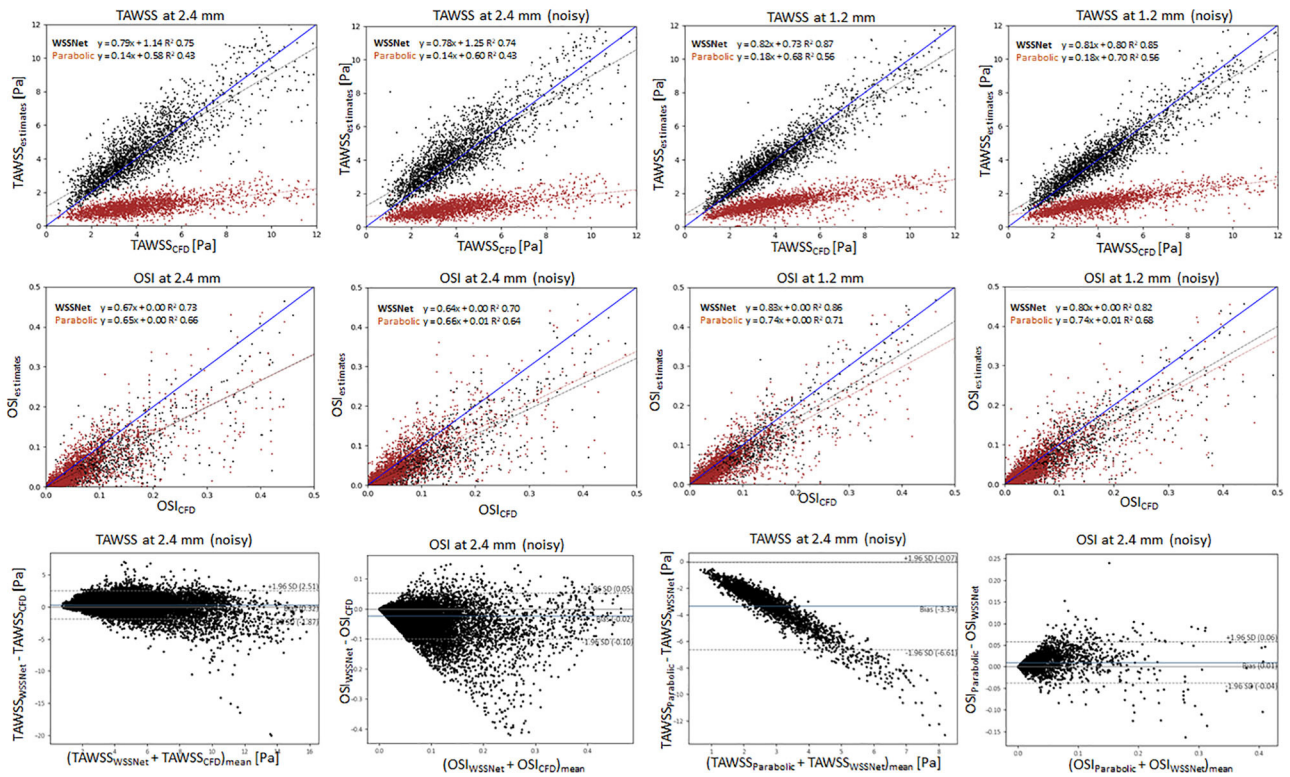


FIGURE 10 | Top: Linear regression plots of TAWSS for synthetic MRI using WSSNet (black) and parabolic fitting method (brown) at different resolutions (dx = 2.4 and 1.2 mm) with and without noise, compared with ground truth CFD. Middle: linear regression plots for OSI at different resolutions (dx = 2.4 and 1.2 mm) with and without noise, compared with ground truth CFD. Bottom left: Bland-Altman plots of TAWSS and OSI between WSSNet and ground truth CFD at 2.4 mm noisy synthetic MRI. Bottom right: Bland-Altman plots of TAWSS and OSI between parabolic fitting method and WSSNet at noisy 2.4 mm synthetic MRI. The plots show 20% of the data points, randomly selected.

with different longitudinal axis lengths can be used (i.e., 48 × N template). As previously mentioned, the coarse mesh and subdivision surfaces were optional and were used to speed up the registration process. Because the network was trained on a patch-basis, a template that contains an extension of the aorta (e.g., more distal part of the descending aorta) may still be predicted by the network, and simply assumed as another patch. With this approach, the network is not fixated on specific markers across the aortic vessels but is more general in predicting WSS, as long as the vessel can be unwrapped into a UV map with the specified circumference size (U = 48).

In this study, we extended on these previous works by adding velocity sheets and coordinate flatmaps, which are crucial to calculate spatial velocity gradients at the vessel wall. We were motivated by previous widely used velocity-based methods, such as the linear extrapolation method, velocity-based-with-wall-position method, and the parabolic fitting method (11, 31). The idea of these methods is to calculate spatial velocity gradients from several inward distance velocities normal to the wall. Potters et al. (10) further implemented the velocity-based method using spline fitting for volumetric image with a similar approach using 3 or more velocity points (including wall point, which is assumed to be 0). While it is possible to use more than 3 points, Potters et al. showed that using 3 points resulted in more accuracy, given enough voxels were available across the diameter.

To ensure the network learns different distances of velocity sheets, the training data contained various predefined distances from the wall surface. To mitigate zero values at the velocity sheets caused by the registration error, the first velocity sheets were extracted at a 0.3 mm inward distance from the surface.

From the aforementioned studies, spatial hemodynamic variables (i.e., WSS, ECAP, pressure) can be derived from geometric features alone. Conversely, conventional velocity-based methods are already used for *in vivo* cases resulting in spatiotemporal WSS. By combining these two concepts, we were able to train a network capable of estimating spatiotemporal WSS more accurately, by using geometry and velocity information.

It is noteworthy that WSSNet returns the WSS vectors at the vessel surface, which is useful for deriving other variables, such as TAWSS, OSI, and different WSS components (i.e., circumferential and longitudinal WSS). Previous studies explored the importance of this directional WSS (2, 32) and WSS angle (33). Increased axial WSS can be an indicator of the presence of high-risk plaque (32) and another study suggested that axial WSS might explain different morphologies in ascending aorta dilatation (2). WSS angle was suggested to be an independent

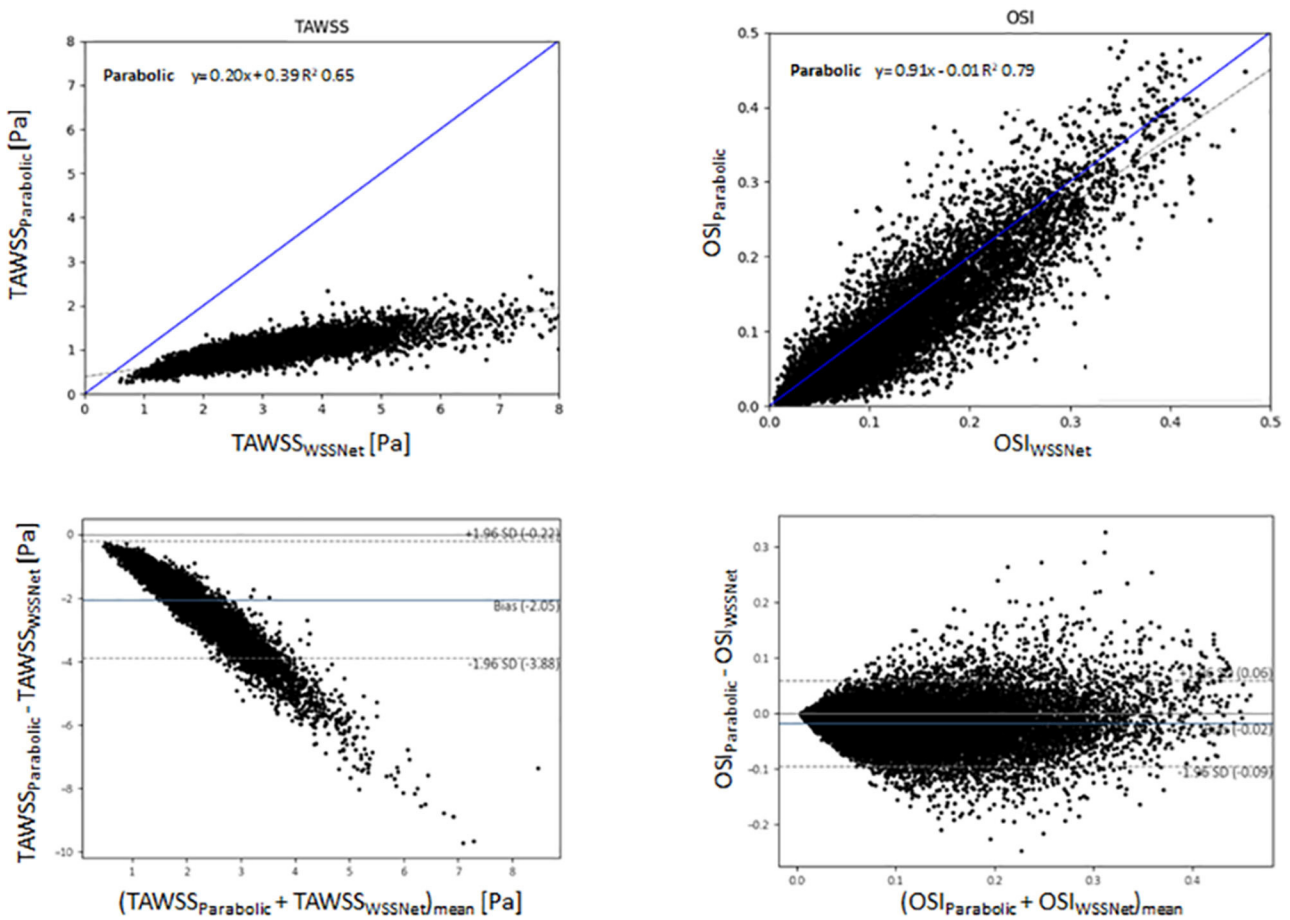


FIGURE 11 | Top: Linear regression comparing the TAWSS and OSI derived from the estimation of WSSNet and parabolic fitting method from *in vivo* 4D Flow MRI. WSSNet estimates are used as the reference values. Bottom: Bland-Altman plots of TAWSS and OSI between the parabolic fitting method and WSSNet for *in vivo* cases (n = 43). The plots show 5% of the data points, randomly selected.

predictor for proximal aortic dilatation for patients with bicuspid aortic valves (33).

Validation of WSS in CFD Point Clouds vs. Uniform Grid

Similar to other studies using deep learning, our training dataset was generated using CFD. While our target data is represented using a grid structure, we opted to train the network from the velocity sheets extracted directly from point clouds. Extracting velocity sheets from CFD point clouds allowed us to obtain smoother and velocity-rich information at flexible inward distances, unaffected by spatial sampling and any partial volume effect.

To showcase the generalizability of the network, we validated the network using a synthetic MRI dataset. The synthetic dataset was generated by sampling the CFD point clouds into MRI grid resolutions (dx = 2.4 and 1.2 mm). Additionally, we evaluated the robustness of the network in the presence of noise in the dataset. Subsequently, velocity sheets were extracted from the synthetic MRI, to simulate partial volume and discretization effects within the velocity sheets. Using WSSNet, reduced accuracy was identified at the velocity sheets acquired at a low resolution grid (dx = 2.4 mm). However, similar accuracy was observed with noise, showing the network is robust to noise. On the other hand, the parabolic fitting method shows much lower WSS values, and the values increase slightly with the presence of noise, which has been described previously (34). Increased accuracy was observed at higher spatial resolution (dx = 1.2 mm) for both methods. This is in agreement with previous studies that higher spatial resolutions lead to a higher average WSS (9–11). Nevertheless, WSSNet performance on low resolution synthetic MRI (dx = 2.4 mm) shows good accuracy and robustness to noise. Based on this validation on the synthetic MRI dataset, we can expect similar performance on the *in vivo* dataset with similar resolution.

To demonstrate the viability of our approach for 4D Flow MRI, we further tested the network in *in vivo* cases on base resolution (dx = 2.4 mm), similar to the test performed in the synthetic MRI. While there are no reference WSS values for *in*

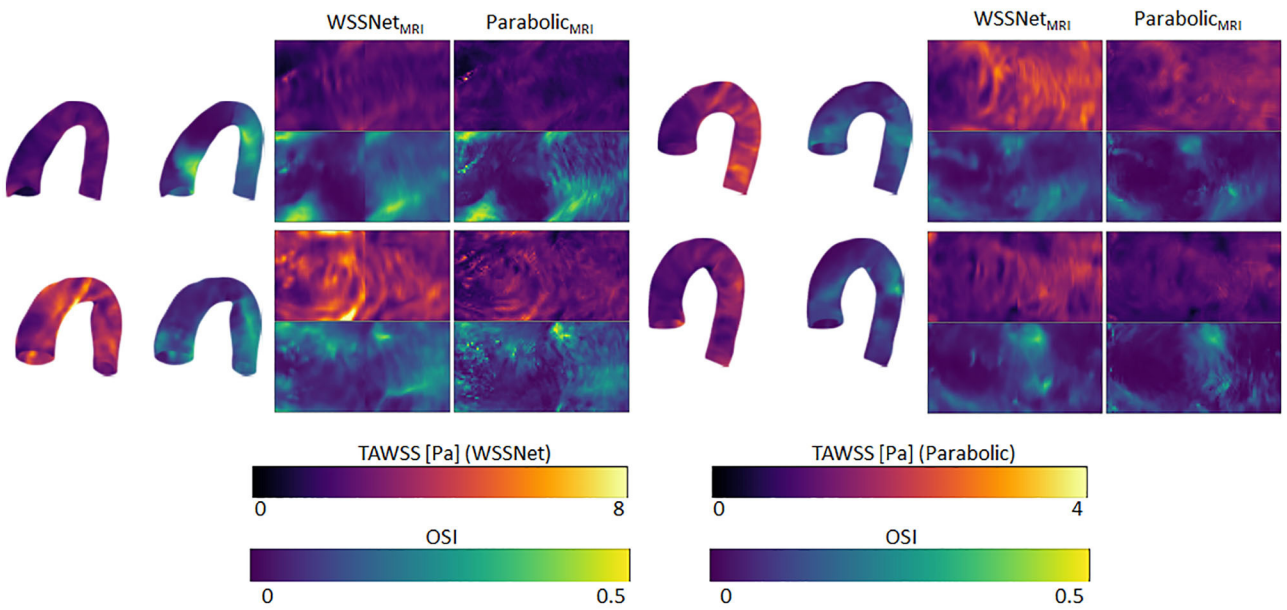


FIGURE 12 | Time-averaged WSS and OSI comparison between WSSNet and parabolic fitting method in 4 cases of *in vivo* 4D Flow MRI. TAWSS derived from the parabolic method were much lower and shown using different dynamic ranges (0–4 Pa) to highlight pattern similarity between methods 3D representations of the TAWSS and OSI from WSSNet are shown on the left side of each flatmap.

vivo cases, based on the previous validation on the synthetic MRI data (**Figure 10**), the regression and Bland-Altman plots in **Figure 11** show similar trends. Moreover, visual observation and structural similarity (Pearson correlation) also show adequate results in terms of the WSS distribution.

In addition, Bland-Altman plots in **Figures 10**, **11** show similar trends with a previous study conducted by Cibis et al. (14), with ours showing a much larger bias. While their comparison shows the WSS difference between MRI and CFD resolution at the carotid arteries using the smoothing-spline fitting method, we show the difference between parabolic fitting and WSSNet at MRI resolution. This result demonstrates the network's capability to perform at a similar level of accuracy as CFD, with evaluation performed at MRI resolution only.

Compared with other factors (i.e., segmentation accuracy, venc), the spatial resolution had the most significant impact on WSS estimation, as shown by Petersson et al. (11), with WSS in MRI typically underestimating true WSS. In their study, the relationship between WSS estimation methods with voxel size, venc, and segmentation accuracy has been assessed extensively. Other non-velocity-based methods were also assessed but were outside the scope of this study.

Our average WSS in *in vivo* cases, using the parabolic fitting method (0.95 ± 0.46 Pa), are relatively similar to previous studies (1, 10). The differences are probably related to the different fitting methods (parabolic vs spline fitting). WSSNet shows a higher spatiotemporal average WSS of 2.95 ± 1.57 Pa. As we have shown in the synthetic MRI dataset, the accuracy of WSSNet is similar to CFD.

For the parabolic fitting method, despite the regression coefficients for TAWSS being low (at 2.4 and 1.2 mm), the correlations for OSI are much higher. It can be observed from **Figure 10** that the systematic WSS underestimations (9) lead to a low correlation. However, OSI is a dimensionless metric measuring the changes in WSS direction, relative to its own magnitude. Therefore, it is independent of the WSS magnitude. As a result, the correlation of OSI between different methods can be compared independently from the magnitudes. This can be observed from the visualization of TAWSS and OSI (**Figure 9**) where both methods show TAWSS in different scales, but show OSI at the same scale and have a similar distribution.

Consistent regression coefficients for both methods compared to CFD reference values, in terms of TAWSS and OSI can be seen on the synthetic MRI dataset, with the higher resolution data showing a slight increase in correlation. The addition of noise to the data only affected the results slightly. Similar to the synthetic MRI, we also observed a similar correlation between the two methods for both TAWSS and OSI (**Figure 11**). This further verifies that WSSNet can be applied to *in vivo* cases, and exhibited similar behavior as when it was applied to synthetic MRI cases.

Additionally, **Figures 10**, **11** also show that the TAWSS estimates were higher in synthetic MRI (0–12 Pa) compared to the *in vivo* cases (0–8 Pa). These differences are likely caused by the choice of simplified boundary conditions (constant pressure at the outlet, outflow ratio) which produced different flow characteristics. However, despite the fact that the *in vivo* predictions are lower, the network can adapt to different flow patterns and still estimate WSS values accurately, since it is

trained on a variety of patient-specific geometries and boundary conditions with the time-varying flow.

The choice of turbulence model also affects the CFD simulations and estimated WSS as ground truth data. Different turbulence models might produce different results, but the effect is likely to be small. The use of a laminar model is also an option, however, the WSS computed by a laminar model is known to be underestimated in the turbulent regime. Also, as observed in previous studies (35, 36), laminar models give rise to lower WSS estimates than turbulence models, which are less accurate for peak WSS estimates. Additionally, while the estimates might introduce differences in values, the WSS patterns are similar.

Importance of WSS Distribution Pattern

Time-averaged WSS (TAWSS) and OSI patterns have been suggested as disease risk indicators, such as for atherosclerosis and aortic dilatation (1, 5, 37). Callaghan et al. (1) presented a normal pattern and range of WSS from 4D Flow MRI across a large population. Higher WSS was observed in the descending aorta compared with the aortic arch. In addition, this study suggested that WSS values are highly dependent on velocity, vessel diameter, and the aortic arch curvature. As previously mentioned, other studies have shown high accuracy in estimating hemodynamic variables, such as ECAP, based only on geometric information (spatial coordinates and curvature) (16, 17). While the presented results are accurate in *in silico* cases, pattern distribution and similarity were moderate. Liang et al. (18) also leveraged a geometric approach, with the predicted results showing similar aortic stress distributions, though it was not quantified and was only tested against finite element models. Another study combined vessel diameter and curvature information, showing good time-averaged WSS predictions in coronary arteries, with pattern similarities compared through visual inspection (15). Pattern matching is typically performed by checking areas of overlap using bins, categorizing WSS magnitudes from low to high (14).

In our study, we chose to use both velocity and geometric information, which is not only able to predict averaged or spatial results but can also predict spatiotemporal WSS. Moreover, our method is optimized using SSIM loss to ensure the network is able to recognize patterns based on spatial and velocity information. SSIM is a commonly used metric for image processing and computer vision, which can be applied to image data representation. Our results showed that WSSNet is able to recover finer WSS pattern details only by a two-fold resolution increase, which by no means is sufficient to accurately estimate WSS values. A much higher resolution is typically necessary to resolve the high gradient changes near the wall.

Overall, recent studies have suggested that WSS is a potential biomarker for atherosclerosis, aortic dilatation, and aneurysm (1, 5, 37). WSS measurement also has improved risk stratification in patients with carotid and coronary artery disease (38). Despite its clinical relevance, the available methods mostly rely on CFD analysis, which is computationally costly and not practical for a clinical setting (39). Therefore, by developing this method, we aim to extract the implicit knowledge from the CFD simulations, reducing a great deal of computation cost, while improving the applicability to a clinical setting. However, a more rigorous evaluation of patients is needed to ensure this approach can be applied for clinical applications.

Limitations and Future Study

Our study has several limitations. First, a modest number of geometries were used to run the CFD simulations. Additional geometries could be used to improve the generalizability of the network. Furthermore, different boundary conditions may help to generate more variations in the training dataset.

Second, WSS computed using CFD are dependent on several assumptions, which may introduce errors. The choice of a turbulence model may also affect the calculation of WSS, due to how the CFD software approaches the calculation of wall viscosity (with the contribution of turbulence viscosity). Although there may be differences in WSS magnitudes introduced by different turbulent models, these differences are small and the WSS patterns are relatively similar. Furthermore, our approach relies on the dataset, where physical properties were inherently derived from the CFD simulations. The simplification of boundary conditions (e.g., plug profile, constant pressure) might have a significant impact on the patient-specific flow and WSS values (40). However, the impact on WSSNet results is likely to be small, as it learns from local velocities and spatial features.

Third, WSSNet requires a large amount of data to train. While it is possible to generate more data through CFD simulations, a more sustainable solution is needed. An alternative for CFD, such as Physics Informed Neural Network (PINN), enables physics-informed solutions to generate surrogate solutions faster than CFD (41). This approach may speed up the data generation process tremendously. Additionally, this method also allows direct estimation of WSS, which may solve this problem in one single step. However, this method requires retraining for each new geometry. In our case, where 4D Flow MRI datasets are already available, an algorithm like WSSNet offers a direct estimation of WSS using the available measurements and geometry information.

Finally, while representing the data as flatmaps saves computation power, and can be considered as a strength of this model, it is not a flexible representation. Using this representation, more complex geometries (including the aortic branches) cannot be represented as a rectangular grid using UV mapping. A more flexible data representation (i.e., mesh or graph) or a different network structure (i.e., SplineCNN, Graph Convolutional Networks) may open a lot more possibilities for more complex geometries (17, 42, 43). Graph data representations have the potential to remove completely the registration step from this workflow, which will improve the time and may be more readily applicable for a clinical setting.

To summarize, several future directions can be taken to extend the capability of this method, namely the addition of data to expand the network generalization capability, a more flexible data representation, and robustness to noise. Nevertheless, our study highlights the potential of combining geometric and velocity information in training deep neural networks to infer hemodynamic variables for 4D Flow MRI.

CONCLUSION

In conclusion, we have presented a method to estimate WSS from 4D Flow MRI, with accuracy close to CFD. Our method is based on principles of similar previous WSS estimation methods, without being constrained by spatial resolution. More importantly, it is applicable to existing clinical MRI without any adjustments. We have shown accurate estimations for both CFD and *in vivo* cases regarding WSS magnitude and distribution throughout the aortic vessel.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Health and Disability Ethics Committees of New Zealand (HDEC 17/CEN/226). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EF wrote the first draft, designed the network architecture, and performed the data preparation and analysis. CM contributed to the subdivision surfaces implementation. All authors participated in the conception and design of the study, interpretation of data, revision of the manuscript, and final approval of the submitted manuscript.

FUNDING

This research has been funded by New Zealand Heart Foundation Scholarship, Grant No. 1786, the Health Research Council (HRC) of New Zealand, Programme Grants 17/608 and 21/144, as well as a grant from Siemens Healthineers, Erlangen, Germany.

ACKNOWLEDGMENTS

We thank the New Zealand eScience Infrastructure (NeSI) for high performance computing facilities, consulting support, and/or training services as part of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation and Employment's Research Infrastructure programme (https://www.nesi.org.nz). We thank the Center for eResearch at the University of Auckland for their help in facilitating this research (http://www.eresearch.auckland.ac.nz). We thank Dr. Ning Jin (Siemens Medical Solutions, Chicago, IL, USA) for providing the prototype 4D Flow sequence. We thank the Center for Advanced MRI, University of Auckland, where all MRI evaluations were performed.

REFERENCES

1. Callaghan FM, Grieve SM. Normal patterns of thoracic aortic wall shear stress measured using four-dimensional flow MRI in a large population. *Am J Physiol Hear Circ Physiol.* (2018) 315:H1174–81. doi: 10.1152/ajpheart.00017.2018
2. Rodríguez-Palomares JF, Dux-Santoy L, Guala A, Kale R, Maldonado G, Teixidó-Turà G, et al. Aortic flow patterns and wall shear stress maps by 4D-flow cardiovascular magnetic resonance in the assessment of aortic dilatation in bicuspid aortic valve disease. *J Cardiovasc Magn Reson*. (2018) 20:28. doi: 10.1186/s12968-018-0451-1
3. van Ooij P, Markl M, Collins JD, Carr JC, Rigsby C, Bonow RO, et al. Aortic valve stenosis alters expression of regional aortic wall shear stress: new insights from a 4-dimensional flow magnetic resonance imaging study of 571 subjects. *J Am Heart Assoc.* (2017) 6:e005959. doi: 10.1161/JAHA.117.005959
4. Jordanski M, Radovic M, Milosevic Z, Filipovic N, Obradovic Z. Machine learning approach for predicting wall shear distribution for abdominal aortic aneurysm and carotid bifurcation models. *IEEE J Biomed Heal Informatics.* (2018) 22:537–44. doi: 10.1109/JBHI.2016.2639818
5. Malek AM, Alper SL, Izumo S. Hemodynamic shear stress and its role in atherosclerosis. *J Am Med Assoc.* (1999) 282:2035–42. doi: 10.1001/jama.282.21.2035
6. Frydrychowicz A, Stalder AF, Russe MF, Bock J, Bauer S, Harloff A, et al. Three-dimensional analysis of segmental wall shear stress in the aorta by flow-sensitive four-dimensional-MRI. *J Magn Reson Imaging.* (2009) 30:77–84. doi: 10.1002/jmri.21790
7. Stankovic Z, Allen BD, Garcia J, Jarvis KB, Markl M. 4D flow imaging with MRI. *Cardiovasc Diagn Ther.* (2014) 4:173–92.
8. Markl M, Frydrychowicz A, Kozerke S, Hope M, Wieben O. 4D flow MRI. *J Magn Reson Imag.* (2012) 36:1015–36. doi: 10.1002/jmri.23632
9. Stalder AF, Russe MF, Frydrychowicz A, Bock J, Hennig J, Markl M. Quantitative 2D and 3D phase contrast MRI: optimized analysis of blood flow and vessel wall parameters. *Magn Reson Med.* (2008) 60:1218–31. doi: 10.1002/mrm.21778
10. Potters WV, Van Ooij P, Marquering H, VanBavel E, Nederveen AJ. Volumetric arterial wall shear stress calculation based on cine phase contrast MRI. *J. Magn. Reson Imaging.* (2015) 41:505–16. doi: 10.1002/jmri.24560
11. Petersson S, Dyverfeldt P, Ebbers T. Assessment of the accuracy of MRI wall shear stress estimation using numerical simulations. *J Magn Reson Imaging.* (2012) 36:128–38. doi: 10.1002/jmri.23610
12. Petersson S, Dyverfeldt P, Ebbers T. Accuracy of MRI wall shear stress estimation using numerical simulations. *J Cardiovasc Magn Reson.* (2012) 14:1–2. doi: 10.1186/1532-429X-14-S1-W6
13. Szajer J, Ho-Shon K. A comparison of 4D flow MRI-derived wall shear stress with computational fluid dynamics methods for intracranial aneurysms and carotid bifurcations—a review. *Magn Reson Imaging.* (2018) 48:62–9. doi: 10.1016/j.mri.2017.12.005
14. Cibis M, Potters WV, Gijsen FJ, Marquering H, VanBavel E, van der Steen AF, et al. Wall shear stress calculations based on 3D cine phase contrast MRI and computational fluid dynamics: a comparison study in healthy carotid arteries. *NMR Biomed.* (2014) 27:826–34. doi: 10.1002/nbm.3126
15. Gharleghi R, Samarasinghe G, Sowmya A, Beier S. Deep learning for time averaged wall shear stress prediction in left main coronary bifurcations. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI).* (2020), p. 1–4
16. Acebes C, Morales X, Camara O. A cartesian grid representation of left atrial appendages for a deep learning estimation of thrombogenic risk predictors. In: Puyol Anton E., et al. (eds) *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges. STACOM 2020.* Lecture Notes in Computer Science, vol 12592 (2021)

17. Morales Ferez X, Mill J, Juhl KA, Acebes C, Iriart X, Legghe B, et al. Deep learning framework for real-time estimation of in-silico thrombotic risk indices in the left atrial appendage. *Front Physiol.* (2021) 12:694945. doi: 10.3389/fphys.2021.694945
18. Liang L, Liu M, Martin C, Sun W. A deep learning approach to estimate stress distribution: a fast and accurate surrogate of finite-element analysis. *J R Soc Interface.* (2018) 15:20170844. doi: 10.1098/rsif.2017.0844
19. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage.* (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015
20. CBO. (2021) Blender—a 3D modelling and rendering package. Stichting Blender Foundation, Amsterdam. Available online at: http://www.blender.org. (accessed November, 2020).
21. Jakob W, Tarini M, Panozzo D, Sorkine-Hornung O. Instant field-aligned meshes. *ACM Trans Graph.* (2015) 34:1–15. doi: 10.1145/2816795.2818078
22. Stein PD, Sabbah HN. Turbulent blood flow in the ascending aorta of humans with normal and diseased aortic valves. *Circ Res.* (1976) 39:58–65. doi: 10.1161/01.RES.39.1.58
23. Miyazaki S, Itatani K, Furusawa T, Nishino T, Sugiyama M, Takehara Y, et al. Validation of numerical simulation methods in aortic arch using 4D Flow MRI. *Heart Vessels.* (2017) 32:1032–44. doi: 10.1007/s00380-017-0979-2
24. Shih TH, Liou WW, Shabbir A, Yang Z, Zhu J. A new K-epsilon Eddy viscosity model for high Reynolds number turbulent flows: model development and validation. *Comput Fluids.* (1995) 24:227–38. doi: 10.1016/0045-7930(94)00032-T
25. Liang L, Liu M, Martin C, Elefteriades JA, Sun W. A machine learning approach to investigate the relationship between shape features and numerically predicted risk of ascending aortic aneurysm. *Biomech Model Mechanobiol.* (2017) 16:1519–33. doi: 10.1007/s10237-017-0903-9
26. PixarAnimationStudios. OpenSubdiv." Available online at: https://github.com/PixarAnimationStudios/OpenSubdiv (accessed Jun 01, 2021).
27. Myronenko A, Song X. Point set registration: coherent point drifts. *IEEE Trans Pattern Anal Mach Intell.* (2010) 32:2262–75. doi: 10.1109/TPAMI.2010.46
28. Khallaghi S. *Python-CPD.* (2017). Available online at: https://github.com/siavashk/pycpd. (accessed April, 2021).
29. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. (2015) *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.* Available online at: https://www.tensorflow.org/. (accessed February, 2021).
30. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process.* (2004) 13:600–12. doi: 10.1109/TIP.2003.819861
31. Oshinski JN, Ku DN, Mukundan S, Loth F, Pettigrew RI. Determination of wall shear stress in the aorta with the use of MR phase velocity mapping. *J Magn Reson Imaging.* (1995) 5:640–7. doi: 10.1002/jmri.1880050605
32. Zhang G, Zhang S, Qin Y, Fang J, Tang X, Li L, et al. Differences in wall shear stress between high-risk and low-risk plaques in patients with moderate carotid artery stenosis: A 4D flow MRI study. *Front Neurosci.* (2021) 15:678358. doi: 10.3389/fnins.2021.678358
33. Minderhoud SCS, Roos-Hesselink JW, Chelu RG, Bons LR, Van Den Hoven AT, Korteland SA, et al. Wall shear stress angle determines aortic growth in patients with bicuspid aortic valves. *Eur Heart J Cardiovasc Imaging.* (2021) 22 (Supplement_2):jeab090. doi: 10.1093/ehjci/jeab090.120
34. Callaghan FM, Grieve SM. Spatial resolution and velocity field improvement of 4D-flow MRI. *Magn Reson Med.* (2017) 78:1959–68. doi: 10.1002/mrm.26557
35. Cheng Z, Tan FPP, Riga CV, Bicknell CD, Hamady MS, Gibbs RGJ, et al. Analysis of flow patterns in a patient-specific aortic dissection model. *J Biomech Eng.* (2010) 132:051007. doi: 10.1115/1.4000964
36. Khanafer KM, Bull JL, Upchurch GR, Berguer R. Turbulence significantly increases pressure and fluid shear stress in an aortic aneurysm model under resting and exercise flow conditions. *Ann Vasc Surg.* (2007) 21:67–74. doi: 10.1016/j.avsg.2006.10.009
37. Takehara Y, Isoda H, Takahashi M, Unno N, Shiiya N, Ushio T, et al. Abnormal flow dynamics result in low wall shear stress and high oscillatory shear index in abdominal aortic dilatation: Initial in vivo assessment with 4D-flow MRI. *Magn Reson Med Sci.* (2020) 19:235–46. doi: 10.2463/mrms.mp.2019-0188
38. Urschel K, Tauchi M, Achenbach S, Dietel B. Investigation of wall shear stress in cardiovascular research and in clinical practice—from bench to bedside. *Int J Mol Sci.* (2021) 22:5635. doi: 10.3390/ijms22115635
39. Gijsen F, Katagiri Y, Barlis P, Bourantas C, Collet C, Coskun U, et al. Expert recommendations on the assessment of wall shear stress in human coronary arteries: Existing methodologies, technical considerations, and clinical applications. *Eur Heart J.* (2019) 40:3421–33. doi: 10.1093/eurheartj/ehz551
40. Madhavan S, Kemmerling EMC. The effect of inlet and outlet boundary conditions in image-based CFD modeling of aortic flow. *Biomed Eng Online.* (2018) 17:1–20. doi: 10.1186/s12938-018-0497-1
41. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys.* (2019) 378:686–707. doi: 10.1016/j.jcp.2018.10.045
42. Sanchez-Gonzalez A, Godwin J, Pfaff T, Ying R, Leskovec J, Battaglia PW. Learning to simulate complex physics with graph networks. In: *37th International Conference Machine Learning ICML* 2020 (2020).
43. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: *5th Int Conf Learn Represent ICLR 2017—Conf Track Proc.* (2017) 2017.

Conflict of Interest: Working expenses and a partial stipend for EF were provided by Siemens Healthineers, Erlangen, Germany.

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.

frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 03 February 2022
doi: 10.3389/fcvm.2021.768419

EP-PINNs: Cardiac Electrophysiology Characterisation Using Physics-Informed Neural Networks

Clara Herrero Martin[1,2], *Alon Oved*[3], *Rasheda A. Chowdhury*[4], *Elisabeth Ullmann*[5], *Nicholas S. Peters*[4], *Anil A. Bharath*[1] *and Marta Varela*[4]*

[1] Department of Bioengineering, Imperial College London, London, United Kingdom, [2] ITACA Institute, Universitat Politècnica de València, Valencia, Spain, [3] Department of Computing, Imperial College London, London, United Kingdom, [4] National Heart and Lung Institute, Imperial College London, London, United Kingdom, [5] Department of Mathematics, Technical University of Munich, Munich, Germany

Accurately inferring underlying electrophysiological (EP) tissue properties from action potential recordings is expected to be clinically useful in the diagnosis and treatment of arrhythmias such as atrial fibrillation. It is, however, notoriously difficult to perform. We present EP-PINNs (Physics Informed Neural Networks), a novel tool for accurate action potential simulation and EP parameter estimation from sparse amounts of EP data. We demonstrate, using 1D and 2D *in silico* data, how EP-PINNs are able to reconstruct the spatio-temporal evolution of action potentials, whilst predicting parameters related to action potential duration (APD), excitability and diffusion coefficients. EP-PINNs are additionally able to identify heterogeneities in EP properties, making them potentially useful for the detection of fibrosis and other localised pathology linked to arrhythmias. Finally, we show EP-PINNs effectiveness on biological *in vitro* preparations, by characterising the effect of anti-arrhythmic drugs on APD using optical mapping data. EP-PINNs are a promising clinical tool for the characterisation and potential treatment guidance of arrhythmias.

Keywords: cardiac electrophysiology, arrhythmia (any), Physics Informed Neural Network (PINN), atrial fibrillation, parameter estimation, optical mapping, biophysical modelling, artificial intelligence

OPEN ACCESS

Edited by:
Paul Leeson,
University of Oxford, United Kingdom

Reviewed by:
Jordi Heijman,
Maastricht University, Netherlands
Alfonso Bueno-Orovio,
University of Oxford, United Kingdom

***Correspondence:**
Marta Varela
marta.varela@imperial.ac.uk

Specialty section:
This article was submitted to Cardiac Rhythmology, a section of the journal Frontiers in Cardiovascular Medicine

Received: *31 August 2021*
Accepted: *22 December 2021*
Published: *03 February 2022*

Citation:
Herrero Martin C, Oved A, Chowdhury RA, Ullmann E, Peters NS, Bharath AA and Varela M (2022) EP-PINNs: Cardiac Electrophysiology Characterisation Using Physics-Informed Neural Networks. Front. Cardiovasc. Med. 8:768419. doi: 10.3389/fcvm.2021.768419

1. INTRODUCTION

Cardiac arrhythmias are extremely common pathologies caused by disturbances in the generation or propagation of electrical signals across the heart. Atrial fibrillation (AF), the most common sustained arrhythmia, affects 0.5% of the world's population and accounts for 1% of the NHS's total budget through its large impact on patient mortality and morbidity, especially stroke (1). Catheter ablation of atrial myocardium believed to host the sources of the arrhythmia is the mainstay of AF treatment, but its long-term efficacy is disappointing (54%), especially in patients with persistent forms of the disease (43%) (2).

The mechanisms behind AF are very complex, involving the interplay of several factors at different scales, from changes in membrane proteins to alterations in cardiac tissue composition and organ shape (3). To characterise the arrhythmia, information about cardiac activity can be acquired by recording electrical potentials using electrodes placed on the chest (electrocardiogram, ECG) or, in a catheter lab, placed in direct contact with the myocardium (contact electrograms, EGMs). Expert analysis of these signals is extremely successful in the clinical diagnosis of

arrhythmias and other types of cardiovascular disease (1). However, as sparse measurements of the combined electrical activity of large areas of the myocardium, electrical signals provide little direct information about local EP properties.

The ability to perform a detailed EP characterisation in the clinical setting could lead to improved treatments for arrhythmias. For example, evidence suggests that areas with abnormal EP properties (such as fibrotic or ischaemic regions) and their border-zone are often the sites of the abnormal electrical activity driving arrhythmias (3). Cardiac regions characterised by EP changes such as low conduction velocity, heightened excitability or shortened action potential duration (APD) could be prime targets for localised therapies such as catheter ablation, likely improving their efficacy. So far, ablation strategies that target these EP heterogeneities have not been successful (4), partly due to the difficulty in identifying suitable ablation sites.

In this study, we present EP-PINNs, a Physics-Informed Neural Network, as an artificial intelligence tool capable of inferring EP properties from sparse measurements of transmembrane potential, V, in cardiac tissue. We test EP-PINNs using *in silico* data from EP biophysical simulations in several conditions and also *in vitro* optical mapping data. EP-PINNs are deployed in forward mode, as high-resolution solvers of the biophysical equations that control EP systems, and also in inverse mode, as estimators of EP parameters. Tests are performed in 1D and 2D for single waves and spiral waves in homogeneous and heterogeneous conditions. We further demonstrate a pharmacological application of EP-PINNs, as a tool to characterise the effect of two different channel blockers in *in vitro* optical mapping data.

In the next section, we will introduce the EP biophysical model used and the PINNs technique, as applied to the EP problem. We will contextualise our work within the available techniques for parameter estimation in EP and other cardiovascular applications of PINNs.

2. BACKGROUND

2.1. Biophysical Models of Cardiac Electrophysiology

Biophysical models of cardiac electrophysiology (5) are an important tool to understand how cardiac tissue properties affect the generation and propagation of cardiac electrical signals (action potentials, APs). They also offer an ideal means for the training and development of computational tools that may aim to infer EP properties from electrical and optical mapping measurements, such as EP-PINNs.

Several biophysical EP models have been proposed (see models.cellml.org/electrophysiology), each with varying degrees of detail aiming to reproduce different EP features, cardiac regions or animal/human experimental findings. Mathematically, these EP models usually take the form of a reaction-diffusion system where a diffusion term or equivalent (6) models the propagation of the electrical signal across the cardiac tissue. In the monodomain formulation, a partial differential equation (PDE) describes the spatio-temporal variations in the electrical potential across a myocyte cell membrane (V). This PDE is usually coupled to one or more ordinary differential equations (ODEs) describing how, at each point in time and space, V and other local state variables both determine and are determined by the flux of ions across the cell membrane (5).

The most parsimonious model of the action potential describes it as a travelling excitation wave followed by a non-excitable (refractory) region. This representation requires at least two state variables: V, which spreads (diffuses) across neighbouring regions, and a non-observable, non-diffusible recovery variable W which effectively controls the refractoriness and restitution properties of the model. One of the simplest models that captures these properties is the 6-parameter canine ventricular Aliev-Panfilov model (7), which models the transmembrane ionic currents ($V - W$ relationship) using smooth, differentiable functions. Furthermore, the diffusion of V across the cardiac tissue can be described by the monodomain equation (5), which, when combined with the Aliev-Panfilov model gives:

$$\frac{\partial V}{\partial t} = \vec{\nabla}.(D\vec{\nabla}V) - kV(V-a)(V-1) - VW \tag{1}$$

$$\frac{\mathrm{d}W}{\mathrm{d}t} = (\epsilon + \frac{\mu_1 W}{V+\mu_2})(-W - kV(V-b-1)) \tag{2}$$

The diffusion term $\vec{\nabla}.(D\vec{\nabla}V)$ reduces to $D\nabla^2 V$ in the case of homogeneous and isotropic conduction, i.e., when the diffusion tensor D is approximated by the same scalar throughout. Intuitively, this term quantifies how fast V is able to spread to its immediate neighbourhood to become more spatially homogeneous. D is determined mostly by the electrical conductivity of the myocardium and is a strong determinant of the propagation velocity of the AP. To prevent a leakage of V to regions outside the heart domain, the system described by Equations (1, 2) usually obeys no flux Neumann boundary conditions: $\frac{\partial V}{\partial \vec{n}} = 0$ in the boundary of the heart tissue.

The $-kV(V-a)(V-1) - VW$ term in Equations (1, 2) models the rapid changes in V caused by ionic fluxes across the cell membrane. a is related to the excitation threshold (i.e., the minimum V value that leads to the onset of an AP). The model's APD and refractoriness can, in turn, be controlled using b. The values for each of the model parameters are typically chosen empirically to reproduce observed electrical signals - we use the values listed in **Supplementary Table 1**. The Aliev-Panfilov model uses rescaled units: V is adimensional (typically in the $[0, 1]$ AU interval) and time is measured in temporal units, referred to as TU throughout this study. 1 TU corresponds to approximately 13 ms (7).

Other more complex EP models exist, through which it is possible to model individual membrane ionic currents and other biological components relevant for the AP and its propagation. One example used in the current study is a 14-current 30-variable canine atrial model that incorporates different degrees of EP remodelling caused by atrial fibrillation (8).

By assigning different sets of parameters and/or initial conditions to EP mathematical models, they can represent the

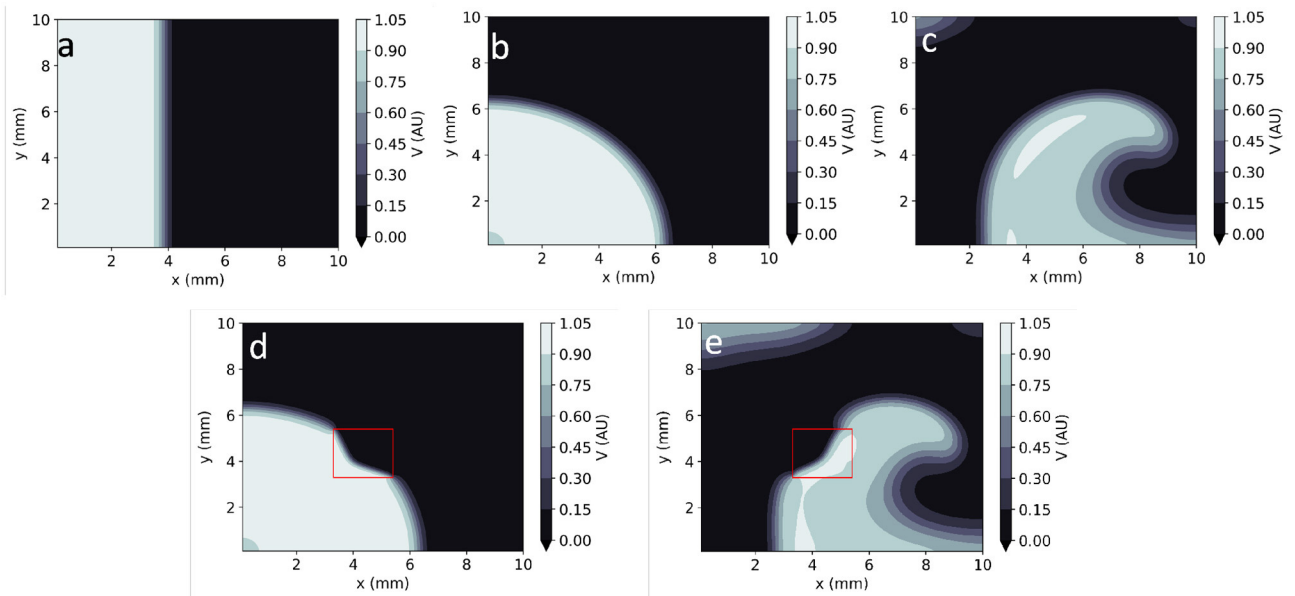


FIGURE 1 | Numerical solutions to the Aliev-Panfilov monodomain system for: **(a)** Planar wave. **(b)** Centrifugal wave. **(c)** Spiral wave. **(d)** Centrifugal wave in the presence of a square heterogeneity in *D*. **(e)** Spiral wave in the presence of a square heterogeneity in *D*.

electrical behaviour of the heart in both healthy and arrhythmic conditions. Healthy conditions are usually represented as unidirectional smooth propagation from a single source (**Figures 1a,b**). Large sources produce wave fronts which are close to planar, whereas point-like wavefronts lead to centrifugal (convex) wavefronts. Arrhythmias are usually modelled as one or more re-entrant waves (called spiral waves or rotors in 2D–**Figure 1c**). Moreover, the wavefronts and/or wavebacks of spiral waves can, in some instances, fragment (break-up), leading to complex activation patterns (9). These models can also consider localised pathology such as fibrosis, scar or ischaemia as heterogeneities in one or more model parameters. Localised reductions in *D*, for instance, can be used to reproduce the slow-down of AP propagation in fibrotic lesions (10). These heterogeneities can lead to local changes in the curvature of the activation wavefront (**Figures 1d,e**).

2.2. Parameter Estimation in Cardiac Electrophysiology

Biophysical models are usually employed in forward mode, with the aim of reproducing the system's behaviour assuming that all model parameters are perfectly known. In many circumstances, such as the identification of pathology from EP measurements, it is more desirable to use these biophysical models in inverse mode, inferring the tissue parameters that underlie an observed system behaviour. This task, often called parameter estimation (or systems identification), is extremely challenging for several reasons.

The observed data are typically insufficient to identify a unique parameter value, since the observations are usually sparse, incomplete and polluted by noise. This can be handled by optimisation methods, such as least squares, that fit data and model in an optimal way, combined with problem-dependent regularisation terms to stabilise the estimate (11). Typically, the optimisation requires many forward runs of the underlying model. Unfortunately, in physical systems that are described by PDEs such as in EP, the forward runs are very computationally expensive in realistic 2D and 3D settings. These difficulties are exacerbated by the fact that the parameters in many EP models are heterogeneous. Hence, we often do not infer a scalar quantity but a (discrete) function in space and/or time. This increases the cost and complexity of the inverse mode even further.

Modern inverse problem solvers such as ensemble Kalman filters (12) sequential Monte Carlo (13) and parameter estimation based on Markov chain Monte Carlo (14) are often combined with reduced order models (15) or multi-fidelity approaches (16) to decrease the computational complexity of parameter estimation. These very complex methods typically make strong assumptions about the statistical distribution of model parameters and require dedicated problem-specific parameterisation. It is not clear how well they can generalise when applied to a different EP model or task.

2.3. Physics Informed Neural Networks (PINNs)

Physics Informed Neural Networks (PINNs) (17) are an exciting new tool for the study of physical systems modelled by PDEs and/or ODEs. PINNs have been shown to both efficiently find high-resolution solutions (forward problem) and perform parameter estimation (inverse problem) in a variety of systems (18). As opposed to most types of neural networks (NNs), whose inputs are exclusively empirical data, PINNs incorporate explicit

knowledge about the physical laws that govern a system. This allows PINNs to compute solutions to initial and/or boundary value problems with comparatively less training data than conventional NNs.

Very briefly, PINNs make use of automatic differentiation (19) to rewrite the differential equations that a system obeys as the minimisation of a functional f. For example, for Equations (1, 2), this functional would be defined as:

$$f_V = -\frac{\partial V}{\partial t} + \vec{\nabla}.(D\vec{\nabla}V) - kV(V-a)(V-1) - VW \quad (3)$$

$$f_W = -\frac{\mathrm{d}W}{\mathrm{d}t} + (\epsilon + \frac{\mu_1 W}{V+\mu_2})(-W - kV(V-b-1)) \quad (4)$$

PINNs are trained to minimise a hybrid loss function, L, which ensures the system obeys known physical laws (described by f_V and f_W), whilst simultaneously fitting known empirical system measurements. L thus includes terms to account for:

- agreement with the experimental measurements, L_{data};
- consistency with the physical laws of the system, $L_{f_V} + L_{f_W}$;
- consistency with boundary $L_{V_{BC}}$ and initial value conditions $L_{V_{IC}}$.

Mathematically:

$$L = L_{data} + L_{f_V} + L_{f_W} + L_{V_{BC}} + L_{V_{IC}} \quad (5)$$

$$L = \frac{1}{N}\sum_{i=1}^{N}(V(x_i,t_i) - V_{GT\,i})^2 + \frac{1}{N_f}\sum_{j=1}^{N_f}(f_V(x_j,t_j)^2 + f_W(x_j,t_j)^2) + \frac{1}{N_b}\sum_{k=1}^{N_b}(\frac{\partial V}{\partial \vec{n}}(x_k,t_k))^2 + \frac{1}{N_0}\sum_{l=1}^{N_0}(V(x_l,t_0) - V_0)^2 \quad (6)$$

Each of the terms of the loss function is typically computed in different domains:

- (x_i, t_i) are the N measurement points, where ground truth (GT) experimental measurements, V_{GT}, are known and should be reproduced as closely as possible.
- (x_j, t_j) are the N_f residual points, where the fulfilment of the biophysical equations is tested.
- (x_k, t_k) are the N_b boundary points, where the network aims to fulfil the Neumann boundary condition for V.
- (x_i, t_0) are the N_0 initial points, where the known initial condition, V_0, is replicated as closely as possible.

There is an asymmetry in L between measurable (V) and latent (W) variables: only experimental measurements (and initial conditions) for V are usually available. Moreover, as W does not diffuse across the tissue, it does not obey any boundary conditions.

PINNs have typically been used for two main purposes (18). In the so-called forward mode, the NN's parameters are optimised to provide a representation of the physical system of interest consistent with observations. Using this representation, the system's differential equations can subsequently be solved at an arbitrarily high spatial or temporal resolution, bypassing the constraints (e.g., small temporal and spatial steps) of traditional numerical solvers. In inverse mode, PINNs additionally perform parameter inference (systems identification) by having the NN optimise one or more of the equation parameters (which here represent tissue EP properties) during the training process.

2.4. Cardiovascular Applications of PINNs

PINNs have recently been used in several areas of cardiovascular medicine, especially for applications related to blood flow. Examples include the estimation of myocardial perfusion and related physiological parameters from dynamic contrast enhanced MRI (20) and the estimation of haemodynamic parameters from microscopic images of aneurysms-on-a-chip (21).

In the field of cardiac EP, Sahli-Costabal et al. (22) used PINNs in forward mode to estimate activation time maps (ATs, i.e., the arrival times of the action potential) and conduction velocity (CV) maps in the left atrium at high spatial resolution. Sahli-Costabal's method uses PINNs to solve the (isotropic diffusion) eikonal equation, a simple relationship between ATs and the spatial gradient of CV. This effectively interpolates AT and CV across the left atrial surface. Although their PINNs implementation was exclusively deployed on simulated data, the proposed application is very clinically relevant, as it aims to mitigate the low spatial resolution of clinical AT measurements. Grandits et al. (23) subsequently extended this PINNs-eikonal equation framework to anisotropic conduction, using it to estimate high-resolution AT maps and fibre directions from *in silico* and patient data. The PINNs method performance was nevertheless lower than that of a traditional (variational) inverse solver (24). As they rely on the eikonal equation, these tools are not well suited to the study of arrhythmic conditions or to the inference of EP parameters other than AT and CV.

2.5. Optical Mapping for Experimental EP-PINNs Testing

Maps of transmural electrical potential (V), similar to those simulated using the Aliev-Panfilov model, can be experimentally recorded using optical mapping. Optical mapping is a technique in which voltage-sensitive fluorescent dyes are added to cardiomyocyte preparations before imaging at high spatio-temporal resolution (25). It can be used to effectively obtain uncalibrated measurements of $V(\vec{x}, t)$ in cardiac tissue across time. Although optical mapping can be challenging *in vivo*, *in vitro* experiments can provide very detailed insights into AP properties and cellular-level EP properties and gain insights into arrhythmic mechanisms (26). In particular, optical mapping can be used to study the effect of anti-arrhythmic drugs on cardiomyocyte preparations (27). These data were used to test EP-PINNs in an experimental setting.

3. MATERIALS AND METHODS

In this section, we provide details about the finite differences (FD) model used, in a variety of settings, to generate training and test data for EP-PINNs. We then introduce the EP-PINNs architecture, before giving details about each of the *in silico* experiments in which EP-PINNs were deployed. We end this section by introducing the experimental data (*in vitro* optical

mapping) used for testing EP-PINNs. The code used for EP-PINNs implementation and the generation of *in silico* GT data is freely available from github.com/martavarela/EP-PINNs.

3.1. Generation of *in silico* EP Data

We use an FD solver to generate *in silico* cardiac EP data, V_{GT}, which we use to train and evaluate the performance of EP-PINNs. Simulations were carried out in two different geometries: in 1D, in a 2-cm 1D domain cable; and in 2D, in a square with a 1-cm side. We use in-house code written in Matlab 2020b (Mathworks, Natick, MA, USA) that relies on central FD and an explicit 4-stage Runge-Kutta method to solve the isotropic monodomain Aliev-Panfilov model, as defined in Equations (1, 2) and model parameters listed in **Supplementary Table 1**. All simulations use Neumann boundary conditions and set $dt = 5 \times 10^{-3}$ *TU* and $dx = 100\mu m$ as temporal and spatial steps. All simulations were run for 300 *TU*, with the calculated $V_{GT}(\vec{x}, t)$ field saved at every 1 *TU* and at every spatial step (every $100\mu m$). We thus generate in total 1.4×10^4 and 7.0×10^5 data points for $V_{GT}(\vec{x}, t)$, respectively in 1D and 2D.

Figure 1 shows example time frames of the generated V maps in 2D. All APs were initialised by adding an external stimulus current $V_{stim} = 0.12$ *AU* to the right-hand side of Equation (1) for 1 *TU* in a sub-domain of the studied geometry (**Figure 1**). In 2D simulations, this includes both planar (**Figure 1a**) waves (emanating from a rectangular stimulus) and centrifugal waves (from a point-like excitation) (**Figure 1b**). Spiral waves were also created using the cross-field protocol: as an initial planar excitation propagates, a second planar excitation wave, orthogonal to the first, is initiated. When timed appropriately (42 *TU* after the first stimulus, in our model), the second planar wave continuously curves as it moves to non-refractory tissue, giving rise to a sustained spiral wave (**Figure 1c**).

The data generated by the FD model is treated as GT data and is used to both train and test EP-PINNs. As in past studies (18, 28), we train EP-PINNs on a case-by-case basis, with a small subset of the data for which parameter inference is going to be performed. This is in contrast to most supervised NNs, which are usually trained with large amounts of data acquired in varied circumstances.

Training data for EP-PINNs, $V_{GT}(\vec{x}, t)$, are provided to the network as an input and used to minimise the data discrepancy loss function, L_{data} in Equation (6). They amount to 10–20% of the total generated data, corresponding to a variable number of data points as detailed below and in **Supplementary Table 2**. The remaining data are withheld from EP-PINNs and used, in a post-processing step, to assess EP-PINNs' ability to reproduce cardiac APs, as detailed below.

3.2. Architecture and Training of EP-PINNs

EP-PINNs are designed, trained and deployed using the Python DeepXDE library (28). As in past successful implementations of PINNs (18, 28), we use a fully connected network architecture.

As detailed in **Figure 2**, EP-PINNs take as inputs the spatio-temporal points, $(\vec{x}, t)$, where they will estimate the main outputs: V (and W). Experimental measurements of V, $V_{GT}(\vec{x}, t)$, are also provided to EP-PINNs as inputs in a (training) subset of $(\vec{x}, t)$. EP-PINNs minimise the hybrid physics-informed loss function described before (see Equations 5 and 6), by adjusting the network's weights and biases (collectively named θ in **Figure 2**). In inverse mode, EP-PINNs additionally adjust one or more parameters of the Aliev-Panfilov model (generically λ in **Figure 2**). The number of layers and neurons used by EP-PINNs is adjusted to the domain size and the complexity of the problem at hand, as described below and detailed in **Supplementary Table 2**.

The optimisation approach used for EP-PINNs also depends on problem size. For 1D problems, we use Adam optimisation (29)—see **Figure 2C**. We empirically determined that, in 2D, EP-PINNs' performance improved when initially using Adam optimisation for only the data agreement term (L_{data} in Equation 6), followed by Adam optimisation for the full loss function and ending in a final phase of L-BFGS optimisation (30)—see **Figure 2C**. The initial Adam training phases are used to rapidly approach the desired minimum and the final L-BFGS optimisation phase helps the network converge faster towards it (28).

In the presence of spatially-varying EP parameters (D in the current study), we use network architecture B (see **Figure 2B**). Here, $D(\vec{x}, t)$ is estimated by a parallel NN, NN_D, with the same number of layers and neurons as the main NN. In this setup, D is treated as a system variable (on par with V and W) and its estimates directly contribute to the loss term that ensures the agreement with the EP equations: L_{f_V} in Equation (6).

The hyperbolic tangent function (*tanh*) is used throughout as the differentiable activation function and Glorot initialisation from a uniform distribution is used for all weights (31). Additionally, to minimise convergence problems caused by explosive gradients (28) and enhance the NN's stability, we implemented an automatic reset of the training process when the losses at the first epoch of the training exceeded a predefined threshold.

EP-PINNs were trained on a high performance machine with 1 RTX6000 GPU and 4 AMD EPYC 7742 CPUs. Typical training times varied between 15 min (for 1D problems) and 16h (for heterogeneous spiral wave problems), as detailed in **Supplementary Table 2**.

3.2.1. Assessment of EP-PINNs Performance

To assess the performance of EP-PINNs, we calculate, across all test points N_{test}, the root mean squared error (*RMSE*) for estimates of V:

$$RMSE = \sqrt{\frac{1}{N_{test}} \sum_{i=1}^{N_{test}} (V(i) - V_{GT}(i))^2} \tag{7}$$

The RMSE is, by construction, adimensional and in the same scale range as V. All experiments were repeated at least 5 times to probe the variability in RMSE. In inverse mode, we additionally calculate the precision of the estimated model parameters using the standard deviation of the parameters estimated by EP-PINNs in these 5 different runs.

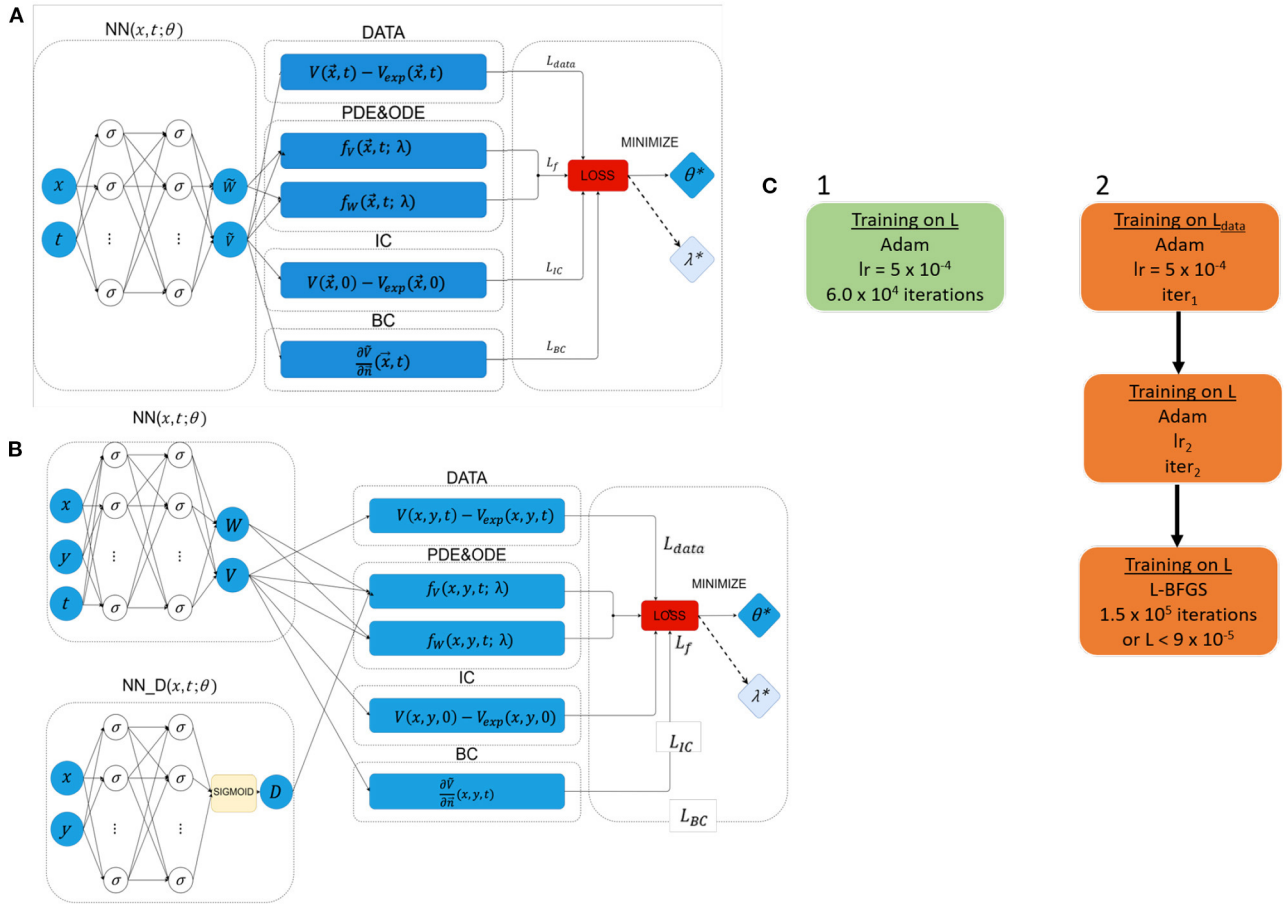


FIGURE 2 | Network architectures and training schemes for EP-PINNs used in this study. Neurons are represented by σ. In forward simulations only network parameters θ were estimated, whereas in the inverse setting one or more EP parameters (λ) were also calculated. **(A)** Architecture used in all forward simulations and all inverse simulations (1D and 2D) with homogeneous EP parameters. **(B)** Architecture used in all 2D forward simulations and all inverse simulations where D was a spatially-varying field. **(C)** Used training schemes. 1 was used in *in silico* 1D problems, whereas training scheme 2 was used in 2D problems and for optical mapping experimental data. The number of NN neurons and layers varied across different experiments, as did the learning rates (lr) and number of iterations (iter) in training scheme 2. Further details about the architecture and parameterisation of the NNs and training schemes can be found in **Supplementary Table 2**.

3.3. Forward Solution of EP Models

3.3.1. 1D Cable Geometry

We assessed the accuracy of EP-PINNs when solving the monodomain equation with the Aliev-Panfilov ionic model (forward problem) in the 20-cm cable during 70 TU (corresponding to 903 ms). We divided the GT data from the FD solver, V_{GT}, into test and training datasets using a 90–10% split. This corresponds to 140 randomly chosen points across the temporal and spatial domains for training EP-PINNs and 1,260 points for testing. As for inverse 1D problems, EP-PINNs were implemented in this instance using architecture A and training scheme 1 (see **Supplementary Table 2**).

We used two different training setups:

1. using only *in silico* experimental V_{GT} measurements as GT, as in Equation (6).
2. using *in silico* experimental points for V_{GT} and W_{GT} in the loss function, by adding an extra term: $L_{W_{GT}} = \frac{1}{N}\sum_{i=1}^{N}(W(x_i,t_i) - W_{GTi})^2$ to Equation (5).

Setup 1 more closely resembles an experimental setup, as the latent variable W_{GT} is not usually measurable.

To gauge whether EP-PINNs performance depended on model parameter choice, we used EP-PINNs on GT data synthesised with two different sets of model parameters, as detailed in **Supplementary Table 1**.

We additionally evaluated the performance of EP-PINNs in *in silico* data corrupted by noise. For this, we added to V_{GT} zero-mean Gaussian noise with standard deviations of 0.05, 0.10, 0.50 or 1.00 *AU* (with 1 AU being the approximate amplitude of an AP).

We also tested the performance of EP-PINNs in the presence of a reduced number of training data points. For this, we provided the network with V_{GT} at 1×10^4, 5×10^3, 1×10^3 or 100 random training points within the 1D space-time domain (compared to 1.40×10^7 in usual conditions). We used architecture A with training scheme 1 in all 1D problems (see **Supplementary Table 2**).

3.3.2. 2D Rectangular Geometry

We solved the forward problem in 2D, using the Aliev-Panfilov model in an isotropic and homogeneous 10-cm side square over 70 TU. We used 1.4×10^5 randomly chosen data points (across the temporal and spatial domains) from the FD solver for the training of EP-PINNs (corresponding to 20% of the total generated points) and reserved the remaining 5.6×10^5 points to assess EP-PINNs' performance. Simulations were carried out in three scenarios: planar wave (**Figure 1a**), centrifugal wave, emanating from point-like excitation in a corner (**Figure 1b**) and a spiral wave (**Figure 1c**). Here and in the equivalent inverse setup, EP-PINNs were implemented using architecture A and training scheme 2 (see **Supplementary Table 2**).

The point-like excitation and spiral wave scenarios were also simulated in heterogeneous conditions, where a 2-mm side square within the spatial domain was assigned a permanent diffusion coefficient ($D_{lesion} = 0.02\ mm^2/TU$) lower than that of background tissue ($D_0 = 0.1\ mm^2/TU$). Architecture B with training scheme 2 was used for all (forward and inverse) heterogeneous problems (see **Supplementary Table 2**). A sigmoid function was used in the NN dedicated to estimating $D(\vec{x}, t)$ (NN_D in **Figure 2B**) to account for the fact that $D(\vec{x}, t)$ follows a binomial distribution: D_0 in healthy tissue and D_{lesion} otherwise.

3.4. Inverse Estimation of EP Parameters

We withheld the value of one or more of the model parameters from EP-PINNs, which were instead estimated by it. These parameters were chosen for their comparatively simple biophysical interpretation, known susceptibility to both disease remodelling and pharmacological action, and the limited degree of mathematical coupling between them. They were:

- a, which is related to the tissue excitation threshold (see Equation 1);
- D, the scalar diffusion coefficient (proportional to the electrical conductivity of the tissue, see Equation 1);
- b, which controls APD, see Equation 2.
- a and D simultaneously;
- b and D simultaneously.

3.4.1. Homogeneous 1D and 2D Geometries

We solved the inverse problem in the same (1D or 2D) setup, architecture and training scheme as for the forward problems and using a similar division of randomly chosen data points for training and testing. We assumed that none of the selected parameters varied across time or space (except for D, in the heterogeneous problem described below) and used the values listed in **Supplementary Table 1** as GT values. In 1D, we additionally investigated how EP parameter estimation was affected by experimental noise. For this, Gaussian noise ($\sigma = 0.05$ or $0.10\ AU$) was added to the *in silico* data as described for the forward mode in section 3.3.

In addition to EP-PINNs robustness in the presence of experimental noise, we are also interested in its ability to cope with model uncertainty. Therefore, to assess EP-PINNs' ability to generalise beyond the model it is trained on, we additionally tested it on APs generated on a much more complex canine atrial EP model (8). These atrial APs are markedly different from those of the Aliev Panfilov model the EP-PINNs assumes, both in morphology and restitution properties. The canine atrial model data were synthesised using Matlab (https://models.cellml.org/workspace/47c) with central FD and explicit forward Euler schemes ($dt = 5\mu s$ and $dx = 100\mu m$), with data saved at every spatial step and at every ms. Using this model, we created GT data for left atrial cells at 3 stages of AF-induced remodelling, which differed in APD. We tested, using the 1D model in inverse mode, whether EP-PINNs could identify the reduction in APD (detected as an increase in parameter b) in left atrial APs caused by increasing amounts of AF remodelling.

3.4.2. Estimation of EP Parameter Heterogeneities

We assessed EP-PINNs' ability to recognise spatial heterogeneities in model parameters in 2D, as a test for EP-PINNs potential for identifying spatially-varying lesions such as fibrosis. For this, we used the same setup as in section 3.3.2, with $D_0 = 0.1\ mm^2/TU$ reduced to $D_{lesion} = 0.02\ mm^2/TU$ in a similar square region. As before, we estimated $D(\vec{x}, t)$ on its own and simultaneously with either the a or b global model parameters.

3.5. Parameter Estimation Using Optical Mapping Data

We tested EP-PINNs performance on *in vitro* datasets using optical mapping data from neonatal ventricular rat myocyte preparations stained with a voltage-sensitive dye, as described in detail by Chowdhury et al. (27). Briefly, we used four time series (movies) of optical mapping images (field of view: 4.1 × 0.1 mm, spatial resolution: 1.2 μm, temporal resolution: 2 ms, duration: 300 ms). In two of these image series, ionic channel modulating drugs (E-4031 or nifedipine) had been administered at half maximal inhibitory concentration (IC_{50}): 772.2 nM for nifedipine and 243.4 nM for E-4031. The other two temporal image series consisted of matched control (baseline) preparations, to which no drug had been given.

We manually selected two square regions of interest (ROIs) with a side of 2.3 μm and with their centres 0.7 mm apart, in the same image location for each time series. We spatially averaged the optical signal over each ROI to obtain a signal trace across time. From this signal, we manually selected two consecutive APs and normalised the signal to the [0, 1] interval for consistency with the Aliev-Panfilov model. To improve the signal to noise ratio (SNR) of this trace, we applied a mean average filter twice, aligned and averaged the two APs over time to obtain a single higher-SNR AP. These pairs of post-processed APs were used as inputs to 1D EP-PINNs in inverse mode, with b as the variable to be estimated. All model parameters were unchanged from those in **Supplementary Table 1**. We used EP-PINNs' architecture A with training scheme 2 to estimate b 10 times for each setting. 116-200 points were used for training EP-PINNs and 29–50 to test it, as detailed in **Supplementary Table 2**.

We investigated in particular whether EP-PINNs could detect the effect on APD of E-4031 and nifedipine, which, respectively, block the hERG voltage-gated potassium channel (I_{Kr} current)

and the L-type calcium channel (I_{CaL} current). Whereas, E-4031 will extend APD (and thus shorten b in the Aliev-Panfilov model—see Equations 1, 2), nifedipine will have the opposite effect, decreasing APD (increasing b). We compared EP-PINNs' b estimates for i) E-4031 vs. baseline and ii) nifedipine vs. baseline and assessed, using a t-test, whether the administered drugs significantly changed b values.

4. RESULTS

EP-PINNs successfully solved the monodomain equation coupled with the Aliev-Panfilov model in 1D and 2D, including in the presence of heterogeneities in D. In inverse mode, the proposed setup could also successfully perform EP parameter estimation from *in silico* and *in vitro* data, as described below.

4.1. Forward Solution of EP Models

4.1.1. 1D Cable Geometry

EP-PINNs accurately reproduced the features, morphology and conduction properties of the APs generated by the Aliev-Panfilov model (**Figure 3A**). We found that EP-PINNs could accurately simulate APs even in the absence of GT values for the latent variable W. Although the error was slightly increased in the absence of W_{GT} ($RMSE = 6.0 \times 10^{-3} \pm 2.0 \times 10^{-3}$ vs. $9.0 \times 10^{-3} \pm 4.0 \times 10^{-3}$), the RMSE was still minimal (see top left inset in **Figure 3A**) and the estimated $V(\vec{x}, t)$ were visually indistinguishable from GT traces in both cases. As a consequence, in all other experiments described in this study, GT data for W was not used to train EP-PINNs, whose training relied solely on V_{GT}.

We found that EP-PINNs could solve the model accurately in the presence of even small numbers of V_{GT} points for training, with $RMSE \leq 2.5 \times 10^{-2}$ even when trained with only 100 points (**Figure 3A**). As expected, increasing noise in V_{GT} led to increasing levels of error in V estimates (**Figure 3B**), but EP-PINNs were able to converge in the presence of Gaussian noise with a standard deviation below 0.5 *AU* (approximately 0.5 times the amplitude of an AP). When using a different set of parameters for the Aliev-Panfilov model (see **Supplementary Table 1**), EP-PINNs' accuracy was comparable, suggesting that the obtained EP-PINNs performance is robust to different biophysical model settings.

4.1.2. 2D Rectangular Geometry

In homogeneous conditions, EP-PINNs were also able to reproduce AP propagation in 2D for planar, centrifugal and spiral waves (see **Figures 4a–c**), with excellent accuracy ($RMSE < 3.0 \times 10^{-2}$ throughout). In the presence of heterogeneities in D, EP-PINNs were also able to accurately simulate APs with an RMSE of $5.6 \times 10^{-3} \pm 6.6 \times 10^{-4}$ for centrifugal waves, which increased slightly to $2.7 \times 10^{-2} \pm 4.3 \times 10^{-3}$ for spiral ones (**Figures 4a–c**). In the spiral wave scenario, EP-PINNs found it most difficult to reproduce V in the high wavefront curvature regions close to the spiral wave tip.

Movies showing the propagation of APs in the 2D rectangular domain across time (for both GT and EP-PINNs solvers) can be seen in **Supplementary Videos 1–5**.

4.2. Inverse Estimation of EP Parameters

4.2.1. Homogeneous 1D and 2D Geometries

EP-PINNs were able to estimate global model parameters in 1D the presence of varying degrees of noise, as detailed in **Figure 5**. As for the forward problem (**Figure 3**), AP morphology and main properties were well reproduced even in the presence of large amounts of noise, with $RMSE < 9.0 \times 10^{-3}$ overall. Example plots of $V(t)$ in inverse mode in 1D are shown in **Supplementary Figure 1**.

When estimating only one model parameter, relative errors (RE) did not exceed on average 27%, even in the presence of noise. Estimates of b, which determines AP duration, were the most accurate ($|RE| < 3\%$), followed by D ($|RE| < 35\%$), which EP-PINNs tended to underestimate. a estimation was considerably more difficult. Joint estimation of two parameters led in general to less accurate parameter estimates (**Figure 5**), with an error as high as 100% for a when estimated jointly with D in the presence of noise (see **Figure 5**). When performing simultaneous estimation of two parameters, no evidence of coupling between them was observed.

EP-PINNs were additionally able to perform robust parameter estimation on synthetic experimental data generated by a different EP model (8). When estimating b in APs generated by a different atrial EP model, it correctly inferred that APD is reduced (i.e., b is increased) for increasing degrees of AF remodelling, as shown in **Figure 6**. Moreover, the main AP features were well reproduced, with small discrepancies related to the differences between the two models (**Figure 6**). Interestingly, the solution proposed by EP-PINNs consistently shows a less steep depolarisation than expected from either Aliev-Panfilov model and the detailed canine model. This mismatch is likely to be a consequence of the EP-PINNs' adjustment to slightly different AP morphologies from those in the Aliev-Panfilov model in its loss function. This suggests that the cross-model estimation of parameters related to excitability (e.g., a) may not be as successful as parameters related to APD (such as b).

Global parameter estimation in 2D was again successful, for both unidirectional propagation and spiral wave conditions, as demonstrated in **Figure 7**. As in 1D, V was also correctly reproduced in the different analysed conditions (**Figures 7b–d**), with $RMSE < 2.3 \times 10^{-2}$ throughout. Across all experiments, parameter estimation and V reconstruction were most successful for planar wave conditions followed by centrifugal waves and less accurate for spiral waves, as shown in **Figure 7a**. The comparatively worse performance of EP-PINNs in spiral wave conditions may be caused by the spatial dependency of wave front curvature in this setting.

Errors were largest when estimating two parameters simultaneously, with EP-PINNs again struggling to estimate a, especially when in conjunction with D ($|RE|_a < 100\%$). Estimates of b were once again the most accurate ($|RE|_b < 8\%$) and, as in 1D, EP-PINNs tended to underestimate D across all settings and to underestimate all parameters in the spiral wave scenario.

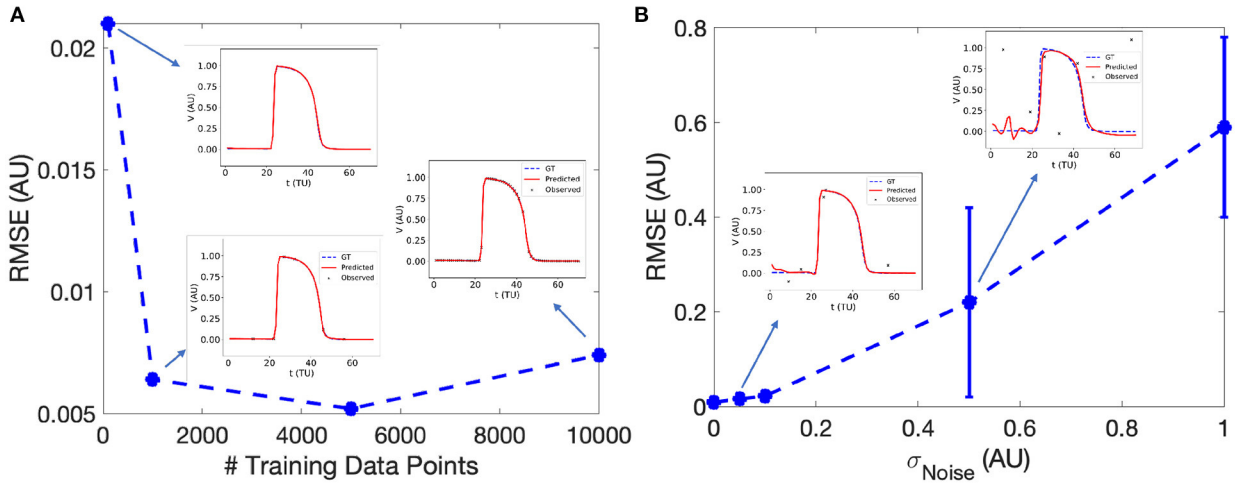


FIGURE 3 | Effect of the number of sampled experimental points **(A)** and experimental noise **(B)** on EP-PINNs *V* estimates. The error in estimation is measured using the root mean square error (RMSE, see Equation 7) in forward mode in 1D. Representative *V*(*t*) plots sampled at a random spatial location are shown as insets for some of the probed conditions.

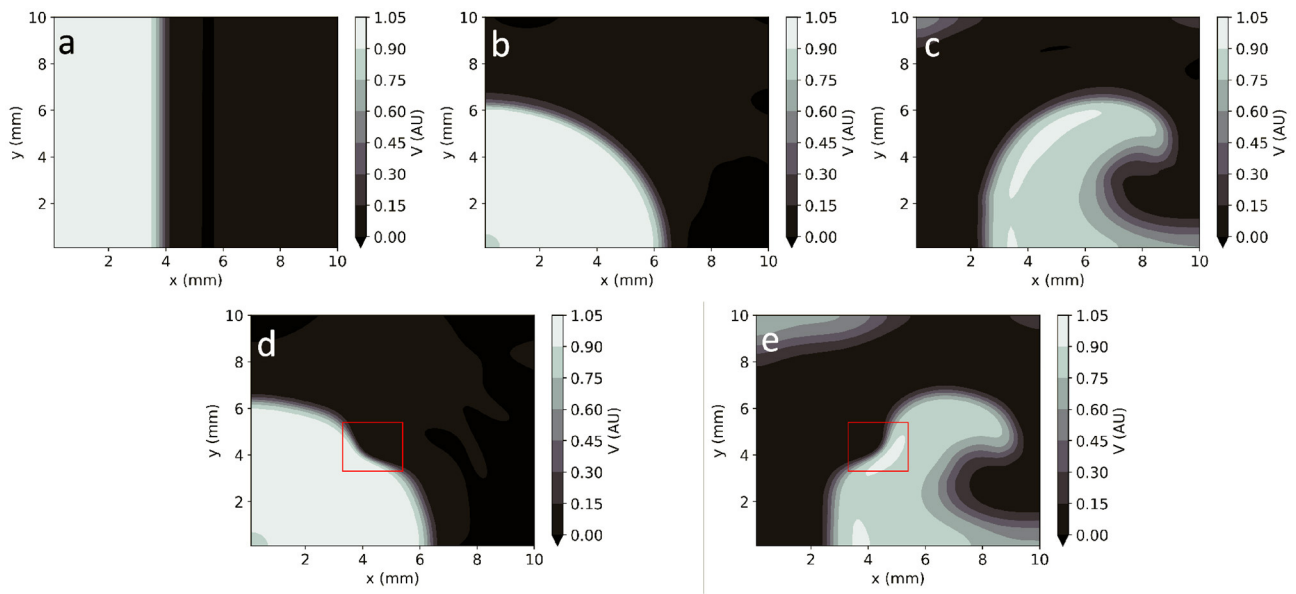


FIGURE 4 | EP-PINNs 2D forward solutions to the Aliev-Panfilov monodomain system for the same conditions as the GT data depicted in **Figure 1**. **(a)** Planar wave. **(b)** Centrifugal wave. **(c)** Spiral wave. **(d)** Centrifugal wave in the presence of a square heterogeneity in *D*. **(e)** Spiral wave in the presence of a square heterogeneity in *D*.

4.2.2. Estimation of EP Parameter Heterogeneities

EP-PINNs were able to estimate *D* on a pixel-by-pixel basis with remarkable accuracy, as demonstrated in **Figures 8c,d**, accurately identifying the low *D* region. $RMSE_D$ was consistently below 3.5×10^{-2} (**Figure 8e**) and was lower for the centrifugal wave case than the spiral wave one. As before, *V* was similarly well reproduced (**Figures 8a,b**), especially in the centrifugal wave scenario (see **Figures 8b,d**), with $RMSE < 3.0 \times 10^{-2}$. In the spiral wave scenario, EP-PINNs found the estimation of *D* hardest near the spiral tip (see **Figure 8d**), where the high wavefront curvature may resemble the wavefront bending caused by low *D* regions.

Using architecture B, global estimation of *a* and *b* was also possible in the presence of the heterogeneous *D* field, keeping the same trends as in the 1D and 2D homogeneous cases (**Figure 8e**). As before, no evidence of coupling between the simultaneously estimated parameters was found. In detail:

- *b* was very accurately estimated ($|RE| < 10\%$), whereas *a* estimates had a larger error ($|RE| < 100\%$);

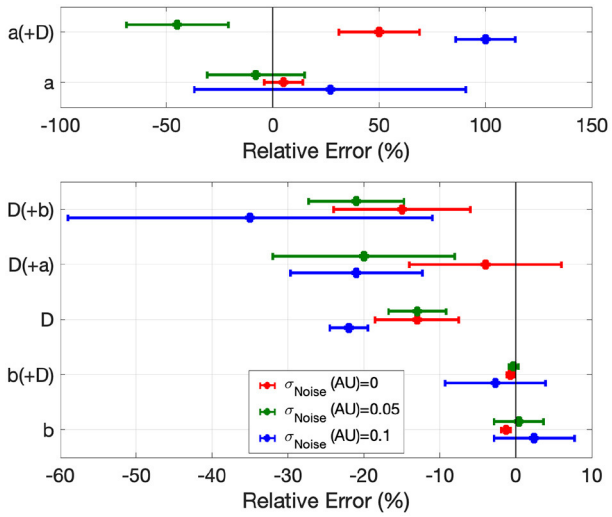


FIGURE 5 | Error in global EP parameter estimates by EP-PINNs in the inverse setting in 1D in the presence of different amounts of experimental noise. We show the relative error for *a*, *b* and *D*, when estimated separately and in pairs.

- estimating *a* and *b* simultaneously with the *D* field led to small decreases in accuracy overall;
- estimating *V* and *D* in the spiral wave setting was more difficult than in the presence of a centrifugal wave.

4.3. Parameter Estimation using Optical Mapping Data

Using optical mapping signals, EP-PINNs were able to accurately reproduce experimental APs and identify the actions of nifedipine and E-4031, correctly estimating that they, respectively, reduce and increase APD (**Figure 9**). The reduction in APD caused by nifedipine, an I_{CaL} blocker, was detected by EP-PINNs as a significant increase in b in the Aliev-Panfilov model [$\Delta b = (2.7 \pm 0.9) \times 10^{-2}, p < 10^{-5}$]. The E-4031-induced increase in APD was more subtle [$\Delta b = (-2.1 \pm 3.0) \times 10^{-2}$] and non-significant ($p = 0.38$). The reduced effect of E-4031 in these data is consistent with the modest role I_{Kr}, the current blocked by E-4031, is expected to play in rodent APs (32).

5. DISCUSSION

We present EP-PINNs, a successful framework to estimate EP parameters from measurements of trans-membrane potential V. We demonstrate EP-PINNs ability to accurately reproduce AP propagation in 1D and 2D in the presence of very sparse experimental measurements, experimental noise and model uncertainty. EP-PINNs can also estimate, for 1D and 2D *in silico* and *in vitro* data, global markers of APD, excitation threshold and/or conductivity (diffusion coefficient, D). We additionally show that EP-PINNs are further capable of identifying heterogeneities in EP parameters, such as D, even in arrhythmic conditions, showcasing their potential for clinically useful applications.

5.1. Forward Solution of EP Models

EP-PINNs offer a flexible and easy to implement framework for parameter estimation in EP. Sahli-Costabal et al. (22) and Grandits et al. (23) had already demonstrated PINNs' potential in cardiac EP by estimating high-resolution left atrial AT and CV maps in sinus rhythm conditions using a simple activation-only biophysical model. We extend PINNs' applications in EP by applying them to a more complex biophysical model, the monodomain Aliev-Panfilov model (7), which also captures restitution properties through the inclusion of a latent (non-measurable) variable, W. For the first time, we use a PINNs framework for the simulation of arrhythmic conditions, such as spiral waves, and for the estimation of parameters unrelated to the AT of V.

Importantly, we show EP-PINNs' are able to reproduce APs and perform parameter estimation in the absence of any data for W, which is not available experimentally. This bodes well for the deployment of PINNs for even more complex EP models, which use a higher number of latent variables to model individual ionic channels. This possibility is also supported by the work of Yazdani et al. (33), who successfully used PINNs for parameter estimation across several biological systems described by large sets of coupled ODEs.

We demonstrate PINNs' ability to describe AP dynamics in several circumstances. In 1D, EP-PINNs were able to reproduce APs even in the presence of very reduced amounts of experimental data (**Figure 3A**) and large amounts of noise (**Figure 3B**). PINNs' incorporation of explicit biophysical equations in the NN's loss function acts as an effective regulariser in EP problems, as demonstrated before in many other physical systems (17, 18). We note that these inherent regularisation properties allow PINNs to be trained with much lower amounts of training data than conventional NNs. The main drawback is the need for a comparatively time-intensive training on a case-by-case basis, compared to the global training usually employed with supervised NNs.

In 2D, we were able to accurately replicate AP dynamics for planar, centrifugal and planar waves, even in the presence of heterogeneities in the diffusion coefficient (see **Figures 4**, **8**). We found that a more sophisticated training scheme and, for spiral waves, an increased NN capacity (5 layers of 64 neurons vs. 4 layers of 32 neurons, see **Supplementary Table 2**) were necessary for convergence in these large and complex problems. This more complex setup could, of course, have been used to solve the simpler 1D problems, through a trade-off between computational time and the convenience of a one-size-fits-all EP-PINNs approach.

5.2. Inverse Estimation of EP Parameters

It is in inverse mode, when estimating model properties from sparse measurements of V, that the EP-PINNs framework showcases its usefulness. Parameter estimation is an important topic in EP, as it is essential for both the personalisation of models and for the understanding of the effect of pathology and drugs on APs. Although parameter estimation in EP has been extensively discussed as a means of reducing the uncertainty associated with current biophysical models (34), NNs had not yet been used

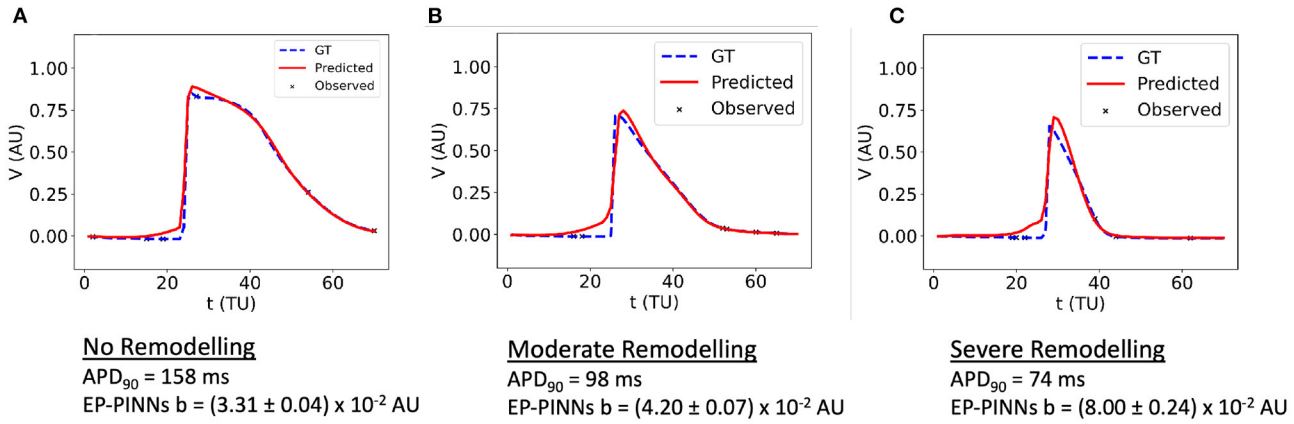


FIGURE 6 | 1D inverse EP-PINNs solution for a detailed canine left atrial model in the conditions of: **(A)** no remodelling. **(B)** moderate AF remodelling. **(C)** severe AF remodelling. Representative $V(t)$ plots are shown throughout, accompanied by the models 90% APD and EP-PINNS estimates for b, a parameter inversely proportional to APD.

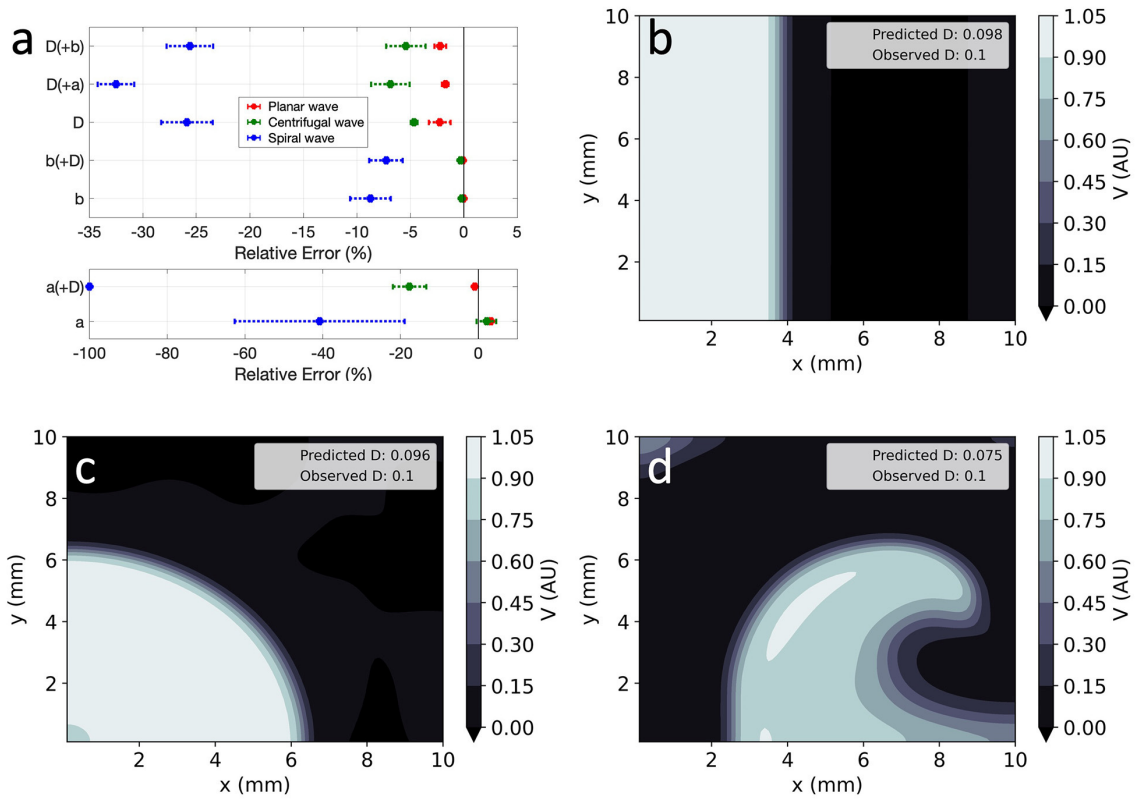


FIGURE 7 | EP-PINNs inverse solution in homogeneous conditions in 2D. **(a)** Relative error for global estimates of: a and D and b and D, when estimated separately or simultaneously. **(b–d)** Corresponding representative V maps for: planar wave **(b)**, centrifugal wave **(c)**, and spiral wave **(d)**. Compare **(b–d)** to the corresponding GT in **Figures 1a-c** and the forward solutions in **Figures 4a–c**.

for dedicated model parameter estimation in EP. This contrasts with the more common use of NNs as efficient solvers of EP systems (in a similar fashion to the EP-PINNs forward mode in the current study) (15, 35, 36).

Using EP-PINNs, we were able to estimate, with different degrees of accuracy, three different biophysical parameters, each controlling, in an almost uncoupled manner, different observable properties of the system: APD (through b), excitability

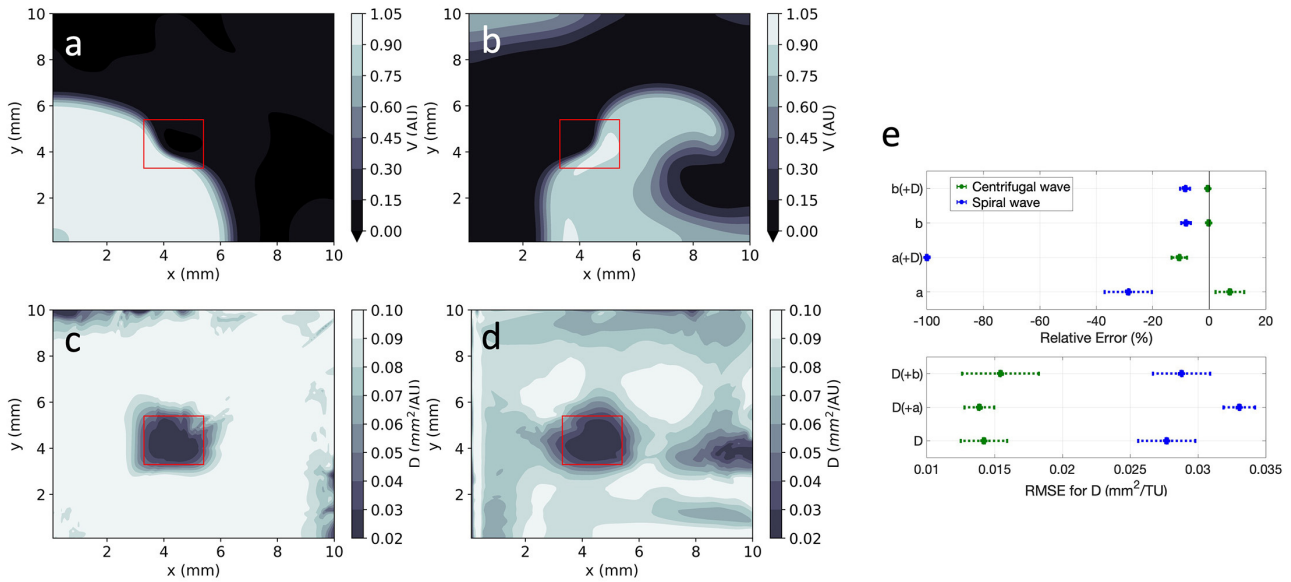


FIGURE 8 | EP-PINNs 2D inverse solution in the presence of heterogeneities in *D*. Maps showing representative *V* and *D* estimates for: **(a,c)** centrifugal wave and **(b,d)** spiral wave. **(e)** Error for global estimates of *a* and *b* and RMSE for estimates of *D* across the 2D domain, for all estimated parameter combinations. Compare **(a,b)** to the corresponding GT in **Figures 1d,e** and the forward solutions in **Figures 4d,e**.

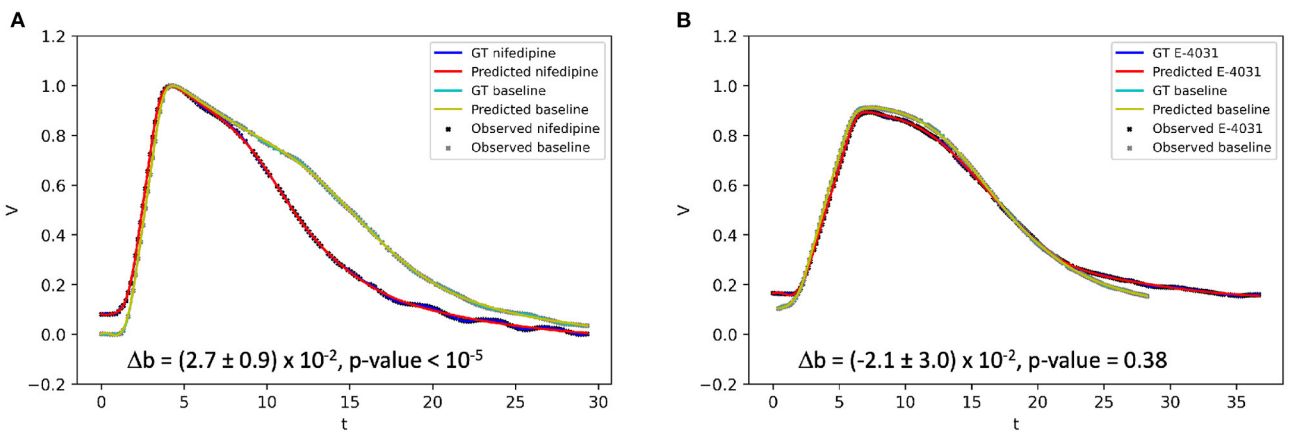


FIGURE 9 | 1D inverse EP-PINNs solution for experimental optical mapping data, in the presence of **(A)** I_{CaL} channel blocker nifedipine and **(B)** I_{Kr} channel blocker E-4031. (Δb refers to the change in model parameter *b* in the presence of the drug when compared to the baseline value. Each AP was acquired in separate datasets and juxtaposed in the figure to allow visual comparisons.

(through *a*) and conduction velocity (through *D*). In both 1D and 2D (homogeneous and heterogeneous) problems, the network was highly successful at estimating *b*, but struggled with *a*, especially when estimating it in tandem with *D*. This is likely to reflect a dependency between EP-PINNs' inverse mode effectiveness and the solution type that is probed experimentally. Indeed, when compared to *b*, the experimental inputs to the network ($V_{GT}(\vec{x},t)$), depend little on *a* providing the initial stimulus is supra-threshold. An exception could have been the spiral wave scenario, whose properties (e.g., the distance between spiral arms) depend strongly on model parameters such as *a*. Model parameters are more strongly coupled in the properties of spiral solutions, however, explaining the consistently lower accuracy of EP-PINNs' estimates in this scenario (**Figures 7a,d**). Parameter estimation in the spiral wave scenario may be improved when EP-PINNs are trained in longer time series, in which the spiral wave tip samples more of the spatial domain.

These issues underlie the difficulties of finding a single experimental design that allows for simultaneous accurate estimation of several EP parameters. A solution for this problem may be the training of PINNs using data from the same system acquired in different experimental conditions, perhaps by training separate NNs in parallel with a combined loss function, in an analogous manner to architecture B in this study (see **Figure 2**).

EP-PINNs were additionally able to identify heterogeneities in D across the 2D domain (see **Figure 8**), by estimating D in a separate parallel NN which shared terms of the loss function (L_{data}, see Equation 5) with the main NN. Characterising heterogeneities in EP parameters from electrical measurements is an interesting problem from a clinical point of view, as arrhythmias such as AF are often accompanied by heterogeneous changes in EP properties. Of these, regions of dense fibrosis [often modelled as areas with a reduced D (10)] are a promising candidate for personalised ablation sites (37), which may increase the overall efficacy of these procedures. EP-PINNs are thus well placed to help locate these putative ablation sites by identifying spatial heterogeneities in EP parameters such as D.

The inputs to the current EP-PINNs implementation, $V(\vec{x}, t)$ are not, however, measurable clinically. Future work will modify EP-PINNs to instead perform parameter inference from extracellular electrical potentials, ϕ_e, which are regularly measured during clinical procedures using contact electrodes. ϕ_e can be interpreted as a weighted spatial convolution of the $\vec{\nabla}.(D\vec{\nabla}V)$ term in Equation (2) (38, 39), making the identification of localised EP changes more difficult. To take this into account, EP-PINNs designed for ϕ_e analysis may benefit from a move away from the current fully-connected architecture to incorporate, for example, convolutional layers.

5.3. Parameter Estimation Using Optical Mapping Data

An important point for future clinical applications of EP-PINNs is its ability to generalise beyond the details of the setup it is trained on. We showed that the proposed EP-PINNs implementation is model-agnostic, as it was able to perform robust parameter inference on *in silico* data generated by a much more complex atrial EP model (8) than the 2-variable ventricular one used in its loss function. In particular, EP-PINNs were able to correctly identify the decrease in APD (manifest as an increase in b) that is associated with increasing degrees of AF-induced remodelling in this model (see **Figure 6**).

In contrast to most previous PINN studies (17, 22, 28, 33, 40), we complemented the *in silico* studies with an assessment of the EP-PINNs performance on experimental biological data. Despite requiring a proportionally higher amount of training data than *in silico* experiments, EP-PINNs were able to cope well with the noise and artefacts unavoidably present in experimental data to identify the effect on APD of two different drugs: an I_{Kr} blocker and an I_{CaL} blocker. As for the data generated by a different mathematical model (**Figure 6**), EP-PINNs coped well with differences between the experimental data and the Aliev-Panfilov model, namely in resting membrane potential (see **Figure 9**). This ability to generalise well to data with different characteristics could be due to the use of PDE as a soft constraint (a term in the loss function) in the EP-PINNs framework, as well as the lack of assumptions about the distributions from which data come from.

As demonstrated for the *in silico* tests, the EP-PINNs framework can easily be extended to simultaneously infer the effect of drugs on more than one EP parameter, which may be useful for the characterisation and safety assessments of anti-arrhythmic drugs. These applications may further benefit from the training of EP-PINNs on more complex biophysical models, to obtain a more fine-grained characterisation of potential pharmacological (or pathological) effects.

5.4. Limitations and Future Plans

The current study aims to demonstrate the potential of PINNs within EP, as an initial necessary step towards clinical applications of this method. As such, we only trained EP-PINNs using one comparatively simple EP model, in 1D and 2D scenarios and for the estimation of a small number of EP parameters. We additionally did not test EP-PINNs in the chaotic or pseudo-chaotic scenarios of spiral wave break-up (9), which may be relevant for some arrhythmias.

Generalisations of the proposed framework to 2D/3D geometries representative of cardiac chambers, anisotropic conditions and more detailed EP biophysical models can be achieved by further increasing the capacity of the deployed EP-PINNs, with a concurrent increase in computational resources. However, promising and less resource-intensive applications for EP-PINNs may be the characterisation of pharmacological effects on AP or the identification of heterogeneities in EP properties from EGM signals.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

CH and AO wrote the EP-PINNs software, conducted the experiments, and analysed the results. RC acquired the experimental data and oversaw its analysis. AB, EU, RC, and NP contributed to the study design and the revision of the results. MV designed the study, wrote the GT data generation software, critically analysed the results, and drafted the manuscript. All authors critically reviewed the manuscript and gave final approval for publication.

FUNDING

This work was supported by the British Heart Foundation (RE/18/4/34215, RG/16/3/32175, PG/16/17/32069, and Centre of Research Excellence), the Imperial-TUM Seed Funding, the National Institute for Health Research (UK) Biomedical Research Centre and the Rosetrees Trust through the interdisciplinary award Atrial Fibrillation: A Major Clinical Challenge and the Generalitat Valenciana Conselleria d'Educació, Investigació, Cultura i Esport (ACIF/2021/205).

ACKNOWLEDGMENTS

We thank the Imperial College Research Computing Service (DOI: 10.14469/hpc/2232) for help with the computational setup of the project and technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2021.768419/full#supplementary-material

REFERENCES

1. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020. ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). *Eur Heart J*. (2020). 42:373–498. doi: 10.1093/eurheartj/ehaa945
2. Ganesan AN, Shipp NJ, Brooks AG, Kuklik P, Lau DH, Lim H, et al. Long-term outcomes of catheter ablation of atrial fibrillation: a systematic review and meta-analysis. *J Am Heart Assoc*. (2013) 2:e004549. doi: 10.1161/JAHA.112.004549
3. Nattel S, Dobrev D. Controversies about atrial fibrillation mechanisms. *Circ Res*. (2017) 120:1396–9. doi: 10.1161/CIRCRESAHA.116.310489
4. Calkins H, Kuck KH, Cappato R, Brugada J, John Camm A, Chen SA, et al. 2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation. *J Intervent Cardiac Electrophysiol*. (2012). 14:171–257.
5. Clayton RH, Bernus O, Cherry E, Dierckx H, Fenton F, Mirabella L, et al. Models of cardiac tissue electrophysiology: progress, challenges and open questions. *Prog Biophys Mol Biol*. (2011) 104:22–48. doi: 10.1016/j.pbiomolbio.2010.05.008
6. Bueno-Orovio A, Kay D, Grau V, Rodriguez B, Burrage K. Fractional diffusion models of cardiac electrical propagation: role of structural heterogeneity in dispersion of repolarization. *J R Soc Interface*. (2014) 11:20140352. doi: 10.1098/rsif.2014.0352
7. Aliev RR, Panfilov A. A simple two-variable model of cardiac excitation. *Chaos Solitons Fractals*. (1996) 7:293–301. doi: 10.1016/0960-0779(95)00089-5
8. Varela M, Colman M, Hancox J, Aslanidi O. atrial heterogeneity generates re-entrant substrate during atrial fibrillation and anti-arrhythmic drug action: mechanistic insights from canine atrial models. *PLoS Comput Biol*. (2016) 12:e1005245. doi: 10.1371/journal.pcbi.1005245
9. Fenton F, Karma A. Vortex dynamics in three-dimensional continuous myocardium with fiber rotation: filament instability and fibrillation. *Chaos*. (1998) 8:879. doi: 10.1063/1.166311
10. Roy A, Varela M, Aslanidi O. Image-Based computational evaluation of the effects of atrial wall thickness and fibrosis on re-entrant drivers for atrial fibrillation. *Front Physiol*. (2018) 9:1352. doi: 10.3389/fphys.2018.01352
11. Nelles O. *Nonlinear System Identification From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes. 2nd ed*. Cham: Springer International Publishing (2020).
12. Hoffman MJ, Cherry EM. Sensitivity of a data-assimilation system for reconstructing three-dimensional cardiac electrical dynamics. *Philos Trans R SocA*. (2020) 378:20190388. doi: 10.1098/rsta.2019.0388
13. Drovandi CC, Cusimano N, Psaltis S, Lawson BAJ, Pettitt AN, Burrage P, et al. Sampling methods for exploring between-subject variability in cardiac electrophysiology experiments. *J R Soc Interface*. (2016) 13:20160214. doi: 10.1098/rsif.2016.0214
14. Dhamala J, Arevalo HJ, Sapp J, Horácek BM, Wu KC, Trayanova NA, et al. Quantifying the uncertainty in model parameters using Gaussian process-based Markov chain Monte Carlo in cardiac electrophysiology. *Med Image Anal*. (2018) 48:43–57. doi: 10.1016/j.media.2018.05.007
15. Fresca S, Manzoni A, Dedè L, Quarteroni A. Deep learning-based reduced order models in cardiac electrophysiology. *PLoS ONE*. (2020) 15:e0239416. doi: 10.1371/journal.pone.0239416
16. Sahli Costabal F, Perdikaris P, Kuhl E, Hurtado DE. Multi-fidelity classification using Gaussian processes: Accelerating the prediction of large-scale computational models. *Comput Methods Appl Mech Eng*. (2019) 357:112602. doi: 10.1016/j.cma.2019.112602
17. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys*. (2019) 378:686–707. doi: 10.1016/j.jcp.2018.10.045
18. Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. *Nat Rev Phys*. (2021) 3:422–40. doi: 10.1038/s42254-021-00314-5
19. Baydin AG, Pearlmutter BA, Radul AA, Siskind JM. Automatic differentiation in machine learning: a survey. *J Mach Learn Res*. (2018) 18:1–43. Available online at: https://w.jmlr.org/papers/volume18/17-468/17-468.pdf
20. van Herten RLM, Chiribiri A, Breeuwer M, Veta M, Scannell CM. Physics-informed neural networks for myocardial perfusion MRI quantification. (2020).
21. Cai S, Li H, Zheng F, Kong F, Dao M, Karniadakis GE, et al. Artificial intelligence velocimetry and microaneurysm-on-a-chip for three-dimensional analysis of blood flow in physiology and disease. *Proc Natl Acad Sci USA*. (2021) 118:e2100697118. doi: 10.1073/pnas.2100697118
22. Sahli Costabal F, Yang Y, Perdikaris P, Hurtado DE, Kuhl E. Physics-informed neural networks for cardiac activation mapping. *Front Phys*. (2020) 8:42. doi: 10.3389/fphy.2020.00042
23. Grandits T, Pezzuto S, Costabal FS, Perdikaris P, Pock T, Plank G, et al. Learning atrial fiber orientations and conductivity tensors from intracardiac maps using physics-informed neural networks. *Lecture Notes Compu Sci*. (2021) 12738:650–8. doi: 10.1007/978-3-030-78710-3_62
24. Grandits T, Pezzuto S, Lubrecht JM, Pock T, Plank G, Krause R. PIEMAP: personalized inverse eikonal model from cardiac electro-anatomical maps. *Lecture Notes Comput Sci*. (2020) 12592:76–86. doi: 10.1007/978-3-030-68107-4_8
25. Efimov IR, Nikolski VP, Salama G. Optical imaging of the heart. *Circ Res*. (2004) 95:21–33. doi: 10.1161/01.RES.0000130529.18016.35
26. Hansen B, Zhao J, Li N, Zolotarev A, Zakharkin S, Wang Y, et al. Human atrial fibrillation drivers resolved with integrated functional and structural imaging to benefit clinical mapping. *JACC: Clin Electrophysiol*. (2018) 4:1501–15. doi: 10.1016/j.jacep.2018.08.024
27. Chowdhury RA, Tzortzis KN, Dupont E, Selvadurai S, Perbellini F, Cantwell CD, et al. Concurrent micro-to macro-cardiac electrophysiology in myocyte cultures and human heart slices. *Sci Rep*. (2018) 8:1–13. doi: 10.1038/s41598-018-25170-9
28. Lu L, Meng X, Mao Z, Karniadakis GE. DeepXDE: a deep learning library for solving differential equations. *SIAM Rev*. (2021) 63:208–28. doi: 10.1137/19M1274067
29. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: *3rd International Conference on Learning Representations*. San Diego, CA: ICLR. (2015).
30. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. *Math Programm*. (1989) 45:503–28. doi: 10.1007/BF01589116
31. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Teh YW, Titterington M, editors. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Sardinia: PMLR (2010). p. 246–56. Available Online at: http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf
32. Odening KE, Gomez AM, Dobrev D, Fabritz L, Heinzel FR, Mangoni ME, et al. ESC working group on cardiac cellular electrophysiology position paper: relevance, opportunities, and limitations of experimental models for cardiac electrophysiology research. *Europace*. (2021) 23:1795–814. doi: 10.1093/europace/euab142
33. Yazdani A, Lu L, Raissi M, Karniadakis GE. Systems biology informed deep learning for inferring parameters and hidden dynamics. *PLoS Comput Biol*. (2020) 16:e1007575. doi: 10.1371/journal.pcbi.1007575

34. Clayton RH, Aboelkassem Y, Cantwell CD, Corrado C, Delhaas T, Huberts W, et al. An audit of uncertainty in multi-scale cardiac electrophysiology models. *PhilosTrans R Soc A*. (2020) 378: 20190335. doi: 10.1098/rsta.2019.0335
35. Cantwell CD, Mohamied Y, Tzortzis KN, Garasto S, Houston C, Chowdhury RA, et al. Rethinking multiscale cardiac electrophysiology with machine learning and predictive modelling. *Comput Biol Med*. (2019) 104:339–51. doi: 10.1016/j.compbiomed.2018.10.015
36. Kashtanova V, Ayed I, Cedilnik N, Gallinari P, Sermesant M. EP-Net 2.0: Out-of-domain generalisation for deep learning models of cardiac electrophysiology. In: Ennis DB, Perotti LE, Wang VY, editors. *Functional Imaging and Modeling of the Heart*. Cham: Springer International Publishing (2021). p. 482–92. doi: 10.1007/978-3-030-78710-3_46
37. Roy A, Varela M, Chubb H, MacLeod R, Hancox JC, Schaeffter T, et al. Identifying locations of re-entrant drivers from patient-specific distribution of fibrosis in the left atrium. *PLoS Comput Biol*. (2020) 16:e1008086. doi: 10.1371/journal.pcbi.1008086
38. Varela M, Aslanidi O. Role of atrial tissue substrate and electrical activation pattern in fractionation of atrial electrograms: a computational study. *IEEE Eng Med Biol Soc Ann Conf*. (2014) 2014:1587–90. doi: 10.1109/EMBC.2014.6943907
39. Plonsey R, Barr R. *Bioelectricity: a Quantitative Approach*. 2nd ed. Berlin: Springer Science Business Media (2007).
40. Arthurs CJ, King AP. Active training of physics-informed neural networks to aggregate and interpolate parametric solutions to the Navier-Stokes equations. *J Comput Phys*. (2021) 438:110364. doi: 10.1016/j.jcp.2021.110364

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 10 March 2022
doi: 10.3389/fcvm.2022.829512

A Collaborative Approach for the Development and Application of Machine Learning Solutions for CMR-Based Cardiac Disease Classification

Markus Huellebrand[1,2*†], *Matthias Ivantsits*[1†], *Lennart Tautz*[2], *Sebastian Kelle*[1,3,4] *and Anja Hennemuth*[1,2,3,5]

[1] Institute of Cardiovascular Computer-Assisted Medicine, Charité—Universitätsmedizin Berlin, Berlin, Germany, [2] Cardiovascular Research and Development, Fraunhofer MEVIS, Bremen, Germany, [3] German Centre for Cardiovascular Research (DZHK), Berlin, Germany, [4] German Heart Center Berlin (DHZB), Berlin, Germany, [5] Department of Diagnostic and Interventional Radiology and Nuclear Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

OPEN ACCESS

Edited by:
Paul Leeson,
University of Oxford, United Kingdom

Reviewed by:
Filippo Cademartiri,
Gabriele Monasterio Tuscany Foundation (CNR), Italy
Anca-Rezeda Florian,
Diagnostikum Berlin, Germany
Johan Reiber,
Leiden University, Netherlands
Paul Schoenhagen,
Case Western Reserve University, United States

***Correspondence:**
Markus Huellebrand
markus.huellebrand@charite.de

[†]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: 05 December 2021
Accepted: 07 February 2022
Published: 10 March 2022

Citation:
Huellebrand M, Ivantsits M, Tautz L, Kelle S and Hennemuth A (2022) A Collaborative Approach for the Development and Application of Machine Learning Solutions for CMR-Based Cardiac Disease Classification. Front. Cardiovasc. Med. 9:829512. doi: 10.3389/fcvm.2022.829512

The quality and acceptance of machine learning (ML) approaches in cardiovascular data interpretation depends strongly on model design and training and the interaction with the clinical experts. We hypothesize that a software infrastructure for the training and application of ML models can support the improvement of the model training and provide relevant information for understanding the classification-relevant data features. The presented solution supports an iterative training, evaluation, and exploration of machine-learning-based multimodal data interpretation methods considering cardiac MRI data. Correction, annotation, and exploration of clinical data and interpretation of results are supported through dedicated interactive visual analytics tools. We test the presented concept with two use cases from the ACDC and EMIDEC cardiac MRI image analysis challenges. In both applications, pre-trained 2D U-Nets are used for segmentation, and classifiers are trained for diagnostic tasks using radiomics features of the segmented anatomical structures. The solution was successfully used to identify outliers in automatic segmentation and image acquisition. The targeted curation and addition of expert annotations improved the performance of the machine learning models. Clinical experts were supported in understanding specific anatomical and functional characteristics of the assigned disease classes.

Keywords: visual analytics, co-learning, machine learning, CMR, human in the loop (HITL), cardiovascular phenotyping, artificial intelligence, classification

1. INTRODUCTION

In recent years publications and product developments have shown the potential of artificial intelligence systems in cardiovascular medicine (1–4). Especially data-driven machine learning models can support automatic interpretation of complex spatio-temporal information such as ECG or image data, and the integrated analysis of complementary data from electronic health records, sensor systems, etc. Two factors that are essential for the successful deployment of AI solutions for image-based and multi-modal data interpretation are the model design and training and the interaction with the users (3, 5).

1.1. Integration of Image-Based Information in Multi-Modal Cardiac Disease Classification

Integrating complementary data of different types such as demographic information and laboratory and image data requires complex models that filter the densely sampled image information appropriately. Many approaches for phenotyping or predictive modeling using multi-modal data integrate image information *via* conventional clinical parameters such as the stenosis degree or the ejection fraction (6, 7). Thereby valuable feature information of contained in the comprehensive image data might be neglected. In contrast to the traditional features, which describe the heart chamber volumes and myocardial motion patterns of the left and right ventricle, so-called radiomics features describe shape and texture properties of segmented regions context-independently based on image intensities and voxel classification (8). Radiomics features extracted from non-contrast cine MRI have successfully been used to differentiate between patients with myocardial infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and an abnormal right ventricle (RV) (9–12). Further approaches used features describing the myocardial texture in MRI-based T1 and T2 maps (13) or delayed enhancement MRI (14) to differentiate myocardial pathologies.

Standard radiomics features, which can be calculated with freely available libraries such as *pyradiomics* were designed for the assessment of compact structures such as tumors (15). To better consider the complex structure of the heart, further features have been suggested. The Minkowski-Bouligand dimensions, as described by Captur et al. (16) assesses how the length or complexity of a contour increases while increasing the scale or detail at which it is measured and is used to assess the trabecularization of myocardium. Further features specifically describing the cardiac anatomy such as the septum thickness have been suggested by Tautz et al. (17, 18).

1.2. Expert Annotations for Cardiac Image Interpretation

Quantitative and radiomics analysis of cardiac MRI image data usually requires segmentation of the relevant anatomical structures (19). Recent publications demonstrate the potential of deep learning models such as the U-net for the segmentation and interpretation of typical imaging sequences such as short-axis cine MRI (20). However, the performance of these models depends on the quality of the training data, and previous studies showed that the annotation performance of clinical experts is influenced by the annotation framework (21). The "Society for Cardiovascular Magnetic Resonance" (SCMR) recommends analyzing image frames in end-diastole (ED) and end-systole (ES) (19) for assessing the global cardiac function. Therefore, clinical datasets are often only sparsely annotated, and interactive intelligent annotation and correction tools are required to extend and improve the data so that they can be used to train machine learning models. Commercial medical products might be used if the software offers the export of the expert segmentation in an open format. Open-source application such as 3D Slicer (22), MITK Workbench (23) also provide generic tools for interactive (24, 25), semi- and fully-automated segmentation algorithms. These software tools integrate open-source libraries such as *MONAI Label*[1] to support an efficient interaction between the annotation and machine learning environment (26). Specialized research software tools such as *Segment* (27) and *CAIPI* (28) provide dedicated solutions for the annotation and processing of four-dimensional cardiac data, which can be used to generate training data. The International Radiomics Platform (IRP) (29) supported by the German Radiological Society[2] further enables the combination of annotated image data with clinical data and questionnaires.

1.3. Clinical Integration of AI-Based Solutions for Cardiac Image Interpretation

Modern deep learning models can classify several cardiac diseases directly from image data (30, 31), but the inference process is hardly understandable for most clinical experts. Explainability approaches for convolutional neural networks support the identification of image regions, which contribute to classification results (32) and provide information for plausibility checks as demonstrated for the interpretation of echocardiograms (33). Explainability methods have been suggested for enhancing the classification of cardiac diseases. Interpretability methods such as *Discovering and Testing with Concept Activation Vectors (D-TCAV)* can be used to show underlying features of the classification (34). Especially in cardiovascular research, it can be highly beneficial for hypothesis generation to understand the shape and tissue characteristics, which determine the assignment of a patient to a particular class. Working with well-defined features, as suggested in Radiomics (8), might enable a compromise between the optimal consideration of the complex image information and a classification that is understandable for clinical experts (35). However, the complex multi-modal data used in phenotyping are difficult to interpret for humans with classical approaches such as heatmaps and two-dimensional diagrams (36, 37). When omics or image data is involved there is a lack of backtracking within these tools, which links the classification to specific relevant locations or time frames of the underlying data.

Integrating AI training setups into clinical environments faces several ethical and legal challenges. The management of health record data is defined by the General Data Protection Regulation (GDPR)[3]. These regulations define how and for what purpose health data can be accessed. Platforms for federated AI training such as JIP (38) and QuantMed (39) provide interfaces for loading data from the Picture archiving and communication system (PACS) and sharing fully trained models in a secure and compliant way. Moreover, JIP implements an interface to connect open-source deep learning libraries and permits the integration

[1]Medical Open Network Artificial Intelligence. Available online at: https://monai.io

[2]https://www.drg.de/de-DE/3601/radiomics/

[3]https://www.eu-patient.eu/globalassets/policy/data-protection/data-protection-guide-for-patients-organisations.pdf

of other image processing frameworks like MITK or platforms like IRP (29).

The FDA guidelines address the problem of the need for AI model adaptation and improvement through retraining, and suggest an efficient dynamic process for the development and quality assurance of DL/ML methods in medical image processing (40). The document describes how to manage data, re-train, evaluate, and update AI methods in clinical settings in such a way that newly trained models fulfill the regulatory requirements for a medical product.

1.4. Goals

We hypothesize that a dedicated setup for the training and application of machine learning methods with an expert-in-the-loop approach can speed up and improve the training of the AI models for image processing and multi-modal classification. Furthermore, it can support the clinical expert in exploring and understanding the analyzed datasets.

The existing infrastructure and tool solutions presented in the previous paragraphs address all aspects required to set up an environment that supports the development and application of machine learning methods for the integrated usage of cardiac MRI data in multi-modal data classification. Based on these building blocks, we present a concept for a central environment that supports dynamic machine learning with experts in the loop. This central infrastructure should manage data and the training and inference of machine learning models for multi-modal cardiac data interpretation. We envision central modules for cardiac structures segmentation, an automated pre-processing and features extraction process, and a multi-modal cardiovascular disease classification. To integrate clinical experts into the loop we suggest an interactive exploration of the extracted data and a disease hypotheses generation method. Furthermore, this module should provide an interactive data correction and data integrity check, as well as dynamic updates of machine learning models. We test the presented setup with two use-cases and publicly available data from MICCAI challenges on image-based disease classification: the ACDC challenge for the classification of cardiomyopathies (41) and the EMIDEC challenge for the detection pathologies (49) using cardiac MRI and non-image information.

2. MATERIALS AND METHODS

We propose a modular web-based software environment to support co-learning and comprehensive analysis of cardiovascular imaging data (**Figure 1**). The architecture of our solution contains the following main components: a data model; a semi-automated tool for efficient labeling; extraction of cardiovascular and radiomics features; visual analytics interface.

For integration into the clinical infrastructure, DICOM network services (42) are used to receive imaging data from PACS systems. On arrival of new data, automated processes import, classify, and, depending on the type of data, automatically pre-process, segment datasets, and extract radiomics features. A web-based application is provided for semi-automated segmentation correction as described in Section Data Correction, Data Integrity, and Dynamic Updates of Machine Learning Models. Cases that the users correct can be directly used to improve the segmentation algorithm by re-training. **Figure 1** shows the workflow for refining the segmentation and classification solution. Study data can be analyzed in a web-based visual analytics application Section Interactive Multi-modal Data Exploration with Visual Analytics.

2.1. Data Model

The data model is essential for the traceability of the origin of classification results. **Figure 2** shows the major entities and their hierarchical organization. Our data model follows a similar structure described by the DICOM standard (42) using patients and studies as entities to describe a patient cohort. Each patient entity can have one or more studies. Each study can contain several cases containing one or more 3- or 4-dimensional images. For each case, deformation fields, clinical parameters,

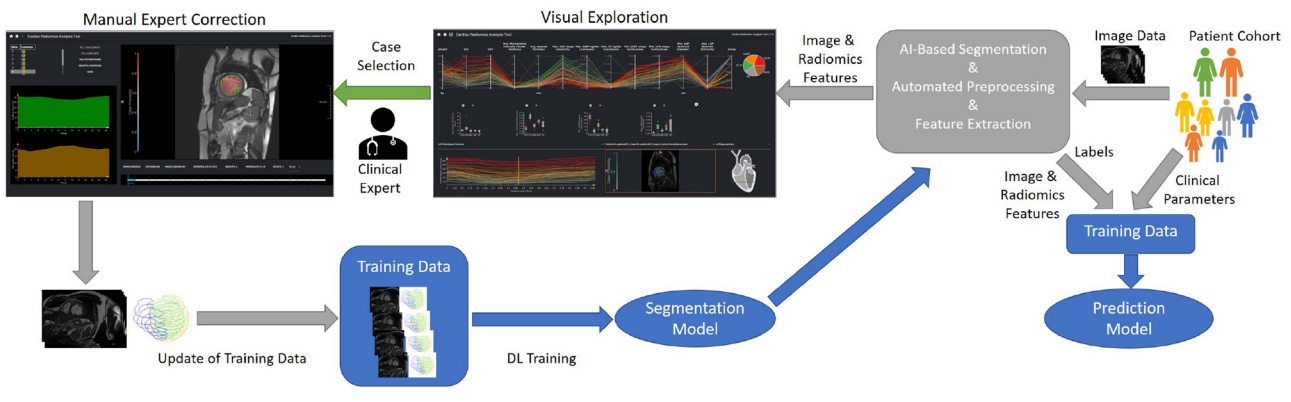


FIGURE 1 | Concept of an iterative process for the training and evaluation of the ML-based segmentation and classification models. On import, image data is automatically segmented, pre-processed, and features are extracted. The results of the automatic segmentation and classification is displayed in the visual exploration interface. In addition clinical experts can manually correct the segmentation results of detected outliers. These corrections can be used to refine the segmentation model.

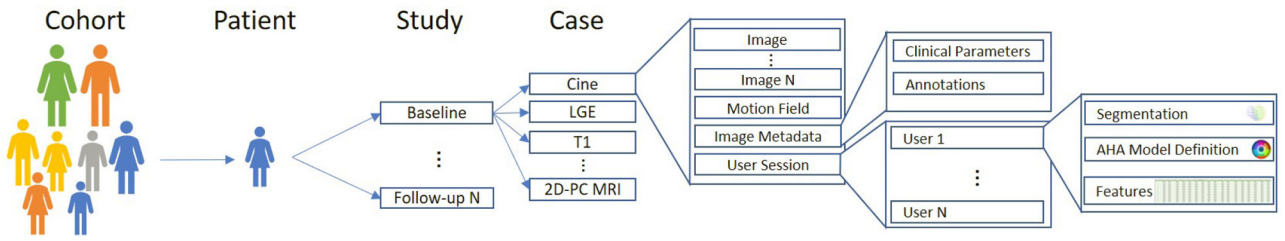


FIGURE 2 | Schematic representation of the data model.

and annotations, such as image type, or classification labels are stored. Automatic segmentation results and landmarks for the definition the 17-segment model defined by the American Heart Association (AHA) are stored in sessions so that it is always traceable on which data the calculated parameters are based. Furthermore, this data handling allows an evaluation of several readers or repeated measurements and thus supports inter- and intra-observer or ML model comparisons.

2.2. Machine Learning for Multi-Modal Disease Classification

2.2.1. Automatic Segmentation, Pre-processing, and Feature Extraction

The image processing part of the application is developed with MeVisLab (43). For data pre-processing and the training of models for the slice-wise segmentation of cardiac structures, we use the Redleaf framework, which allows the integration of inference methods directly in the MeVisLab based applications. U-nets are trained for the segmentation of the relevant structures such as RV, LV, and myocardium (44). These segmentations form the basis for the extraction of typical radiomics features and image-based cardiac biomarkers as suggested in Section Introduction. For 4D image data we generate radiomics feature curves that provide dynamic changes and motion patterns. These time-resolved features are aggregated using minimum, maximum, median, and (arithmetic) mean.

2.2.2. Multi-Modal Cardiovascular Disease Classification

The classifications are based on features describing local and global cardiac function, and radiomics features. We apply eight-fold nested cross-validation to select relevant feature classes. Moreover, during the cross-validation, we perform a model selection of five classifiers and their respective hyperparameters and train a classifier as described in Ivantsits et al. (45).

We perform a feature importance analysis as (46) proposed. This analysis can be performed on any fitted model by calculating a base score produced by the training or test set model. This is followed by a random shuffle to one of the features and compared to the baseline's predictive power. This procedure is then repeated and applied to all features to come up with an importance score. This importance analysis gives insights into the decision made by a classifier and can further be used to discover potential data integrity issues.

2.3. Interactive Multi-Modal Data Exploration With Visual Analytics

The interactive visualization is designed to support the evaluation, validation, and hypothesis generation. It is provided as a web-based tool for the clinical experts (**Figure 3**).

The interface provides an overview of relevant features for a given patient cohort. These features are identified by the feature importance analysis of the machine learning module. In order to be sure that features such as, e.g., gender are always considered, users can also select features to be included in the exploration view. The parallel coordinate plots (PCP) visualize the multi-dimensional data as line sets with points representing the datasets' parameters. Each y-axis represents the relevant value range of one parameter. Each line corresponds to one patient dataset. Time-resolved parameters are represented by the aggregated minimum, maximum, or mean values.

Further chart types support an advanced exploration of relationships between different parameters. Scatter plots with regression lines visualize linear relationships between parameters. Histograms show the distribution of different parameters. Box plots give a standardized overview of the data set. Pie charts visualize how frequently individual values or cases of the disease class occur in the study. This can also be used to visually identify unbalanced data sets, for which appropriate measures can be taken in the case of subsequent training.

The exploration tool is designed as a hierarchical tool with different interlinked views. The linking of the data is based on the data model described in Section Data Model. Cases can be selected interactively in the PCP by a technique called brushing (47). The selected parameter range specifies the subset of patients considered for the dependent diagram and curve views. Images can be selected in the 2D viewer from this subset by a drop-down menu above the viewer. The line corresponding to the image selected in the 2D viewer is highlighted in the PCP and the curve diagram by changing thickness and alpha value.

A curve diagram enables the analysis of temporal dynamics of individual parameters. As shown in **Figure 3**, the curve color corresponds to the class assigned to the underlying dataset to enable a comparison of feature dynamics.

The image viewer shows the image data of a selected case with segmentation contours and an overlay of the segmentation uncertainty. Schematic visualization of the heart shows the position of the displayed image slice with regard to the cardiac anatomy. This approach supports the identification and

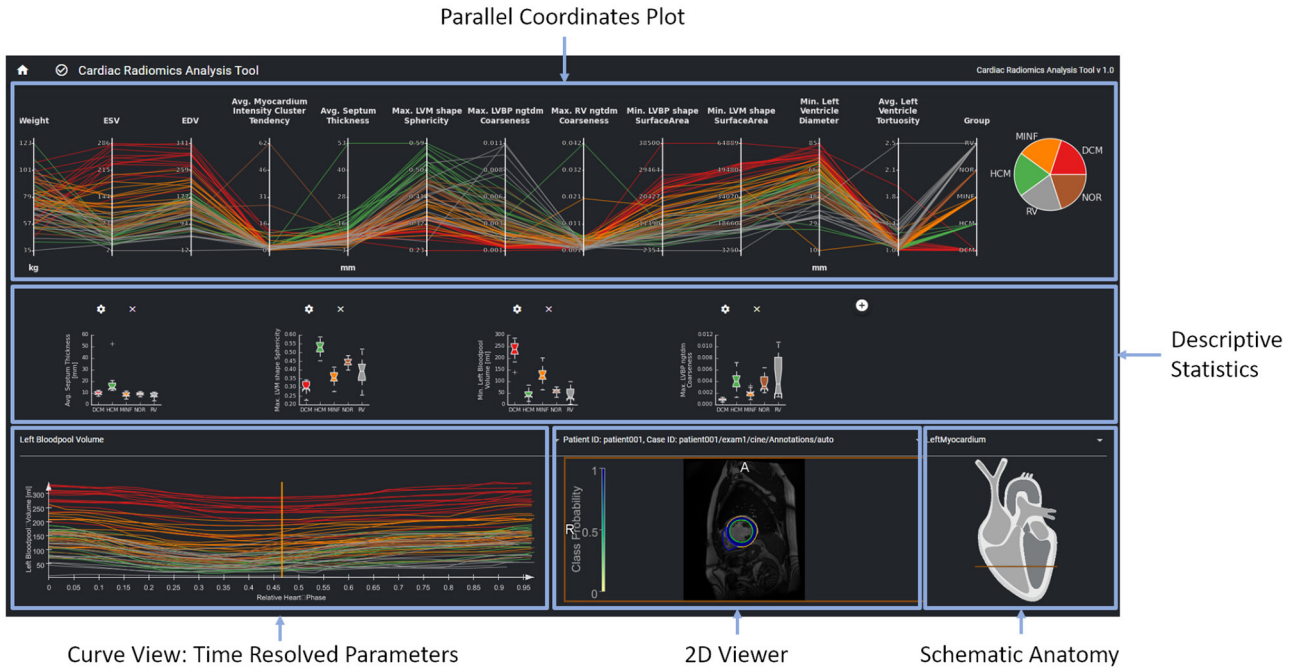


FIGURE 3 | The parallel coordinates plot in the top row shows multi-dimensional data as line sets. Each y-axis represents a specific parameter. The diagrams in the center row show distributions and correlations of selected parameters. The curve view on the bottom left supports the exploration of the temporal dynamics of a selected parameter over the cardiac cycle The image data, the curve parameter of the selected timepoint is based on is displayed in the image viewer on the bottom right. Contours and overlay show the segmentation as well as the local certainty of the segmentation model.

exploration of outliers as shown in **Figure 4** and thereby the data curation. Furthermore, *via* the aggregation of time-resolved data in the PCP the user can backtrack information from a specific feature to the relevant regions within the underlying image. This can further be used to identify any data integrity issues.

2.3.1. Data Correction, Data Integrity, and Dynamic Updates of Machine Learning Models

A second tool supports the correction and extension of ML-based segmentation results (**Figure 5**). The image viewer allows to delineate and correct contours defining the anatomical structures (left and right endocardium, left epicardium) using spline, freehand, and brush tools. The overview table indicates the segmentation status of the image slices. When analyzing time-resolved data, such as cine MRI, area and volume as segmentation certainty curve diagrams support the identification of mis-segmented timeframes to reduce the manual interactions. Sparse corrections by the users can be transferred in 3D using shape-based interpolation (25). To transfer segmentation results motion-compensated onto adjacent time frames, we use the deformation field generated by a Morphon-based method (48). In the **timepoint widget** users can specify which timepoints to consider (**Figure 6**). To help the users to generate consistent segmentation results, an optional tool can enforce that the LV epicardial contour encloses the LV endocardial contour, and that left epicardial and right endocardial contour do not intersect, using spatial set-theoretical boolean operations (**Figure 6C**). For each individual contour, we store whether it was manually corrected. Thereby, the quality of the automatic segmentation algorithms can be assessed. This information can also identify new cases to improve the AI-based segmentation approach *via* fine-tuning or re-training.

3. RESULTS/APPLICATION

We test the presented setup with two use-cases and publicly available data from MICCAI challenges on image-based disease classification: the ACDC challenge for the classification of cardiomyopathies (41) and the EMIDEC challenge for the detection of myocardial pathologies (49) using cardiac MRI and non-image information.

3.1. Classification of Cardiovascular Disease Based on Cardiac Cine MRI

Cardiac cine MRI provides information on the anatomy and the function of the heart and can help to differentiate between cardiovascular diseases. In this study, we use the freely available dataset from the ACDC challenge (41) to demonstrate how our software environment can be used to support experts in improvement and understanding of cardiomyopathy classification. The dataset comprises normal subjects and patients with one of the following cardiovascular diseases (CVD): previous myocardial infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM),

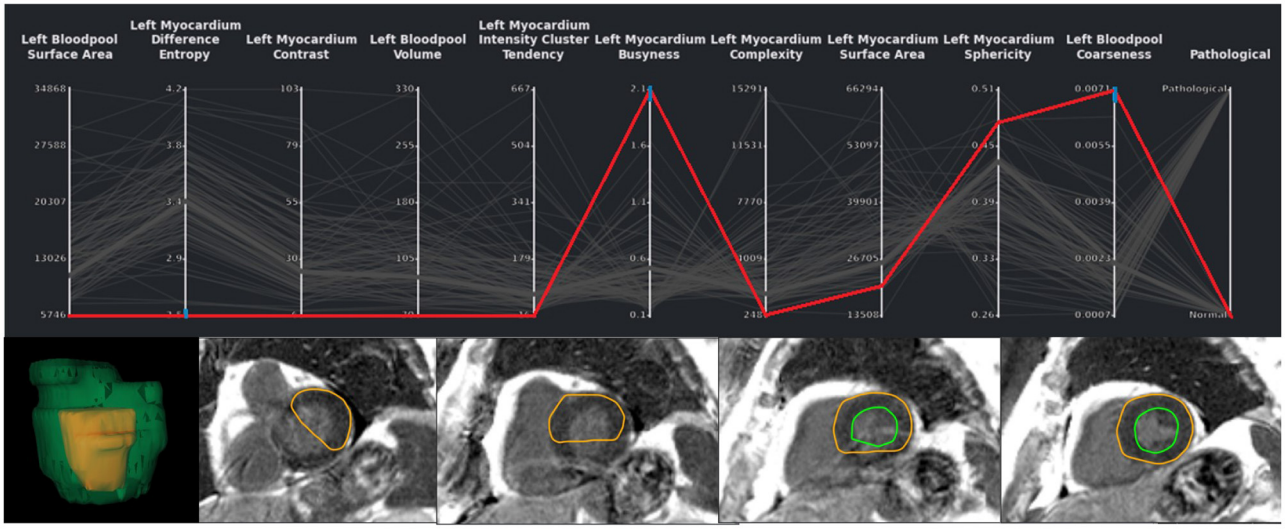


FIGURE 4 | Exemplary case of a segmentation outlier (represented by the red line) in the total EMIDEC study consisting of 100 cases. It can be identified in the PCP by three blue brushes. In the image viewer the erroneous segmentation is clearly visible.

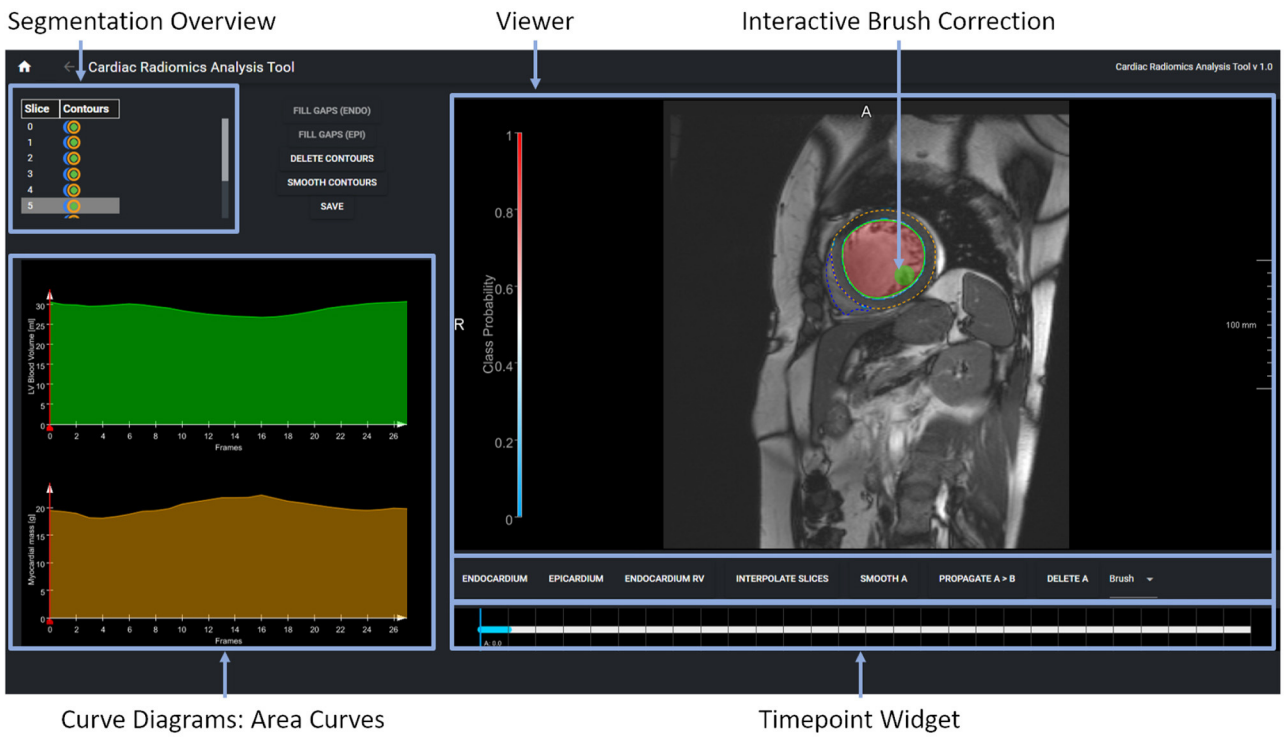


FIGURE 5 | The expert segmentation correction tool shows the segmentation results as contour overlay in the image viewer. The brush tool is visualized as a circle. The table in the upper displays the segmentation status per image slice and anatomical structure. The curve diagrams on the bottom left show the area curves and the segmentation algorithms' probability of the segmentation results on the current slice.

and an abnormal right ventricle (RV). The dataset contains the same number of cases in each subgroup. Clinical experts delineated the left epicardial and endocardial borders in end-systole and end-diastole and assigned the CVD class. The data was acquired on Siemens MRI scanners on 1.5T (Aera) and 3T (Trio Trim); the in-plane resolution was between 1.37×1.37 and $1.68 \times 1.68 mm^2$, the slice thickness was between 5 and 8 mm, distance between slices was 5–10 mm, and 28–40 phases

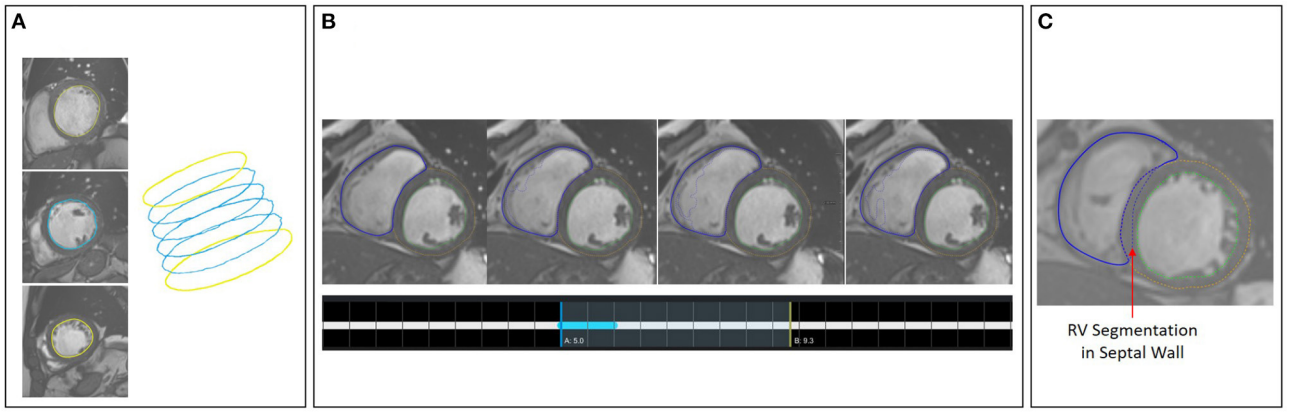


FIGURE 6 | (A) Shape-based interpolation: manual contours (yellow) are defined on two slices. The blue contours represent the interpolated results. **(B)** Temporal correction: top row shows a series of time frames of a cine MRI dataset. The solid contour shows the manually corrected contours. The dotted contours represent the incorrect segmentation results. The time-point widgets show the range the user selected for correction *via* contour propagation. **(C)** Correction with logical conditions: the contour of the right endocardium was originally drawn into the septal wall (blue dotted contour). The solid contour represents the corrected segmentation.

covered each cardiac cycle. In this study a pre-trained 2D U-Net was retrained on 75 cases from the challenge' training set. The optimal classifier obtained by the described grid search on the ACDC challenge dataset yields a random forest classifier with 230 estimators, the Gini criterion, a maximum depth of six, a minimum of samples per leaf of six, and a minimum sample split of nine. This classifier is built with 112 shape- and texture-based features plus the patients height and weight. For the random forest classifier 80 cases were used for training and 20 cases for testing. **Figure 10A** shows the confusion matrix of this classifier before the correction. This results in an overall accuracy of 0.85, with a precision of 1.0 and recall of 0.75 on RV cases, a precision of 0.75 and recall of 0.75 on normal cases, a precision of 0.67 and recall of 1.0 on MINF cases, a precision of 1.0 and recall of 1.0 on HCM cases, and a precision of 1.0 and recall of 0.75 on DCM cases. Additionally, **Figures 10C,D** illustrate the AUC scores for each individual class plus the macro AUC score of 0.94 before the correction and an AUC score of 0.98 after the correction. After correcting segmentation outliers that were identified *via* the PCP, the accuracy of the classifier improved from 0.85 to 0.9 (**Figure 10B**). Furthermore, **Figure 10E** exemplifies the feature importance of the random forest classifier. We observe the patients' **left myocardium sphericity** to be the most crucial variable in detecting pathological cases, closely followed by the **left blood pool volume** and the **interventricular septum thickness** parameter.

Figure 7 shows the visual exploration interface for the complete ACDC dataset. Cases of all patients are shown in the PCP. The rightmost y-axis shows the patients' CVD, which is also visualized in the pie chart on the top right. One can see an equal distribution of the diseases in the dataset. Multiple clusters and outliers can be observed. As a first step, clear outliers were removed by deselecting outliers by the average **myocardium intensity cluster tendency**, average **septum thickness**, and average **left ventricle tortuosity**. After removing these outliers, one can differentiate between HCM and DCM patients based on the left blood pool coarseness and relative septum thickness. This can also be seen in the corresponding box plots in the second row. When only selecting patients with HCM and DCM, this becomes even more prominent as shown in **Figure 8**.

While outliers were excluded in this first analysis of the study cohort, it is also possible to select and analyze these individual outlier cases. **Figure 9** shows cases with strong motion artifacts that could be detected by the left blood pool surface area parameter. The crosshair in the images in the second row shows the center of the left blood pool in the basal slice. The misalignments in the slices can also be depicted in the 3D visualization of the segmentation.

In outlier cases where the segmentation had failed, it was manually corrected using our labeling interface. After correcting the outlier cases, parameters were extracted again, and the cases were classified again. **Figure 10B** highlights the classification performance after the correction of the cases as described in Section Interactive Multi-modal Data Exploration with Visual Analytics. We observer an improvement in accuracy from 0.72 to 0.8.

3.2. Classification of Normal and Pathological Cases From Late-Gadolinium Enhanced MRI in the Left Myocardium

The EMIDEC challenge (49) provided benchmarking data to assess the performance of segmentation and classification algorithms using clinical parameters and late gadolinium enhancement (LGE) MRI data (50). The dataset consists of 150 cases: 100 diseased patients and 50 normal cases. Patients were split into 100 training and 50 testing sets, containing 1/3 normal and 2/3 pathological cases, which roughly corresponds to real-life observations in the clinical settings. The data was acquired on Siemens MRI scanners on 1.5T (Aera) and 3T (Skyra); the in-plane resolution was 1.25×1.25 and $2 \times 2mm^2$, the slice thickness was $8mm$ and the distance between slices 8–13 mm. In

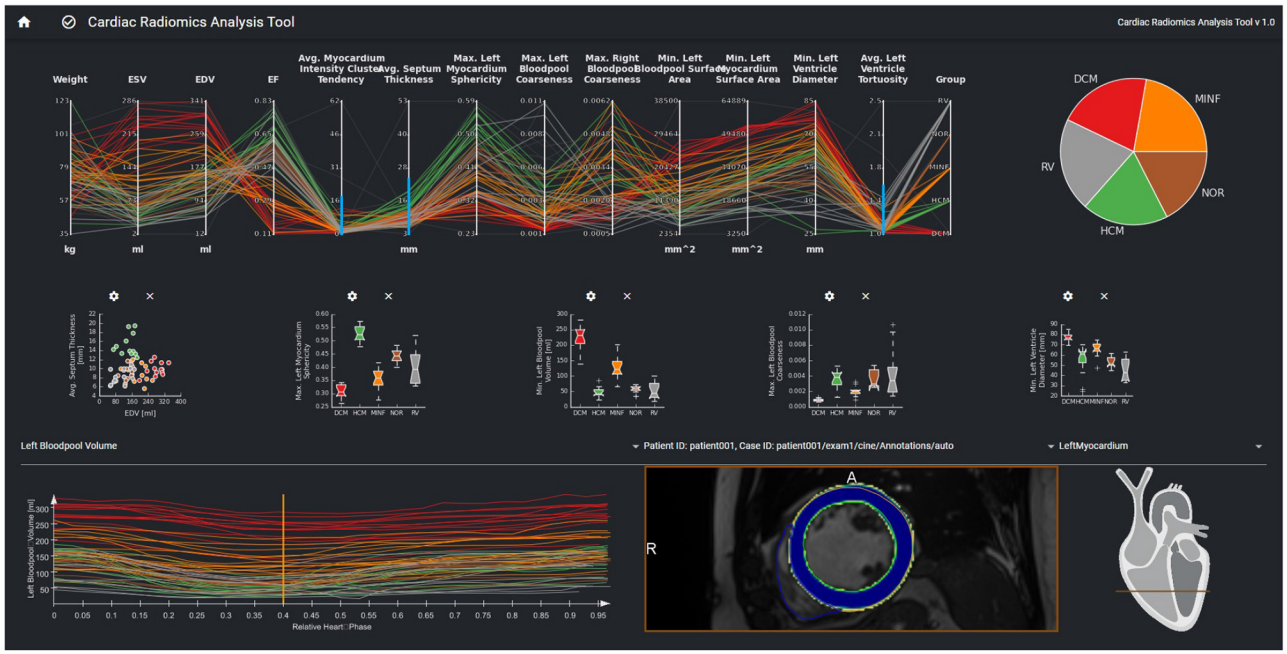


FIGURE 7 | Visual analytics of complete ACDC dataset, which consists of 100 cases. The lines in the PCP represent the patients of the underlying cohort. The color of the lines visualizes the classification result and corresponds to class color in the pie chart. The selection *via* the blue brushes in the PCP includes patients without outliers for features such as septum thickness, myocardium intensity cluster tendency, and left ventricle tortuosity. The scatter plot in the second row shows the EDV vs. avg. myocardium intensity cluster tendency for this cohort. In the box-plots, the distributions of max. myocardial sphericity, max. LV volume and left ventricular diameters in each disease subgroup are shown. The curve diagram shows the LV blood pool curve of each patient over the cardiac cycle.

a post-processing step, the image slices were realigned to prevent any effects of breathing motions.

Analogously to the previous case study, the dataset from the EMIDEC challenge was integrated into our classification and exploration environment. A 2D U-Net was used to generate segmentations of the LV endocardial and epicardial border trained on data from 100 patients with myocarditis and cardiomyopathy to segment the myocardium in the LGE MRI data. As no cases with no-reflow areas were included in the patient data, we also included the 50 unlabeled cases from the challenge's test set. Initial segmentations for these cases were generated by a pre-trained model. An expert used the tools described in Section Data Correction, Data Integrity, and Dynamic Updates of Machine Learning Models to correct these segmentations. Next, the segmentation results were added to the internal cases with expert segmentation, and the final segmentation model was trained. This final model segmented the LV endocardial and epicardial contours and extracted the radiomics parameters on the 80 cases from the challenge's training set. A classifier was trained to differentiate between normal and pathological cases with 25 shape- and texture-based features. We applied a similar strategy as in the ACDC case study. An eight-fold nested CV was used for model and hyperparameters selection. Essential image-based features for the classifiers are shown in the PCP.

The optimal classifier identified by our grid-search on the EMIDEC challenge dataset turned out to be an extra tree classifier with 190 estimators, the Gini criterion, a maximum depth of six, a minimum of samples per leaf of six, and a minimum sample split of nine. **Figure 11A** illustrates the confusion matrix of this classifier before the interactive correction. This results in an overall accuracy of 0.75, with a precision of 0.79 and a recall of 0.85 on pathological cases, a precision of 0.67, and a recall of 0.57 on normal cases. Additionally, **Figures 11C,D** illustrate the AUC scores for each individual class of 0.85 before the correction and an AUC score of 0.87 after the correction. **Figure 11E** represents the feature importance, where the importance is defined by the difference of the models' baseline and the score after a feature permutation. We observe the patients **left blood pool surface area** to be the most crucial variable in detecting pathological cases, closely followed by **left myocardium difference entropy** and **left myocardium contrast** parameter.

Outliers could be disabled *via* brushing on **myocardial contrast, myocardial cluster tendency, myocardial complexity, left blood pool coarseness** in the PCP analogous to the analysis of the ACDC dataset in **Figure 7**. After correcting segmentation outliers that were identified *via* the PCP, the accuracy of the classifier improved from 0.75 to 0.8 (**Figure 11B**).

4. DISCUSSION

To test our hypotheses, we applied the presented software environment to two multi-modal machine learning tasks: the

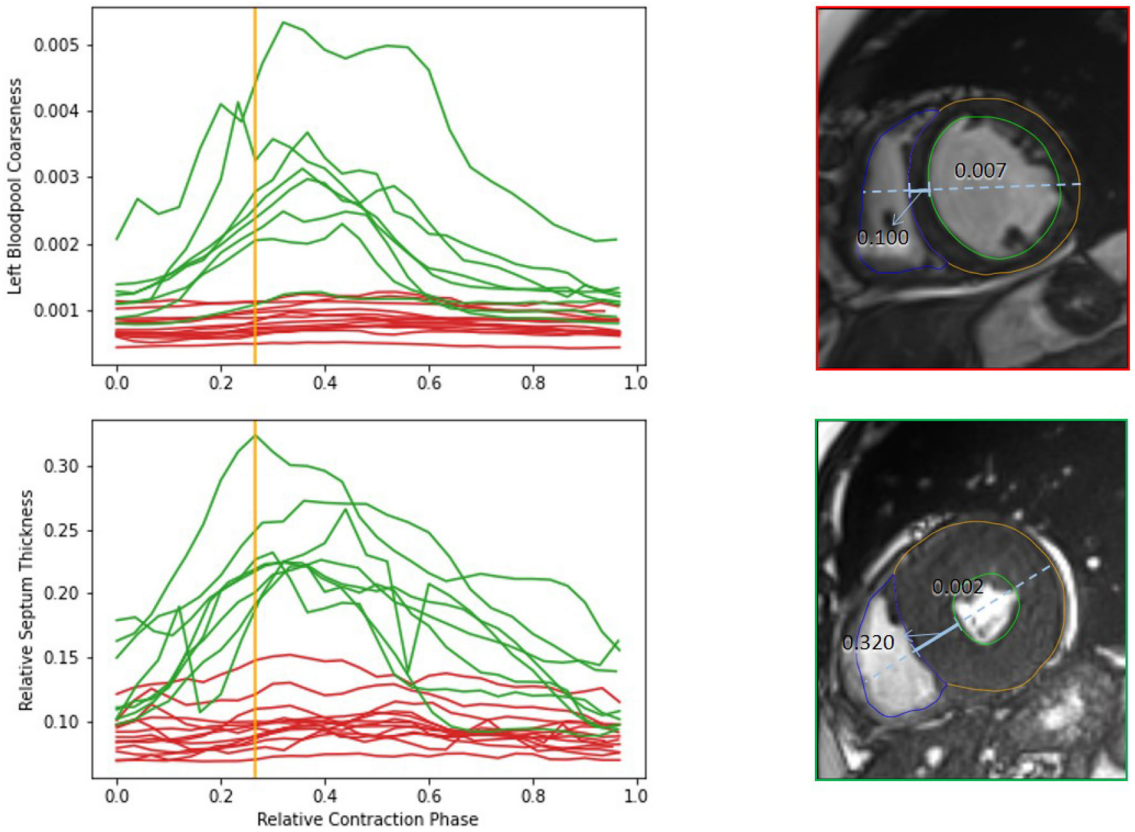


FIGURE 8 | Comparison of features of HCM (green) and DCM (red) patients from the complete dataset. The upper diagram shows dynamics of the left bloodpool coarseness over the cardiac cycle. The lower diagram display the changes of the relative septum thickness. The viewers on the right show the time frames corresponding to the orange line for an example case of a DCM and HCM patient. The solid blue line indicates the septum thickness. The dashed line shows the diameter heart diameter, that is used for normalization of the septum thickness. The coarseness of the selected timepoint is shown on top of the left blood pool. The example cases differ strongly in the anatomical relations as well as in the blood pool intensity distribution.

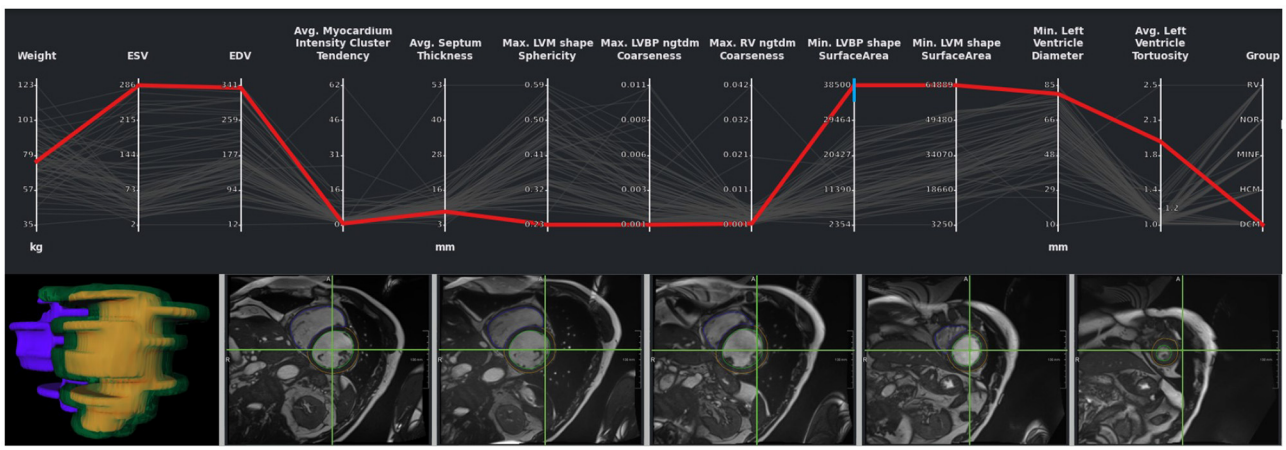


FIGURE 9 | Outlier with motion artifacts. Changes in the left endocardial and epicardial surface area indicate the presence of an outlier. The 3D rendering of the segmentation surface highlights the misaligned slices. The green crosshair in the image viewers was placed in the center of the left blood pool in a basal slice, giving an impression of the motion of the blood pool center.

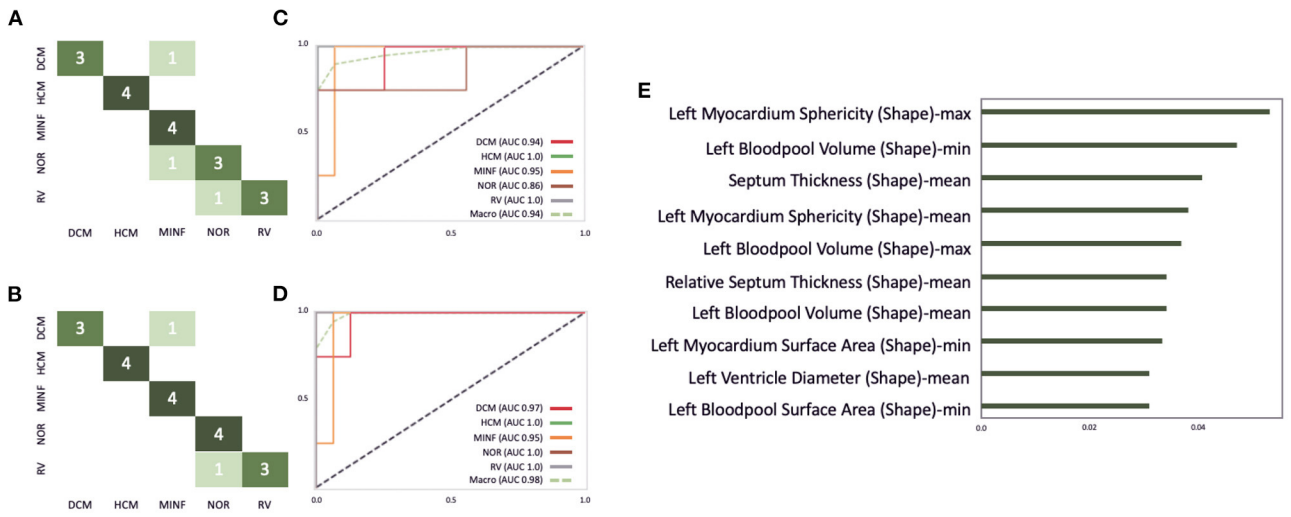


FIGURE 10 | Illustration of the ACDC classifier performance on the test dataset consisting of 20 cases. **(A,B)** Highlights the confusion matrix for the ACDC classifier before and after the correction of image segmentations *via* our described visual outlier detection. Similarly, **(C,D)** show the AUC score improvement of the classifier. **(E)** Depicts the 10 most important features for this classifier.

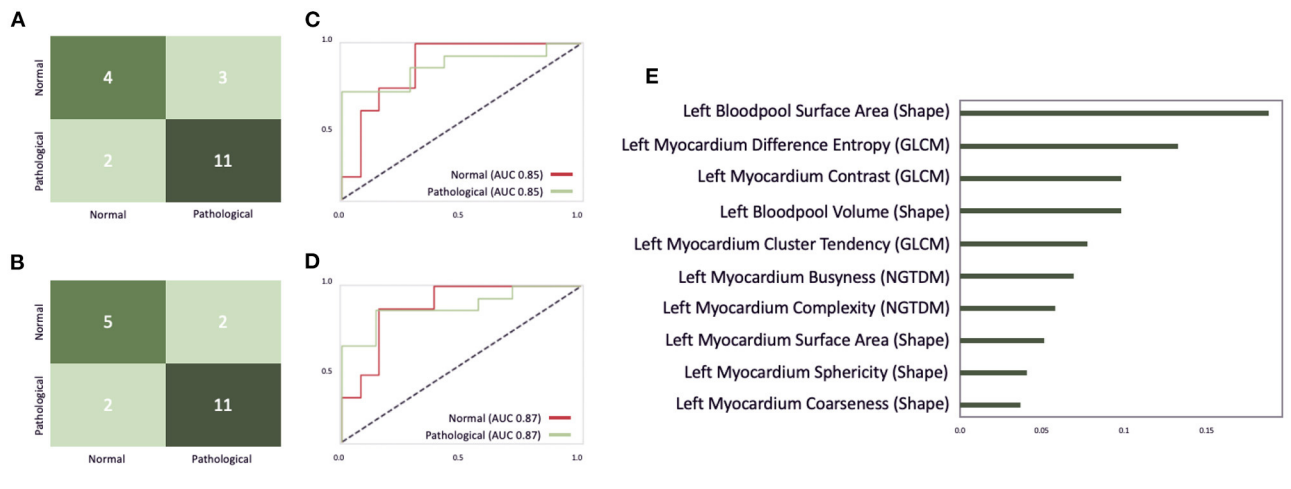


FIGURE 11 | An illustration of the EMIDEC classifier performance on the test dataset consisting of 20 cases. **(A,B)** Show the confusion matrix for the EMIDEC classifier before and after the correction of image segmentations *via* our described visual outlier detection. **(C,D)** Illustrate the AUC score before and after the correction. **(E)** Depicts the importance of the top 10 features for this classifier.

classification of patients with cardiomyopathies considering cardiac cine MRI data and the detection of myocardial pathologies considering late gadolinium MRI data. *Via* our visual analytics tools experts could identify erroneous segmentations as outliers in the PCP as illustrated in **Figure 10**. In both use cases, the performance of the classifier could be improved. The classification accuracy on the test set was improved for pathology detection from 0.75 to 0.8. The accuracy for cardiomyopathies was improved from 0.85 to 0.9. This was achieved by correcting the training dataset and thus the input parameters for the classifier. Consistent with Demirer et al. (21), we found that providing annotation tools with familiar interactions to their routine clinical tools assisted experts with manual corrections. The expert corrections improved the input parameters of single cases, which could then be classified correctly. The suggested visual analytics interface can thus be used to extend approaches, which support retraining models with user-corrected image annotations such as the setup suggested by Dikici et al. (51). Related approaches for the application of visual analytics tools in the exploration of multimodal study data including image information as suggested by Bannach et al. (52) and Angulo et al. (53) strongly focus on the visualization of parameter distributions and have not been applied in a data curation context. However, our solution could also be used for cohort

exploration and enhanced by more context-specific visualizations of the cardiac anatomy as suggested e.g., by Meuschke et al. (54). Furthermore, the classification and visualization environment could be used not only for the interpretation of the patient data, but also for a comparison with characteristic cohorts to support the assessment of the certainty and underlying features of the disease classification.

The visualization of the certainty maps produced by the DL-based segmentation, which were displayed as overlays in the image viewers (see **Figure 7**), did not influence the expert corrections. In future work, we could introduce more sophisticated DL-based outlier detections *via* Bayesian inference as proposed by Gal and Ghahramani (55). This method casts dropout training in DL methods and produces a distribution of outcomes. Unfortunately, this method increases in inference time and can hardly be included in real-time applications. As an additional layer for outlier detection, these certainty maps can be aggregated and displayed in the PCP to give medical experts insights into corrupt results produced by DL models.

The combination of the feature importance analysis and the link to the underlying data model enabled the exploration essential anatomical and functional disease properties in both use cases. **Figure 10C** illustrates that the most significant parameters for the classification of cardiomyopathies were mostly shape-based parameters. Whereas, the classification of myocardial infarctions was a combination of shape- and texture-based features, as highlighted in **Figure 9C**. Furthermore, **Figure 7** shows that the temporal dynamics of features can also be important for the classification and the understanding of disease types.

Our proposed local and specialized cardiovascular software environment could be successfully applied within a clinical software environment and was used collaboratively by three experts. In order to support multi-centric collaborations, it could be integrated into federated learning platforms such as JIP *via* Docker. This leverages the capabilities of our proposed solution to be applied to federated learning environments that are compliant with GDPR suggestions on health records.

The organization of the training setup follows the suggestions by the FDA (40). However, quality assured model development requires a private validation set to detect model degeneration. This could be added for future applications.

4.1. Limitations

The datasets used to test our solution are publicly available and thereby readers can reproduce the described machine learning setup. However, both datasets are relatively small, and the available clinical information is limited. The labels of the EMIDEC dataset are solely based on the inspection of the image data, so that they mean *infarction visible in LGE MRI* and *no infarction visible in LGE MRI* (49). The second label does exclude myocardial pathologies. Therefore the clinical parameters were not included for the interactive optimization of the classifier as described in Section Classification of Normal and Pathological Cases from Late-Gadolinium Enhanced MRI in the Left Myocardium and only integrated for the dataset exploration. Future work with larger datasets will help to further evaluate and improve the presented solution.

5. CONCLUSIONS AND OUTLOOK

We have presented a conceptual design for a software environment that supports the development and application of machine learning methods for multi-modal disease classification using MRI data. We tested the potential of an expert-in-the-loop approach based on visual analysis tools for accelerating algorithm training and for making the learned features understandable with promising results. In future work, we will further quantify the potential of our solution for improving the usage of multi-modal imaging and proteomics data. In addition, we plan to add the monitoring module for an FDA-compliant training setup to offer quality-assured AI solutions. Further clinical studies will have to assess whether an improved disease classification achieved through our setup will have and impact patient outcomes through improved treatment personalization.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: ACDC: https://acdc.creatis.insa-lyon.fr/description/databases.html and EMIDEC: http://emidec.com/dataset.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

AH and SK: funding acquisition. MH, MI, and LT: basic software development and user interface concept. MI and MH: machine learning implementation and validation of machine learning methods. MH, MI, and AH: writing—original draft. All authors helped in conceptualization and writing—review and editing. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was partially funded by the BMBF project Berlin Institute for the Foundations of Learning and Data (Grant Number 01IS18037E) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB-1470 – B06.

REFERENCES

1. Weintraub WS. Role of big data in cardiovascular research. *J Am Heart Assoc*. (2019) 8:e012791. doi: 10.1161/JAHA.119.012791
2. Leopold J, Maron B, Loscalzo J. The application of big data to cardiovascular disease: paths to precision medicine. *J Clin Invest*. (2020) 130:29–38. doi: 10.1172/JCI129203
3. Romiti S, Vinciguerra M, Saade W, Cortajarena I, Greco E. Artificial intelligence (AI) and cardiovascular diseases: an unexpected alliance. *Cardiol Res Pract*. (2020) 2020:4972346. doi: 10.1155/2020/4972346
4. Lopez-Jimenez F, Attia Z, Arruda-Olson A, Carter R, Chareonthaitawee P, Jouni H, et al. Artificial intelligence in cardiology: present and future. *Mayo Clin Proc*. (2020) 95:1015–39. doi: 10.1016/j.mayocp.2020.01.038
5. Sermesant M, Delingette H, Cochet H, Jaïs P, Ayache N. Applications of artificial intelligence in cardiovascular imaging. *Nat Rev Cardiol*. (2021) 18:600–9. doi: 10.1038/s41569-021-00527-2
6. 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*. (2016) 38:500–7. doi: 10.1093/eurheartj/ehw188
7. Banerjee A, Chen S, Fatemifar G, Zeina M, Lumbers RT, Mielke J, et al. Machine learning for subtype definition and risk prediction in heart failure, acute coronary syndromes and atrial fibrillation: systematic review of validity and clinical utility. *BMC Med*. (2021) 19:85. doi: 10.1186/s12916-021-01940-7
8. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169
9. 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*. Lecture Notes in Computer Science. Quebec: Springer International Publishing (2018). p. 82–90. doi: 10.1007/978-3-319-75541-0_9
10. Wolterink JM, Leiner T, Viergever MA, Išgum I. *Automatic Segmentation and Disease Classification Using Cardiac Cine MR Images*. Lecture Notes in Computer Science. Quebec: Springer International Publishing (2018). p. 101–10. doi: 10.1007/978-3-319-75541-0_11
11. 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*. Lecture Notes in Computer Science. Quebec: Springer International Publishing (2018). p. 140–51. doi: 10.1007/978-3-319-75541-0_15
12. 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*. Lecture Notes in Computer Science. Quebec: Springer International Publishing (2018). p. 120–9. doi: 10.1007/978-3-319-75541-0_13
13. Baessler B, Luecke C, Lurz J, Klingel K, von Roeder M, de Waha S, et al. Cardiac MRI texture analysis of T1 and T2 maps in patients with infarctlike acute myocarditis. *Radiology*. (2018) 289:357–65. doi: 10.1148/radiol.2018180411
14. Larroza A, Materka A, López-Lereu M, Monmeneu J, BodíV, Moratal D. Differentiation between acute and chronic myocardial infarction by means of texture analysis of late gadolinium enhancement and cine cardiac magnetic resonance imaging. *Eur J Radiol*. (2017) 92:78–83. doi: 10.1016/j.ejrad.2017.04.024
15. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
16. Captur G, Karperien A, Li C, Zemrak F, Tobon-Gomez C, Gao X, et al. Fractal frontiers in cardiovascular magnetic resonance: towards clinical implementation. *J Cardiovasc Magn Reson*. (2015) 17:80. doi: 10.1186/s12968-015-0179-0
17. Tautz L, Zhang H, Hüllebrand M, Ivantsits M, Kelle S, Kuehne T, et al. Cardiac radiomics: an interactive approach for 4D data exploration. *Curr Direct Biomed Eng*. Hamburg (2020) 6:20200008. doi: 10.1515/cdbme-2020-0008
18. Tautz L, Hüllebrand M, Steinmetz M, Voit D, Frahm J, Hennemuth A. Exploration of interventricular septum motion in multi-cycle cardiac MRI. In: Bruckner S, Hennemuth A, Kainz B, Hotz I, Merhof D, Rieder C, editors. *Eurographics Workshop on Visual Computing for Biology and Medicine*. Hamburg: The Eurographics Association (2017).
19. Schulz-Menger J, Bluemke D, Bremerich J, Flamm S, Fogel M, Friedrich M, et al. Standardized image interpretation and post-processing in cardiovascular magnetic resonance - 2020 update. *J Cardiovasc Magn Reson*. (2020) 22:19. doi: 10.1186/s12968-020-00610-6
20. Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med*. (2020) 7:25. doi: 10.3389/fcvm.2020.00025
21. Demirer M, Candemir S, Bigelow MT, Yu SM, Gupta V, Prevedello LM, et al. A user interface for optimizing radiologist engagement in image data curation for artificial intelligence. (2019) *Radiol Artif Intell*. 1:e180095. doi: 10.1148/ryai.2019180095
22. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging*. (2012) 30:1323–41. doi: 10.1016/j.mri.2012.05.001
23. Wolf I, Vetter M, Wegner I, Nolden M, Böttger T, Hastenteufel M, et al. The medical imaging interaction toolkit (MITK): a toolkit facilitating the creation of interactive software by extending VTK and ITK. In: *Proceedings of SPIE - The International Society for Optical Engineering*, San Diego, CA (2004). doi: 10.1117/12.535112
24. Falcao AX, Udupa JK. Segmentation of 3D objects using live wire. *Proc. SPIE*, Vol. 3034. Newport Beach, CA: SPIE (1997). doi: 10.1117/12.274112
25. Herman GT, Zheng J, Bucholtz CA. Shape-based interpolation. *IEEE Comput Graph Appl*. (1992) 12:69–79. doi: 10.1109/38.135915
26. Diaz O, Kushibar K, Osuala R, Linardos A, Garrucho L, Igual L, et al. Data preparation for artificial intelligence in medical imaging: a comprehensive guide to open-access platforms and tools. *Phys Med*. (2021) 83:25–37. doi: 10.1016/j.ejmp.2021.02.007
27. Heiberg E, Sjögren J, Ugander M, Carlsson M, Engblom H, Arheden H. Design and validation of Segment - freely available software for cardiovascular image analysis. *BMC Med Imaging*. (2010) 10:1. doi: 10.1186/1471-2342-10-1
28. Huellebrand M, Messroghli D, Tautz L, Kuehne T, Hennemuth A. An extensible software platform for interdisciplinary cardiovascular imaging research. *Comput Methods Prog Biomed*. (2020) 184:105277. doi: 10.1016/j.cmpb.2019.105277
29. Overhoff D, Kohlmann P, Frydrychowicz A, Gatidis S, Loewe C, Moltz J, et al. The international radiomics platform - an initiative of the german and austrian radiological societies - first application examples. *RoFo Fortsch Gebiet Rontgenstr Bildgebenden Verfahren*. (2021) 193:276–87. doi: 10.1055/a-1244-2775
30. Martin-Isla C, Campello VM, Izquierdo C, Raisi-Estabragh Z, Baeßler B, Petersen SE, et al. Image-based cardiac diagnosis with machine learning: a review. *Front Cardiovasc Med*. (2020) 7:1. doi: 10.3389/fcvm.2020.00001
31. Cao Y, Liu Z, Zhang P, Zheng Y, Song Y, Cui L. Deep learning methods for cardiovascular image. *J Artif Intell Syst*. (2019) 1:96–109. doi: 10.33969/AIS.2019.11006
32. Singh A, Sengupta S, Lakshminarayanan V. Explainable deep learning models in medical image analysis. *J Imaging*. (2020) 6:52. doi: 10.3390/jimaging6060052
33. Ghorbani A, Ouyang D, Abid A, He B, Chen JH, Harrington RA, et al. Deep learning interpretation of echocardiograms. *NPJ Digit Med*. (2020) 3:10. doi: 10.1038/s41746-019-0216-8
34. Janik A, Dodd J, Ifrim G, Sankaran K, Curran K. Interpretability of a deep learning model in the application of cardiac MRI segmentation with an ACDC challenge dataset. In: *Proceedings of SPIE - The International Society for Optical Engineering* (2021). doi: 10.1117/12.2582227
35. Cetin I, Raisi-Estabragh Z, Petersen SE, Napel S, Piechnik SK, Neubauer S, et al. Radiomics signatures of cardiovascular risk factors in cardiac MRI: results from the UK biobank. *Front Cardiovasc Med*. (2020) 7:232. doi: 10.3389/fcvm.2020.591368
36. Wiens V, Stocker M, Auer S. Towards customizable chart visualizations of tabular data using knowledge graphs. In: Ishita E, Pang NLS, Zhou L, editors. *Digital Libraries at Times of Massive Societal Transition*. Cham: Springer (2020). p. 71–80. doi: 10.1007/978-3-030-64452-9
37. Shah SJ. Precision medicine for heart failure with preserved ejection fraction: an overview. *J Cardiovasc Transl Res*. (2017) 10:233–44. doi: 10.1007/s12265-017-9756-y

38. Scherer J, Nolden M, Kleesiek J, Metzger J, Kades K, Schneider V, et al. Joint imaging platform for federated clinical data analytics. *JCO Clin Cancer Inform*. (2020) 4:1027–38. doi: 10.1200/CCI.20.00045
39. Klein J, Wenzel M, Romberg D, Köhn A, Kohlmann P, Link F, et al. QuantMed: component-based deep learning platform for translational research. In: Chen PH, Deserno TM, editors. *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*. Vol. 11318. Houston, TX: International Society for Optics and Photonics (2020). p. 229–36. doi: 10.1117/12.2549582
40. The U.S. Food and Drug Administration. *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback*. The U.S. Food and Drug Administration (2020).
41. 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
42. NEMA PI. *Digital Imaging and Communications in Medicine (DICOM) Standard. Supplement 97: CT/MR Cardiovascular Analysis Report*. National Electrical Manufacturers Association (2015).
43. Ritter F, Boskamp T, Homeyer A, Laue H, Schwier M, Link F, et al. Medical image analysis. *IEEE Pulse*. (2011) 2:60–70. doi: 10.1109/MPUL.2011.942929
44. Hüllebrand M, Ivantsits M, Zhang H, Kohlmann P, Kuhnigk J, Kühne T, et al. Comparison of a hybrid mixture model and a cnn for the segmentation of myocardial pathologies in delayed enhancement MRI. In: Puyol-Antón E, Pop M, Sermesant M, Campello VM, Lalande A, Lekadir K, et al., editors. *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges - 11th International Workshop, STACOM 2020, Held in Conjunction With MICCAI 2020*. Vol. 12592 of Lecture Notes in Computer Science. Lima: Springer (2020). p. 319–27.
45. Ivantsits M, Huellebrand M, Kelle S, Schonberg SO, Kuehne T, Hennemuth A. *Deep-Learning-Based Myocardial Pathology Detection*. (2021). Available online at: http://emidec.com/downloads/papers/paper-26.pdf
46. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324
47. Martin AR, Ward MO. High dimensional brushing for interactive exploration of multivariate data. In: *Proceedings of the IEEE Visualization Conference*, Atlanta, GA (1995). p. 271–8. doi: 10.1109/VISUAL.1995.485139
48. Tautz L, Hennemuth A, Peitgen HO. Motion analysis with quadrature filter based registration of tagged MRI sequences. In: Camara O, Konukoglu E, Pop M, Rhode K, Sermesant M, Young A, editors. *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*. Berlin; Heidelberg: Springer Berlin Heidelberg (2012). p. 78–87. doi: 10.1007/978-3-642-28326-0_8
49. Lalande A, Chen Z, Decourselle T, Qayyum A, Pommier T, Lorgis L, et al. Emidec: a database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac MRI. *Data*. (2020) 5:89. doi: 10.3390/data5040089
50. Vogel-Claussen J, Rochitte CE, Wu KC, Kamel IR, Foo TK, Lima JAC, et al. Delayed enhancement MR imaging: utility in myocardial assessment. *RadioGraphics*. (2006) 26:795–810. doi: 10.1148/rg.263055047
51. Dikici E, Bigelow M, Prevedello LM, White RD, Erdal BS. Integrating AI into radiology workflow: levels of research, production, and feedback maturity. *J Med Imaging*. (2020) 7:1–12. doi: 10.1117/1.JMI.7.1.016502
52. Bannach A, Bernard J, Jung F, Kohlhammer J, May T, Scheckenbach K, et al. Visual analytics for radiomics: combining medical imaging with patient data for clinical research. In: *2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC)*, Phoenix, AZ (2017). p. 84–91. doi: 10.1109/VAHC.2017.8387545
53. Angulo DA, Schneider C, Oliver JH, Charpak N, Hernandez JT. A multi-facetted visual analytics tool for exploratory analysis of human brain and function datasets. *Front Neuroinformatics*. (2016) 10:36. doi: 10.3389/fninf.2016.00036
54. Meuschke M, Niemann U, Behrendt B, Gutberlet M, Preim B, Lawonn K. GUCCI - guided cardiac cohort investigation of blood flow data. *IEEE Trans Visual Comput Graph*. (2021) 1–1. doi: 10.1109/TVCG.2021.3134083
55. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: *Proceedings of the 33rd International Conference on Machine Learning*. New York, NY (2016).

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 12 May 2022
doi: 10.3389/fcvm.2022.764629

Predicting the Prognosis of Patients in the Coronary Care Unit: A Novel Multi-Category Machine Learning Model Using XGBoost

Xingchen Wang[1†], Tianqi Zhu[2†], Minghong Xia[1], Yu Liu[1], Yao Wang[1], Xizhi Wang[1], Lenan Zhuang[1], Danfeng Zhong[1], Jun Zhu[1], Hong He[1], Shaoxiang Weng[1], Junhui Zhu[1] and Dongwu Lai[1*]*

[1] Key Laboratory of Cardiovascular Intervention and Regenerative Medicine of Zhejiang Province, Department of Cardiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2] College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:
Shane Nanayakkara,
Alfred Hospital, Australia

Reviewed by:
Jiali Deng,
Shanghai University, China
Fabrizio D'Ascenzo,
University of Turin, Italy

***Correspondence:**
Dongwu Lai
laidw@zju.edu.cn
Junhui Zhu
zhujhsrrsh@zju.edu.cn

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

Specialty section:
This article was submitted to General Cardiovascular Medicine, a section of the journal Frontiers in Cardiovascular Medicine

Received: 25 August 2021
Accepted: 20 April 2022
Published: 12 May 2022

Citation:
Wang X, Zhu T, Xia M, Liu Y, Wang Y, Wang X, Zhuang L, Zhong D, Zhu J, He H, Weng S, Zhu J and Lai D (2022) Predicting the Prognosis of Patients in the Coronary Care Unit: A Novel Multi-Category Machine Learning Model Using XGBoost. Front. Cardiovasc. Med. 9:764629. doi: 10.3389/fcvm.2022.764629

Background: Early prediction and classification of prognosis is essential for patients in the coronary care unit (CCU). We applied a machine learning (ML) model using the eXtreme Gradient Boosting (XGBoost) algorithm to prognosticate CCU patients and compared XGBoost with traditional classification models.

Methods: CCU patients' data were extracted from the MIMIC-III v1.4 clinical database, and divided into four groups based on the time to death: <30 days, 30 days−1 year, 1–5 years, and ≥5 years. Four classification models, including XGBoost, naïve Bayes (NB), logistic regression (LR), and support vector machine (SVM) were constructed using the Python software. These four models were tested and compared for accuracy, F1 score, Matthews correlation coefficient (MCC), and area under the curve (AUC) of the receiver operating characteristic curves. Subsequently, Local Interpretable Model-Agnostic Explanations method was performed to improve XGBoost model interpretability. We also constructed sub-models of each model based on the different categories of death time and compared the differences by decision curve analysis. The optimal model was further analyzed using a clinical impact curve. At last, feature ablation curves of the XGBoost model were conducted to obtain the simplified model.

Results: Overall, 5360 CCU patients were included. Compared to NB, LR, and SVM, the XGBoost model showed better accuracy (0.663, 0.605, 0.632, and 0.622), micro-AUCs (0.873, 0.811, 0.841, and 0.818), and MCC (0.337, 0.317, 0.250, and 0.182). In subgroup analysis, the XGBoost model had a better predictive performance in acute myocardial infarction subgroup. The decision curve and clinical impact curve analyses verified the clinical utility of the XGBoost model for different categories of patients. Finally, we obtained a simplified model with thirty features.

Conclusions: For CCU physicians, the ML technique by XGBoost is a potential predictive tool in patients with different conditions, and it may contribute to improvements in prognosis.

Keywords: MIMIC-III, coronary care unit (CCU), machine learning, multi-category, prognosis, XGBoost

INTRODUCTION

Cardiovascular disease (CVD), the leading cause of global mortality and disability, causes ∼18.6 million deaths annually. China has the highest mortality worldwide (1). Coronary care units (CCU), which concentrate on the treatment of patients with critical cardiovascular diseases, reduce mortality and prolong life expectancy in patients (2–6). To further improve survival outcomes, early evaluation and classification of prognosis are vital, as this can provide significant information for evaluating a patient's condition and deciding on appropriate treatments in advance. However, despite the availability of many clinical indicators such as the anion gap (AG) and serum osmolarity (7, 8), for assessing prognosis, the modest prognostic value of a single indicator and individual differences in the curative effect and toxicity of treatments make it difficult for clinicians to estimate the prognosis of CCU patients accurately and quickly.

The rapid development of medical artificial intelligence (AI) supported by big data and cloud computing makes it possible to improve the efficiency and accuracy of individual prognosis evaluation (9). AI has good adaptability in assessing disease prognosis given its abilities, including non-linear processing, high tolerance, intelligence, and self-learning. Machine learning (ML) has been widely applied in the field of disease prognosis assessment in recent years (10–13). The traditional ML models mainly include logistic regression (LR), naïve Bayes (NB), and support vector machine (SVM). Compared with serum indicators or clinical scores, these models can comprehensively evaluate patient status for accurate prognosis classification. However, these models still have many limitations. Recently, novel ML models have demonstrated improved performance compared to traditional ML models.

The eXtreme Gradient Boosting (XGBoost) model is an ML algorithm with excellent features, such as the efficient processing of missing data, flexibility, and assembly of weak prediction models to build an accurate model (14). It is an up-and-coming, widely favored algorithm in the field of ML. Besides, the establishment of specialized medical databases, such as the Medical Information Mart for Intensive Care III (MIMIC-III database), helps ML models extract data easily and enables further analysis. XGBoost (15), submitted by Tianqi Chen in 2016, is an integrated learning algorithm based on gradient boosting. It has been improved on the basis of the gradient boosting decision tree algorithm (16), with inclusion of the ability to customize the loss function, normalize the regular term, sparse feature processing, missing data processing, and parallel algorithm design, to name a few. These features allow the model to use variables with different degrees of flexibility in different areas of the output space, thereby realizing automatic feature selection and fitting of high-order interactions.

ML has made breakthroughs in the prognostic evaluation of diseases, and ML prediction models established for different diseases have achieved good prediction results. Hou et al. (17) used 4,559 sepsis patients from the MIMIC-III database and constructed XGBoost, LR, and SAPS-II score models to predict the 30-day mortality after admission in the intensive care unit (ICU). The areas under the curve (AUCs) of the three models were 0.857, 0.819, and 0.797, respectively. Li et al. (18) extracted 1,244 acute myocardial infarction (AMI) patients and built Gaussian naïve Bayes, LR, K-nearest neighbor, decision tree, random forest, and XGBoost models to predict 1-year mortality. The AUCs of the six models ranged from 0.709 to 0.942. Similarly, D'Ascenzo et al. (19) enrolled 19,826 patients diagnosed with acute coronary syndrome and constructed a risk prediction model based on ML algorithm to predict the 1-year mortality, recurrent acute myocardial infarction and bleeding risk of patients. However, most existing prognostic evaluation models use only two categories to predict the prognosis of patients, by prediction of 30-day morality and 1-year morality, which have limited clinical applications due to the lack of precision.

Therefore, we extracted CCU patients' data from the MIMIC-III database. Fifty-six clinical features were selected as inputs for the model, based on clinical experience and the completeness of prognostic indicators. Further, we attempted to construct an XGBoost model to prognosticate the time to death of CCU patients and used traditional ML models, such as LR, BN, and SVM, as benchmark comparisons. Finally, we established sub-models of each model to assess the clinical value and utility of the models. To our knowledge, this is the first study to apply a multi-category prediction approach in prognostic evaluation of CCU patients, and its findings will be of great significance to clinicians and patients.

MATERIALS AND METHODS

Data Source

We used the MIMIC III version 1.4 for the study. MIMIC-III, an openly usable critical care database, includes data of 46,520 patients admitted to multifarious ICUs of the Beth Israel Deaconess Medical Center (BIDMC) in Boston, Massachusetts, from 2001 to 2012 (20, 21). The database contains general

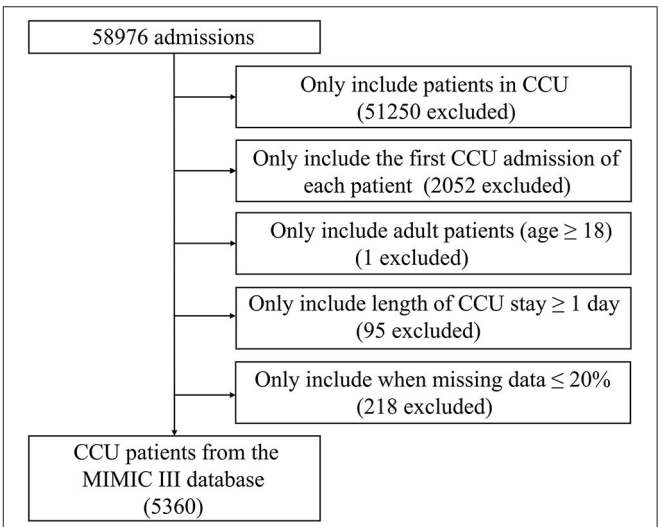


FIGURE 1 | Screening flowsheet of the study population. CCU, coronary care unit; MIMIC-III, Medical Information Mart for Intensive Care III.

information (such as demographics, the dates of birth and death, ICU admission, and discharge information), laboratory parameters, vital signs, body fluid analyses, medication use, and nursing records. Permission to use the MIMIC-III database was acquired from the institutional review boards of BIDMC and the Massachusetts Institute of Technology. Moreover, the user must pass an examination to gain access to the database and be authorized by the MIMIC-III institute. Our certificate number is 9648065. All patient data from MIMIC-III were extracted using Structured Query Language (SQL).

TABLE 1 | Predictor variables used in this study.

Predictor	
Demographic data	Serum osmolarity*
Age	Urine output*
Gender	**Diagnosis of heart disease**
Ethnicity	Coronary heart disease (CHD)
Height	Acute myocardial infarction (AMI)
Weight	AMI_ anterior wall
Body mass index (BMI)	Atrial fibrillation
Acute physiology (first 24 h in the CCU)	Ventricular arrhythmias†
Vital signs	Third-degree atrioventricular block
Heart rate* (HR)	Congestive heart failure (CHF)
Systolic blood pressure* (SBP)	Primary cardiomyopathy‡
Diastolic blood pressure* (DBP)	Valve disease
Mean blood pressure*(MAP)	Endocarditis
Respiratory rate* (RR)	Cardiogenic shock
Temperature* (TEMP)	**Comorbidity and medical history**
blood oxygen saturation* (SpO2)	Diabetes
Laboratory parameters	Chronic obstructive pulmonary diseases (COPD)
Anion gap* (AG)	Hypertension
Bicarbonate*	Respiratory failure
Glucose*	Hypercholesterolemia
Sodium*	Chronic liver disease
Potassium*	Chronic kidney disease
Calcium*	Prior myocardial infarction
Chloride*	**Medication use**
Creatinine*	Antiplatelet
Blood urea nitrogen* (BUN)	Anticoagulants
White blood cell* (WBC)	Beta-blocks
Hemoglobin*	ACEI/ARB
Platelet*	Statin
Mean corpuscular volume* (MCV)	Vasopressin
Mean corpuscular hemoglobin* (MCH)	**Other**
Red blood cell volume distribution width* (RDW)	Sequential organ failure score (SOFA)

*Each predictor marked with * means that it is a time-stamped variable, and its corresponding average values within the first 24 h in the CCU were used as inputs in model development.*

†*Includes ventricular tachycardia, ventricular flutter, and ventricular fibrillation.*

‡*Includes disorders of mitral, aortic, pulmonary, and tricuspid valve; rheumatic diseases of valves and congenital diseases of valve. CCU, coronary care unit; ACEI/ARB, angiotensin-converting enzyme inhibitor/angiotensin receptor blocker.*

Study Population

CCU patients registered in the MIMIC-III database were included. Only the first admission of each patient was included. The exclusion criteria were (a) age <18 years, (b) ≥20% missing individual data, and (c) length of CCU stay <1 day. Eventually, 5,360 patients were included (**Figure 1**).

Data Collection

All data were extracted from the MIMIC-III database using SQL. The following data were extracted: demographics, acute physiology (vital signs and laboratory parameters), diagnoses of heart disease, comorbidities and prior myocardial infarction, medication use, and sequential organ failure score. As shown in **Table 1**, 56 clinical features were selected as inputs for the model, based on clinical experience and the completeness of prognostic indicators. In addition, although viral myocarditis may lead to heart failure or cardiac arrest (22), this indicator was excluded because of the lack of sufficient samples in the MIMIC-III database. Serum osmolarity was calculated using the equation $(2 \times Na^+ + K^+) + (\text{glucose}/18) + (\text{urea nitrogen}/2.8)$ (7). Only values of the three variables measured at the same time were used for calculations. All laboratory parameters and vital signs were extracted within 24 h of CCU admission; we calculated the mean of each indicator separately.

Outcome and Statistical Analysis

The outcome was time to death, defined as the time from CCU admission to death. Fatality information was extracted from the file named "Patients" in the MIMIC-III database. Based on the time to death, we divided patients in this study into four groups: <30 days (class 0), 30 days−1 year (class 1), 1–5 years (class 2), ≥5 years (class 3), and variables were displayed and compared between the groups. Normally and non-normally distributed continuous variables were, respectively, summarized as the mean ± SD and the median (interquartile ranges, IQR). One-way analysis of variance or the Kruskal-Wallis test was used to analyse differences. Categorical variables were summarized as a number (percentage) and were compared between groups using the chi-square test or Fisher' exact test. All analyses were performed using the STATA 15 software, and statistical difference was defined as p-value < 0.05.

Model and Metrics

In the model-construction phase, we employed an ML model using XGBoost to predict the time to death, while using LR, NB, and SVM models as benchmark comparisons. For XGBoost, we set the reduction rate to 0.3, the maximum tree depth as 2, while other parameters were set to the default parameters of the scikit-learn library. In the model-comparison phase, we tested and compared the performances of the four predictive models for

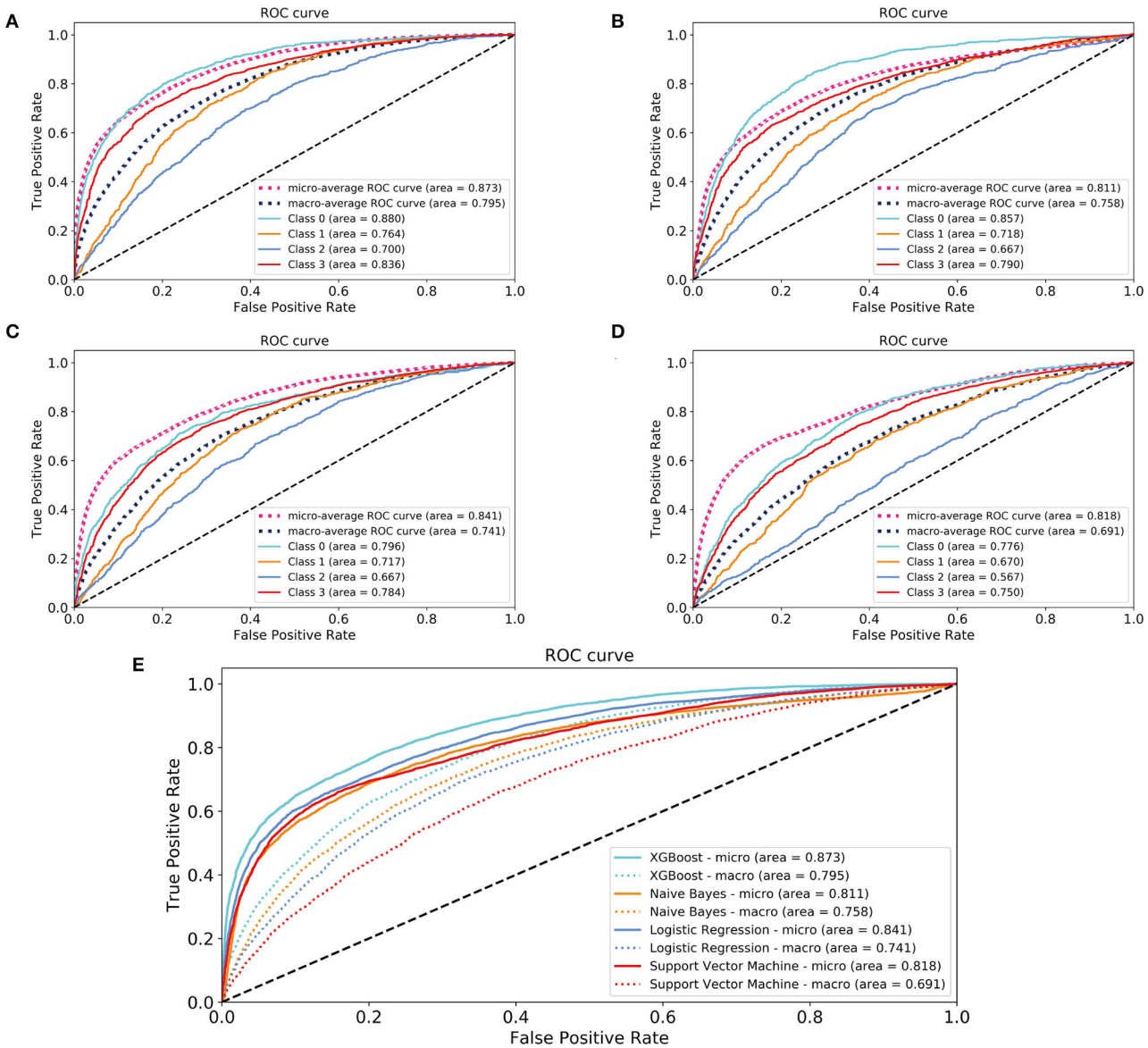


FIGURE 2 | Areas under the receiver operating characteristic curves for evaluating the classification performance of the different models. **(A)** XGBoost model; **(B)** naïve Bayes model; **(C)** logistic regression model; **(D)** support vector machine model; **(E)** a comparison of four models. Class 0: time to death < 30 days; Class 1: 30 days ≤ time to death < 1 year; Class 2: 1 year ≤ time to death < 5 years; Class 3: time to death ≥5 years. XGBoost, eXtreme Gradient Boosting.

their accuracy (ACC), F1 score, Matthews correlation coefficient (MCC), and AUCs of the receiver operating characteristic curves (ROC). The AUCs and F1 score were calculated by micro-average and macro-average methods (23). For classification tasks with imbalanced data, AUCs, the F1 score, and MCC have better adaptability (24). Thus, these three indicators were included in the performance evaluation of the model. Subsequently, we performed the Local Interpretable Model-Agnostic Explanations (LIME) algorithm to obtain the direction in which the features change. LIME places emphasis on training local surrogate models to explain individual predictions (25). Besides, to further assess the clinical practicability of the model, we divided patients into four two-class data sets according to the time of death (for example, class 1 is for one group, classes 0, 2, and 3 are for another group) and sequentially established sub-models of each model. Decision curve analysis (DCA) was used to calculate the net benefit and compare differences between these four sub-models. The optimal model was further analyzed using a clinical impact curve (CIC) to assess the clinical practicability and net benefit of the model with the best prognostic predictive value. Finally, the feature ablation curves (excluded one by one according to the feature importance score from low to

TABLE 2 | Performance of the four prediction models.

	XGBoost	NB	LR	SVM
Accuracy (%), 95% CI	0.663 (0.655–0.671)	0.605 (0.594–0.617)	0.632 (0.621–0.642)	0.622 (0.618–0.627)
AUC-micro, 95% CI	0.873 (0.867–0.879)	0.811 (0.802–0.820)	0.841 (0.833–0.849)	0.818 (0.812–0.825)
AUC-macro, 95% CI	0.795 (0.782–0.808)	0.758 (0.745–0.772)	0.741 (0.727–0.756)	0.691 (0.678–0.711)
F1-micro, 95% CI	0.663 (0.655–0.671)	0.605 (0.594–0.617)	0.632 (0.621–0.642)	0.622 (0.618–0.627)
F1-macro, 95% CI	0.416 (0.395–0.434)	0.441 (0.424–0.455)	0.343 (0.326–0.360)	0.275 (0.262–0.287)
MCC, 95% CI	0.337 (0.318–0.357)	0.317 (0.295–0.340)	0.250 (0.224–0.276)	0.182 (0.166–0.198)

NB, Naive Bayes; LR, logistic regression; SVM, support vector machine; CI, confidence interval; AUC, area under the curve; MCC, Matthews correlation coefficient.

high) of the XGBoost model were conducted to obtain the simplified model.

For all the models above, we used a 10-fold cross validation method to obtain the performance of the model for the whole data set. For cross validation, the dataset was divided into 10-folds, of which 1-fold was used as the test set and the remaining were used as the training set; all results of the 10 repetitions were averaged as the overall performance. In the XGBoost model, we used 20% of the training set as the validation set to perform the early stopping strategy. All experiments of the XGBoost model and other models were constructed using the scikit-learn of the standard ML software package in the Python 3.8 software.

RESULTS

Baseline Characteristics

In total, 5,360 patients admitted to the CCU were included (**Figure 1**). The baseline characteristics of patients stratified by time to death are displayed in **Supplementary Table S1**. Except for the ethnicity, third-degree atrioventricular block, primary cardiomyopathy, chronic liver disease, prior myocardial infarction, and blood oxygen saturation (SpO_2), other clinical features showed a statistically significant difference between the groups ($p < 0.05$).

Model Comparisons and Validations

In the model-construction and validation phase, ML models had different recognition and classification capabilities for different classes. These capabilities had some consistency, that is, the model had better classification capabilities for classes 0 and 3, while the classification performance for classes 1 and 2 was poor. If the XGBoost model is taken as an example, the micro-AUCs of classes 0 and 3 were 0.88 and 0.836, respectively; those of classes 1 and 2 were 0.764 and 0.7, respectively (**Figures 2A–D**). All four models (XGBoost model, NB model, LR model, and SVM model) showed good discriminatory power with micro-AUCs of 0.873 (95% CI 0.867–0.879), 0.811 (95% CI 0.802–0.820), 0.841 (95% CI 0.833–0.849), and 0.818 (95% CI 0.812–0.825), respectively, and macro-AUCs of 0.795 (95% CI 0.782–0.808), 0.758 (95% CI 0.745–0.772), 0.741 (95% CI 0.727–0.756), and 0.691 (95% CI 0.678–0.711), respectively. The accuracy and F1-micro of the models were 0.663 (95% CI 0.655–0.671), 0.605 (95% CI 0.594–0.617), 0.632 (95% CI 0.621–0.642), and 0.622 (95% CI 0.618–0.627), respectively. The MCCs of models were 0.337 (95% CI 0.318–0.357), 0.317 (95% CI 0.295–0.340), 0.250 (95% CI 0.224–0.276), and 0.182 (95% CI 0.166–0.198), respectively. These indicators showed that the XGBoost model was the most optimal option, although its F1-macro was not the largest among the four models (**Figure 2E**; **Table 2**).

The performance indicators of the validation set and test set under the 10-fold cross validation test of the XGBoost model are shown in **Table 3**. There was no significant difference between the two sets. Subsequently, we applied the XGBoost model to the subgroup analysis of the four major heart diseases. The results are shown in **Table 4**, and the models all showed good performance. The coronary heart disease subgroup was the best.

Features Assessed Using XGBoost

As shown in **Figure 3**, according to the results of each feature's analysis in the XGBoost model, age was most important feature of the data set. The remaining top 10 features were temperature, mean arterial pressure (MAP), SpO_2, systolic blood pressure (SBP), chloride, red blood cell volume distribution width urine_24 h, hemoglobin, and body mass index, in that order. However, traditional prognostic-related indicators, such as diabetes and hypercholesterolemia, showed poor importance contribution scores.

Interpretability of the Prediction Model

Figure 4 shows the decision process for the single-sample prediction of class 0, which is a local interpretation of the XGBoost model based on LIME. This sample was correctly classified as class 0 by the model, where the features in green allowed the model to identify the sample as class 0, and the features in red allowed the model to identify the sample as not class 0. The LIME results of classes 1–3 are shown in **Supplementary Figure S1**.

Sub-models Comparisons

According to DCA of four prediction sub-models, the net benefit for the XGBoost model was all greater than that of the traditional models for the threshold probabilities of different outcomes, meaning that the XGBoost model was the most optimal (**Figures 5A–D**). Thus, the XGBoost model was further analyzed using CIC. The CIC is shown in **Figures 6A–D**, and clearly shows that the XGBoost model had an excellent clinical net benefit within the general range of threshold probabilities and impacted patient outcomes, which verifies that the XGBoost model had better clinical decision-making performance than

the other models for different categories of patients. **Table 5** shows the quantitative results of the DCA and CIC curves of the XGBoost sub-model. For example, for class 0, the risk probability threshold of 0.15 (cost-benefit ratio 15:85) corresponds to at least 75% of the population. This means that <25% of patients were classified as positive by the model (245 patients). Among these, 99 patients had a positive outcome. The calculated net benefit is as follows: 99/1,000–(245–99)/1,000 × 0.15/(1–0.15) = 0.07.

Simplified Model

As presented in **Figure 7**, When the number of features of the XGBoost model is reduced from 56 to 30 one by one, the MCC remains basically constant, that is, TOP30 in **Figure 3** is the input feature of the simplified model. The detailed information of **Figure 7** is listed in **Supplementary Table S2**.

DISCUSSION

In this study based on the need for clinical applications, we pioneered the multi-category ML model for predicting time to death, rather than the traditional two-category model, for the first time. We found that the XGBoost model, when compared with some traditional classification models, showed obvious superiority in classification performance and clinical utility for different categories of patients.

In previous ML model studies, predictive performance was evaluated and compared. However, clinical applicability and clinical consequences were not investigated. These models, including XGBoost (17, 18, 26, 27), were applied to patients diagnosed with AMI, those who underwent open-heart surgery, or those admitted to the ICU. All of them had a two-category pattern, that is, they focused on identifying high-risk populations for early intervention. However, this type of model is limited by the fact that patients outside the high-risk category are not accurately classified. Therefore, some intermediate-risk groups of patients do not receive clinicians' full attention. In our study, we divided patients into four categories according to the time to death. We further optimized the XGBoost micro-parameters, making it more suitable for multi-category prediction. Consequently, our ML model could assist CCU physicians in developing treatment strategies and determining the follow-up intensity according to different risk levels. For example, when patients enter the CCU, their baseline data, vital signs, and laboratory data on the first day will be inputted into the model for analysis. According to the analysis results, the patient's prognosis can be stratified. Those predicted to die within 30 days are classified in the high-risk group. Improved vital sign monitoring and continue hospitalization are recommended for such patients. Those predicted to die between 30 days and 1 year belong to the medium-risk group. Increased follow-up frequency after discharge is recommended, and the attending physician should pay attention to these patients' potential risks. Those predicted to die 1 year later are categorized as the low-risk group and should be followed up regularly after discharge. However, these are only approximate clinical decisions. Moreover, specific treatment measures also depend on the immediate state of the individual. This study has not discussed this in depth.

An important finding of this study was that the ML model had a different classification performance for different classes, but still had a few commonalities. All models had better classification capabilities for classes 0 and 3, and classification performance for classes 1 and 2 was poor. This phenomenon may be explained as follows. First, the grouped data set was unbalanced. More than 3,000 patients survived for more than 5 years. The models we built were all supervised ML models. When the samples are unbalanced, the model tends to ignore the small sample loss to reduce the overall loss (28). Due to the model's intrinsic characteristics, the classification results are often influenced several categories, resulting in the overestimation of classification performance. Therefore, the model had a higher classification accuracy for survival over 5 years. Further, the scales of classes

TABLE 3 | Performance of the validation set and test set under the 10-fold cross-validation test of the XGBoost model.

	Validation set	Test set
Accuracy, 95% CI	0.661 (0.654–0.667)	0.663 (0.655–0.671)
AUC-micro, 95% CI	0.870 (0.867–0.874)	0.873 (0.867–0.879)
AUC-macro, 95% CI	0.789 (0.783–0.795)	0.795 (0.782–0.808)
F1-micro, 95% CI	0.661 (0.654–0.667)	0.663 (0.655–0.671)
F1-macro, 95% CI	0.420 (0.410–0.430)	0.416 (0.395–0.434)
MCC, 95% CI	0.336 (0.320–0.352)	0.337 (0.318–0.357)

CI, confidence interval; AUC, area under the curve; MCC, Matthews correlation coefficient.

TABLE 4 | Performance of the XGBoost model in the four major types of heart disease.

	CHD	AMI	CHF	VA
Accuracy (%), 95% CI	0.703 (0.691–0.715)	0.751 (0.725–0.777)	0.571 (0.551–0.592)	0.666 (0.628–0.704)
AUC-micro, 95% CI	0.897 (0.888–0.906)	0.917 (0.903–0.931)	0.811 (0.797–0.826)	0.873 (0.855–0.891)
AUC-macro, 95% CI	0.812 (0.791–0.834)	0.815 (0.782–0.849)	0.755 (0.733–0.777)	0.806 (0.773–0.838)
F1-micro, 95% CI	0.703 (0.691–0.715)	0.751 (0.725–0.777)	0.571 (0.551–0.592)	0.666 (0.628–0.704)
F1-macro, 95% CI	0.428 (0.401–0.456)	0.394 (0.348–0.439)	0.414 (0.392–0.435)	0.409 (0.357–0.461)
MCC, 95% CI	0.343 (0.310–0.375)	0.371 (0.294–0.448)	0.306 (0.279–0.333)	0.384 (0.310–0.458)

CHD, coronary heart disease; AMI, acute myocardial infarction; CHF, congestive heart failure; VA, ventricular arrhythmia; CI, confidence interval; AUC, area under the curve; MCC, Matthews correlation coefficient.

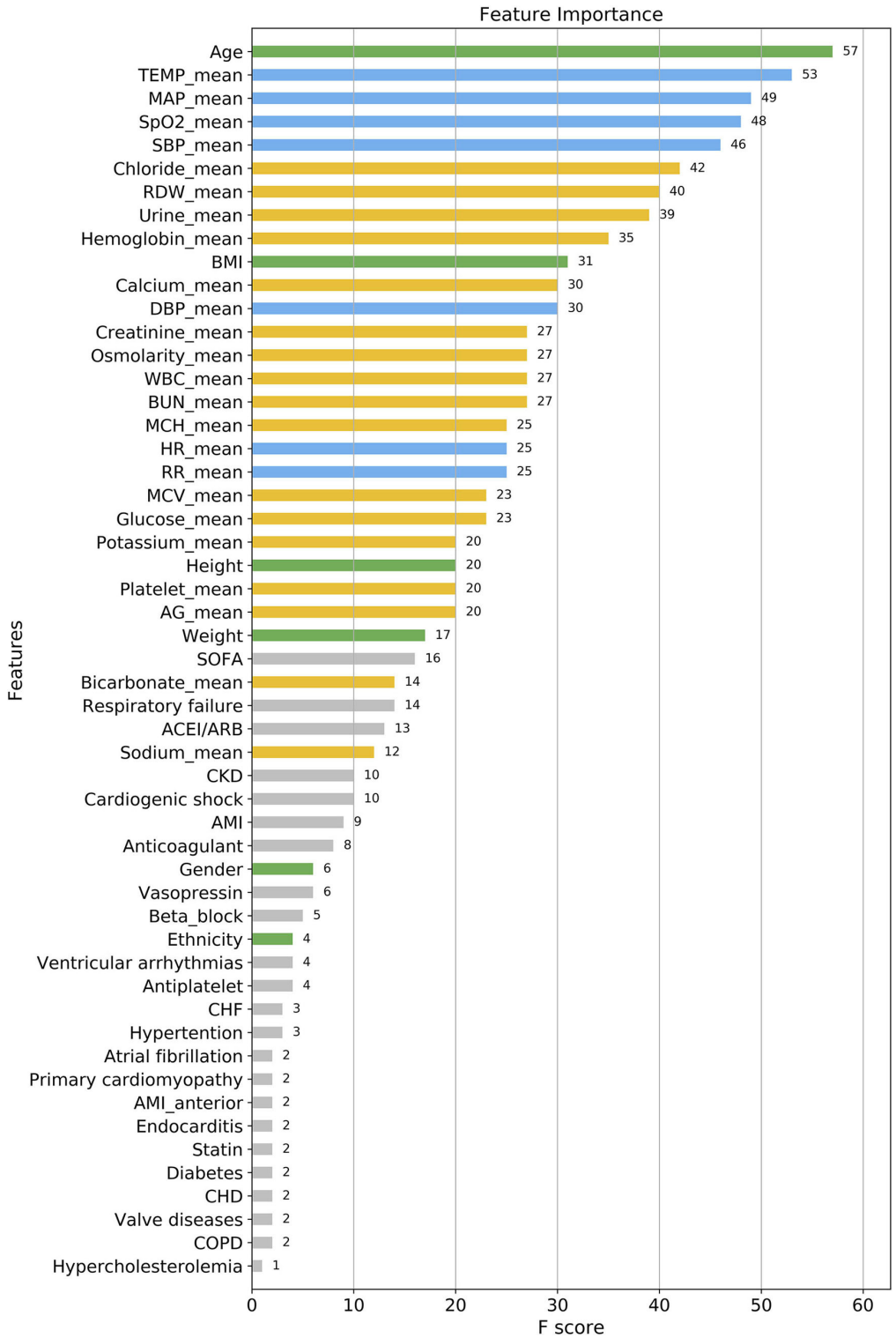


FIGURE 3 | Feature importance score ranking for 56 clinical features of the four-group XGBoost predictor. The information reflects the contribution of different features to the classification performance of XGBoost model (from top to bottom). **Green**, demographic data; **Blue**, vital signs; **Yellow**, laboratory parameters; **Gray**, others. TEMP, temperature; MAP, mean arterial pressure; SpO2, oxygen saturation; SBP, systolic blood pressure; RDW, red blood cell volume distribution width; BMI, body mass index; DBP, diastolic blood pressure; WBC, white blood cell; BUN, blood urea nitrogen; MCH, mean corpuscular hemoglobin; HR, heart rate; RR, respiratory rate; MCV, mean corpuscular volume; AG, anion gap; SOFA, sequential organ failure score; CKD, chronic kidney disease; AMI, acute myocardial infarction; CHF, congestive heart failure; CHD, coronary heart disease; COPD, chronic obstructive pulmonary diseases.

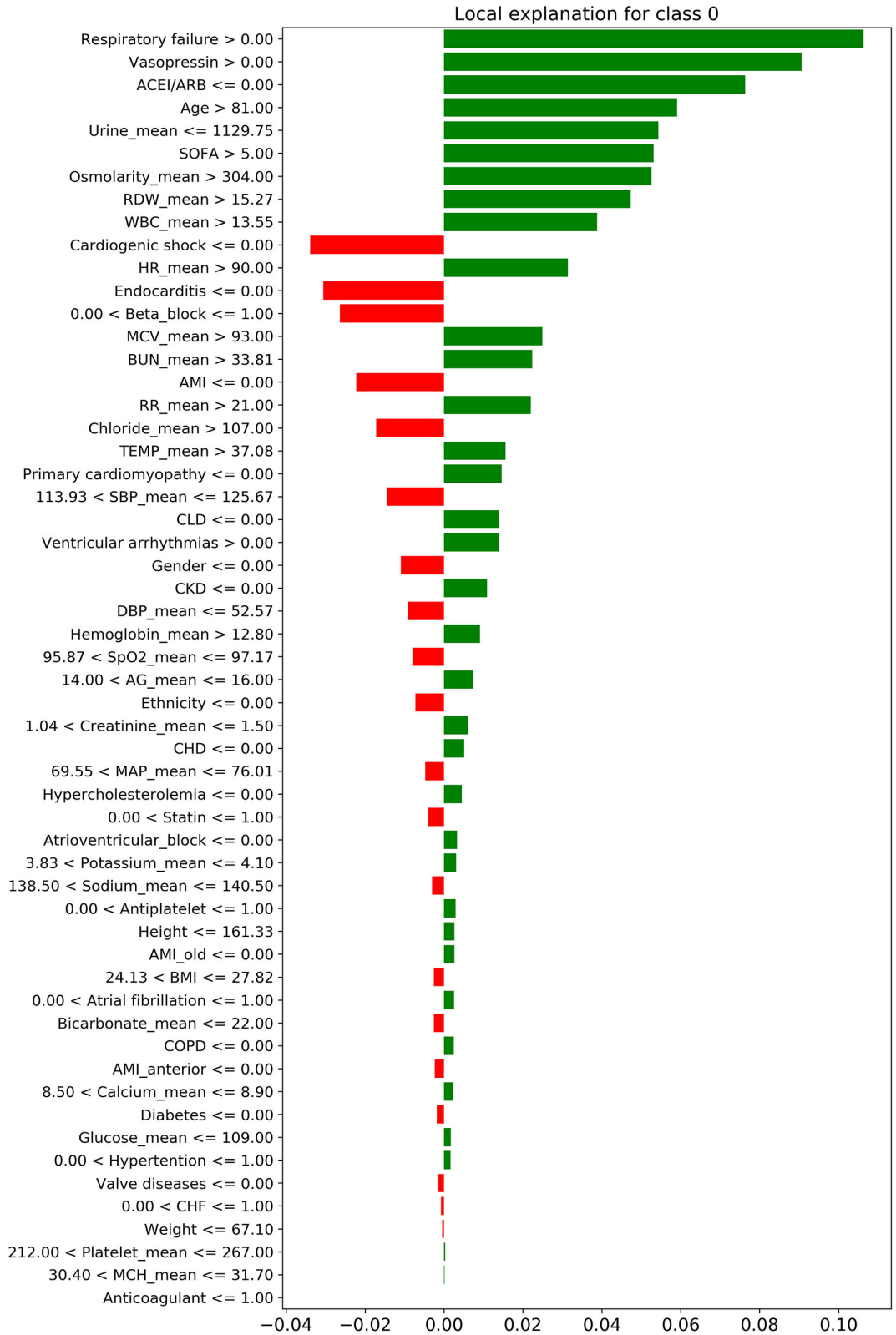


FIGURE 4 | The local interpretation of the XGBoost model for class 0. The features of the green column make the model identify the sample as class 0, while the features of the red column allow the model to identify the sample as non-class 0. Since the sum of green column score exceeded red, the model finally identifies this sample as class 0.

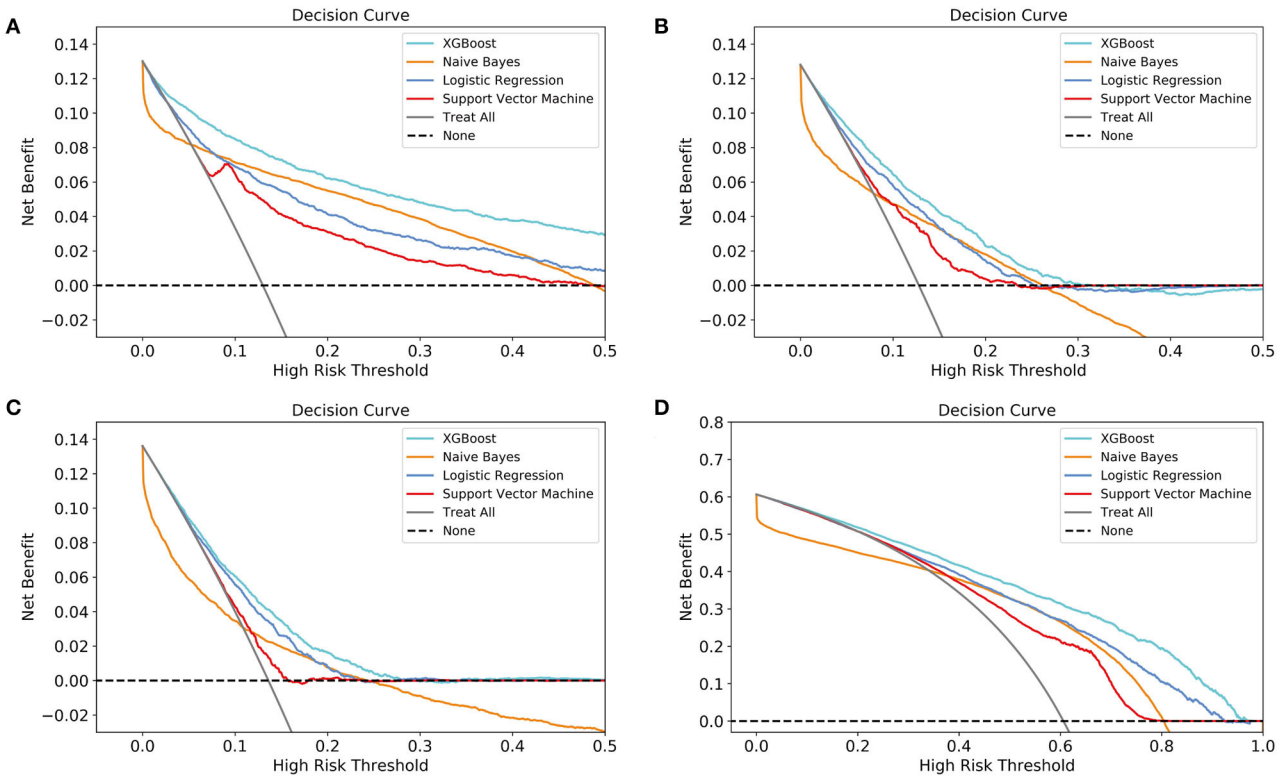


FIGURE 5 | Decision curve analysis (DCA) of the four prediction sub-models. **(A)** class 0 (<30 days); **(B)** class 1 (30 days−1 year); **(C)** class 2 (1–5 years); **(D)** class 3 (≥5 years). The net benefit curves for the four prognostic sub-models are shown. The lateral-axis shows the threshold probability for different class outcomes, and the direct-axis shows the net benefit. The horizontal dashed line represents no intervention in all patients, with a net benefit of 0, and the sloping gray line represents intervention in all patients. The four colored curves represent the four schemes (prediction models) with a larger net benefit to XGBoost compared to the other models for the threshold probabilities of different outcomes.

0, 1, and 2 were very similar and they all had about 700 patients. This may have attributed to the immediacy of the clinical indicators. The patients' physiological state changes all the time. Over time, the predictive performance of the indicator collected early may decline. Three studies published previously, in fact, have predicted the short-term prognosis of patients based on this feature (17, 18, 26).

Additionally, many studies applied statistical methods to initially screen predictors and then, incorporated the screened factors into the model. However, we did not pre-process the input factors for the following reasons: firstly, in the initial selection process, we screened out these 56 predictors from hundreds of clinical factors in the database based on the literature and clinical practical applications. Secondly, for the first time, we innovatively divided the patient's time to death into multiple intervals for prediction. We do not know whether predictors with or without significant differences in traditional two-category studies are applicable to multi-category situations. Furthermore, traditional statistical screening methods may have limitations in case of multiple classifications. This may have led to over-screening or meaningless screening of predictors. Finally, the XGBOOST algorithm model used can automatically screen the importance of predictive variables while ignoring the interference of irrelevant variables, which greatly improved the effectiveness of our research.

The AUCs, accuracy, F1 score, and MCC testified for the excellent performance of the XGBoost model. The XGBoost model builds a host of sub-models for classification, and finally assembles the classification results. Since the sub-model only uses a few indicators for model construction, some of the outliers and missing values will have a smaller impact on the performance of the model, thus making the model more robust (15). This feature has good suitability for the MIMIC-III database. Moreover, the XGBoost algorithm can standardize the regularization term to prevent the model from overfitting. Thus, these features enable the model to have a stronger classification performance for retrospective data. However, accuracy, AUCs, and the F1 score focus solely on the predictive accuracy of the model, without the results caused by the prediction information. For improvement in purely mathematical metrics, DCA is widely used in clinical analysis (17, 29). DCA is based on a decision-making theoretical framework that considers both the benefit of the intervention and the cost of the intervention for patients who cannot benefit (30). Therefore, DCA can compare the clinical application value of different models and tell us which model is worth using.

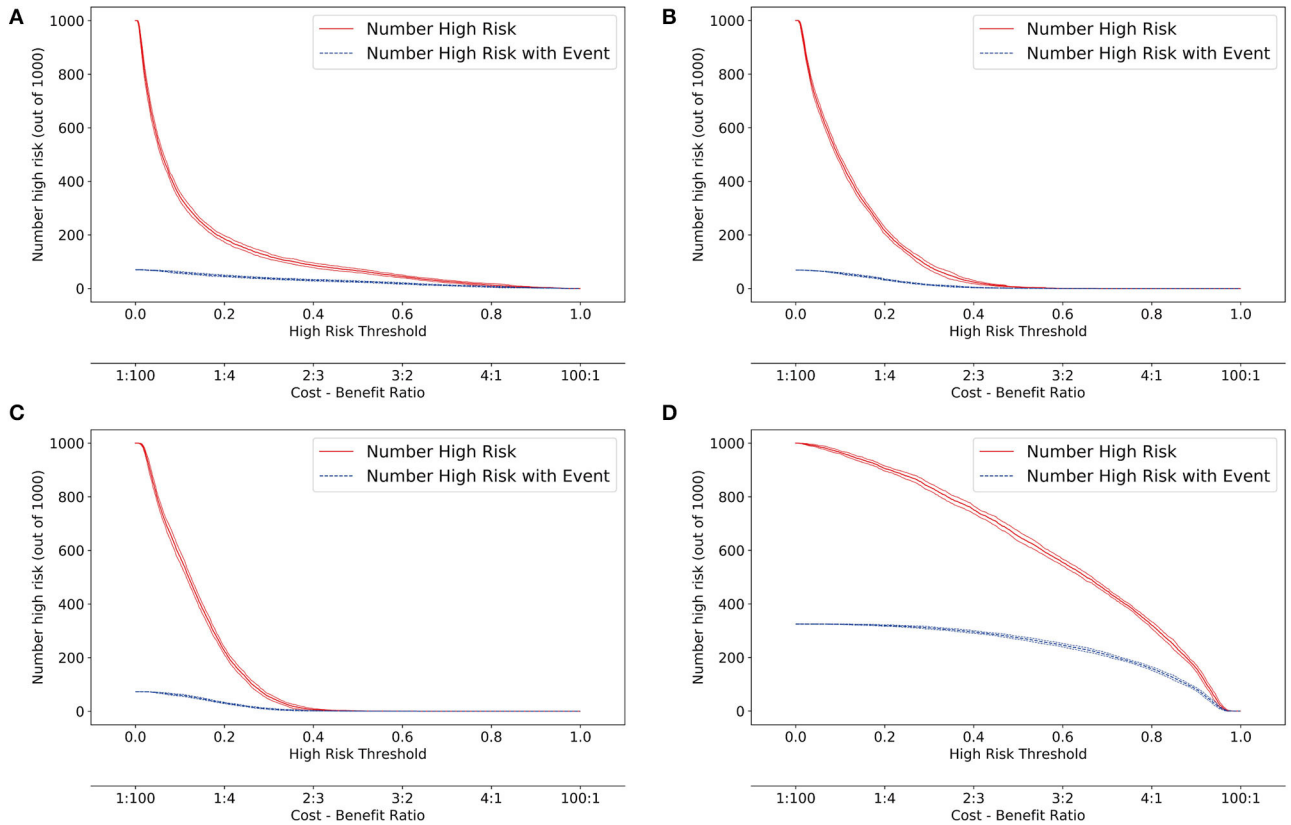


FIGURE 6 | Clinical impact curve (CIC) of the XGBoost model. **(A)** class 0 (<30 days); **(B)** class 1 (30 days−1 year); **(C)** class 2 (1–5 years); **(D)** class 3 (≥5 years). The red curve (number of high-risk individuals) indicates the number of people classified as positive (high risk) by the model at each threshold probability; the blue curve (number of high-risk individuals with outcome) is the number of true positives at each threshold probability.

However, DCA is used to evaluate the clinical value of the two-category model. To make it suitable for multi-category models, we divided the patients into binary data sets in turn, for example, group 1: class 2 and group 2: classes 0, 1, and 3, and built sub-model of each model. Then, DCA was used to evaluate the clinical practicability and decision-making performance of different models for patients with different outcomes.

The relationship between the contribution features of the XGBoost model and death cannot be fully explained. Thus, further research is needed to investigate the specific relationship between these features and death. The following is a brief summary of the important results obtained by the XGBoost model. Among these features, the weight of age was the greatest, meaning that it was the most significant predictor for the time to death of CCU patients. This result is consistent with those of previous clinical studies. Albanese et al. (31) reported that for CCU patients after percutaneous coronary intervention, older age was associated with major endpoints such as ventricular fibrillation, tachycardia, and sudden cardiac or arrhythmic death. Al-Ghamdi et al. (32) concluded that age >50 years was an independent predictor of death in CCU patients. Ruiz-Bailén et al. (33) enrolled 17 761 CCU/ICU patients with AMI, and indicated that age was an important independent predictive variable for mortality. This may be due to the following potential mechanisms: first, older patients tend to have more complications and infection risks (33); second, older patients, despite the higher mortality risk, are treated with less aggressive therapies than younger patients (34); finally, older patients show poor adaptability and tolerance under stressful conditions such as hypoxia, myocardial ischaemia, and so on. Besides, we find that the top 2–4 important features of the XGBoost model are temperature, MAP, and SpO_2, which are all clinically vital signs. This reminds clinicians to focus on the modest change in patients' vital signs at an early stage. Vital signs have been shown to be the most accurate predictors of clinical deterioration (35). In the CCU, hyperthermia often indicates infection and hypothermia indicates shock, both of which are predictors of poor prognosis; body temperature is thus a good prognostic factor. Similarly, MAP, SBP, and SpO_2 reflect the respiratory and circulatory state of the patients, and their abnormalities may indicate early physiological duress. However, traditional prognostic indicators, such as diabetes, hypercholesterolemia, had poor contribution scores. On one hand, these indicators may display lower performance in predicting death in multi-class classification. In contrast, due to them being categorical variables, the model may reduce its prognostic classification weight while simultaneously dealing with categorical and continuous data.

TABLE 5 | Critical prediction accuracy under different XGBoost sub-model risk thresholds.

Model risk percentile	RPT	Cost-benefit ratio	NHR (out of 1,000)	NHR with event (out of 1,000)	Sensitivity (%)	Specificity (%)	NB
<30 days (class 0)							
≥0	0.00	1: ∞	1,000	130	100	0	0.13
≥25	0.03	3: 97	710	127	97	33	0.11
≥50	0.06	6: 94	497	121	93	57	0.10
≥75	0.15	15: 85	245	99	76	83	0.07
≥90	0.36	36: 64	99	62	48	96	0.04
30 days–1 year (class 1)							
≥0	0.00	1: ∞	1,000	128	100	0	0.13
≥25	0.04	4: 96	738	124	97	30	0.10
≥50	0.10	1: 9	480	105	83	57	0.06
≥75	0.19	19: 81	239	70	55	81	0.03
≥90	0.29	29: 71	91	28	22	93	0.00
1–5 years (class 2)							
≥0	0.00	1: ∞	1,000	136	100	0	0.14
≥25	0.06	6: 94	740	126	93	29	0.09
≥50	0.12	12: 88	498	103	76	54	0.05
≥75	0.20	20: 80	225	58	43	81	0.02
≥90	0.27	27: 73	90	26	19	93	0.00
≥5 years (class 3)							
≥0	0.00	1: ∞	1,000	606	100	0	0.61
≥25	0.40	2: 3	750	550	91	49	0.42
≥50	0.66	66: 34	494	422	70	82	0.28
≥75	0.85	85: 15	246	230	38	96	0.14
≥90	0.93	93: 7	84	81	13	99	0.04

RPT, risk probability thresholds; NHR, number high risk; NB, net benefit.

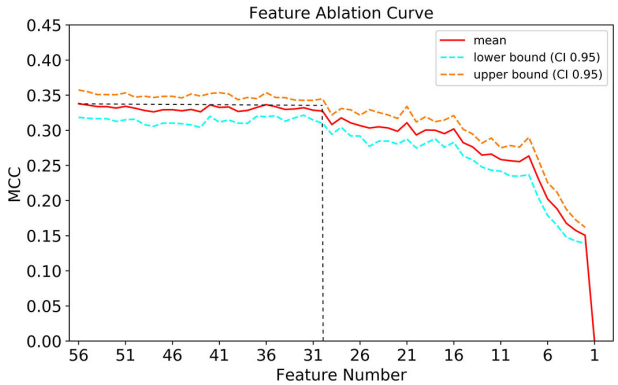


FIGURE 7 | Feature ablation curves of XGBoost model. MCC, Matthews correlation coefficient; CI, confidence interval.

Our study has several limitations due to its retrospective design. First, a few patients had small amounts of missing data. Although statistical methods were used to compensate, they could also have led to data bias and inaccurate prediction results. Second, measurement bias within calculations is possible, as the methods were based on specialists' individual opinion. Finally, as patient data were extracted from the MIMIC-III database, clinically common prognostic indicators of cardiovascular disease, such as troponin, creatine kinase-MB, and lactate, were excluded because the measurement volume was too small. Nonetheless, the XGBoost model is an efficient and robust method for multi-categorically predicting patients' time to death.

CONCLUSIONS

In summary, our study indicates that the XGBoost model does outperform traditional models. It has the potential to assist physicians in the CCU to perform optimal clinical interventions quickly and accurately, and may thus improve the prognosis of CCU patients.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

YW, XizW, LZ, DZ, JunZ, HH, and SW collected and interpreted the data. XinW, TZ, and YL analyzed the data. XinW, TZ, and MX generated the figures and

wrote the manuscript. DL and JunhZ designed study and revised the manuscript. All authors approved the final manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (Nos. 81974025 and 81941003), Natural Science Foundation of Zhejiang Province (Nos. LY19H020007 and LY18H020001), and Medical and Health Science Program of Zhejiang Province (Nos. 2020RC016 and 2019RC183).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2022.764629/full#supplementary-material

REFERENCES

1. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the Gbd 2019 Study. *J Am Coll Cardiol.* (2020) 76:2982–3021. doi: 10.1016/j.jacc.2020.11.010
2. Fye WB. Resuscitating a circulation abstract to celebrate the 50th anniversary of the coronary care unit concept. *Circulation.* (2011) 124:1886–93. doi: 10.1161/CIRCULATIONAHA.111.033597
3. Julian DG. The history of coronary care units. *Br Heart J.* (1987) 57:497–502. doi: 10.1136/hrt.57.6.497
4. Killip T, Kimball JT. Treatment of myocardial infarction in a coronary care unit: a two year experience with 250 patients. *Am J Cardiol.* (1967) 20:457–64. doi: 10.1016/0002-9149(67)90023-9
5. Loughran J, Puthawala T, Sutton BS, Brown LE, Pronovost PJ, DeFilippis AP. The cardiovascular intensive care unit-an evolving model for health care delivery. *J Intens Care Med.* (2017) 32:116–23. doi: 10.1177/0885066615624664
6. MacMillan RL, Brown KW. Comparison of the effects of treatment of acute myocardial infarction in a coronary unit and on a general medical ward. *Can Med Assoc J.* (1971) 105:1037–40.
7. Shen Y, Cheng X, Ying M, Chang HT, Zhang W. Association between serum osmolarity and mortality in patients who are critically ill: a retrospective cohort study. *BMJ Open.* (2017) 7:e015729. doi: 10.1136/bmjopen-2016-015729
8. Sun T, Cai C, Shen H, Yang J, Guo Q, Zhang J, et al. Anion gap was associated with inhospital mortality and adverse clinical outcomes of coronary care unit patients. *Biomed Res Int.* (2020) 2020:4598462. doi: 10.1155/2020/4598462
9. Musib M, Wang F, Tarselli MA, Yoho R, Yu KH, Andrés RM, et al. Artificial intelligence in research. *Science.* (2017) 357:28–30. doi: 10.1126/science.357.6346.28
10. Alsinglawi B, Alnajjar F, Mubin O, Novoa M, Darwish O. Predicting length of stay for cardiovascular hospitalizations in the intensive care unit: Machine learning approach. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society*. IEEE (2020).
11. 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
12. Sevakula RK, Au-Yeung WM, Singh JP, Heist EK, Isselbacher EM, Armoundas AA. State-of-the-art machine learning techniques aiming to improve patient outcomes pertaining to the cardiovascular system. *J Am Heart Assoc.* (2020) 9:e013924. doi: 10.1161/JAHA.119.013924
13. Zack CJ, Senecal C, Kinar Y, Metzger Y, Bar-Sinai Y, Widmer RJ, et al. Leveraging machine learning techniques to forecast patient prognosis after percutaneous coronary intervention. *JACC Cardiovasc Interv.* (2019) 12:1304–11. doi: 10.1016/j.jcin.2019.02.035
14. Yuan KC, Tsai LW, Lee KH, Cheng YW, Hsu SC, Lo YS, et al. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. *Int J Med Inform.* (2020) 141:104176. doi: 10.1016/j.ijmedinf.2020.104176
15. Chen T, Guestrin C. XGBoost : A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA; New York, NY: Association for Computing Machinery (2016). p. 785–94.
16. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat.* (2001) 29:1189–232. doi: 10.1214/aos/1013203451
17. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for mimic-iii patients with sepsis-3: a machine learning approach using Xgboost. *J Transl Med.* (2020) 18:462. doi: 10.1186/s12967-020-02620-5
18. Li YM, Jiang LC, He JJ, Jia KY, Peng Y, Chen M. Machine learning to predict the 1-year mortality rate after acute anterior myocardial infarction in Chinese patients. *Ther Clin Risk Manag.* (2020) 16:1–6. doi: 10.2147/TCRM.S236498
19. D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (praise): a modelling study of pooled datasets. *Lancet.* (2021) 397:199–207. doi: 10.1016/S0140-6736(20)32519-8
20. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. Mimic-Iii, a freely accessible critical care database. *Sci Data.* (2016) 3:160035. doi: 10.1038/sdata.2016.35
21. Oweira H, Schmidt J, Mehrabi A, Kulaksiz H, Schneider P, Schöb O. et al. Comparison of three prognostic models for predicting cancer-specific survival among patients with gastrointestinal stromal tumors. *Fut Oncol.* (2018) 14:379–89. doi: 10.2217/fon-2017-0450
22. Ma Z, Shen Z, Gong Y, Zhou J, Chen X, Lv Q, et al. Weighted gene co-expression network analysis identified underlying hub genes and mechanisms in the occurrence and development of viral myocarditis. *Ann Transl Med.* (2020) 8:1348. doi: 10.21037/atm-20-3337
23. Liu C, Wang W, Wang M, Lv F, Konan M. An efficient instance selection algorithm to reconstruct training set for support vector machine. *Knowl Based Syst.* (2017) 116:58–73. doi: 10.1016/j.knosys.2016.10.031
24. Russo DP, Zorn KM, Clark AM, Zhu H, Ekins S. Comparing multiple machine learning algorithms and metrics for estrogen receptor binding prediction. *Mol Pharm.* (2018) 15:4361–70. doi: 10.1021/acs.molpharmaceut.8b00546
25. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?": Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA; New York, NY: Association for Computing Machinery (2016). p. 1135–44.
26. Zhai Q, Lin Z, Ge H, Liang Y, Li N, Ma Q, et al. Using machine learning tools to predict outcomes for emergency department intensive care unit patients. *Sci Rep*. (2020) 10:20919. doi: 10.1038/s41598-020-77548-3
27. Zhong Z, Yuan X, Liu S, Yang Y, Liu F. Machine learning prediction models for prognosis of critically ill patients after open-heart surgery. *Sci Rep.* (2021) 11:3384. doi: 10.1038/s41598-021-83020-7

28. Lin WJ, Chen JJ. Class-imbalanced classifiers for high-dimensional data. *Brief Bioinform.* (2013) 14:13–26. doi: 10.1093/bib/bbs006
29. Cadrin-Tourigny J, Bosman LP, Nozza A, Wang W, Tadros R, Bhonsale A, et al. A new prediction model for ventricular arrhythmias in arrhythmogenic right ventricular cardiomyopathy. *Eur Heart J.* (2019) 40:1850–8. doi: 10.1093/eurheartj/ehz103
30. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* (2006) 26:565–74. doi: 10.1177/0272989X06295361
31. Albanese M, Alpaslan K, Ouarrak T, Merguet P, Schneider S, Schols W. In-hospital major arrhythmias, arrhythmic death and resuscitation after successful primary percutaneous intervention for acute transmural infarction: a retrospective single-centre cohort study. *BMC Cardiovasc Disord.* (2018) 18:116. doi: 10.1186/s12872-018-0851-z
32. Al-Ghamdi MA. Morbidity pattern and outcome of patients admitted in a coronary care unit: a report from a secondary hospital in southern region, Saudi Arabia. *J Commun Hosp Intern Med Perspect.* (2018) 8:191–4. doi: 10.1080/20009666.2018.1500421
33. Ruiz-Bailén M, Aguayo de Hoyos E, Ramos-Cuadra JÁ, Diaz-Castellanos MÁ, Issa-Khozouz Z, Reina-Toral A, et al. Influence of age on clinical course, management and mortality of acute myocardial infarction in the Spanish Population. *Int J Cardiol.* (2002) 85:285–96. doi: 10.1016/S0167-5273(02)00187-0
34. Barakat K, Wilkinson P, Deaner A, Fluck D, Ranjadayalan K, Timmis A. How should age affect management of acute myocardial infarction? A prospective cohort study. *Lancet.* (1999) 353:955–9. doi: 10.1016/S0140-6736(98)07114-1
35. Churpek MM, Yuen TC, Winslow C, Robicsek AA, Meltzer DO, Gibbons RD, et al. Multicenter development and validation of a risk stratification tool for ward patients. *Am J Respir Crit Care Med.* (2014) 190:649–55. doi: 10.1164/rccm.201406-1022OC

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.

frontiers | Frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 25 May 2022
doi: 10.3389/fcvm.2022.763048

Left Ventricular Pressure Estimation Using Machine Learning-Based Heart Sound Classification

Philip Westphal[1,2], *Hongxing Luo*[1], *Mehrdad Shahmohammadi*[3], *Luuk I. B. Heckman*[1], *Marion Kuiper*[1], *Frits W. Prinzen*[1], *Tammo Delhaas*[3] *and Richard N. Cornelussen*[1,2*]

[1] *Department of Physiology, Cardiovascular Research Institute Maastricht (CARIM), Maastricht, Netherlands,* [2] *Bakken Research Center, Medtronic, plc, Maastricht, Netherlands,* [3] *Department of Biomedical Engineering, Cardiovascular Research Institute Maastricht (CARIM), Maastricht, Netherlands*

Objective: A method to estimate absolute left ventricular (LV) pressure and its maximum rate of rise (LV dP/dtmax) from epicardial accelerometer data and machine learning is proposed.

Methods: Five acute experiments were performed on pigs. Custom-made accelerometers were sutured epicardially onto the right ventricle, LV, and right atrium. Different pacing configurations and contractility modulations, using isoflurane and dobutamine infusions, were performed to create a wide variety of hemodynamic conditions. Automated beat-by-beat analysis was performed on the acceleration signals to evaluate amplitude, time, and energy-based features. For each sensing location, bootstrap aggregated classification tree ensembles were trained to estimate absolute maximum LV pressure (LVPmax) and LV dP/dtmax using amplitude, time, and energy-based features. After extraction of acceleration and pressure-based features, location specific, bootstrap aggregated classification ensembles were trained to estimate absolute values of LVPmax and its maximum rate of rise (LV dP/dtmax) from acceleration data.

Results: With a dataset of over 6,000 beats, the algorithm narrowed the selection of 17 predefined features to the most suitable 3 for each sensor location. Validation tests showed the minimal estimation accuracies to be 93% and 86% for LVPmax at estimation intervals of 20 and 10 mmHg, respectively. Models estimating LV dP/dtmax achieved an accuracy of minimal 93 and 87% at estimation intervals of 100 and 200 mmHg/s, respectively. Accuracies were similar for all sensor locations used.

Conclusion: Under pre-clinical conditions, the developed estimation method, employing epicardial accelerometers in conjunction with machine learning, can reliably estimate absolute LV pressure and its first derivative.

Keywords: heart sound, hemodynamics, cardiac resynchronization therapy, artificial intelligence, machine learning, animal, epicardial acceleration

OPEN ACCESS

Edited by:
Shane Nanayakkara, Alfred Hospital, Australia

Reviewed by:
Palani Thanaraj Krishnan, St. Joseph's College of Engineering, India
Espen W. Remme, Oslo University Hospital, Norway

***Correspondence:**
Richard N. Cornelussen
richard.cornelussen@medtronic.com

Specialty section:
This article was submitted to Cardiac Rhythmology, a section of the journal Frontiers in Cardiovascular Medicine

Received: *23 August 2021*
Accepted: *20 April 2022*
Published: *25 May 2022*

Citation:
Westphal P, Luo H, Shahmohammadi M, Heckman LIB, Kuiper M, Prinzen FW, Delhaas T and Cornelussen RN (2022) Left Ventricular Pressure Estimation Using Machine Learning-Based Heart Sound Classification. Front. Cardiovasc. Med. 9:763048. doi: 10.3389/fcvm.2022.763048

INTRODUCTION

Heart failure is a major public health concern for healthcare systems that struggle to treat the ∼37 million patients worldwide (1). The Western world alone experiences more than 1 million hospitalizations each year, a number that rapidly increases due to the aging population groups (2–4). Cardiac decompensation results in more frequent/prolonged hospitalization of patients, causing treatment costs to increase while reducing the quality of life and life expectancy (5, 6). Therefore, tools that may prevent hospitalizations in patients with heart failure would be beneficial (7).

Continuous hemodynamic monitoring improves the conventionally static behavior of current treatment methods, reducing the need for follow-up visits or hospitalization. Several implantable devices have been developed to optimize therapy and identify decompensation episodes in an early stage (7, 8).

For this purpose, the measurement of left ventricular pressure (LVP) or its first derivative would be the first choice. However, the required invasive intervention and potential complications like drift, sensor overgrowth, leakage, and embolization make this approach less suitable for chronic applications (9). As a surrogate for LV function, the Chronicle^{TM} system measures RV pressure. Clinical studies show promise in reducing readmission rates due to congestive heart failure (10). Similarly, the CardioMEMS^{TM} system consists of a small pressure-sensing device that is implanted directly into the pulmonary artery (11). However, these systems are expensive, stand-alone devices and it is unclear how right-sided measurements are related to LV function (10).

An alternative method is the use of accelerometers. These are small mechano-sensors that can easily be integrated with, for example, devices and catheters. The best-known example is currently the Peak Endocardial Acceleration (PEA, later renamed as SonR) system that uses an accelerometer integrated into an implantable right atrial or RV pacing lead (12) to measure the amplitude of the first heart sound (13, 14). The RESPOND-CRT trial demonstrated that this automatic, SonR sensor-guided optimization of pacemaker therapy was safe and slightly superior to the conventional Echo-guided optimization (15). However, the SonR system does not provide absolute pressures.

In recent years machine learning (ML) has rapidly developed. ML is a computational discipline focused on building algorithms that model or recognize (complex) patterns or characteristics within large amounts of data. It is used increasingly in the heart failure space either prognostically or as in this article diagnostically (16). A previous study has indicated successful classification of heart sounds for valvular diseases *via* machine learning-based methods (17). We hypothesized that machine learning may improve the analysis of accelerometer data to the extent that also absolute values of hemodynamic parameters can be estimated.

Abbreviations: AV, atrio-ventricular; BiV, biventricular; DOB, dobutamine; ECG, electrocardiogram; ISO, isoflurane; LV, left ventricle; LVP (max), left ventricular (maximal) pressure; LV dP/dtmax, maximum rate of rise of LVP; VV, intraventricular.

For this purpose, animal experiments were performed, where gold standard pressure and accelerometer measurements were recorded under widely varying hemodynamic conditions and at different cardiac anatomic positions. Automatic accelerometer classification was facilitated *via* beat-by-beat segmentation of accelerometer and pressure signals. After extraction of acceleration and pressure-based features, a model was trained using machine learning to estimate absolute values of LVPmax and its maximum rate of rise (LV dP/dtmax).

METHODOLOGY

Study Overview

A total of five acute open chest (weighing 60–65 kg) experiments in pigs were performed in accordance with Dutch Law on Animal Experimentation and the European Directive for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes. The protocol was approved by the Central Committee for Animal experiments (CCD) in the Netherlands and the Animal Experimental Committee of Maastricht University.

Experimental Setup

The animals were premedicated with Zoletil (5–8 mg/kg) whereafter anesthesia was induced using thiopenthal (5–15 mg/kg IV). Propofol (2.5–10 mg/kg/h), sufentanyl (4-8 μg/kg/h), and rocuronium (0.1 mg/kg/h) were given at regular intervals to maintain the anesthesia. Heparin was given throughout the experiment as an anticoagulant to suppress blood clotting.

Data acquired during the experiments consisted of electrocardiogram (ECG), LVP, and epicardially measured acceleration signal. An overview of the experiment and the data analysis are depicted in **Figure 1**.

ECG measurements were acquired using the limb-leads. Pressures were measured using 7F catheter-tip manometers (CD-Leycom, Zoetermeer, the Netherlands). Under fluoroscopic guidance, pressure catheters were inserted *via* the carotid artery and the jugular vein into both the right and left ventricles as well as the right atrium. In addition, a pressure catheter was placed in the aorta.

Pacing leads were inserted transvenously into the right atrium and right ventricle (RV), while an LV lead was placed epicardially on the LV free wall, using a plunge electrode introduced into the thorax through a small incision. Pacing thresholds were evaluated on an individual basis for each electrode.

Radiofrequency ablation of the atrioventricular node was used to an create atrio-ventricular block (AVB). The process made use of an ablation catheter (MarinR, Medtronic plc.) and a radio frequency power generator (Atakr, Medtronic plc.)was performed under fluoroscopic guidance. After the creation of AVB, ventricular pacing was initiated for hemodynamic stability.

Mechanical data was acquired *via* custom-made epicardial mechano-sensors, designed to facilitate recordings of high acquisition resolution and sampling rate while keeping the overall size of the sensor to a minimum dimension of 3.3 mm (X) * 5 mm (Y) * 1.6 mm (Z). Each sensor package consisted of 3* Hall-effect-based accelerometers, perpendicularly aligned to each other, sensing at a resolution of 16-bit at a sampling

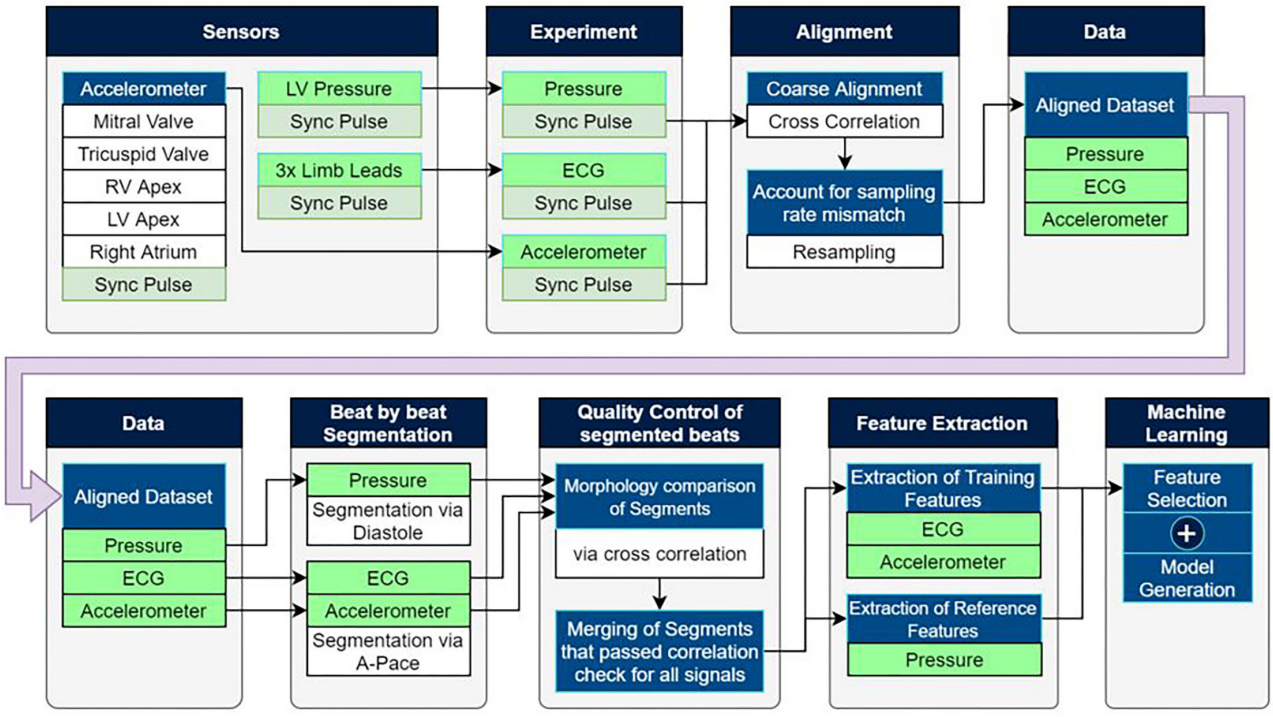


FIGURE 1 | Experimental setup as well as the post-processing pipeline used to extract training data for machine learning based estimation models.

rate of 1,000 Hz. Each sensor was paired with a single analog to digital converter (ADC) (a total of 3), allowing synchronous capture of data. The data were recorded in a range of ±4 g to accommodate a minimum range of ±3 g when accounting for gravity with a 16-bit digitization resolution. A custom device was made to facilitate the simultaneous acquisition of the mechano-sensors and to allow simple integration into existing systems *via* a shared synchronization pulse.

Alignment

Alignment of the datasets recorded from the ECG, pressure and mechano-sensors was performed *via* a shared synchronization pulse that is broadcasted by the acquisition device. Each system/device connected to the synchronization pulse generated a tracing on a separate channel and was contained within each of the datasets at the end of the experiment. Using this signal, dropped/duplicate samples or mixed sampling frequency systems can be recognized, and the section of the signal was resampled automatically.

Embedded in a silicone suturing fixture, the mechano-sensors were sutured onto the epicardium through a small thoracic incision. The dimensions of the fixture allowed only for one-sided mounting of the sensor and enforced correct/consistent orientation to maintain signal uniformity between the experiments by aligning the sensor's Z-axis perpendicular with the tissue surface. The thorax was partially closed to minimize the effect on the animal's hemodynamics after the placement of the sensors. A small hole was retained for the cabling without hermetic sealing and therefore the experiments remained "open chest."

A total of five sensors were attached to the tissue at the locations shown in **Figure 1**, being the LV and RV apex, RV and LV free wall close to the mitral and tricuspid valves, as well as the right atrium.

Experimental Protocol

Pacing protocols consisted of RV and biventricular (BiV) pacing with incrementally increasing atrioventricular (AV) pacing delays, ranging from 50 to 300 ms, and incrementally increasing interventricular (VV) delays ranging from −150 to +150 ms. The pacing protocol was performed in DDD mode and repeated under different hemodynamic loading conditions. Dobutamine (DOB) was used to increase cardiac function and its dose was adjusted to reach approximately twice the baseline LV dP/dt max value. After a sufficient weaning period from the dobutamine, the animals were ventilated with the cardiovascular depressant isoflurane (ISO) to decrease the baseline LV dP/dt(max) value to around half of the baseline value. Pacing protocols were repeated during both dobutamine and isoflurane administration. Each setting maintained a 60 s recording time unless the applied pacing setting appeared detrimental to the animal's hemodynamic state.

Data Analysis

Please note that more details regarding the data analysis can be found in the **Supplementary Material**. The experimental setup made use of multiple standalone recording devices with different sampling rates which shared an auxiliary synchronization pulse

as shown in **Figure 1.** Matching the pulse between devices allowed for precise temporal alignment of the ECG, pressure, and mechanosensory data.

The automated post-processing pipeline segmented the input signals into individual cardiac cycles. Training features were extracted from the ECG and acceleration signals, while reference features were extracted from the LVP signals. A morphological cross-correlation analysis was performed within each pacing setting to identify the largest coherent group of cardiac cycles and remove deviating beats containing artifacts or non-typical paced beats.

ECG analysis employed a band-pass filter to remove the DC-offset and the effects of respiration drift. Individual cardiac cycles were segmented by identification of atrial pacing spikes after which training features were extracted from the signal.

As acceleration signal's energy, during occurrences of the heart sounds, of S1 and S2 was found to be negligible for frequencies above 250 Hz, a (10–250 Hz) bandpass filter was used to remove DC-offset and high-frequency noise.

The atrial pacing spike was used to segment each beat of the acceleration signal post-alignment. Features reflecting amplitude, timing, and energy of S1 and S2 were extracted from the signals for the machine learning process. A selection of the 17 (based upon literature) predefined training features extracted from the acceleration signal are shown in **Table 1**. **Figures 2**, **3** depict examples of the features shown in **Table 1**.

The pressure analysis segmented the pressure signal during the center of the diastolic phase and extracted the largest, morphologically coherent, group of beats to account for morphology changes caused by factors such as respiration, prior systolic pressures, and independency from the ECG annotation algorithm. This allowed complete morphological assessment of the pressure curve during the cross-confirmation stage.

Following the segmentation of each heartbeat, reference features were extracted from the LV signal being LVP max and LV dP/dt max.

Only the acceleration-based features from each Z-axis sensor were retained because this direction is the most reproducible, being perpendicular to the epicardium, resulting in better interpretable signals, and reducing computational overhead.

The automated morphological assessment was performed on all input signals to remove irregular heartbeats. All cardiac cycles within each pacing configuration were resampled to match the most reoccurring number of samples per segment to improve the comparison of the individual beats. For each pacing configuration, the largest group of beats, providing the greatest amount of coherence were evaluated using the cross-correlation coefficient for each beat permutation (18). This comparison was performed on the ECG, pressure, and acceleration signal separately to ensure that each beat's mechanic and electronic response conform with each other. Beats that displayed large morphological deviations from the rest of the beats within each pacing setting were excluded from further processing.

Cardiac Function Modeling

A decision tree (19–22) based machine-learning model was employed to estimate cardiac function in form of absolute LVP and/or its first derivative. The model structure consists of a multi-class classification system facilitated by decision trees. To reduce the high amount of variance demonstrated by individual decision trees, a bootstrap aggregated ensemble was used, for which several subsets of training data were used in the training of individual trees.

During the training process, acceleration feature-based rules were generated to allow optimal estimation of the absolute maximum LVP and/or its derivative. The Gini's diversity index (Equation 1) aids in maximizing information gain for each decision tree by identifying splits in the training data that reduce the probability of misclassification and hence maximize estimation performance.

$$Gini\ index\ \left(D_p, f\right) = I\left(D_p\right) - \sum_{j=1}^{m} \frac{N_j}{N} I(D_j) \tag{1}$$

Gini index = identifier used to reduce misclassification probability; f = analyzed feature (subset); D_j = samples at the child node; D_p = samples of the parent node; I = current node;

N_j = total number of samples available at the current node; N = total number of samples.

This process iteratively refined each decision tree by subdividing the training dataset into smaller sub-categories until the maximum number of splits was reached or the remaining data did not require any more subdivisions.

Feature selection consisted of an iterative process that generated multiple competing estimation models. Each of these models was based on a unique permutation of the available feature sets for their training. This result allowed investigation into the estimation potential of individual features and their potential to be complementary with secondary and/or tertiary features.

To prevent over-fitting, the number of features used in each permutation was limited to a total of three. To reduce the number of permutations, the process started with a single feature and optimized estimation accuracy or loss by selectively adding and/or replacing features until the maximum potential was reached.

Performance metrics applied in the model training process were estimation –accuracy/–loss (Equations 2 and 3). Both metrics were evaluated during 10 × k-fold cross-validation, using the mean value of all folds. The mean accuracy evaluates the number of correctly against incorrectly estimated values. Alternatively, the K-fold validation loss was used. Validation loss penalizes larger discrepancies in misclassifications significantly higher than small discrepancies. While this study primarily focused on estimation accuracy, estimation loss was used as validation to confirm model performance.

$$\text{Loss}\left(y, \hat{y}\right) = \frac{1}{N} \sum_{i=0}^{N} \left(y - \hat{y}_i\right)^2 \tag{2}$$

$$Validation\ accuracy = \frac{1}{N} \sum_{i=0}^{N} y_p \tag{3}$$

N = total number of classification attempts; y_p = correctly estimated classification; Loss$(y, \hat{y})$ = validation loss for single

TABLE 1 | Training features that are extracted from regions of interest (**Figure 3**) located around S1 & S2 of the acceleration and ECG signals.

	Features	Code	Details
Amplitude	Amplitude (max)	A1 or A2	The maximum – minimum amplitude of S1 & S2 derived from the acceleration signal.
	Differential maxima	B1 or B2	The maximum – minimum differential amplitude of S1 & S2 derived from the acceleration signal.
	Envelope	C1 or C2	Integral of the heart sound signal.
Energy	Shannon energy	D1 or D2	Attenuates high amplitude signals and provides higher weight toward low intensity content. $E = -x^2 \log(x^2)$. (E=Energy||x=Signal).
	Shannon energy integral	E1 or E2	Integral of Shannon Energy.
	Shannon entropy	F1 or F2	Emphasizes medium strength amplitudes while attenuating low & high intensity amplitudes $E = -\|x\| \log \|x\|$.
	Shannon entropy integral	G1 or G2	Integral of Shannon Entropy.
Temporal	S1 abs max to S2 max interval	H	Interval between maximum positive rectified S1 & S2 of the acceleration signal.
	S1 & S2 max to Vpace interval	I	The maximum amplitude location with respect to the left ventricular pacing spike was measured.
	S1 & S2 min to Vpace interval	J	The minimum amplitude (-ve peak) location with respect to the left ventricular pacing spike was measured. Interval between maximum negative rectified S1 & S2 of the acceleration signal.

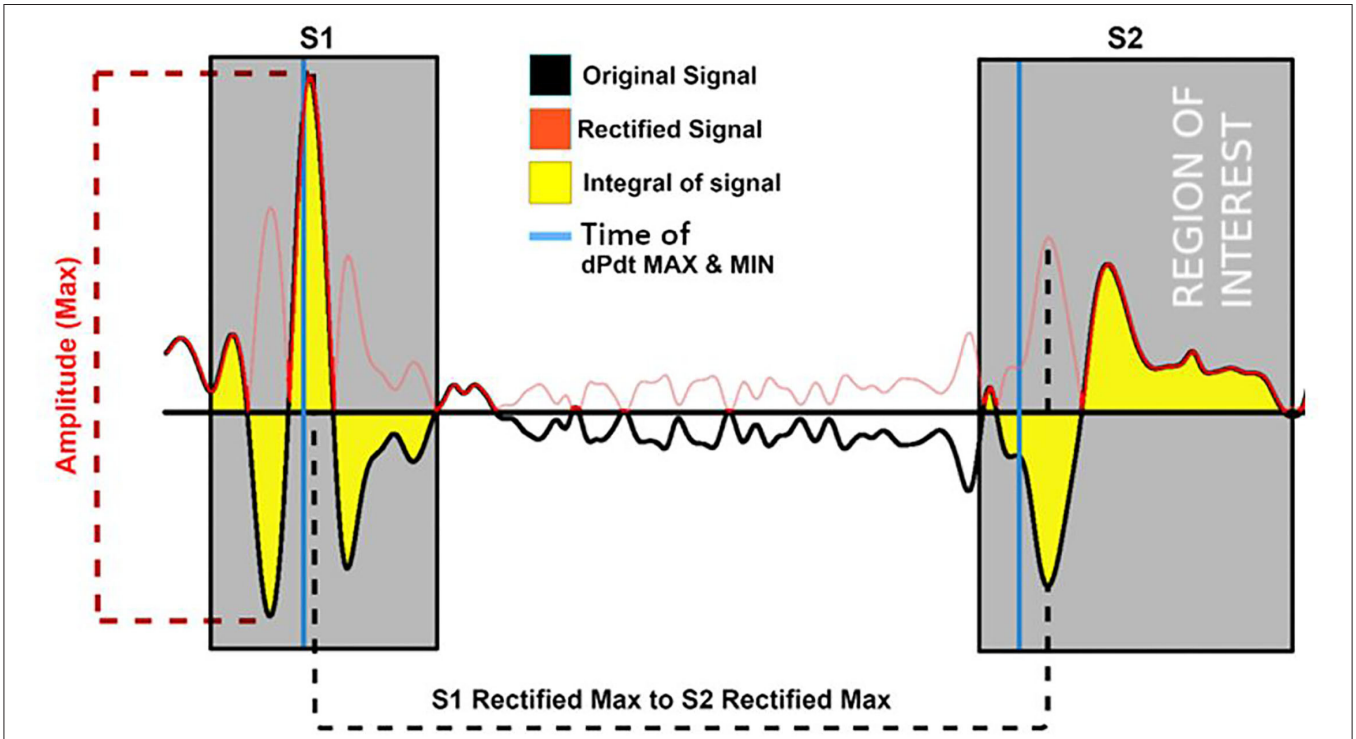


FIGURE 2 | Examples of the training features extracted from the accelerations signal (black) and its rectified version (red). Features depicted are the maximum amplitude of the signal; its integral and the location time duration between the rectified S1 and S2 maximum of the rectified (negative and positive as positive) signals. Gray area = region of feature extraction.

sample; N = total number of classification attempts; y = true classification; $\hat{y}$ = estimated classification.

Model validation was addressed by using the above described k-fold validation method in addition to holdout validation, to address potential issues that may arise from over/underfitting. In addition, each feature permutation used to generate an estimation model was limited to a maximum of three to prevent the chance of overfitting. Before the training of the estimation model, for the personalized model, 5% of the dataset was removed for illustrative purposes *via* holdout validation of the final model. Depending on the availability of data in each pressure category, the selection of validation samples was reduced to retain a robust training dataset. The model's performance was evaluated using the average k-fold

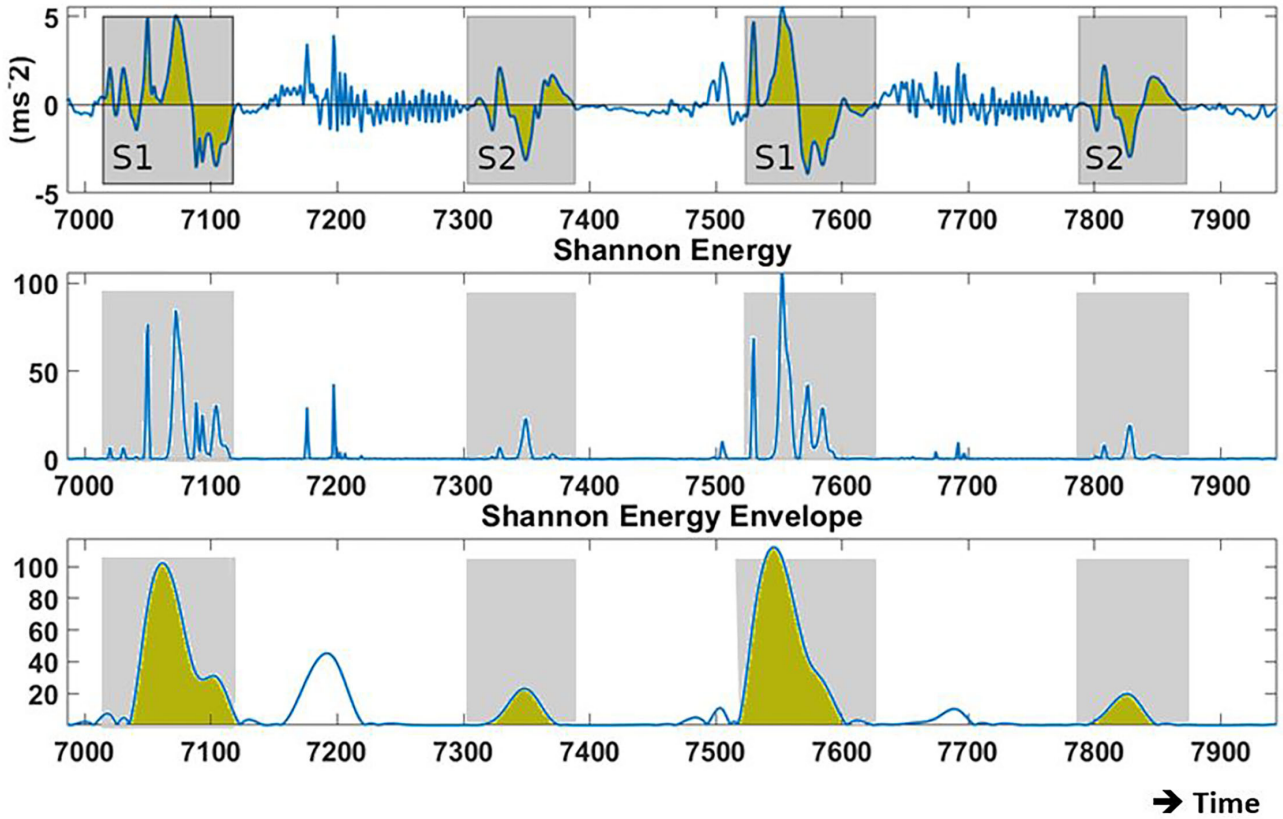


FIGURE 3 | Examples of the recorded baseline acceleration signal (Bi-Ventricular (BIV) pacing, Atrio-Ventricular (AVd): 150 ms) and its Shannon energy from which training features are derived. (Top) acceleration during two cardiac cycles in proximity to the mitral valve. (Center) Shannon energy of the original signal, pronouncing low intensity amplitudes over high intensity amplitudes. (Bottom) Displays the envelope based on the Shannon energy. Blue line = raw signal; Yellow area = signal integral; Gray area = region for feature extraction.

validation whose results are shown in **Table 2**. The "Leave one out"/generalized method used k-fold validation to find the most performant model, while one animal was completely removed from the training set. The model was then validated with holdout estimation whose results are given in **Table 3**.

During the modeling process, each feature permutation underwent k-fold validation which ensured that rules generated on subsets of data were applicable to the remaining dataset.

Statistics

We have performed a one-way analysis of variance (ANOVA) for testing significance ($p < 0.05$) between bin-size (low-medium-high) as well as between personalized and generalized models in both accuracy and loss estimations.

RESULTS

Figure 4 displays an example of the acquired signals during the 3 hemodynamic steady states: baseline, isoflurane, and dobutamine. The profound negative and positive effects on hemodynamics can be seen in isoflurane and dobutamine, respectively. Also, the specific acceleration signals of the 5 anatomic sensors and their response to changing hemodynamics are illustrated in **Figure 4**.

Post-experiment segmentation processing resulted in a dataset of over 6,000 cardiac cycles that consisted of a complete annotated set of training and reference data. The final selection of the best-performing models is listed in **Table 4**. Of all initially proposed features in **Table 1**, acceleration amplitude and/or energy-based features proved to correlate best to the hemodynamic variables; with S1/S2 maximum amplitude and S1/S2 integral appearing most frequently in the final models. In contrast, time-based features were widely neglected and only the feature expressing the duration between the maximum amplitude of S1 and S2 acceleration signals showed any re-occurrence in high-performing models.

Examples of the model performances are depicted in **Figure 5** which shows the result of four models tested under holdout validation with data extracted from the sensor in the accelerometer in the proximity of the mitral valve. Each row represents the holdout validation results of two models, with increasing resolution, estimating LVPmax and LV dP/dt max, respectively.

TABLE 2 | Accuracies for estimating LVP and LV dP/dtmax at different levels of resolution and for all sensor locations.

	Mitral valve (%)		LV apex (%)		RV apex (%)		Right atrium (%)		Tricuspid valve (%)	
Optimization method	**Acc**	**Loss**	**Acc**	**Loss**	**Acc**	**Loss**	**Acc**	**Loss**	**Acc**	**Loss**
Left ventricular absolute maximum pressure (range 60–130 mmHg)										
Resolution (Interval)										
20 mmHg	96	94	93	93	93	93	94	90	94	91
10 mmHg	87	87	88	87	88	87	87	86	89	87
5 mmHg	81	78	83	79	84	80	82	75	83	82
Left ventricular dP/dt max pressure (range 600–1,600 mmHg/s)										
Resolution (Interval)										
200 mmHg/s	93	90	94	94	94	89	94	92	96	94
100 mmHg/s	90	90	90	88	88	88	88	88	87	86
50 mmHg/s	86	85	86	85	85	84	86	85	86	84

Optimization methods: (Acc): accuracy optimization || (Loss): loss optimization. All values are evaluated using average 10×k-fold validation results from the respective estimation models. LV, left ventricular; RV, right ventricular.

TABLE 3 | Accuracy optimization using holdout validation results generated by the "Leave one out" method.

LV dPdt(max) 200 mmHg/s	**Mitral valve (%)**	**Left ventricular apex (%)**	**Right ventricular apex (%)**	**Right atrium (%)**	**Tricuspid valve (%)**
Leave out Animal 1	83	81	80	81	82
Leave out Animal 2	90	85	81	83	87
Leave out Animal 3	82	79	82	80	82
Leave out Animal 4	71	69	67	73	74
Leave out Animal 5	81	84	78	79	77

Each row indicates a model that excludes and was validated on the animal listed in the first column.

Table 2 displays the results obtained for all sensor locations for both estimation accuracy and loss optimization and three levels of resolution for LVP and LV dP/dtmax. Accuracies for LVP with bin-size of 20 mmHg ranged between 90 and 96%, for bin-size of 10 mmHg between 86 and 89%, and for bin-size of 5 mmHg between 75 and 83%, with small non-significant differences between the sensor locations. Similar results were obtained for LV dP/dtmax using differences of 200, 100, and 50 mmHg/s, wherein accuracy ranges of 89–96%, 86–90%, and 84–86%, were achieved respectively. An overview of the results is given in **Table 2**.

Bin-size significantly ($p < 0.001$) affected both accuracy and loss estimations, indicating increased accuracy/loss at larger bin-size.

An additional investigation was performed using the "Leave one out" method, which validates estimation models, that are trained on N-1 participating subjects, against the remaining subject. Using the lowest estimation resolution for the estimation of LV dP/dtmax, an average estimation accuracy evaluated across all sensing locations was 80% with an SD of 5.4% (**Table 3**). "As expected, there was a significantly ($P < 0.01$) lower estimation accuracy in the "leave-one-out"/generalized model than the personalized model, comparing both at a bin-size of 200 mmHg/s."

The holdout validation figures (in **Table 3**) indicated about a 15-percentage point lower accuracy than the k-fold validation results listed in **Table 2**.

DISCUSSION

This study provides the proof of principle for a novel method for estimating absolute LVPmax (and LV contractility) using machine learning analysis of epicardial accelerometer signals. The majority of features that contributed to our prediction models were related to the amplitude and energy of the accelerometer signal and very few related to the timing of them. Accuracies were similar for all five sensor locations.

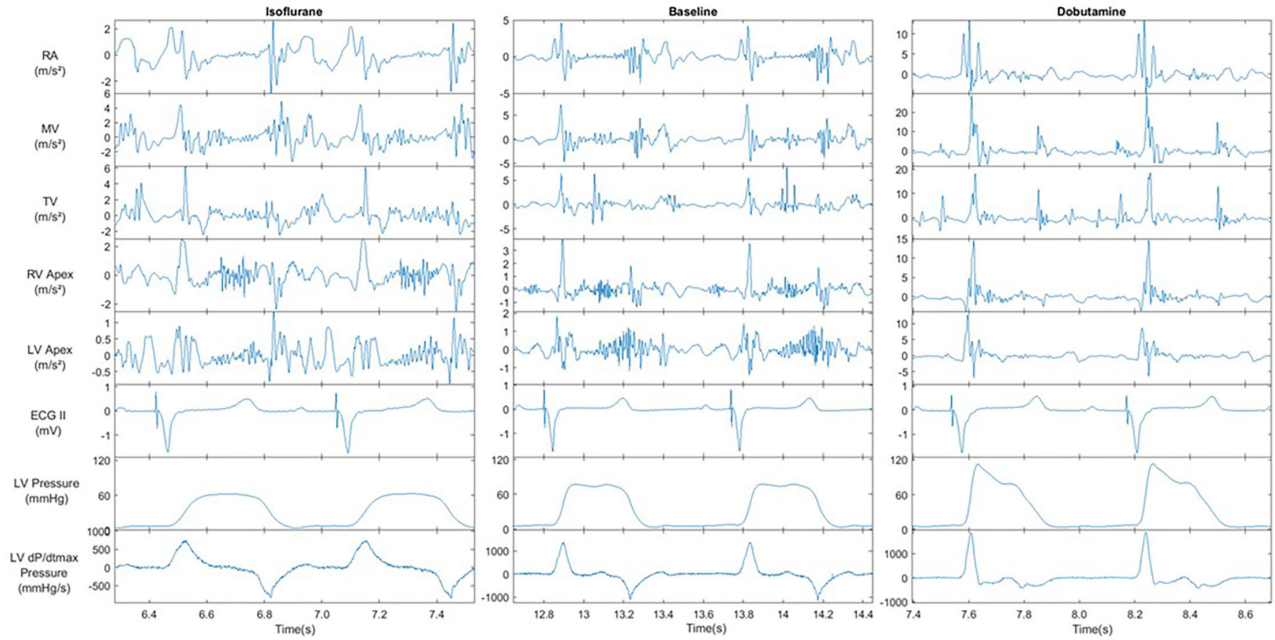


FIGURE 4 | Recorded signals of Left ventricular pressure, ECG and all acceleration recording sites under the influence of cardiovascular modifiers. All signals were recorded @BIV | AVd = 150 ms | VVd = 0. Acceleration signal show the result of the X, Y and Z axis magnitude filtered between 1-150 Hz. AVd = atrio-ventricular delay, VV = intra-ventricular delay.

TABLE 4 | Most relevant features according to their re-occurrence when generating accuracy optimized models.

	LV Pressure (max)			LV dP/dt (max)		
Sensing location	**Feature 1**	**Feature 2**	**Feature 3**	**Feature 1**	**Feature 2**	**Feature 3**
Mitral valve	C2	B1	C1	G1	C1	C2
Left ventricular apex	A1	C1	C2	C2	A1	C1
Right ventricular apex	E2	A1	C1	A1	G1	G2
Right atrium	C1	G2	A1	G2	H	C1
Tricuspid valve	C2	C1	A1	A1	C1	C2

*Feature explanations are given in **Table 1**. 1 and 2 designations refer to the acceleration-based equivalent to the occurrences to S1 and S2 heart sounds.*

Comparison to Other Studies

These results significantly extend the application of mechano-sensors in estimating cardiac function, which is so far largely limited to optimization of pacemaker settings by the SonR system (15, 23), without the knowledge of absolute values of pressures. Another study on pigs showed that an epicardially placed accelerometer can be used to assess changes in preload, and so filling status, using the frequency of myocardial acceleration (24). In preclinical studies, Thakur et al. showed that analysis of S1 and S3 amplitude signals from accelerometers integrated into an implanted pacemaker device was able to monitor the change of absolute left atrial pressure over time (25). While these studies showed the relation between accelerometer signals and hemodynamic status, this study, to the best of our knowledge, is the first to show that absolute LV pressure and contractility levels can be derived from accelerometers. Moreover, the finding that such estimations can be obtained from various (atrial, right, and left ventricular) locations is novel. Potentially, the introduction of additional features derived from, e.g., frequency components or gyroscopic signals (easily imbedded in current accelerometer sensors) may contribute positively to our developed pressure estimation model (25, 26).

Possible Need for Personalization

Two ways of model development were used, namely, a personalized model (including all available subjects' data) and a generalized model (leave-one subject's data out). In the former approach, k-fold validation from all 5 individuals was used while holdout validation was used for the latter. The approach clearly shows the proof of concept. However, when applying this approach to the clinical situation it would require a period of validation/calibration with a gold-standard measurement before continuing with the mechano-sensor information only (see

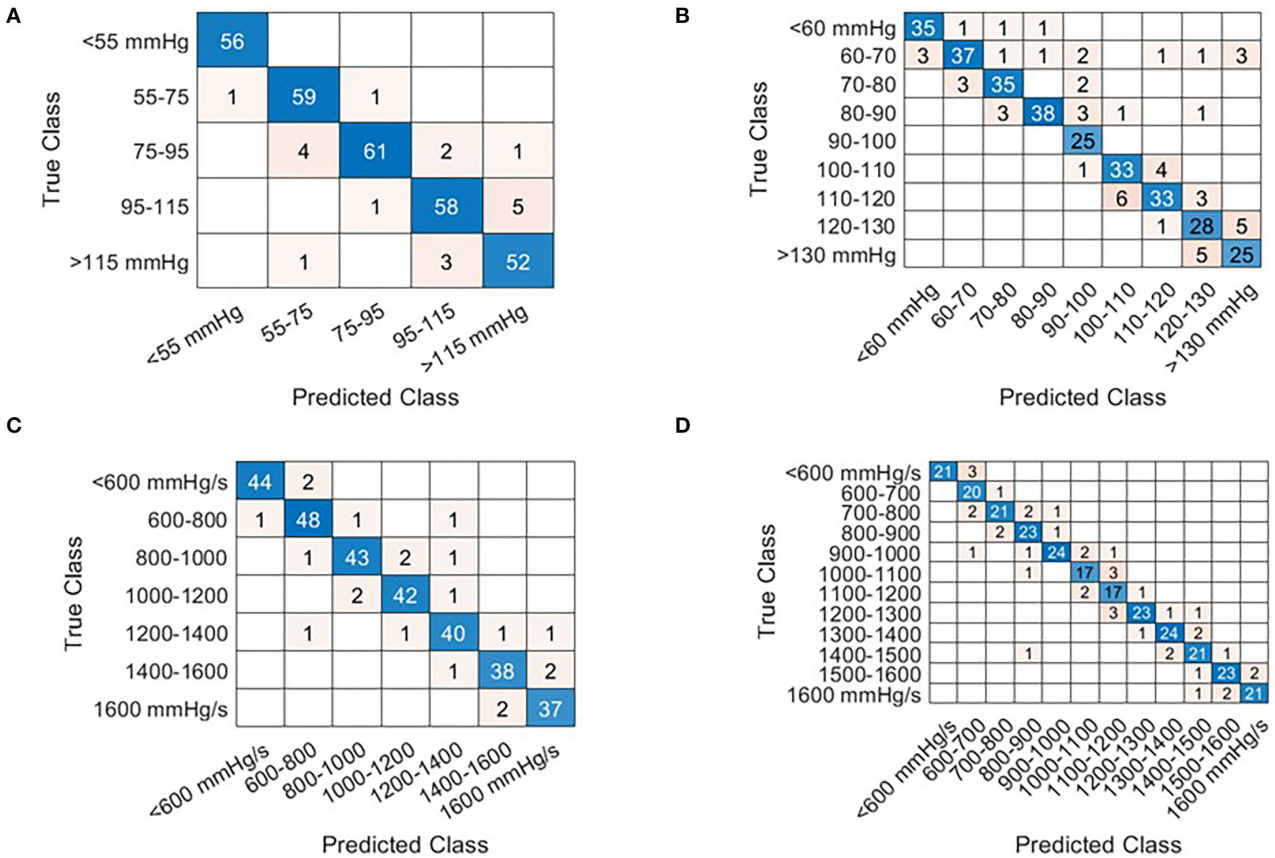


FIGURE 5 | Confusion matrices that illustrate the correctly and incorrectly estimation results for each given beat via holdout validation from the Mitral valve sensor. The row of the matrix corresponds to the true class while columns correspond to the predicted class. Diagonal entries correspond to correct estimates while off-diagonal entries represent incorrect estimates. The beats selected for holdout validation were selected at random. At low prevalence of cardiac cycles in any given category, the number of selected samples is reduced in favor of the training dataset. Examples are shown for LVPmax [**(A,B)** Interval bins of 20 and 10 mmHg respectively] and LV dP/dt(max) **(C,D)** interval bins of 200 mmHg/s, **(D)** 100 mmHg/s.

below). Because such an approach may not always be possible or desirable, the second option shows that using the information of, in our case four individuals, the mechano-sensor information of the fifth individual is still quite reliable. In the clinical situation, such a development-set is likely to be considerably larger than four, which can be expected to significantly increase the accuracy of the leave-out approach in the clinical setting. With either approach, it may also be required to create different prediction models for the various sensor locations. While accuracy was comparably high for all sensor locations, the optimal three features to reach this result differed to some extent. This "location independence" is reflected in the results wherein differences in recording locations showed only minute changes in estimation performance while selecting dissimilar features (see **Table 4**).

Further Possible Applications/Integration

The proposed method may be used in general hemodynamic monitoring applications at low-to-moderate resolution (e.g., intensive care unit). Secondly, it may be used to track cardiac function to detect decompensation in an early stage. The currently developed method allows for simple integration into embedded software of monitoring and other (e.g., pacemaker, ICD) devices (27). With respect to the latter, our method may assist in identifying life-threatening arrhythmias by adding the hemodynamic analysis on top of the current electrophysiological analysis in these devices to prevent unnecessary defibrillatory interventions (28, 29).

While the present study used epicardial sensors, the approach used in this study may also be applied to less invasive acceleration measurements, such as heart sounds measured on the skin, or microphones or accelerometers mounted on or inside implanted devices or pacing/defibrillation leads. Calibration steps can be either procedure-related or added to the procedure and can be either invasive (like in the current study) or non-invasive, e.g., using a common pressure cuff or finger plethysmography as shown for CRT patients (30, 31). Different calibration circumstances can also be envisioned such as rest/exercise/pacing intervention in case of an implantable pulse generator placement

used in CRT. Supplementary data of patients, with similar conditions and equipment setup, can be used to improve the robustness of the model during implantation or any follow-up session.

The processing pipeline, used for an automated beat and feature selection, may be of clinical use. The implemented method for synchronization of acceleration signals and ECG accounts for differences in sampling rates and standardizes the acquisition process before analysis. The prerequisite for this integration was limited to a single auxiliary recording channel with a synchronization pulse. This pipeline may also be useful for the rapid generation of estimation models with new data, including those in patients during a period of monitoring or the implantation of pacemakers or other monitoring devices. Scalability of estimation models allows accommodation of device constraints such as processing overhead, battery life, and cost. A tradeoff exists between model complexity and its estimation accuracy, wherein a reduction in complexity may cause a reduction in estimation accuracy and vice versa [i.e., the use of neural networks (17, 32)].

For specific clinical applications, the method of feature acquisition post-implantation may have to be altered in case ECG, which is used as the reference signal, is not available. Alternative sources of reference can be used post-implantation as long as they include timing information about the cardiac cycle. Preferred alternatives would be the use of "surrogate ECG" provided by a defibrillator, EGM feedback, or timed pacing intervals.

Limitations of the Study

Some limitations of the study should be noted. First of all, the acquisition of mechanosensory data was performed under preclinical conditions in normal porcine hearts with an open chest. Clinical application in patients with compromised heart function should be considered with caution because several differences may apply, like differences in cardiac contractile force, the influence of the open- or closed-chest as well as differences in sources of noise. These noise sources were kept to a minimum in the controlled experimental environment, whereas in the real-life situation, environmental noise and movement artifacts may be higher because in patients with heart failure also lung sounds may become stronger. In addition, while an extensive experimental protocol with a generous variety of interventions was performed, only five animals of similar weight and dimensions were used in this trial. This may reduce the signal variability in comparison to the amount experienced in a clinical setting. Finally, the personalized model may learn to identify the animal instead of the pressure-derived estimates, therefore risk mitigation methods such as k-fold validation and ensembling are employed. In addition, the numerous (pacing) configurations provided a considerable range in LV max pressure and LV dP/dt max values. To further reduce this known phenomenon of machine learning, it is recommended that future investigations increase the number and physiological conditions of the subjects in the training set.

To adhere to the processing limitations of (cardiac) pacing devices, only features requiring little computational overhead were used. This limited feature selection to the temporal domain wherein only amplitude, timing, and energy-based features was considered. Expansion of potential features with more complex acquisition processes may further improve results.

CONCLUSION

The use of epicardial accelerometer data in combination with a bootstrapped decision tree ensemble algorithm can reliably estimate absolute hemodynamic statuses, such as intracardiac left ventricular pressure and its derivative in a controlled preclinical setting. The algorithm is simple enough to be scaled with low computational requirements to be used for monitoring cardiac function by a simple computer, microcontroller, or dedicated integrated circuitry.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because data are owned by Maastricht University and Medronic, plc. Requests to access the datasets should be directed to richard.cornelussen@medtronic.com.

ETHICS STATEMENT

The protocol was approved by the Central Committee for Animal experiments (CCD) in the Netherlands and the Animal Experimental Committee of Maastricht University.

AUTHOR CONTRIBUTIONS

PW, HL, MK, FP, and RC designed the research studies. PW, MK, LH, and HL carried out the studies. PW and LH analyzed the data. PW wrote the paper with help from HL, TD, FP, and RC. RC had primary responsibility for the final content. All authors read and approved the final manuscript.

FUNDING

This project has received funding from the European Union's Horizon 2020 Research and innovation program under the Marie Sklodowska-Curie Grant Agreement: No 764738 (Personalized *In-silico* Cardiology (PIC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2022.763048/full#supplementary-material

REFERENCES

1. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* (2012) 380:2163–96. doi: 10.1016/S0140-6736(12)61729-2
2. Davie AP, Francis CM, Caruana L, Sutherland GR, Mcmurray JJ. Assessing diagnosis in heart failure: which features are any use? *QJM.* (1997) 90:335–9. doi: 10.1093/qjmed/90.5.335
3. O'connell, J.E., Brown, J., Hildreth, A.J., and Gray, C.S. Optimizing management of congestive heart failure in older people. *Age Ageing.* (2000) 29:371–2. doi: 10.1093/ageing/29.4.371
4. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *J Am Coll Cardiol.* (2014) 63:1123–33. doi: 10.1016/j.jacc.2013.11.053
5. Martirosyan M, Caliskan K, Theuns D, Szili-Torok T. Remote monitoring of heart failure: benefits for therapeutic decision making. *Expert Rev Cardiovasc Ther.* (2017) 15:503–15. doi: 10.1080/14779072.2017.1348229
6. Jackson SL, Tong X, King RJ, Loustalot F, Hong Y, Ritchey MD. National Burden of Heart Failure Events in the United States, 2006 to 2014. *Circ Heart Fail.* (2018) 11:e004873. doi: 10.1161/CIRCHEARTFAILURE.117.004873
7. Klersy C, De Silvestri A, Gabutti G, Raisaro A, Curti M, Regoli F, et al. Economic impact of remote patient monitoring: an integrated economic model derived from a meta-analysis of randomized controlled trials in heart failure. *Eur J Heart Fail.* (2011) 13:450–9. doi: 10.1093/eurjhf/hfq232
8. Pekmezaris R, Mitzner I, Pecinka KR, Nouryan CN, Lesser ML, Siegel M, et al. The impact of remote patient monitoring (telehealth) upon Medicare beneficiaries with heart failure. *Telemed J E Health.* (2012) 18:101–8. doi: 10.1089/tmj.2011.0095
9. Kypta A, Blessberger H, Lichtenauer M, Steinwender C. Complete encapsulation of a leadless cardiac pacemaker. *Clin Res Cardiol.* (2016) 105:94. doi: 10.1007/s00392-015-0929-x
10. Teerlink JR. Learning the points of COMPASS-HF: assessing implantable hemodynamic monitoring in heart failure patients. *J Am Coll Cardiol.* (2008) 51:1080–2. doi: 10.1016/j.jacc.2007.12.009
11. Ayyadurai P, Alkhawam H, Saad M, Al-Sadawi MA, Shah NN, Kosmas CE, et al. An update on the CardioMEMS pulmonary artery pressure sensor. *Ther Adv Cardiovasc Dis.* (2019) 13:1753944719826826. doi: 10.1177/1753944719826826
12. Duncker D, Delnoy PP, Nagele H, Mansourati J, Mont L, Anselme F, et al. First clinical evaluation of an atrial haemodynamic sensor lead for automatic optimization of cardiac resynchronization therapy. *Europace.* (2016) 18:755–61. doi: 10.1093/europace/euv114
13. Bordachar P, Labrousse L, Ploux S, Thambo JB, Lafitte S, Reant P, et al. Validation of a new noninvasive device for the monitoring of peak endocardial acceleration in pigs: implications for optimization of pacing site and configuration. *J Cardiovasc Electrophysiol.* (2008) 19:725–9. doi: 10.1111/j.1540-8167.2008.01105.x
14. Delnoy PP, Marcelli E, Oudeluttikhuis H, Nicastia D, Renesto F, Cercenelli L, et al. Validation of a peak endocardial acceleration-based algorithm to optimize cardiac resynchronization: early clinical results. *Europace.* (2008) 10:801–8. doi: 10.1093/europace/eun125
15. Brugada J, Delnoy PP, Brachmann J, Reynolds D, Padeletti L, Noelker G, et al. Contractility sensor-guided optimization of cardiac resynchronization therapy: results from the RESPOND-CRT trial. *Eur Heart J.* (2017) 38:730–8. doi: 10.1093/eurheartj/ehw526
16. Plati DK, Tripoliti EE, Bechlioulis A, Rammos A, Dimou I, Lakkas L, et al. A Machine Learning Approach for Chronic Heart Failure Diagnosis. *Diagnostics (Basel).* (2021) 11:1863. doi: 10.3390/diagnostics11101863
17. Malik A, Barin S, Yuksel M. Accurate classification of heart sound signals for cardiovascular disease diagnosis by wavelet analysis and convolutional neural network: preliminary results. In: *2020 28th Signal Processing and Communications Applications Conference (SIU)*. Istanbul: Istanbul Medipol University (2020). p. 1–4.
18. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. *"Linear Correlation." §14.5 in Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press. (1992). p. 630–33.
19. Breiman L. Bagging predictors. *Mach Learn.* (1996) 26:123–40. doi: 10.1007/BF00058655
20. Breiman L. Prediction games and arcing algorithms. *Neural Comput.* (1999) 11:1493–517. doi: 10.1162/089976699300016106
21. Loh W. Regression trees with unbiased variable selection and interaction detection. *Stat Sin.* (2002) 12:361–86.
22. Tu M, Shin D. Effective diagnosis of heart disease through bagging approach. In: *2009 2nd International Conference on Biomedical Engineering and Informatics*. Tainjin (2009). p. 1–4.
23. Donal E, Giorgis L, Cazeau S, Leclercq C, Senhadji L, Amblard A, et al. Endocardial acceleration (sonR) vs. ultrasound-derived time intervals in recipients of cardiac resynchronization therapy systems. *Europace.* (2011) 13:402–8. doi: 10.1093/europace/euq411
24. Krogh MR, Halvorsen PS, Grymyr OHN, Bergsland J, Elle OJ, Fosse E, et al. Continuous estimation of acute changes in preload using epicardially attached accelerometers. *IEEE Trans Biomed Eng.* (2020) 68:2067–75. doi: 10.1109/TBME.2020.3020358
25. Thakur PH, An Q, Swanson L, Zhang Y, Gardner RS. Haemodynamic monitoring of cardiac status using heart sounds from an implanted cardiac device. *ESC Heart Fail.* (2017) 4:605–13. doi: 10.1002/ehf2.12171
26. Siecinski S, Kostka PS, Tkacz EJ. Gyrocardiography: a review of the definition, history, waveform description, and applications. *Sensors (Basel).* (2020) 20:6675. doi: 10.3390/s20226675
27. Siejko KZ, Thakur PH, Maile K, Patangay A, Olivari MT. Feasibility of heart sounds measurements from an accelerometer within an ICD pulse generator. *Pacing Clin Electrophysiol.* (2013) 36:334–46. doi: 10.1111/pace.12059
28. Luciani M, Saccocci M, Kuwata S, Cesarovic N, Lipiski M, Arand P, et al. Reintroducing heart sounds for early detection of acute myocardial ischemia in a porcine model - correlation of acoustic cardiography with gold standard of pressure-volume analysis. *Front Physiol.* (2019) 10:1090. doi: 10.3389/fphys.2019.01090
29. Covino G, Volpicelli M, Ciardiello C, Capogrosso P. Usefulness of Hemodynamic Device-Based Optimization in Heterogeneous Patients Implanted with Cardiac Resynchronization Therapy Defibrillator. *J Cardiovasc Transl Res.* (2020) 13:938–43. doi: 10.1007/s12265-020-10004-9
30. Whinnett ZI, Davies JE, Willson K, Chow AW, Foale RA, Davies DW, et al. Determination of optimal atrioventricular delay for cardiac resynchronization therapy using acute non-invasive blood pressure. *Europace.* (2006) 8:358–66. doi: 10.1093/europace/eul017
31. Whinnett ZI, Francis DP, Denis A, Willson K, Pascale P, Van Geldorp I, et al. Comparison of different invasive hemodynamic methods for AV delay optimization in patients with cardiac resynchronization therapy: implications for clinical trial design and clinical practice. *Int J Cardiol.* (2013) 168:2228–37. doi: 10.1016/j.ijcard.2013.01.216
32. Chen TE, Yang SI, Ho LT, Tsai KH, Chen YH, Chang YF, et al. S1 and S2 Heart Sound Recognition Using Deep Neural Networks. *IEEE Trans Biomed Eng.* (2017) 64:372–80. doi: 10.1109/TBME.2016.2559800

Conflict of Interest: PW and RC are Medtronic employees.

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.

The reviewer ER declared a past co-authorship with the RC to the handling editor.

frontiers | Frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 08 July 2022
doi: 10.3389/fcvm.2022.735906

Development of an Electronic Frailty Index for Predicting Mortality and Complications Analysis in Pulmonary Hypertension Using Random Survival Forest Model

Jiandong Zhou[1†], Oscar Hou In Chou[2,3†], Ka Hei Gabriel Wong[2], Sharen Lee[2], Keith Sai Kit Leung[4], Tong Liu[5], Bernard Man Yung Cheung[3], Ian Chi Kei Wong[6,7], Gary Tse[2,8] and Qingpeng Zhang[9*]*

[1] Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, [2] Frailty Assessment Unit, Cardiovascular Analytics Group, Hong Kong, Hong Kong SAR, China, [3] Division of Clincal Pharmacology, Department of Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, Hong Kong SAR, China, [4] Aston Medical School, Aston University, Birmingham, United Kingdom, [5] Tianjin Key Laboratory of Ionic-Molecular Function of Cardiovascular Disease, Department of Cardiology, Tianjin Institute of Cardiology, Second Hospital of Tianjin Medical University, Tianjin, China, [6] Department of Pharmacology and Pharmacy, The University of Hong Kong, Hong Kong, Hong Kong SAR, China, [7] Medicines Optimisation Research and Education, UCL School of Pharmacy, London, United Kingdom, [8] Kent and Medway Medical School, Canterbury, United Kingdom, [9] School of Data Science, City University of Hong Kong, Hong Kong, Hong Kong SAR, China

OPEN ACCESS

Edited by:
Shane Nanayakkara, Alfred Hospital, Australia

Reviewed by:
Roberto Carlos Castrejón Pérez, Instituto Nacional de Geriatría, Mexico
Paul Leeson, University of Oxford, United Kingdom

***Correspondence:**
Qingpeng Zhang
qingpeng.zhang@cityu.edu.hk
Gary Tse
gary.tse@kmms.ac.uk;
garytse@tmu.edu.cn

[†]These authors share first authorship

Specialty section:
This article was submitted to Atherosclerosis and Vascular Medicine, a section of the journal Frontiers in Cardiovascular Medicine

Received: *03 July 2021*
Accepted: *20 April 2022*
Published: *08 July 2022*

Citation:
Zhou J, Chou OHI, Wong KHG, Lee S, Leung KSK, Liu T, Cheung BMY, Wong ICK, Tse G and Zhang Q (2022) Development of an Electronic Frailty Index for Predicting Mortality and Complications Analysis in Pulmonary Hypertension Using Random Survival Forest Model. Front. Cardiovasc. Med. 9:735906. doi: 10.3389/fcvm.2022.735906

Background: The long-term prognosis of the cardio-metabolic and renal complications, in addition to mortality in patients with newly diagnosed pulmonary hypertension, are unclear. This study aims to develop a scalable predictive model in the form of an electronic frailty index (eFI) to predict different adverse outcomes.

Methods: This was a population-based cohort study of patients diagnosed with pulmonary hypertension between January 1st, 2000 and December 31st, 2017, in Hong Kong public hospitals. The primary outcomes were mortality, cardiovascular complications, renal diseases, and diabetes mellitus. The univariable and multivariable Cox regression analyses were applied to identify the significant risk factors, which were fed into the non-parametric random survival forest (RSF) model to develop an eFI.

Results: A total of 2,560 patients with a mean age of 63.4 years old (interquartile range: 38.0–79.0) were included. Over a follow-up, 1,347 died and 1,878, 437, and 684 patients developed cardiovascular complications, diabetes mellitus, and renal disease, respectively. The RSF-model-identified age, average readmission, anti-hypertensive drugs, cumulative length of stay, and total bilirubin were among the most important risk factors for predicting mortality. Pair-wise interactions of factors including diagnosis age, average readmission interval, and cumulative hospital stay were also crucial for the mortality prediction. Patients who developed all-cause mortality had higher values of the eFI compared to those who survived ($P < 0.0001$). An eFI $\geq$ 9.5 was associated with increased risks of mortality [hazard ratio (HR): 1.90; 95% confidence interval [CI]: 1.70–2.12; $P < 0.0001$]. The cumulative hazards were higher among patients who were 65 years old or above with eFI $\geq$ 9.5. Using the same cut-off point, the eFI predicted a long-term mortality over

10 years (HR: 1.71; 95% CI: 1.53–1.90; $P < 0.0001$). Compared to the multivariable Cox regression, the precision, recall, area under the curve (AUC), and C-index were significantly higher for RSF in the prediction of outcomes.

Conclusion: The RSF models identified the novel risk factors and interactions for the development of complications and mortality. The eFI constructed by RSF accurately predicts the complications and mortality of patients with pulmonary hypertension, especially among the elderly.

Keywords: pulmonary hypertension, electronic frailty index, random survival forest (RSF), diabetes mellitus, cardiovascular disease, renal complications

INTRODUCTION

Pulmonary hypertension (PHTN) was defined in the First World Symposium on Pulmonary Hypertension as having a mean pulmonary arterial pressure >25 mmHg at resting by right heart catheterization (1, 2). PHTN was traditionally classified as either primary or secondary PHTN (3). The epidemiology and prognosis of the PHTN vary with different causes, but generally, PHTN can progress to a severe stage and ultimately cause death if left untreated (4). Therefore, it is essential to evaluate the prognosis of patients with PHTN as early as possible. Nevertheless, the prognostic risk factors to predict the risks of complication development and mortality are unclear.

Currently, the PHTN mortality is predicted using the dynamic risk stratification strategy suggested in the European Society of Cardiology/European Respiratory Society pulmonary hypertension guidelines. The strategy utilizes the clinical features and laboratory results to predict the mortalities and stratify the risk of death according to the one-year mortality expectations (5, 6). However, the model requires extensive clinical investigations, such as echocardiography and cardiopulmonary exercise testing (7). Calculating the risks of death by simply using the demographics and the laboratory testing results allows the determination of the treatment objectives to be more readily accessible.

PHTN can result in serious complications involving the cardiovascular, renal, and metabolic systems (8). Conversely, patients with pre-existing comorbidities have a poorer prognosis (9, 10). For example, acute right heart failure is one of the most important causes of mortality among patients with PHTN (11). PHTN can also co-exist with chronic renal disease owing to altered fibroblast growth factor-23 signaling (12, 13). The presence of diabetes mellitus can induce pulmonary endothelial dysfunction, and patients with PHTN may also develop diabetes mellitus and metabolic syndrome as complications due to the chronic pro-inflammatory states (14, 15). The concurrence of the above conditions increases the complexity of the PHTN clinical profile.

Frailty is a geriatric syndrome that results in age-associated functional limitations across multiple systems. Older people with frailty are prone to poorer health outcomes such as falls and disability. This contributes to frequent hospitalization and premature death (16–19). The development of an electronic frailty index (eFI) through the random survival forest (RSF) model allows the analysis of the survival data using electronic health data (20). RSF can approximate complex survival functions while maintaining low prediction error (21). This study aims to construct a scalable eFI with improved predictability for complications and short-term mortality among patients with PHTN through the application of the RSF model.

METHODS

Study Design and Population

The retrospective population-based cohort was designed to investigate long-term clinical prognostic risk factors that predict the survival of patients with newly diagnosed PHTN. This cohort included patients diagnosed with PHTN between January 1st, 2000 and December 31st, 2017 at centers managed by the Hong Kong Hospital Authority. The patients were identified from the Clinical Data Analysis and Reporting System (CDARS), a territory-wide database that centralizes patient information from individual local hospitals to establish comprehensive medical data. This system has previously been used by local teams to conduct population-based epidemiological studies (22, 23), including the development of eFIs (24, 25).

Demographics, comorbidities, hospitalization characteristics, drug prescriptions, and laboratory examinations at the baseline were extracted. Drug prescriptions following the diagnosis of PHTN were determined. The calculated mean daily prescribed drug dosage of medications was noted. The mean daily dose of each drug class is derived from multiplying the daily dose frequency by the drug dose then averaged by the cumulative duration. Details regarding the International Classification of Diseases, Nineth Edition (ICD-9) codes for identifying the comorbidities and PHTN drugs are provided (**Supplementary Tables 1, 2**).

Statistical Analysis

The study outcomes were the development of cardiovascular and renal complications, diabetes mellitus, and mortality after diagnosing PHTN. The mortality data were obtained from the Hong Kong Death Registry, a population-based official government registry with the registered death records of all Hong Kong citizens. Continuous variables were represented as median (95% confidence interval [CI] or interquartile range [IQR]), and categorical variables were presented count (%).

Continuous variables were compared using the Mann-Whitney U test. The $\chi 2$ test with Yates' correction was utilized for 2×2 contingency data. The Pearson's $\chi 2$ test was applied for variables with over two categories of contingency data.

The univariable Cox model was used to uncover the significant prognostic risk factors associated with the outcomes *via* adjustments based on baseline characteristics. Significantly predictive factors are used as input of the multivariable Cox regression and the RSF analysis model for the complication prediction. Hazard ratios (HRs) with corresponding 95% CIs and *P*-values were reported. All significance tests were two tailed and considered significant if $P < 0.05$. Data analyses were performed using the RStudio software (Version: 1.1.456) and Python (Version: 3.6).

Development of a Tree-Based Mortality Prediction Model

The eFI was developed to use the primary electronic health record to predict the frailty status of the patients, based on the principle that it reflects the health deficit accumulation (26, 27). Our team has previously developed an eFI for heart failure (24). A survival analysis was utilized to estimate the probability of mortality after diagnosing PHTN. It identifies the most influential prognostic risk factors that efficiently predict mortality outcomes. The RSF model was used to conduct a supervised survival learning analysis with the electronic health data. We used 20% of the data as a test set for a model performance evaluation and comparisons while using the remaining 80% for a model training. The parameters that gave the highest value of the Concordance Index (C-index) on the test set were chosen as the final model. The RSF was employed to learn the distribution of survival times based on the observed preoperative symptom data.

Variable Importance Measure

The variable importance value of each factor was calculated to investigate the predictive strength. We leave out about 40% of instances whenever a bootstrap sample is down with replacement from the training data set. These left-out instances are referred to as out-of-bag (OOB) ones and the instances in the bootstrap sample as in bag ones. To calculate the variable importance value, we dropped each OOB instance down its in-bag competing risk tree and assigned a child node randomly whenever a split for the variable is encountered. The event-specific cumulative probability function from each such tree is calculated and averaged. The importance value is the prediction error for the original ensemble event-specific cumulative probability function (obtained when each OOB instance is dropped down its in-bag competing risks tree) subtracted from the prediction error for the new ensemble obtained using randomizing assignments of the variable (28, 29). The prediction errors are computed using squared loss. A higher importance value indicates higher predictive strength of the variable, whereas zero or negative values indicate non-predictive variables.

Minimal Depth Approach

Minimal depth (30) ranks the variables through the inspection of the forest construction process under the assumption that variables with high impact on the prediction are those that most frequently split nodes nearest to the root node, where they partition the largest samples. Within each split tree, minimal depth approaches numbers of node levels according to their relative distance to the root of the tree (with the root node at 0). In such a way, the minimal depth approach can identify important variables by averaging the depth of the first split for each variable over all trees within the final forest.

The minimal depth approach was used to capture the variable interactions. We calculate the importance measures of pairwise interactions among variables since the minimal depth measure is defined by averaging the tree depth of the variable of interest relative to the root node. To compute the interaction strength for prediction, this calculation is modified to measure the minimal depth of a variable xi with respect to the maximal subtree for variable xj.

In general, to select the most influential variables with a variable importance approach, we examine the calculated variable importance values. The minimal depth approach is a non-event-specific criterion, whereas the variable importance approach can be both event specific and non-event specific. In the following analysis, we used the mortality-specific and time-to-event-specific variable importance.

Performance Evaluation

The 5-fold cross validation approach is adopted to evaluate the prediction performance of the RSF model. The metrics of precision and recall are as follows:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{Tp}{Tp + FN},$$

and the area under the receiver operating characteristic curve (AUC) and the C index are used for performance evaluation. *TN*, *TP*, *FP*, and *FN* represent true positive, true negative, false positive, and false negative rates in the confusion matrix, respectively. Conventional non-parametric survival models primarily select the best analysis model that maximizes the C index (31). C index quantifies the degree to which the predicted outcomes in the pairwise orderings are consistent with the observed outcomes and can be regarded as a generalization of the AUC. That is, any survival analysis model that properly estimates the ordered but proportional event times can score high in terms of C index. C index can be used as a global assessment of the model's discrimination power: the ability to correctly provide a reliable ranking of the survival times based on the individual risk scores. In this study, we compare the RSF model with the traditional Cox analysis to predict the mortality and complications after diagnosis of PHTN.

RESULTS

Basic Characteristics

The study cohort included 2,560 patients (37% men) with an age of 63.4 (IQR: 38.0–79.0) (**Figure 1**). The baseline characteristics of the PHTN patients are detailed in **Tables 1**, **2**. Over the follow-up, 38% of patients died; those who died were older (median:

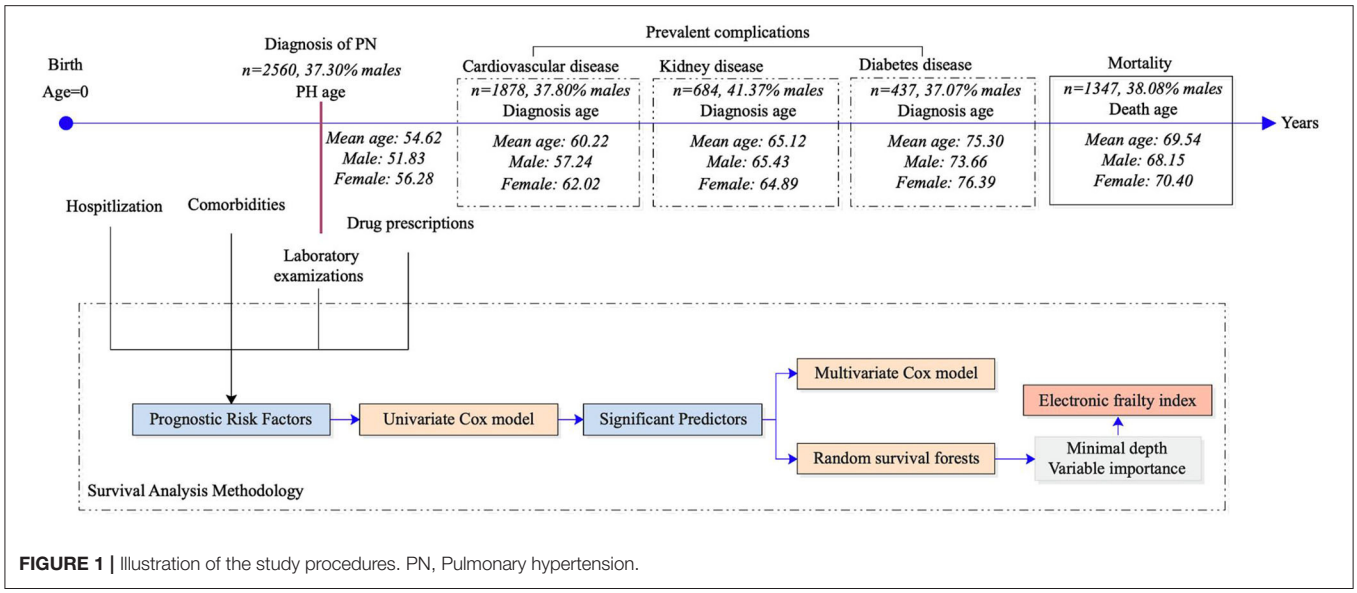


FIGURE 1 | Illustration of the study procedures. PN, Pulmonary hypertension.

75.2, IQR: 60.0–83.0) than those who survived (median: 44.6, IQR: 1.0–65.0).

The Risk of Adverse Outcomes With Cox Analysis

A univariable Cox regression analysis identified the predictors of mortality after the diagnosis of PHTN (**Supplementary Table 3**). Patients with past comorbidities, such as cardiovascular diseases (HR: 1.947, 95% CI: 0.8132–2.102), respiratory diseases (HR: 1.295, 95% CI: 0.646–2.595), endocrine disease (HR: 1.524, 95% CI: 1.127–2.059), and hypertension (HR: 1.362, 95% CI: 1.172–2.7618) were associated with higher mortality risks. The univariable Cox regression analysis also identified the predictors for the complications (**Supplementary Table 4**). The diagnosis age was an important predictor (all $P < 0.0001$). The multivariable Cox regression predictors of mortality were also identified (**Supplementary Table 5A**). The age of diagnosis (HR: 1.822; 95% CI: 1.815–1.829), cumulative hospital length of stay (LOS) (HR: 1.0007; 95% CI: 1.00040–1.0009), prior cardiovascular diseases (HR: 1.266; 95% CI: 1.064–1.507), kidney diseases (HR: 1.279; 95% CI: 1.125–1.454), diabetes mellitus (HR: 1.208; 95% CI: 1.032–1.415), and hypertension (HR: 1.549; 95% CI: 0.665–3.608) were predictive of mortality. The Kaplan-Meier survival curves demonstrated the mortality of the patients in **Figure 2A**.

Machine Learning Survival Analysis

The RSF model identified the influential prognostic risk factors by capturing the non-linearity and interactions. The predicted events for all-cause mortality with the RSF model are shown in **Figure 2B**. Most deaths occurred at age equal or above 75. The predictions for the time to event of the complications are shown in **Figure 3**. The RSF provided better discrimination performance in predicting the mortality risk, given its ability to capture the non-linearity and the interaction patterns (**Supplementary Figures 4–7**).

Age was the most important for the mortality risk prediction, followed by average readmission, antihypertension drugs, cumulative LOS, and total bilirubin level (**Supplementary Figure 8**). The risk factors that predicted the three complications were also identified. Age was the most important predictor for cardiovascular complications, followed by lymphocyte count and mean readmission interval. Age also showed the highest prediction strength for kidney and diabetic complications.

Important interactions with the demographics included the interactions of PHTN age and sex with average readmission interval, cumulative hospital stay, and total bilirubin level (**Supplementary Figure 4**). The interactions of the variables with past comorbidities (**Supplementary Figure 5**), drug prescriptions (**Supplementary Figure 6**), and laboratory examinations (**Supplementary Figure 7**) were also derived. An interaction importance ranking pattern could be observed: interactions formed by a variable with the influential individual risk factors demonstrated high predictive strength.

Predictors of Adverse Events Risk With eFI

Correspondingly, an eFI was developed for predicting the all-cause mortality (**Supplementary Table 5B**). The calculated eFI for patients with/without mortality was significantly different (median: 9.0, IQR: 8.0–10.0 vs. median: 8.0, IQR: 6.0–9.0, $P < 0.0001$) (**Table 3A**, **Supplementary Figure 9**). The eFI was significantly associated with higher mortality risk (HR: 1.25, 95% CI: 1.22–1.29, $P < 0.0001$) (**Table 3B**). The marginal effects of eFI also demonstrated that higher eFI was associated with higher risks of mortality (**Supplementary Figure 10**).

The binarized eFI also predicted mortality risk based on the Youden cut-off of 9.5 (HR: 1.90, 95% CI: 1.70–2.12, $P < 0.0001$) (**Table 3B**). Patients with eFI $>$ 9.5 had a higher cumulative

TABLE 1 | The characteristics of patients with pulmonary hypertension stratified by mortality outcomes.

Characteristics	All patients (n = 2,560) Median (IQR) or count (%)	Mortality (n = 1,347) Median (IQR) or count (%)	No mortality (n = 1,213) Median (IQR) or count (%)	P-value
Demographics				
Male sex	955 (37.30%)	513 (38.08%)	442 (36.43%)	<0.0001***
Age at diagnosis	63.4 (38.0–79.0)	75.2 (60.0–83.0)	44.6 (1.0–65.0)	<0.0001***
Hospitalization before PHTN				
Total number of hospital admissions	13.0 (7.0–23.0)	16.0 (8.0–26.0)	11.0 (5.0–20.0)	<0.0001***
Number of emergency readmissions	3.0 (1.0–6.0)	4.0 (2.0–8.0)	2.0 (1.0–4.0)	0.0115*
Mean readmission interval (days)	223.7 (105.0–435.0)	221.3 (115.0–387.0)	225.4 (95.0–509.0)	0.1725
Cumulative length-of-stay	80.0 (34.0–162.0)	110.0 (58.0–191.0)	48.0 (21.0–118.0)	<0.0001***
Comorbidity before PHTN				
Respiratory disease	2,537 (99.10%)	1,330 (98.73%)	1,207 (99.50%)	0.2782
Hypertension	2,511 (98.08%)	1,313 (97.47%)	1,198 (98.76%)	0.0617
Cardiovascular disease	1,916 (74.84%)	1,150 (85.37%)	766(63.14%)	<0.0001***
Gastrointestinal disease	1,014 (39.60%)	601 (44.61%)	413 (34.04%)	0.0016**
Kidney disease	768 (30.00%)	534 (39.64%)	234 (19.29%)	<0.0001***
Endocrine disease	88 (3.43%)	44 (3.26%)	44 (3.62%)	0.0916
Diabetes mellitus	337 (13.16%)	218 (16.18%)	119 (9.81%)	<0.0001***
Obesity	71 (2.77%)	29 (2.15%)	42 (3.46%)	<0.0001***
Drug prescriptions after PHTN				
Alpha blockers	***n* = 224**	***n* = 159**	***n* = 65**	
Daily dosage, mg/day	1.7 (1.0–3.0)	1.7 (0.9, 3.0)	1.9 (1.1, 3.5)	p = 0.2443***
Cumulative dosage, mg	377.5 (84.0, 1165.0)	258.5 (49.5, 924.5)	543.5 (177.0–2319.5)	0.0016**
Cumulative duration, days	260.5 (62.0–753.5)	237.0 (40.0–696.0)	396.0 (112.0–1141.0)	0.0648
Anti-arrhythmias drugs	***n* = 238**	***n* = 179**	***n* = 59**	
Daily dosage, mg/day	225.0 (46.5–636.5)	225.0 (48.0–661.5)	117.0 (20.0–584.5)	0.2261
Cumulative dosage, mg	227.86 (36.0–1524.11)	255.0 (37.4–1660.55)	182.81 (24.0–1023.75)	0.1277
Cumulative duration, days	4.0 (2.0–17.5)	4.0 (2.0–12.5)	5.0 (2.0–35.0)	0.582
Beta blockers	***n* = 670**	***n* = 402**	***n* = 268**	
Daily dosage, mg/day	15.76 (6.25–31.41)	13.26 (6.23–26.36)	18.99 (6.26–41.97)	0.0611
Cumulative dosage, mg	784.0 (125.0–4948.44)	575.0 (76.5–3966.0)	1599.5 (187.0–8473.44)	0.0001***
Cumulative duration, days	178.0 (37.0–499.0)	133.0 (26.0–397.0)	280.0 (54.0–867.0)	<0.0001***
Cardiac glycosides	***n* = 491**	***n* = 362**	***n* = 129**	
Daily dosage, mg/day	62.5 (42.85–144.34)	63.14 (48.28–148.2)	57.34 (25.79–125.0)	0.0111*
Cumulative dosage, mg	4437.5 (371.0–23625.0)	5181.25 (562.5–19250.0)	2662.5 (155.0–45562.5)	0.7103
Cumulative duration, days	172.0 (24.0–551.5)	154.0 (21.0–473.0)	252.0 (30.0–954.5)	0.0228*
Centrally acting antihypertensive drugs	***n* = 63**	***n* = 48**	***n* = 15**	
Daily dosage, mg/day	75.0 (25.0–524.7)	100.0 (27.81–524.7)	38.5 (25.0–556.46)	0.5393
Cumulative dosage, mg	525.0 (42.0–7945.0)	920.0 (79.5–7945.0)	175.0 (26.5–9175.0)	0.5291
Cumulative duration, days	135.0 (28.0–387.5)	139.5 (37.0–371.5)	28.0 (8.5–459.0)	0.4433
Loop diuretics	***n* = 1,636**	***n* = 1,042**	***n* = 594**	
Daily dosage, mg/day	27.39 (9.15–72.58)	27.76 (11.26–64.57)	25.66 (7.54–144.78)	0.5814
Cumulative dosage, mg	525.0 (76.0–2673.5)	577.0 (114.0–2721.75)	388.5 (48.0–2538.0)	0.0109*
Cumulative duration, days	158.0 (38.0–467.0)	184.0 (45.5–516.5)	128.0 (31.0–372.5)	0.0014**
Phosphodiesterase type-3 inhibitors	***n* = 13**	***n* = 7**	***n* = 6**	
Daily dosage, mg/day	7.7 (1.67–45.0)	16.0 (5.02–60.06)	2.97 (0.9–15.67)	0.1336
Cumulative dosage, mg	74.0 (3.84–171.0)	171.0 (15.05–296.88)	38.92 (2.73–104.5)	0.1336
Cumulative duration, days	3.0 (2.0–6.0)	3.0 (2.0–6.0)	4.5 (2.5–6.0)	0.9416
Potassium-sparing diuretics and aldosterone	***n* = 664**	***n* = 359**	***n* = 305**	
Daily dosage, mg/day	21.37 (8.09–29.17)	25.0 (12.5–35.69)	13.86 (6.36–25.0)	<0.0001***
Cumulative dosage, mg	568.0 (88.0–3676.88)	475.0 (75.0–4625.0)	654.0 (125.0–2975.0)	0.6552
Cumulative duration, days	98.5 (31.0–330.0)	87.0 (24.5–334.5)	117.0 (39.0–322.0)	0.1124

(Continued)

TABLE 1 | Continued

Characteristics	All patients (n = 2,560) Median (IQR) or count (%)	Mortality (n = 1,347) Median (IQR) or count (%)	No mortality (n = 1,213) Median (IQR) or count (%)	P-value
Thiazides and related diuretics	**n = 215**	***n* = 145**	***n* = 70**	
Daily dosage, mg/day	3.26 (1.77–9.65)	2.99 (1.67–5.62)	5.0 (2.5–17.08)	0.0322*
Cumulative dosage, mg	83.0 (15.0–531.25)	42.0 (7.0–231.0)	461.5 (70.75–922.0)	<0.0001***
Cumulative duration, days	43.0 (7.0–219.5)	21.0 (4.0–112.0)	116.0 (35.0–344.0)	<0.0001***
Vasodilator antihypertensive drugs	***n* = 346**	***n* = 202**	***n* = 144**	
Daily dosage, mg/day	35.48 (6.33–62.32)	39.22 (6.87–67.1)	33.85 (6.0–56.46)	0.0144*
Cumulative dosage, mg	474.0 (90.0-3040.5)	285.5 (73.0–1789.0)	966.0 (160.0–4419.0)	0.0033**
Cumulative duration, days	126.0 (22.5–419.0)	58.0 (14.0–269.0)	237.0 (42.5–724.5)	<0.0001***
Laboratory examinations on PHTN				
Hemoglobin, g/dL	12.1 (10.0–14.0); n = 2,515	11.6 (10.0–13.0); n = 1,340	12.4 (11.0–14.0); n = 1,175	0.0115*
Hematocrit, L/L	0.4 (0.0–0.0); n = 2,511	0.4 (0.0–0.0); n = 1,339	0.4 (0.0–0.0); n = 1,172	0.5656
Lymphocyte, x10∧9/L	1.3 (1.0–2.0); n = 2,506	1.0 (1.0–2.0); n = 1,339	1.7 (1.0–3.0); n = 1,167	0.1625
Neutrophil, x10∧9/L	5.3 (4.0–7.0); n = 2,506	5.4 (4.0–8.0); n = 1,339	5.0 (4.0–7.0); n = 1,167	0.0615
Platelet, x10∧9/L	197.0 (147.0–256.0); n = 2,508	178.0 (130.0–232.0); n = 1,336	221.5 (170.0–278.0); n = 1,172	<0.0001***
APTT, secs	32.9 (29.0–38.0); n = 2,352	33.3 (29.0–38.0); n = 1,299	32.7 (29.0–38.0); n = 1,053	0.0016**
INR	1.2 (1.0–1.0); n = 2,364	1.2 (1.0–2.0); n = 1,303	1.2 (1.0–1.0); n = 1,061	0.0166*
Prothrombin time, sec	13.4 (12.0–17.0); n = 2,333	13.8 (12.0–18.0); n = 1,292	13.1 (12.0–16.0); n = 1,041	0.6319
Red cell count, x10∧12/L	4.1 (4.0–5.0); n = 2,510	4.0 (3.0–5.0); n = 1,339	4.2 (4.0–5.0); n = 1,171	0.2716
Total protein, g/L	68.0 (62.0–74.0); n = 2,465	67.7 (62.0–73.0); n = 1,330	69.0 (62.0–75.0); n = 1,135	0.7112
Total bilirubin, umol/L	13.0 (8.0–22.0); n = 2,453	14.6 (9.0–24.0); n = 1,326	11.4 (7.0–20.0); n = 1,127	<0.0001***
Alkaline phosphatase, U/L	49.5 (34.0–73.0); n = 2,463	51.5 (35.0–74.0); n = 1,329	47.0 (32.0–72.0); n = 1,134	0.1625
Red cell distance width, %	15.1 (14.0–17.0); n = 2,472	15.7 (14.0–18.0); n = 1,317	14.5 (13.0–16.0); n = 1,155	0.0317*
Mean cell volume, fL	89.9 (85.0–95.0); n = 2,512	91.0 (86.0–96.0); n = 1,339	89.0 (84.0–93.0); n = 1,173	<0.0001***
Mean cell hemoglobin concentration, g/dL	34.2 (33.0–35.0); n = 2,512	34.2 (33.0–35.0); n = 1,339	34.3 (33.0–35.0); n = 1,173	0.0815

APTT, activated partial thromboplastin time; INR, international normalized ratio; IQR, interquartile range; LOS, length of stay; PHTN, pulmonary hypertension.
** for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$.*

hazard for all-cause mortality. The cumulative hazards were especially higher among patients who were 65 years old or above with eFI $\geq$ 9.5 (**Figure 4**). Within 2 years of follow-up, the eFI predicted the mortality (HR: 1.15, 95% CI: 1.02–1.28; P = 0.0169). The HR of the eFI increased with a longer duration of follow-up. Upon 10 years of follow-up, the eFI was associated with higher risks of mortality (HR: 1.71, 95% CI: 1.53–1.90; $P <$ 0.0001); meanwhile, upon 20 years of follow-up, the HR was even higher (HR: 1.90; 95% CI: 1.70–2.12; $P < 0.0001$) (**Table 3C**).

The RSF model showed a better performance in terms of precision (0.9263 vs. 0.8382), recall (0.9058 vs. 0.8992), AUC (0.9478 vs. 0.9051), and C index (0.9361 vs. 0.9240) compared to the Cox model in the 5-fold cross-validation. Similarly, compared to multivariate Cox regression, the precision, recall, AUC, and C index were significantly higher for RSF in predicting the cardiovascular, kidney, diabetic complications, and mortality (**Table 4**).

DISCUSSION

The Principal Findings of the Study

The main findings of this study include (i) risk factors including admission interval, cumulative LOS, and total admissions times were predictive of the complications and mortality; (ii) the RSF-identified non-linear relationship between the predictors and outcome was predictive of mortality; (iii) the RSF model performed better in mortality and complication predictions than the Cox regression; (iv) the eFI predicted the risks of all-cause mortality accurately, especially among patients who were 65 years old or above.

Strength and Limitations of the Study

To the best of our knowledge, this is the first study using the eFI in predicting the PHTN outcomes. The usage of the cohort from a real-world clinical database to derive the RSF analysis was shown to have performed better than the multivariable logistic regression to predict the PHTN mortality and complications. This would allow better clinical management based on the eFI. However, there are certain limitations to this study. Firstly, given this is a local study conducted in Hong Kong, the PHTN results should be validated using the data from other databases in other countries. Secondly, medical history, such as smoking, asbestos, and family history, and clinical parameters such as partial pressure of oxygen (PaO2) and N-terminal pro-brain natriuretic peptide (NT-proBNP), which are associated with PHTN in other literature, were not included in this predictive model given the lack of the codes in CDARS (32). Thirdly, given the retrospective nature of this study, the results may be subjected

TABLE 2 | The characteristics of patients who developed complications after the diagnosis of pulmonary hypertension.

Characteristics	Cardiovascular complications (n = 1,878) Median (IQR) or count (%)	Renal complications (n = 684) Median (IQR) or count (%)	Diabetes (n = 437) Median (IQR) or count (%)	P-value
Demographics				
Male sex	710 (37.80%)	283 (41.37%)	162 (37.07%)	0.0017**
Age at diagnosis	69.2 (48.0–81.0)	72.0 (54.0–82.0)	75.8 (66.0–83.0)	0.0008***
Baseline hospitalizations				
Total number of hospital admissions	15.0 (8.0–25.0)	19.0 (11.0–34.0)	19.0 (11.0–33.0)	<0.0001***
Number of emergency readmissions	2.0 (0.0–5.0)	3.0 (1.0–7.0)	2.0 (1.0–7.0)	0.0216*
Mean readmission interval (days)	211.3 (102.0–388.0)	188.0 (98.0–315.0)	219.1 (128.0–405.0)	<0.0001***
Cumulative length-of-stay	93.0 (44.0–170.0)	137.0 (74.0–236.0)	120.0 (63.0–226.0)	<0.0001***
Baseline comorbidities				
Respiratory disease	1,878 (100.00%)	684 (100.00%)	437 (100.00%)	0.5816
Hypertension	1,872 (99.68%)	682 (99.70%)	437 (100.00%)	0.2671
Cardiovascular disease	-	593 (86.69%)	382 (87.41%)	<0.0001***
Gastrointestinal disease	792 (42.17%)	377 (55.11%)	253 (57.89%)	0.0002***
Kidney disease	645 (34.34%)	-	242 (55.37%)	<0.0001***
Endocrine disease	56 (2.98%)	24 (3.50%)	14 (3.20%)	0.0017**
Diabetes mellitus	275 (14.64%)	190 (27.77%)	-	<0.0001***
Obesity	55 (2.92%)	30 (4.38%)	34 (7.78%)	<0.0001***
Drug prescriptions for PHTN				
Cardiac glycosides	466 (24.81%)	166 (24.26%)	98 (22.42%)	<0.0001***
Phosphodiesterase type-3 inhibitors	9 (0.47%)	5 (0.73%)	0 (0.00%)	0.0031**
Thiazides and related diuretics	190 (10.11%)	109 (15.93%)	73 (16.70%)	<0.0001***
Loop diuretics	1,377 (73.32%)	525 (76.75%)	323 (73.91%)	0.0211*
Potassium-sparing diuretics and aldosterone	579 (30.83%)	204 (29.82%)	108 (24.71%)	0.0022**
Anti-arrhythmias drugs	224 (11.92%)	98 (14.32%)	48 (10.98%)	<0.0001***
Beta blockers	567 (30.19%)	271 (39.61%)	192 (43.93%)	<0.0001***
Vasodilator antihypertensive drugs	232 (12.35%)	89 (13.01%)	44 (10.06%)	0.0012**
Centrally acting antihypertensive drugs	56 (2.98%)	30 (4.38%)	22 (5.03%)	<0.0001***
Alpha blockers	184 (9.79%); n = 184	116 (16.95%); n = 116	83 (18.99%); n = 83	<0.0001***
Laboratory tests				
Hemoglobin, g/dL	11.9 (10.0–14.0); n = 1,862	10.8 (9.0–13.0); n = 679	11.3 (10.0–13.0); n = 432	0.0035**
Hematocrit, L/L	0.4 (0.0–0.0); n = 1,861	0.3 (0.0–0.0); n = 679	0.3 (0.0–0.0); n = 432	0.6241
Lymphocyte, x10∧9/L	1.3 (1.0–2.0); n = 1,858	1.0 (1.0–2.0); n = 679	1.1 (1.0–2.0); n = 432	0.042*
Neutrophil, x10∧9/L	5.2 (4.0–7.0); n = 1,858	5.3 (4.0–7.0); n = 679	5.4 (4.0–7.0); n = 432	0.0071**
Platelet, x10∧9/L	191.5 (144.0–249.0); n = 1,858	180.5 (135.0–240.0); n = 678	191.0 (144.0–238.0); n = 431	0.0063**
APTT, secs	33.0 (30.0–38.0); n = 1,781	33.8 (30.0–39.0); n = 665	31.7 (28.0–36.0); n = 426	0.0163*
INR	1.2 (1.0–1.0); n = 1,789	1.2 (1.0–2.0); n = 668	1.2 (1.0–1.0); n = 425	0.0166*
Prothrombin time, sec	13.6 (12.0–17.0); n = 1,766	13.8 (12.0–17.0); n = 658	12.9 (12.0–15.0); n = 423	<0.0001***
Red cell count, x10∧12/L	4.1 (4.0–5.0); n = 1,861	3.8 (3.0–4.0); n = 679	4.0 (3.0–5.0); n = 432	0.0714
Total protein, g/L	68.0 (62.0–74.0); n = 1,831	68.0 (62.0–73.0); n = 671	70.0 (64.0–74.0); n = 430	0.5312
Total bilirubin, umol/L	14.0 (9.0–23.0); n = 1,834	13.0 (8.0–22.0); n = 668	11.4 (8.0–18.0); n = 431	<0.0001***
Alkaline phosphatase, U/L	49.0 (34.0–71.0); n = 1,826	53.6 (35.0–78.0); n = 668	48.5 (33.0–68.0); n = 431	<0.0001***
Red cell distribution width, %	15.2 (14.0–17.0); n = 1,844	15.7 (14.0–18.0); n = 670	15.0 (14.0–17.0); n = 431	0.0615
Mean cell volume, fL	90.1 (85.0–95.0); n = 1,861	90.5 (86.0–95.0); n = 679	89.8 (86.0–94.0); n = 432	<0.0001***
Mean cell hemoglobin concentration, g/dL	34.2 (33.0–35.0); n = 1,861	34.2 (33.0–35.0); n = 679	34.0 (33.0–35.0); n = 432	0.0915

APTT, activated partial thromboplastin time; INR, international normalized ratio; IQR, interquartile range; LOS, length of stay; PHTN, pulmonary hypertension.
** for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$.*

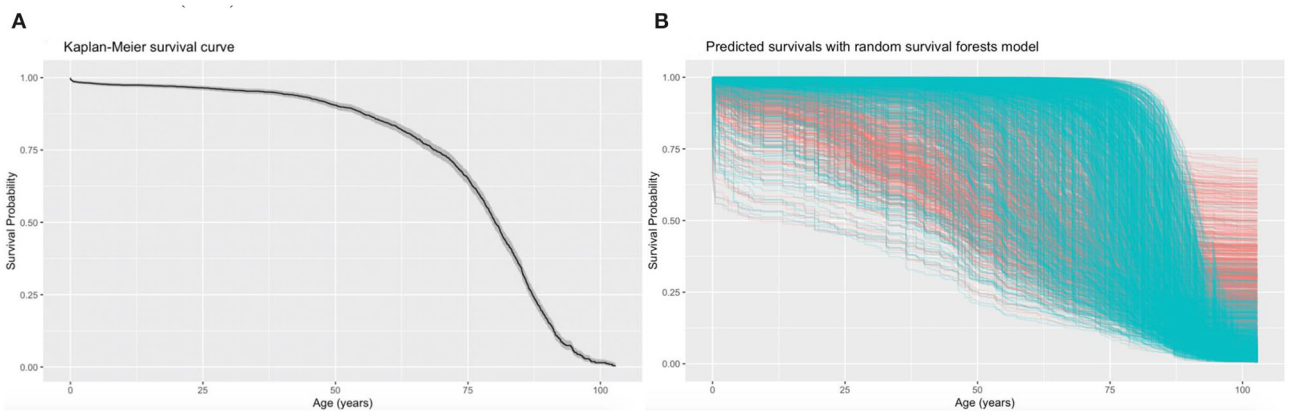


FIGURE 2 | Kaplan-Meier curves for all-cause mortality and the predicted mortality. The Kaplan-Meier survival curves demonstrate the moralities of the patients with pulmonary hypertension (PHTN) **(A)** and the predicted mortalities using the random survival forest (RSF) prediction **(B)**. Each line represents a single patient in the training data set, where censored patients are colored blue, and patients who have experienced the mortality event are colored in red. The median survival (black) with a 95% shaded confidence band (gray) are indicated.

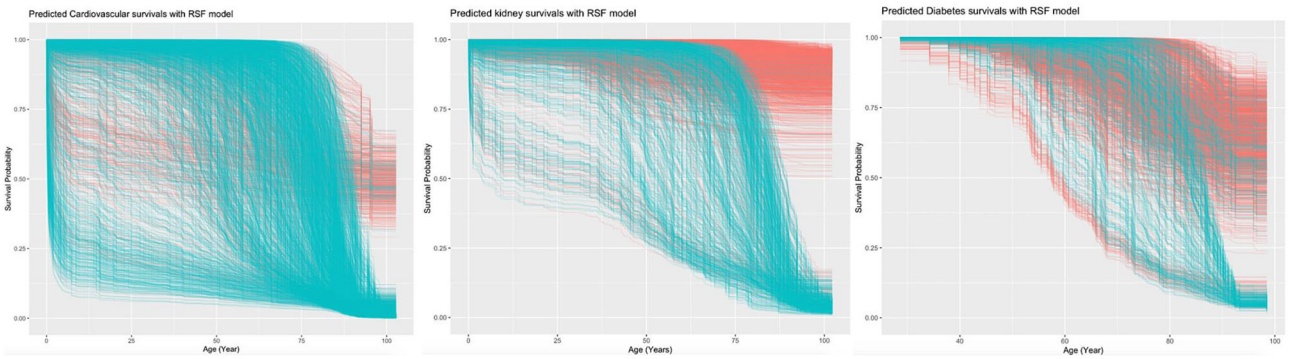


FIGURE 3 | Predicted survivals of cardiovascular, kidney, and diabetes complications with RSF model.

TABLE 3A | Descriptive statistics of electronic frailty index for all-cause mortality risk prediction.

Characteristics	Cut-off	All Median (IQR); (*N* = 2,560) N or Count (%)	All-cause mortality Median (IQR); (*N* = 1,347) N or Count (%)	Alive (*N* = 1,213) Median (IQR); N or Count (%)	*P*-value
Electronic frailty index	9.5	8.0 (7.0–10.0); *n* = 2,560	9.0 (8.0–10.0); *n* = 1,394	8.0 (6.0–9.0); *n* = 1,166	<0.0001***
Electronic frailty index≥9.5	-	722 (28.20%)	527 (39.12%)	195 (16.08%)	<0.0001***

** for P ≤ 0.05, ** for P ≤ 0.01, *** for P ≤ 0.001.*

TABLE 3B | Prediction strength of electronic frailty index for all-cause mortality risk prediction.

Characteristics	HR[CI]; *P*-value	Precision	Recall	AUC	C-index
Electronic frailty index	1.25 [1.22–1.29]; <0.0001***	0.8234	0.8867	0.9015	0.9109
Electronic frailty index≥9.5	1.90 [1.70–2.12]; <0.0001***	0.8351	0.8909	0.9105	0.9202

** for p ≤ 0.05, ** for p ≤ 0.01, *** for p ≤ 0.001.*

to instabilities of the laboratory results, including the equipment modifications and blood samples artifacts, and changes in the clinical criteria for the diagnosis of the PHTN. Furthermore, clinical information suggested in the European PHTN guidelines, such as the clinical courses and syncope episodes, is also lacking. Furthermore, in our predictive model, the number of patients on centrally acting antihypertensive drugs and phosphodiesterase type-3 inhibitors was relatively small. Lastly, due to the lack of

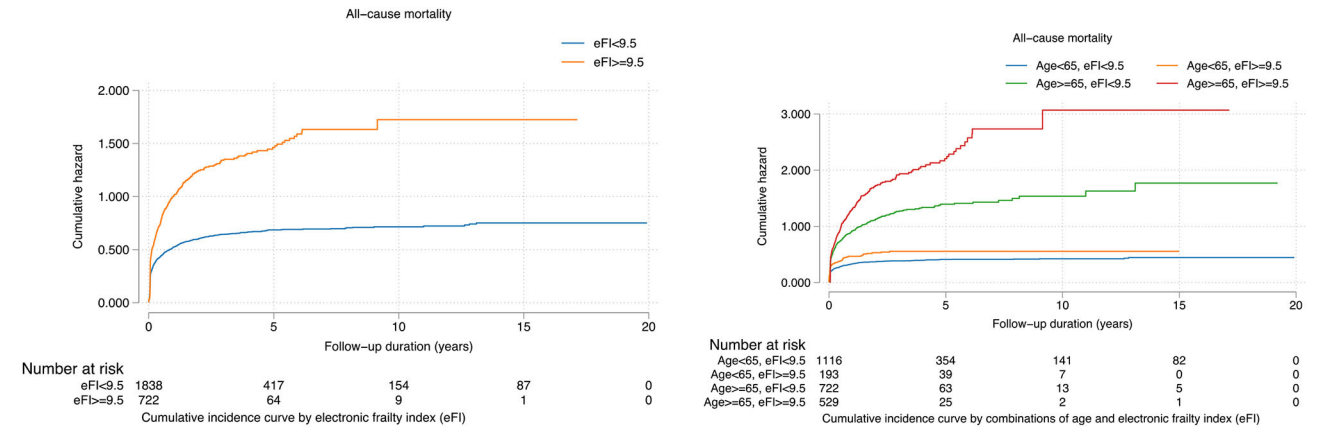


FIGURE 4 | Cumulative incidence curves for all-cause mortality stratified by the constructed electronic frailty index and age.

TABLE 3C | Prediction strength of the constructed electronic frailty index for mortality risks within 2-year follow-up, 5-year follow-up, 10-year follow-up, 15-year follow-up, and 20-year follow-up.

Characteristics	All-cause mortality HR [95% CI]; *P*-value
Within 2-year follow-up	
Electronic frailty index	1.03 [1.00–1.07];0.0459*
<9	1.0 [Reference]
(9, 10)	1.15 [1.02–1.28];0.0183*
>12	1.15 [0.64–2.09];0.6394
High vs. low	1.15 [1.02–1.28];0.0169*
Within 5-year follow-up	
Electronic frailty index	1.15 [1.11–1.18];<0.0001***
<9	1.0 [Reference]
(9, 10)	1.50 [1.34–1.67];<0.0001***
>12	1.86 [1.03–3.38];0.0409*
High vs. low	1.51 [1.35–1.68];<0.0001***
Within 10-year follow-up	
Electronic frailty index	1.21 [1.17–1.24];<0.0001***
<9	1.0 [Reference]
(9, 10)	1.70 [1.52–1.90];<0.0001***
>12	1.98 [1.09–3.59];0.0244*
High vs. low	1.71 [1.53–1.90];<0.0001***
Within 15-year follow-up	
Electronic frailty index	1.22 [1.18–1.25];<0.0001***
<9	1.0 [Reference]
(9, 10)	1.76 [1.58–1.96];<0.0001***
>12	2.09 [1.15–3.79];0.0153*
High v.s. low	1.77 [1.58–1.97];<0.0001***
Within 20-year follow-up	
Electronic frailty index	1.25 [1.22–1.29];<0.0001***
<9	1.0 [Reference]
(9, 10)	1.89 [1.70–2.11];<0.0001***
>12	2.23 [1.23–4.04];0.0085**
High vs. low	1.90 [1.70–2.12];<0.0001***

* *for P ≤ 0.05,* ** *for P ≤ 0.01,* *** *for P ≤ 0.001.*
High: Electronic frailty index≥9.5; Low: Electronic frailty index<9.5.

TABLE 4A | Performance comparisons between multivariable and random survival forest with five-fold cross validation.

Model	Precision	Recall	AUC	C-index
Multivariable Cox analysis	0.8382	0.8992	0.9051	0.9240
Random survival forests	0.9263	0.9058	0.9478	0.9361

the relevant diagnostic codes, the PHTN cases were not classified according to the World Health Organization (WHO) PHTN classification (2). Nevertheless, frailty models with the lack of enlisted information are still strong predictors of mortality and mortality (33).

Comparing the Findings With the Other Studies

Our data presented different risk factors that contributed to the development of the PHTN complications. Older patients and male patients showed a higher mortality risk. Patients with a higher comorbidity burden, including cardiovascular, hypertension, and renal conditions, had a lower survival rate and were more likely to develop complications (34). The variable importance and minimal depth approach indicated that the age of PHTN diagnosis had the highest predictive strength. This is in accordance with the previous predictive model (REVEAL), indicating that old male patients would have a worse prognosis, even though PHTN is a predominantly female disorder. This is worrying given the average age of patients diagnosed is shifting toward an older population (35). In agreement with previous studies, the hospitalization characteristics and drugs are also associated with the survival outcome in a pattern (36, 37). Compared to other predictive models, our study does not involve the use of the hemodynamic data nor lung function and radiological test results (38). However, our study demonstrated a higher AUC compared to the previous studies (39, 40). Nevertheless, the PHTN diagnosis in our study was not classified according to the WHO classification owing to the lack of the CDARS code.

TABLE 4B | Performance comparisons between multivariable Cox and random survival forest with five-fold cross validation.

	Cardiovascular		Kidney		Diabetes		Mortality	
Metric	**Multivariable Cox**	**RSF**	**Multivariable Cox**	**RSF**	**Multivariable Cox**	**RSF**	**Multivariable Cox**	**RSF**
Precision	0.84	0.92	0.83	0.92	0.84	0.92	0.84	0.93
Recall	0.83	0.91	0.78	0.91	0.87	0.91	0.83	0.91
AUC	0.90	0.95	0.85	0.94	0.85	0.95	0.76	0.94
C-index	0.88	0.94	0.87	0.93	0.90	0.94	0.79	0.91

Our results demonstrated that eFI significantly predicts the risks of complications and mortality among elderlies. Compare to the traditional Cox models, prediction models based on RSF have improved prediction performance across diseases such as heart failure (41), Brugada syndrome (42), congenital (43, 44) and acquired (45) long QT syndromes, diabetes mellitus (46, 47), metabolic diseases (48), stroke (49), and cancer (50). It improved the risk prediction in the context of PHTN for disease onset (51) and pressure prediction based on echocardiographic parameters (52). However, few studies have examined the long-term prognosis of patients with PHTN. In our study, the RSF model had a higher predictive accuracy with 5-fold cross-validation than the Cox model since it does not have a strong assumption about individual proportional hazard functions. Furthermore, the model can capture the interactions, reducing the prediction variances and bias (20, 53). The interactions formed by a variable with the influential individual risk factors demonstrated a high predictive strength.

The Implications for the Clinicians and Future Research Directions

The eFI derived from the significant variables allows predicting the risks of mortality and complications of the patients with PHTN, especially among elderlies. Clinicians can make use of the electronic medical records to estimate the outcome of the patients with PHTN without using any hemodynamic and radiological investigation modalities. The eFI can be further translated into a risk diagnosis tool to be deployed in the computer for real-time clinical applications.

CONCLUSION

The RSF model identified the influential prognostic risk factors and their interaction from the clinical accessible data. The usage of this RSF-derived eFI would allow stratifying the risks of complication and mortality and optimizing the PHTN management among elderlies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by NTEC-CUHK, Hong Kong. The Ethics Committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

OC, JZ, and SL: data analysis, data interpretation, statistical analysis, manuscript drafting, and critical revision of manuscript. KW, KL, IW, TL, and BC: project planning, data acquisition, data interpretation, and critical revision of manuscript. QZ and GT: study conception, study supervision, project planning, data interpretation, statistical analysis, manuscript drafting, and critical revision of manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

QZ acknowledges the National Key Research and Development Program of China, Ministry of Science and Technology of China: 2019YFE0198600; National Natural Science Foundation of China (NSFC): 71972164 and 72042018; Health and Medical Research Fund of the Food and Health Bureau of Hong Kong: 16171991; Innovation and Technology Fund of Innovation and Technology Commission of Hong Kong: MHP/081/19.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2022.735906/full#supplementary-material

REFERENCES

1. Hatano S, Strasser T, World Health Organization. *Primary Pulmonary Hypertension: Report on a WHO Meeting, Geneva*. Hatano S, Strasser T, editors. Geneva: World Health Organization (1975).

2. Simonneau G, Montani D, Celermajer DS, Denton CP, Gatzoulis MA, Krowka M, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J.* (2019) 53:1801913. doi: 10.1183/13993003.01913-2018

3. Simonneau G, Galiè N, Rubin LJ, Langleben D, Seeger W, Domenighetti G, et al. Clinical classification of pulmonary hypertension. *J Am Coll Cardiol.* (2004) 43(12 Suppl):S5–12. doi: 10.1016/j.jacc.2004.02.037
4. Corciova FC, Arsenescu-Georgescu C. Prognostic factors in pulmonary hypertension. *Maedica.* (2012) 7:30–7. Available online at: https://www.maedica.ro/articles/2012/1/2012_Vol7(10)_No1_pg30-37.pdf
5. Galie N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, et al. 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension: the joint task force for the diagnosis and treatment of pulmonary hypertension of the european society of cardiology (ESC) and the european respiratory society (ERS): endorsed by: association for european paediatric and congenital cardiology (AEPC), international society for heart and lung transplantation (ISHLT). *Eur Heart J.* (2016) 37:67–119. doi: 10.1093/eurheartj/ehv317
6. Galie N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, et al. 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension: the joint task force for the diagnosis and treatment of pulmonary hypertension of the european society of cardiology (ESC) and the european respiratory society (ERS): endorsed by: association for european paediatric and congenital cardiology (AEPC), international society for heart and lung transplantation (ISHLT). *Eur Respir J.* (2015) 46:903–75. doi: 10.1183/13993003.01032-2015
7. Hoeper MM, Kramer T, Pan Z, Eichstaedt CA, Spiesshoefer J, Benjamin N, et al. Mortality in pulmonary arterial hypertension: prediction by the 2015 European pulmonary hypertension guidelines risk stratification model. *Eur Respir J.* (2017) 50:1700740. doi: 10.1183/13993003.00740-2017
8. Mak SM, Strickland N, Gopalan D. Complications of pulmonary hypertension: a pictorial review. *Br J Radiol.* (2017) 90:20160745. doi: 10.1259/bjr.20160745
9. Lang IM, Palazzini M. The burden of comorbidities in pulmonary arterial hypertension. *Eur Heart J Suppl.* (2019) 21(Suppl. K):K21–8. doi: 10.1093/eurheartj/suz205
10. Parvathy U, Rajan R, Faybushevich A, Zhanna K. Pulmonary dysfunction: a predictor of postoperative outcome in severe mitral stenosis. *Ann Clin Cardiol.* (2020) 2:60–9. doi: 10.4103/ACCJ.ACCJ_11_19
11. Sztrymf B, Souza R, Bertoletti L, Jais X, Sitbon O, Price LC, et al. Prognostic factors of acute heart failure in patients with pulmonary arterial hypertension. *Eur Respir J.* (2010) 35:1286–93. doi: 10.1183/09031936.00070209
12. Haase M, Kellum JA, Ronco C. Subclinical AKI–an emerging syndrome with important consequences. *Nat Rev Nephrol.* (2012) 8:735–9. doi: 10.1038/nrneph.2012.197
13. Bitker L, Sens F, Payet C, Turquier S, Duclos A, Cottin V, et al. Presence of kidney disease as an outcome predictor in patients with pulmonary arterial hypertension. *Am J Nephrol.* (2018) 47:134–43. doi: 10.1159/000487198
14. Rosenkranz S, Howard LS, Gomberg-Maitland M, Hoeper MM. Systemic Consequences of pulmonary hypertension and right-sided heart failure. *Circulation.* (2020) 141:678–93. doi: 10.1161/CIRCULATIONAHA.116.022362
15. Grinnan D, Farr G, Fox A, Sweeney L. The role of hyperglycemia and insulin resistance in the development and progression of pulmonary arterial hypertension. *J Diabetes Res.* (2016) 2016:2481659. doi: 10.1155/2016/2481659
16. Xue QL. The frailty syndrome: definition and natural history. *Clin Geriatr Med.* (2011) 27:1–15. doi: 10.1016/j.cger.2010.08.009
17. Zhang Y, Yuan M, Gong M, Li G, Liu T, Tse G. Associations between prefrailty or frailty components and clinical outcomes in heart failure: a follow-up meta-analysis. *J Am Med Dir Assoc.* (2019) 20:509–10. doi: 10.1016/j.jamda.2018.10.029
18. Zhang Y, Yuan M, Gong M, Tse G, Li G, Liu T. Frailty and clinical outcomes in heart failure: a systematic review and meta-analysis. *J Am Med Dir Assoc.* (2018). 19:1003–8.e1. doi: 10.1016/j.jamda.2018.06.009
19. Tse G, Gong M, Nunez J, Sanchis J, Li G, Ali-Hasan-Al-Saegh S, et al. Frailty and mortality outcomes after percutaneous coronary intervention: a systematic review and meta-analysis. *J Am Med Dir Assoc.* (2017) 18:1097.e1–10. doi: 10.1016/j.jamda.2017.09.002
20. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat.* (2008) 2:841–60. doi: 10.1214/08-AOAS169
21. Ishwaran H, Kogalur UB. Consistency of random survival forests. *Stat Probab Lett.* (2010) 80:1056–64. doi: 10.1016/j.spl.2010.02.020
22. Chou OHI, Zhou J, Lee TTL, Kot T, Lee S, Wai AKC, et al. Comparisons of the risk of myopericarditis between COVID-19 patients and individuals receiving COVID-19 vaccines: a population-based study. *Clin Res Cardiol.* (2022). doi: 10.1007/s00392-022-02007-0. [Epub ahead of print].
23. Zhou J, Zhang G, Chang C, Chou OHI, Lee S, Leung KSK, et al. Metformin versus sulphonylureas for new onset atrial fibrillation and stroke in type 2 diabetes mellitus: a population-based study. *Acta Diabetol.* (2022) 59:697–709. doi: 10.1007/s00592-021-01841-4
24. Ju C, Zhou J, Lee S, Tan MS, Liu T, Bazoukis G, et al. Derivation of an electronic frailty index for predicting short-term mortality in heart failure: a machine learning approach. *ESC Heart Fail.* (2021) 8:2837–45. doi: 10.1002/ehf2.13358
25. Chen Y, Zhou J, Lee S, Liu T, Hothi SS, Wong ICK, et al. Development of an electronic frailty index for predicting mortality in patients undergoing transcatheter aortic valve replacement using machine learning. *medRxiv.* (2020). doi: 10.1101/2020.12.23.20248770
26. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *Sci. World J.* (2001) 1:323–36. doi: 10.1100/tsw.2001.58
27. Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing.* (2016) 45:353–60. doi: 10.1093/ageing/afw039
28. Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324
29. Ishwaran H. Variable importance in binary regression trees and forests. *Electron J Stat.* (2007) 1:519–37. doi: 10.1214/07-EJS039
30. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-dimensional variable selection for survival data. *J Am Stat Assoc.* (2010) 105:205–17. doi: 10.1198/jasa.2009.tm08622
31. Koziol JA, Jia Z. The concordance index C and the mann–whitney parameter Pr(X>Y) with randomly censored data. *Biom J.* (2009) 51 467–74. doi: 10.1002/bimj.200800228
32. Chen Y, Liu C, Lu W, Li M, Hadadi C, Wang EW, et al. Clinical characteristics and risk factors of pulmonary hypertension associated with chronic respiratory diseases: a retrospective study. *J Thorac Dis.* (2016) 8:350–8. doi: 10.21037/jtd.2016.02.58
33. Kojima G, Iliffe S, Walters K. Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age Ageing.* (2018) 47:193–200. doi: 10.1093/ageing/afx162
34. Hjalmarsson C, Radegran G, Kylhammar D, Rundqvist B, Multing J, Nisell MD, et al. Impact of age and comorbidity on risk stratification in idiopathic pulmonary arterial hypertension. *Eur Respir J.* (2018) 51:1702310. doi: 10.1183/13993003.02310-2017
35. Hoeper MM, Simon RGJ. The changing landscape of pulmonary arterial hypertension and implications for patient care. *Eur Respir Rev.* (2014) 23:450–7. doi: 10.1183/09059180.00007814
36. Burger CD, Long PK, Shah MR, McGoon MD, Miller DP, Romero AJ, et al. Characterization of first-time hospitalizations in patients with newly diagnosed pulmonary arterial hypertension in the REVEAL registry. *Chest.* (2014) 146:1263–73. doi: 10.1378/chest.14-0193
37. Campo A, Mathai SC, Le Pavec J, Zaiman AL, Hummers LK, Boyce D, et al. Outcomes of hospitalisation for right heart failure in pulmonary arterial hypertension. *Eur Respir J.* (2011) 38:359–67. doi: 10.1183/09031936.00148310
38. Benza RL, Gomberg-Maitland M, Elliott CG, Farber HW, Foreman AJ, Frost AE, et al. Predicting survival in patients with pulmonary arterial hypertension: the REVEAL risk score calculator 20 and comparison With ESC/ERS-based risk assessment strategies. *Chest*. (2019) 156:323–37. doi: 10.1016/j.chest.2019.02.004
39. Benza RL, Miller DP, Gomberg-Maitland M, Frantz RP, Foreman AJ, Coffey CS, et al. Predicting survival in pulmonary arterial hypertension: insights from the registry to evaluate early and long-term pulmonary arterial hypertension disease management (REVEAL). *Circulation.* (2010) 22:164–72. doi: 10.1161/CIRCULATIONAHA.109.898122
40. Wernhart S, Hedderich J. Prediction of pulmonary hypertension in older adults based on vital capacity and systolic pulmonary artery pressure. *JRSM Cardiovasc Dis.* (2020). doi: 10.1177/2048004020973834
41. Hsich E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS. Identifying important risk factors for survival in patient with systolic heart failure using

random survival forests. *Circ Cardiovasc Qual Outcomes.* (2011) 4:39–45. doi: 10.1161/CIRCOUTCOMES.110.939371
42. Lee S, Zhou J, Li KHC, Leung KSK, Lakhani I, Liu T, et al. Territory-wide cohort study of Brugada syndrome in Hong Kong: predictors of long-term outcomes using random survival forests and non-negative matrix factorisation. *Open Heart*. (2021) 8:e001505. doi: 10.1136/openhrt-2020-001505
43. Lee S, Zhou J, Jeevaratnam K, Wong WT, Wong ICK, Mak C, et al. Paediatric/young versus adult patients with long QT syndrome. *Open Heart*. (2021) 8:e001671. doi: 10.1136/openhrt-2021-001671
44. Tse G, Lee S, Zhou J, Liu T, Wong ICK, Mak C, et al. Territory-wide chinese cohort of long QT syndrome: random survival forest and cox analyses. *Front Cardiovasc Med.* (2021) 8:608592. doi: 10.3389/fcvm.2021.608592
45. Chen C, Zhou J, Yu H, Zhang Q, Gao L, Yin X, et al. Identification of important risk factors for all-cause mortality of acquired long QT syndrome patients using random survival forests and non-negative matrix factorization. *Heart Rhythm.* (2021) 18:426–33. doi: 10.1016/j.hrthm.2020.10.022
46. Lee S, Zhou J, Guo CL, Wong WT, Liu T, Wong ICK, et al. Predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death. *Endocrinol Diabetes Metab.* (2021) 4:e00240. doi: 10.1002/edm2.240
47. Lee S, Zhou J, Leung KSK, Wu WKK, Wong WT, Liu T, et al. Development of a predictive risk model for all-cause mortality in patients with diabetes in Hong Kong. *BMJ Open Diabetes Res Care*. (2021) 9:e001950. doi: 10.1136/bmjdrc-2020-001950
48. Cure-Cure CAC, Gu P, Tian Y, Patel X, Wu T, Sviglin CO,et al. Predictors of all cause mortality and their gender differences in a hispanic population from barranquilla-colombia using machine learning with random survival forests. *Circulation*. (2018) 138(Suppl_1):A16252. doi: 10.1161/circ.138.suppl_1.16252
49. Ntaios G, Sagris D, Kallipolitis A, Karagkiozi E, Korompoki E, Manios E, et al. Machine-learning-derived model for the stratification of cardiovascular risk in patients with ischemic stroke. *J Stroke Cerebrovasc Dis.* (2021) 30:106018. doi: 10.1016/j.jstrokecerebrovasdis.2021.106018
50. Jung SY, Papp JC, Sobel EM Yu H, Zhang Z-F. Breast cancer risk and insulin resistance: post genome-wide gene–environment interaction study using a random survival forest. *Cancer Res.* (2019) 79:2784–94. doi: 10.1158/0008-5472.CAN-18-3688
51. Bauer Y, de Bernard S, Hickey P, Ballard K, Cruz J, Cornelisse P, et al. Identifying early pulmonary arterial hypertension biomarkers in systemic sclerosis: machine learning on proteomics from the DETECT cohort. *Eur Respir J.* (2021) 57:2002591. doi: 10.1183/13993003.02591-2020
52. Leha A, Hellenkamp K, Unsold B, Mushemi-Blake S, Shah AM, Hasenfuss G, et al. A machine learning approach for the prediction of pulmonary hypertension. *PLoS ONE.* (2019) 14:e0224453. doi: 10.1371/journal.pone.0224453
53. Ciampi A, Thiffault J, Nakache J-P, Asselain B. Stratification by stepwise regression, correspondence analysis and recursive partition: a comparison of three methods of analysis for survival data with covariates. *Comput Stat Data Anal.* (1986) 4:185–204. doi: 10.1016/0167-9473(86)90033-2

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

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: frontiersin.org/about/contact

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