European Radiology (2023) 33:8203–8213
https://doi.org/10.1007/s00330-023-09785-9

European Radiology
ESR EUROPEAN SOCIETY OF RADIOLOGY

CARDIAC

Deep learning-based prognostic model using non-enhanced cardiac cine MRI for outcome prediction in patients with heart failure

Yifeng Gao[1] · Zhen Zhou[1] · Bing Zhang[2] · Saidi Guo[2] · Kairui Bo[1] · Shuang Li[1] · Nan Zhang[1] · Hui Wang[1] · Guang Yang[3,4] · Heye Zhang[2] · Tong Liu[5] · Lei Xu[1]

Received: 16 November 2022 / Revised: 14 April 2023 / Accepted: 21 April 2023 / Published online: 7 June 2023

Abstract

Objectives To evaluate the performance of a deep learning-based multi-source model for survival prediction and risk stratification in patients with heart failure.

Methods Patients with heart failure with reduced ejection fraction (HFrEF) who underwent cardiac magnetic resonance between January 2015 and April 2020 were retrospectively included in this study. Baseline electronic health record data, including clinical demographic information, laboratory data, and electrocardiographic information, were collected. Short-axis non-contrast cine images of the whole heart were acquired to estimate the cardiac function parameters and the motion features of the left ventricle. Model accuracy was evaluated using the Harrell's concordance index. All patients were followed up for major adverse cardiac events (MACEs), and survival prediction was assessed using Kaplan–Meier curves.

Results A total of 329 patients were evaluated (age 54 ± 14 years; men, 254) in this study. During a median follow-up period of 1041 days, 62 patients experienced MACEs and their median survival time was 495 days. When compared with conventional Cox hazard prediction models, deep learning models showed better survival prediction performance. Multi-data denoising autoencoder (DAE) model reached the concordance index of 0.8546 (95% CI: 0.7902–0.8883). Furthermore, when divided into phenogroups, the multi-data DAE model could significantly discriminate between the survival outcomes of the high-risk and low-risk groups compared with other models ($p < 0.001$).

Conclusions The proposed deep learning (DL) model based on non-contrast cardiac cine magnetic resonance imaging could independently predict the outcome of patients with HFrEF and showed better prediction efficiency than conventional methods.

Clinical relevance statement The proposed multi-source deep learning model based on cardiac magnetic resonance enables survival prediction in patients with heart failure.

Key Points

- *A multi-source deep learning model based on non-contrast cardiovascular magnetic resonance (CMR) cine images was built to make robust survival prediction in patients with heart failure.*
- *The ground truth definition contains electronic health record data as well as DL-based motion data, and cardiac motion information is extracted by optical flow method from non-contrast CMR cine images.*
- *The DL-based model exhibits better prognostic value and stratification performance when compared with conventional prediction models and could aid in the risk stratification in patients with HF.*

Yifeng Gao and Zhen Zhou have contributed equally to this work and share first authorship.

Tong Liu and Lei Xu have contributed equally as the corresponding authors.

✉ Tong Liu
Ltanzhen@126.com

✉ Lei Xu
leixu2001@hotmail.com

[1] Department of Radiology, Beijing Anzhen Hospital, Capital Medical University, No.2, Anzhen Road, Chaoyang District, Beijing 100029, China

[2] School of Biomedical Engineering, Sun Yat-Sen University, Guangzhou, China

[3] Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, UK

[4] National Heart and Lung Institute, Imperial College London, London SW7 2AZ, UK

[5] Department of Cardiology, Beijing Anzhen Hospital, Capital Medical University, No.2, Anzhen Road, Chaoyang District, Beijing 100029, China

Keywords Deep learning · Magnetic resonance imaging · Heart failure · Optical flow · Prognosis

Abbreviation

3D	Three-dimensional
AI	Artificial intelligence
CMR	Cardiovascular magnetic resonance
CNN	Convolutional neural network
DAE	Denoising autoencoder
DL	Deep learning
EHR	Electronic health records
GCS	Global circumference strain
GLS	Global longitudinal strain
HF	Heart failure
HFrEF	Heart failure with reduced ejection fraction
HFpEF	Heart failure with preserved ejection fraction
MACEs	Major adverse cardiac events
MSA	Multi-level semantic adaption
MICE	Multiple imputation by chained equations
NT-proBNP	N-Terminal pro-brain natriuretic peptide
LGE	Late gadolinium enhancement
LV	Left ventricular
LVEF	Left ventricular ejection fraction
LVEDV	Left ventricular end-diastolic volume
LVESV	Left ventricular end-systolic volume
TRIPOD	Transparent Reporting of a multivariable model for Individual Prognosis Or Diagnosis
RV	Right ventricular

Introduction

Heart failure (HF) is a global pandemic in healthcare and affects more than 64 million people worldwide [1–3]. The prevalence of HF is still increasing among adults owing to the aging population and the improved therapies for cardiovascular diseases [4]. The overall prevalence of HF poses a heavy burden on healthcare costs. Meanwhile, patients with HF usually have a poor prognosis despite the therapeutic advances, with 1-year mortality ranging 15–30% and 5-year mortality rate up to 75% [4]. The prognosis prediction and risk stratification of HF can identify patients with a high risk and facilitate the application of robust clinical treatment strategies [5].

Conventional risk prediction models mainly include demographics, clinical biomarkers such as N-terminal pro-brain natriuretic peptide (NT-proBNP), and some other complications like diabetes and hypertension [6–8]. The ability of classical function parameters, including left ventricular ejection fraction (LVEF) and left ventricular end-systolic or end-diastolic volume (LVESV/LVEDV) derived from cardiovascular imaging modalities to predict the prognosis of HF, has been widely reviewed [4, 8–10]. However, the efficiency of conventional prediction models may be limited with wide variations in discrimination, meanwhile constrained by non-informative or random censoring and hazard risk linearity [11, 12].

Cardiac magnetic resonance (CMR) is the gold standard for volume and functional quantification when compared with other imaging modalities, such as echocardiography, owing to its high accuracy and reproducibility [13]. CMR is the tool of choice for the "one-stop" evaluation of cardiac status, including motion, function, perfusion, cardiac viability, and fibrosis of the myocardium [14]. Functional information from non-contrast cine and tissue characteristics from late gadolinium enhancement (LGE) and quantitative T1 mapping can aid in deciphering the cause for cardiac dysfunction in HF and serve as powerful predictors of major adverse cardiovascular outcomes [15–18]. However, classical CMR models often concentrate on single-source parameter. This may be imperfect for adverse outcome prediction especially in patients with moderate to severe impaired cardiac function, for most function parameters are closely related to LVEF [19]. Meanwhile, conventional quantitative CMR parameters are mainly based on manual cardiac segmentation approaches, which are time-consuming and tedious. Therefore, a more robust and convenient risk model is needed to make accurate prognostic predictions in patients with HF, especially in those with severe cardiac dysfunction.

Artificial intelligence (AI) is being increasingly applied in various aspects of cardiovascular imaging and holds immense potential for the analysis of multi-source dataset [20]. AI offers innumerable benefits in earlier disease diagnosis and detection as well as in accurate prediction of mortality and hospitalization risks [21–24]. As one of the most advanced machine learning (ML) subtypes, deep learning (DL) can automatically accomplish multi-source parameter integration using a convolutional neural network (CNN) or U-Net architecture [25, 26] and improve the prediction over classical statistical methods through higher dimensional variables [27].Computational models based on DL methods have been proven to possess good survival prediction and risk stratification values [28–31]. However, current DL prediction models based on cardiovascular imaging modalities are limited to single-dimension parameters, which hinders the process of clinical practice translation. The application of AI in the prognostic assessment of patients with heart failure remains to be explored. To fully exploit the predictive potential of CMR in HF, a new multi-source survival prediction model combined with a DL method was built in this study and the performance of this model to improve the feasibility and accuracy of HF prediction was validated.

Materials and methods

Study population

In this retrospective observational study, 503 patients referred to our Department of Radiology and diagnosed with heart failure with reduced ejection fraction (HFrEF, in which LVEF is ≤40%) according to the ACC/AHA guidelines were consecutively included [4]. All patients underwent routine CMR examinations between January 2015 and April 2020 and were followed up until November 2020 (IQR, 212–1238; median follow-up time of 1041 days) to obtain the outcomes and end events. Those with congenital heart diseases, infiltration myocardiopathy (e.g., amyloidosis and sarcoidosis), atrial fibrillation, and acute myocardial infarction within 1 month were excluded. Patients who have undergone cardiac resynchronization therapy and those with implantable cardiac defibrillator were also excluded. Patient absence or poor quality of CMR data were the other exclusion criteria. Finally, a total of 329 patients with HFrHF were included; 174 patients were excluded because of missing CMR (49 patients), poor image quality (13 patients), or lost to follow-up (67 patients). Otherwise, 45 patients were excluded due to the absence of NT-proBNP, which is an important laboratory marker associated with HF (see Fig. 1). All enrolled patients signed informed consent, and the study was reviewed and approved by the local ethics committee.

CMR protocol

The CMR study was performed using two different 3 T magnetic resonance imaging (MRI) systems (Magnetom Verio; Siemens AG Healthcare, or MR750w, General Electric Healthcare) with retrospective ECG and respiratory gate. Cardiac volume and function parameters were acquired using cardiac cine imaging. Both cine images derived from two different MRI systems were based on a balanced steady-state free protocol (Siemens: True Fast Imaging with Steady-state Precession, True-FI, SP and GE750w: Fats Imaging Employing Steady State Acquisition, FIESTA). The cardiac cine images covered bi-ventricle in short and long axes, with 25 reconstruction phases per cardiac cycle, covering from the base to the apex. The detail of CMR scan parameters is shown in Supplementary Materials.

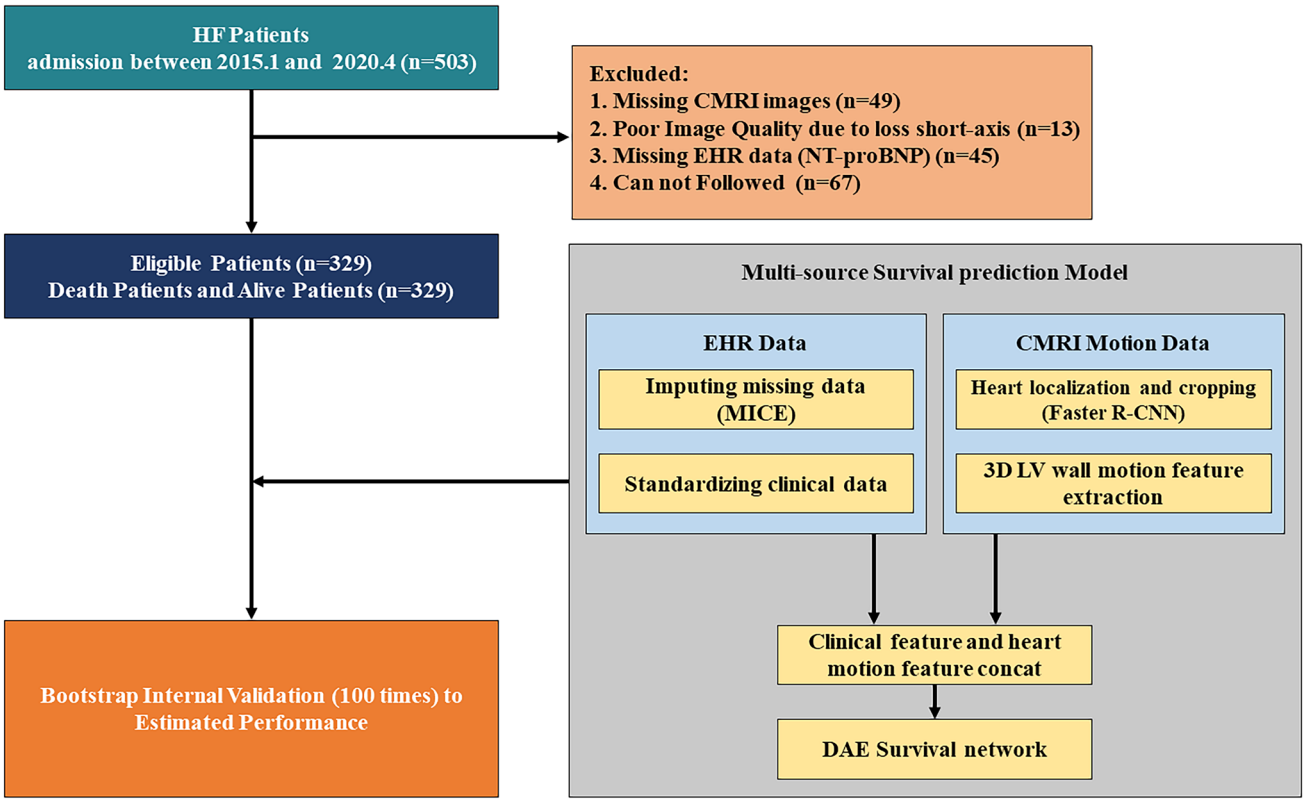

Fig. 1 A flow diagram showing the selection and analysis of patients with heart failure (HF) ($n=329$). HF, heart failure; CMRI, cardiac MRI; EHR, electronic health record; MICE, multivariate imputation by chained equations; R-CNN, region-based convolutional neural network; LV, left ventricle; DAE, denoising autoencoder

EHR datasets

Electronic health records (EHR) datasets consisted of clinical data and functional CMR parameters. The clinical data were numerical diagnostic indicators, including demographic and anthropometric information, clinical history, and electrocardiographic variables. Laboratory data (white blood cell count, creatinine level, etc.) were simultaneously collected from the medical records retrospectively. All the EHR characteristics used to construct the multi-source model are presented in Table 1.

CMR functional and strain analysis was conducted using commercial workstation CVI42 software (Version 5.6.3 Circle Cardiovascular Imaging). Endocardial and epicardial contours at end-systolic and diastolic cardiac phases were automatically delineated in both short- and long-axes cardiac cine by a radiologist (G.Y., with 4 years of experience in CMR) and manually amended by another reviewer (Z.Z., with 7 years of experience in CMR) according to standard operating procedures [8]. The papillary muscle and trabeculae were excluded in volume, and the anterior and inferior insertions in the short axis were manually signed to ensure the accuracy of volume calculation. Functional parameters included biventricular ejection fraction (EF), EDV, and ESV. Three-dimensional (3D) tissue tracking of global circumference and longitudinal strain (GCS and GLS) of both LV and right ventricle (RV) were also automatically performed and manually corrected using CVI42 to reflect the deformation of the myocardium. The CMR parameters used for the construction of the model are also shown in Table 1.

Follow-up

All patients were followed up by attending physicians at an interval of 6 months via a standard phone call questionnaire. The primary endpoints were major adverse cardiac events (MACEs) including cardiovascular death, rehospitalization because of cardiac dysfunction, and cardiac transplantation. Cardiac death was defined as HF death, fatal MI, sudden death, stroke, or cardiovascular hemorrhage. All medical personal adjudicated events were identified according to signs, test findings, and medical therapy based on annual telephone interviews and hospitalization data. The type of outcome was the right-censored time-to-event outcome, where censored patients signified those who did not suffer the outcome of death events during the follow-up period and the non-censored category referred to patients whose survival information was accurate from diagnosis to death. The duration was defined as the time from the start of hospitalization to the endpoint events or the last contact until the beginning of November 2020.

Table 1 Baseline patient characteristics

	All patients ($n=329$)
Clinical characteristics	
Age, mean (years)	54.0 ± 14.0
Male (%)	254 (77.2)
Weight (kg)	73.7 ± 16.0
BMI, mean (kg/m^2)	25.9 ± 5.1
Hypertension (%)	151 (45.8)
Diabetes mellitus (%)	92 (32.9)
LBBB on ECG (%)	36 (10.9)
RBBB on ECG (%)	16 (4.8)
QRS duration, mean (ms)	116.6 ± 30.8
ALT (U/L)	39.0 ± 51.0
AST (U/L)	31.8 ± 29.3
Creatinine, mean (mmol/L)	86.8 ± 23.2
Na^+ (mmol/L)	139.4 ± 3.0
Cl^- (mmol/L)	103.3 ± 4.3
CRP, mean (mmol/L)	3.5 ± 5.4
Homocysteine (μmol/L)	23.3 ± 30.7
White blood cell (G/L)	7.3 ± 2.1
Hemoglobin (G/L)	149.1 ± 19.2
Platelets (G/L)	211.9 ± 66.5
BNP (pg/mL)	564.0 ± 698.8
CMR parameters*	
LVEDV (mL)	271.1 ± 103.9
LVESV (mL)	213.6 ± 98.5
LVEF (%)	22.9 ± 10.5
LVSV (	57.3 ± 25.4
LVMASS (g)	143.6 ± 46.7
RVEDV (mL)	129.6 ± 47.5
RVESV (mL)	88.4 ± 44.6
RVEF (%)	33.9 ± 14.9
RVSV	41.2 ± 19.6
LVGLS (%)	−6.1 ± 3.0
LVGCS (%)	−7.3 ± 3.1
RVGLS (%)	−8.5 ± 4.1
RVGCS (%)	−10.8 ± 6.1
Medication treatment	
Beta-blocker (%)	296 (90.0)
Spironolactone (%)	276 (83.9)
ACEI/ARB (%)	299 (90.9)

*CMR parameters derived from commercial software

BMI body mass index, *LBBB* left bundle brunch block, *RBBB* right bundle brunch block, *ECG* electrocardiogram, *ALT* alanine aminotransferase, *AST* aspartate aminotransferase, *CRP* C-reactive protein, *BNP* brain natriuretic peptide, *LVEDV* left ventricle end diastolic volume, *LVESV* left ventricle end systolic volume, *LVEF* left ventricle ejection fraction, *LVSV* left ventricle stroke volume, *RVEDV* right ventricle end diastolic volume, *RVESV* right ventricle end systolic volume, *RVEF* right ventricle ejection fraction, *RVSV* right ventricle stroke volume, *LVGLS* left ventricle global longitudinal strain, *LVGCS* left ventricle global circumference strain, *RVGLS* right ventricle global longitudinal strain, *RVGCS* right ventricle global circumference strain, *ACEI* angiotensin-converting enzyme inhibitors, *ARB* angiotensin receptor blocker

Deep learning

Our DL model used multi-source data, including EHR data and cardiac motion data, to predict the survival of patients with HF. Heart motion information and clinical information were mapped to the same dimensional space using feature extraction methods. Based on the joint representation feature, a neural network was applied for survival prediction.

Deep learning and conventional model setting

In the training datasets, methods based on our DL framework were applied for denoising autoencoder (DAE) survival prediction, including (a) Motion + DAE: only the 3D heart motion information from CMR data were used along with the DL-DAE methods; (b) EHR + DAE: only CMR function indices from commercial CMR software and clinical information were used together with DL-DAE methods; (c) Multi-data + DAE: a combination of the former two models was used. Other popular survival methods based on EHR data used for model comparison including (c) Cox-Reg, (d) Cox-Net, (e) GBSA, (f) FastKernelSurvSVM, (g) FastSurvSVM, and (h) RSF. The details of the latter six survival prediction methods are provided in the Supplementary Materials.

Deep learning framework construction

The DL frameworks are illustrated in Fig. 2. Briefly, the framework consisted of (a) a clinical feature extraction unit to process the EHR data via a standardized method while imputing the missing data using multiple imputation by chained equations method; (b) a motion feature extraction unit using multi-level semantic adaption [32] and optical flow value method to detect the heart position and 3D LV wall motion status; and (c) a DL architecture that combines the DAE units with Cox proportional hazards to robustly fuse the clinical and motion features for survival prediction. All short-axis CMR cine images were cropped to a size of 80 times 80 pixels using R Fast-convolutional neural network, which included the intact myocardium, and were concatenated from the apex regions to the basest layer for covering the entire 3D myocardium. The motion features were encoded in 3D spaces as time-resolved vectors by intra-myocardial together with endo-and epicardial boundary points. Then the DAE unit was adapted to reconstruct the input from a corrupted version to generate a latent representation with the robust feature by minimizing the mean square error loss. The observed outcome data (follow-up time) were employed for training the prediction branch, and the predicted output was the log hazard ratio. More detailed information on the motion feature extraction and DL frameworks construction is provided in Supplementary Materials.

Statistical analysis

The continuous variables were presented as mean ± SD for normal distribution and as median and interquartile range for non-normal variables. The categorical variables were expressed as percentages. $p < 0.05$ was statistically significant.

For model performance and internal validation, a comprehensive full confusion matrix including the Harrell's concordance index (C-index), precision, recall, and F1-score was used to measure the predictive accuracy based on the bootstrap technique [33]. This technique is suitable for correcting the optimistic estimation of C statistics because of the overfitting of statistics in small datasets [34]. To estimate the C-index, the model was fit to the original data and each bootstrap sample data. When applying the model fitted with the bootstrap dataset to the original data, the predictive accuracy was usually lower than the apparent accuracy evaluated using the model with the same data. The difference in these predictive abilities for each bootstrap sample was calculated and then extended to obtain the average across the 100-time bootstrap samples. Finally, the estimate of optimism was then subtracted from the apparent estimate of the predictive ability. Delong test was performed to evaluate the difference of C-index between the multi-DAE DL model and other DL or traditional cox model.

Kaplan–Meier curves were used to depict the survival probability estimated over time. The Kaplan–Meier plot was generated for the full sample using the out-of-bag predictions. During the internal validation procedure, each bootstrap sample was created by taking 100 random draws (with replacement) from the full sample. According to bootstrap resampling theory, for large n, each bootstrap sample will contain (on average) only 63.2% of the subjects from the full sample. After training $b = 1,...,$ 100 bootstrap samples, we identify the bootstrap samples for which that subject was out-of-bag and average the predictions of the subject across these bootstrap samples. Then patients with HF were stratified into low-risk and high-risk groups according to the median survival time derived from the out-of-bag results with each method. A log-rank test was then performed to discern whether these subgroups exhibited significantly different survival outcomes. SPSS 23.0 (SPSS, Inc.) and R version 3.6.3 (R Foundation for Statistical Computing) were used for statistical analyses. The details of the generation process are provided in the Supplementary Materials.

Results

Population characteristics

A total of 329 eligible patients were enrolled in the study to construct and validate the model. EHR data and CMR

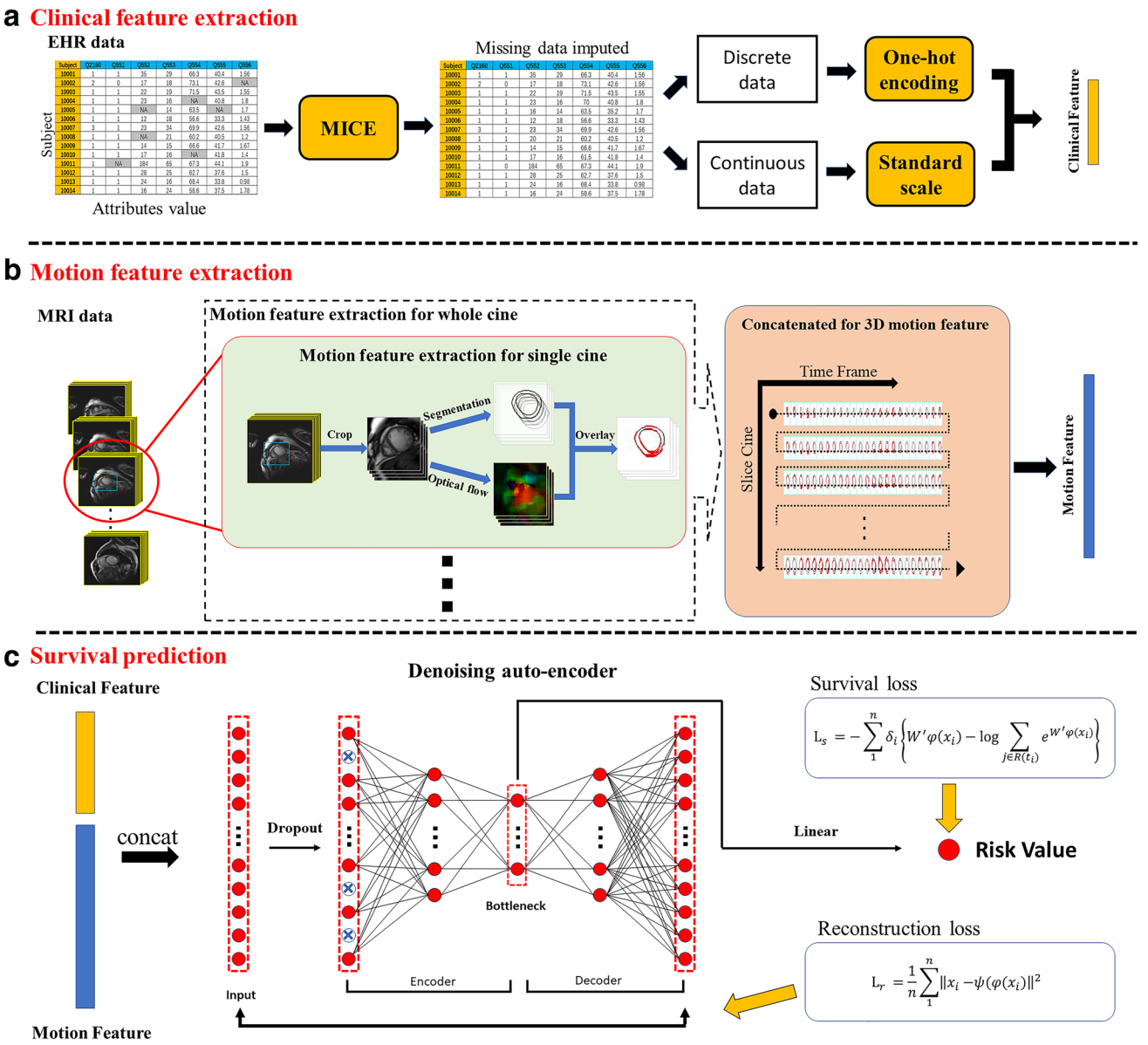

Fig. 2 An overview of the survival prediction network: **a** the feature extraction method for EHR data; **b** the motion feature extraction method using Faster R-CNN, a localization deep network, to detect the heart and a segmentation deep network and optical flow to get the motion feature of LV wall; **c** the deep learning method for survival prediction. EHR, electronic health records; MICE, multivariate imputation by chained equations; CNN, convolutional neural networks

characteristics used to construct the DL model are shown in Table 1. A total of 254 (77.2%) patients were men and the average age was 54 years. After following up, 18.8% (62 of 329) of the patients developed MACEs, with a median survival time of 495 days (IQR, 8–1900). Among these patients, 51 died because of cardiovascular diseases, 8 were re-hospitalized because of the aggravation of the HF symptoms, and 3 underwent cardiac transplantation.

C-index-based model accuracy and internal validation

Bootstrap-based internal validation was used to verify the accuracy and consistency of the prediction model as well as the conventional Cox model according to the guidelines recommended for TRIPOD (Transparent Reporting of a multivariable model for Individual Prognosis Or Diagnosis). These

Table 2 Comparison of different methods of Concordance Index (C-index) for survival prediction

Method	C-index (95%CI)	*p* value*	Precision	Recall	F1-score
DAE survival prediction method with different source data					
Multi-Data + DAE	0.8457 (0.7902–0.8883)	Ref	0.8316	0.9034	0.8405
Motion + DAE	0.8447 (0.8018–0.8808)	5.46e-5	0.8161	0.8822	0.8478
EHR + DAE	0.8299 (0.7751–0.8873)	2.75e-5	0.8017	0.8540	0.8270
Popular survival prediction methods with EHR data					
Cox-reg	0.6394 (0.5736–0.6988)	2.39e-6	0.5467	0.6657	0.6003
Cox-net	0.7073 (0.6418–0.7747)	1.84e-5	0.6778	0.7941	0.7313
GBSA	0.8253 (0.7716–0.8847)	7.49e-5	0.7857	0.8293	0.8069
FastKernelSurvivalSVM	0.6428 (0.5904–0.6915)	5.82e-6	0.6806	0.7305	0.7046
FastSurvivalSVM	0.6588 (0.6114–0.7047)	5.94e-5	0.7631	0.6370	0.6944
RSF	0.8129 (0.7721–0.8520)	1.48e-5	0.7571	0.8391	0.7960

*p value represents discrimination of C-index between Multi-Data + DAE model and other single-source DL model and traditional cox model, evaluated by Delong test

CMR cardiovascular magnetic resonance, *DAE* denoising autoencoder, *GBSA* galaxy-based search algorithm, *SVM* support vector machines, *RSF* random survival forest

guidelines allow optimistic apparent accuracy from the original full dataset [35, 36]. Table 2 shows the C-index obtained for bootstrap internal validation. The DL-based DAE survival prediction method consistently outperformed the comparison methods. Specifically, the multi-data DAE method achieved the best optimism-corrected C-index value (0.8457, 95% CI: 0.7902–0.883), precision (0.8316), and recall (0.9034). This was followed by the DAE methods combined with CMR or EHR data (0.8447 vs. 0.8299, 95% CI: 0.8018–0.8808 vs. 0.7751–0.8873, respectively). Figure 3 shows the Box plots of the C-index with 95% CI, which establishes that our proposed method achieved the best result.

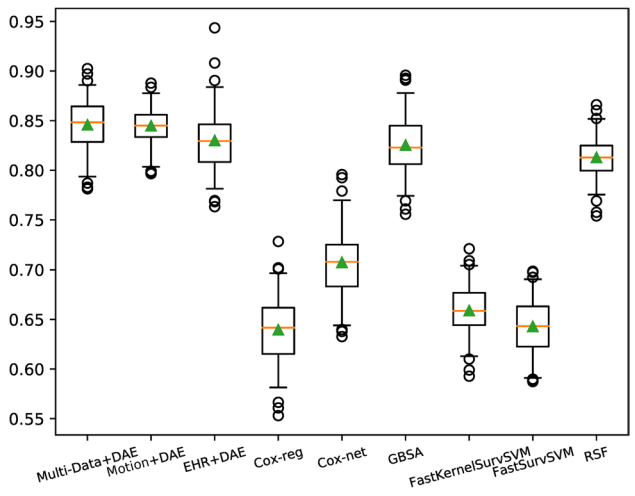

Fig. 3 Comparison of different methods for calculating the Concordance Index (C-index) for survival prediction. The larger C-index indicates better prediction results

When compared with all the popular survival prediction models, our proposed multi-data DAE-DL method yielded better results (Delong test, $p < 0.001$ for all), which shows the high representative ability of the features from multiple data sources. These results demonstrate that considering the enhancing prognostic information from multiple sources of data can help improve the survival prediction performance.

Survival prediction and outcome model validation

Kaplan–Meier survival curves of the different methods are presented in Fig. 4. The figure shows that the proposed multi-data source DL survival prediction method can offer personalized risk scores that can better stratify the patients into two risk groups. Furthermore, a log-rank test was performed to test the difference between the two curves. The proposed multi-data DAE survival prediction method achieved superior stratification performance ($p < 0.001$) when compared with both single-source (EHR or motion) DL method and other conventional prediction models.

Discussion

In this study, a DL survival model based on cardiac non-contrast cine images was developed and validated. The major finding of the study is that the DL-based multi-DAE model exhibits better prognostic value when compared with conventional prediction models and is helpful in the risk stratification of patients with HF.

Given its noninvasive characteristics and high accuracy, CMR has become the modality of choice for the comprehensive assessment of in vivo cardiac status. For patients

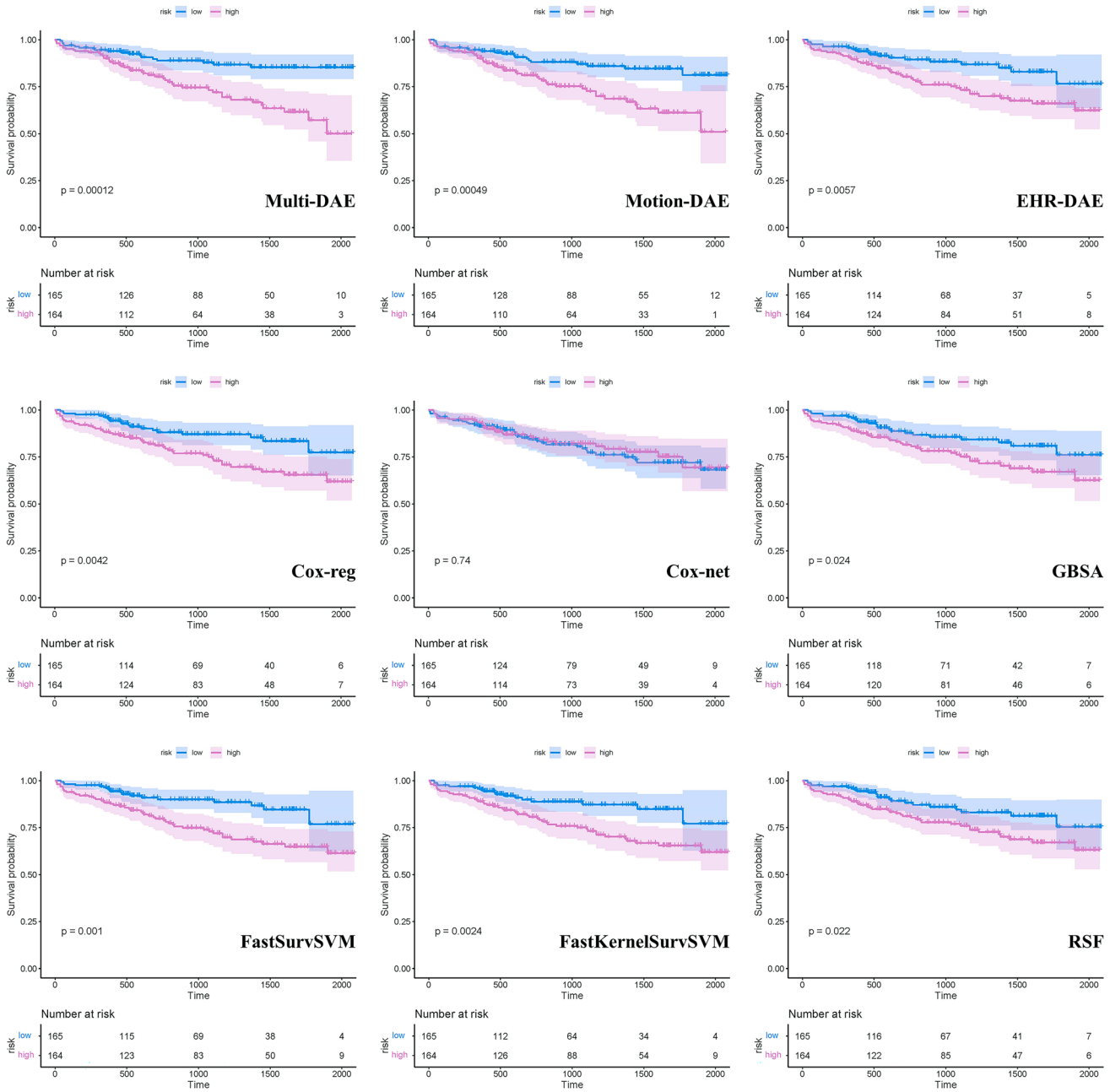

Fig. 4 Kaplan–Meier survives curves of different methods for the full sample using out-of-bag predictions. High-risk groups are plotted in purple, and low-risk groups are plotted in blue. Survival function estimates for each group with 95% CI are shown. The *x*-axis shows the time in days and the *y*-axis shows the probability of overall survival. The log rank *p* value to compare the survival curves between the risk groups is shown in each plot

with HF, conventional CMR can provide function and volume indices, such as EF and ventricle volume, which have been conventionally used to reflect motion conditions for risk stratification. Moreover, the method has a good prognostic value [37]. However, during our advanced-phase preparation study, we found that most patients in the study cohort were in stage C or D with severely impaired cardiac function, which may limit the prediction value of CMR [38]. Although myocardial fibrosis has been wildly employed for mortality prediction and risk stratification in patients with HF [39, 40], the extensive pattern of LGE in this study cohort showed no significant differences between patients who reached endpoints and those who survived. Emerging technical advances in CMR, such as feature

tracking strain analysis, have contributed to the evaluation of geometric cardiac deformation. Many studies have shown that both left ventricular (LV) and right ventricular (RV) strains can serve as independent prognostic factors in patients with HF [18, 41]. But given the inevitable relation between LV myocardial strain and LVEF, LV strain analysis may not offer enough prediction value in patients with severe impaired cardiac function. However, our previous study has confirmed that global 3D RV longitudinal strain is associated with HF [36]. As the earlier study was mainly conducted on single-source data, in this study, we created a novel integrated model that included both imaging and biological datasets for better survival prediction.

AI is the current trend in various cardiovascular imaging modalities against the general background of the popularity of big data [20, 42]. The use of AI has a far-reaching impact on all aspects of cardiovascular magnetic resonance imaging, such as imaging reconstruction, automatic segmentation, and survival prediction. Either supervised or unsupervised learning can yield vital and intricate information for building a prediction model [43]. Machine learning methods have been applied to predict adverse events in patients with HF and have shown better performance than conventional Cox models. The study conducted by Shu et al revealed that the application of the CMR-derived machine learning method to integrate clinical and CMR data can predict adverse events in patients with dilated cardiomyopathy accompanied by HF with reduced ejection fraction [44]. Furthermore, a recently conducted large, multi-cohort study developed a novel machine learning model to help identify race-specific contributors of HF [5]. Dawes et al designed a machine learning survival prediction model combined with 3D right ventricular motion features to predict the outcome in patients with pulmonary hypertension [45]. However, few studies have examined the predictive value of DL methods in patients with HF. For the motion dynamics of the heart in a status of dysfunction may result in a complex disturbance of cardiac motion and geometry deformation, we found that the computational models based on deep-learning methods, which combined cardiac motion with clinical indices, have been proven to have survival prediction and risk stratification values.

Our study has some limitations. First, the model was trained and validated based on data from patients with HFrEF. Therefore, whether the model is suitable for patients with heart failure with preserved ejection fraction (HFpEF) needs to be further verified, for the pathophysiology may much differ between these two kinds of HF [46, 47]. Second, we did not subdivide the patients with HF into ischemic and nonischemic groups because of the limited number of patients with ischemic cardiomyopathy, which may influence the robust of the model. Therefore, the universality of the model should be further tested to determine whether the pathogenesis of HF could affect the performance of the model because the pattern of cardiac motion may differ in these subgroups. Thirdly, for the ventricular segmentation, we focused on the deformation of the LV rather than that of the RV, thus resulting in the omission of some diseases that affect the right-side function, such as pulmonary hypertension, over the whole disease spectrum. Hence, further work is required to add right ventricular segmentation by creating a whole-heart model to augment the diagnostic efficiency. Finally, the study was conducted in single center, the model performance was evaluated using internal validation, and future multicenter prospective studies should be conducted to validate our results.

In conclusion, the multi-source DL model based on clinical EHR datasets, popular CMR characteristics, and cardiac motion derived from non-contrast cine CMR images enables accurate survival prediction and risk stratification in patients with HFrEF. Further studies may be conducted in multiple centers to perform further validation, and right ventricular segmentation may be investigated to improve the prediction efficiency of the model.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-023-09785-9.

Acknowledgements We would like to thank all the staff for their work on this study.

Funding This study was supported by grants from the National Key Research and Development Program of China (2022YFE0209800), the National Natural Science Foundation of China (U1908211), the National Natural Science Foundation of China (82271986), and the Capital's Funds for Health Improvement and Research Foundation of China (2020-1-1052).

Declarations

Guarantor The scientific guarantor of this publication is Lei Xu.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors (Nan Zhang) has significantly statistical expertise. And we would like to appreciate Dr. Zhiguang Liu from Beijing Institute of Heart and Blood Vessel Diseases for guidance on statistical methods.

Informed consent Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval Institutional Review Board was approved. The study protocol was approved by the Human Subjects Review Committee at Beijing Anzhen Hospital (Approval No. 2013007X).

Methodology
- retrospective
- diagnostic or prognostic study
- performed at one institution

References

1. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats A (2022) Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res cvac013. https://doi.org/10.1093/cvr/cvac013
2. McDonagh TA, Metra M, Adamo M et al (2021) ESC Scientific Document Group. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 42(36):3599–3726. https://doi.org/10.1093/eurheartj/ehab368
3. Ambrosy AP, Fonarow GC, Butler J et al (2014) The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. J Am Coll Cardiol 63(12):1123–1133. https://doi.org/10.1016/j.jacc.2013.11.053
4. Heidenreich PA, Bozkurt B, Aguilar D et al (2022) 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 145(18):e895–e1032. https://doi.org/10.1161/CIR.0000000000001063
5. Khan SS, Ning H, Shah SJ et al (2019) 10-Year risk equations for incident heart failure in the general population. J Am Coll Cardiol 73:2388–2397. https://doi.org/10.1016/j.jacc.2019.02.057
6. Coughlin SS, Neaton JD, Sengupta A, Kuller LH (1994) Predictors of mortality from idiopathic dilated cardiomyopathy in 356,222 men screened for the Multiple Risk Factor Intervention Trial. Am J Epidemiol 139(2):166–172. https://doi.org/10.1093/oxfordjournals.aje.a116978
7. Felker GM, Thompson RE, Hare JM et al (2000) Underlying causes and long-term survival in patients with initially unexplained cardiomyopathy. N Engl J Med 342:1077e84. https://doi.org/10.1056/NEJM200004133421502
8. Yancy CW, Jessup M, Bozkurt B et al (2017) 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. J Am Coll Cardiol 70(6):776–803. https://doi.org/10.1016/j.jacc.2017.04.025
9. Ponikowski P, Voors AA, Anker SD et al (2016) 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail 18(8):891–975. https://doi.org/10.1002/ejhf.592
10. Marrow BA, Cook SA, Prasad SK, McCann GP (2020) Emerging techniques for risk stratification in nonischemic dilated cardiomyopathy: JACC review topic of the week. J Am Coll Cardiol 75(10):1196–1207. https://doi.org/10.1016/j.jacc.2019.12.058
11. Segar MW, Vaduganathan M, Patel KV et al (2019) Machine learning to predict the risk of incident heart failure hospitalization among patients with diabetes: The WATCH-DM Risk Score. Diabetes Care 42:2298–2306. https://doi.org/10.2337/dc19-0587
12. Grothues F, Smith GC, Moon JC et al (2002) Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. Am J Cardiol 90(1):29–34. https://doi.org/10.1016/s0002-9149(02)02381-0
13. Patel AR, Kramer CM (2017) Role of cardiac magnetic resonance in the diagnosis and prognosis of nonischemic cardiomyopathy. JACC Cardiovasc Imaging 10(10 Pt A):1180–1193. https://doi.org/10.1016/j.jcmg.2017.08.005
14. Peterzan MA, Rider OJ, Anderson LJ (2016) The role of cardiovascular magnetic resonance imaging in heart failure. Card Fail Rev 2(2):115–122. https://doi.org/10.15420/cfr.2016.2.2.115
15. Becker MAJ, Cornel JH, van de Ven PM, van Rossum AC, Allaart CP, Germans T (2018) The prognostic value of late gadolinium-enhanced cardiac magnetic resonance imaging in nonischemic dilated cardiomyopathy: a review and meta-analysis. JACC Cardiovasc Imaging 11(9):1274–1284. https://doi.org/10.1016/j.jcmg.2018.03.006
16. Pezel T, Hovasse T, Sanguineti F et al (2021) Long-term prognostic value of stress CMR in patients with heart failure and preserved ejection fraction. JACC Cardiovasc Imaging 14(12):2319–2333. https://doi.org/10.1016/j.jcmg.2021.03.010
17. Puntmann VO, Carr-White G, Jabbour A et al (2016) T1-mapping and outcome in nonischemic cardiomyopathy: all-cause mortality and heart failure. JACC Cardiovasc Imaging 9(1):40–50. https://doi.org/10.1016/j.jcmg.2015.12.001
18. Romano S, Judd RM, Kim RJ et al (2018) 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 11(10):1419–1429. https://doi.org/10.1016/j.jcmg.2017.10.024
19. Pi SH, Kim SM, Choi JO et al (2018) Prognostic value of myocardial strain and late gadolinium enhancement on cardiovascular magnetic resonance imaging in patients with idiopathic dilated cardiomyopathy with moderate to severely reduced ejection fraction. J Cardiovasc Magn Reson 20(1):36. Published 2018 Jun 14. https://doi.org/10.1186/s12968-018-0466-7
20. Haq IU, Haq I, Xu B (2021) Artificial intelligence in personalized cardiovascular medicine and cardiovascular imaging. Cardiovasc Diagn Ther 11(3):911–923. https://doi.org/10.21037/cdt.2020.03.09
21. Yasmin F, Shah SMI, Naeem A et al (2021) Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future. Rev Cardiovasc Med 22(4):1095–1113. https://doi.org/10.31083/j.rcm2204121
22. Angraal S, Mortazavi BJ, Gupta A et al (2020) Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. JACC Heart Fail 8(1):12–21. https://doi.org/10.1016/j.jchf.2019.06.013
23. Segar MW, Jaeger BC, Patel KV et al (2021) Development and validation of machine learning-based race-specific models to predict 10-year risk of heart failure: a multicohort analysis. Circulation 143(24):2370–2383. https://doi.org/10.1161/CIRCULATIONAHA.120.053134
24. Leiner T, Rueckert D, Suinesiaputra A et al (2019) Machine learning in cardiovascular magnetic resonance: basic concepts and applications. J Cardiovasc Magn Reson 21(1):61. Published 2019 Oct 7. https://doi.org/10.1186/s12968-019-0575-y
25. Patravali J, Jain S, Chilamkurthy S (2017) 2D-3D fully convolutional neural networks for cardiac MR segmentation. arXiv:1707.09813 [cs.CV]. https://doi.org/10.48550/arXiv.1707.09813
26. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. arXiv:1505.04597 [cs.CV]. https://doi.org/10.48550/arXiv.1505.04597
27. Mortazavi BJ, Downing NS, Bucholz EM et al (2016) Analysis of machine learning techniques for heart failure readmissions. Circ Cardiovasc Qual Outcomes 9(6):629–640. https://doi.org/10.1161/CIRCOUTCOMES.116.003039
28. Pandey A, Kagiyama N, Yanamala N et al (2021) Deep-learning models for the echocardiographic assessment of diastolic dysfunction. JACC Cardiovasc Imaging 14(10):1887–1900. https://doi.org/10.1016/j.jcmg.2021.04.010

29. Liang F, Xie W, Yu Y (2017) Beating heart motion accurate prediction method based on interactive multiple model: an information fusion approach. Biomed Res Int 2017:1279486. https://doi.org/10.1155/2017/1279486BelloGA,DawesTJW,DuanJ,etal
30. Bello GA, Dawes TJW, Duan J et al (2019) Deep learning cardiac motion analysis for human survival prediction. Nat Mach Intell 1:95–104. https://doi.org/10.1038/s42256-019-0019-2
31. Dawes TJW, de Marvao A, Shi W et al (2017) Machine learning of three-dimensional right ventricular motion enables outcome prediction in pulmonary hypertension: a cardiac MR imaging study. Radiology 283(2):381–390. https://doi.org/10.1148/radiol.2016161315
32. Guo S, Xu L, Feng C, Xiong H, Gao Z, Zhang H (2021) Multi-level semantic adaptation for few-shot segmentation on cardiac image sequences. Med Image Anal 73:102170. https://doi.org/10.1016/j.media.2021.102170
33. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA (1982) Evaluating the yield of medical tests. JAMA 247(18):2543–2546
34. Smith GC, Seaman SR, Wood AM, Royston P, White IR (2014) Correcting for optimistic prediction in small data sets. Am J Epidemiol 180(3):318–324. https://doi.org/10.1093/aje/kwu140
35. Moons KG, Altman DG, Reitsma JB et al (2015) Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 162(1):W1-73. https://doi.org/10.7326/M14-0698
36. Harrell FE Jr, Lee KL, Mark DB (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15(4):361–387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4%3c361::AID-SIM168%3e3.0.CO;2-4
37. Rao RA, Jawaid O, Janish C, Raman SV (2021) When to use cardiovascular magnetic resonance in patients with heart failure. Heart Fail Clin 17(1):1–8. https://doi.org/10.1016/j.hfc.2020.09.001
38. Liu T, Gao Y, Wang H et al (2020) Association between right ventricular strain and outcomes in patients with dilated cardiomyopathy Heart 107(15):1233-1239. https://doi.org/10.1136/heartjnl-2020-317949
39. Lehrke S, Lossnitzer D, Schöb M et al (2011) Use of cardiovascular magnetic resonance for risk stratification in chronic heart failure: prognostic value of late gadolinium enhancement in patients with non-ischaemic dilated cardiomyopathy. Heart 97(9):727–732. https://doi.org/10.1136/hrt.2010.205542
40. Becker MAJ, Cornel JH, van de Ven PM, van Rossum AC, Allaart CP, Germans T (2018) The prognostic value of late gadolinium-enhanced cardiac magnetic resonance imaging in nonischemic dilated cardiomyopathy: a review and meta-analysis. JACC Cardiovasc Imaging 11(9):1274–1284. https://doi.org/10.1016/j.jcmg.2018.03.006
41. Buss SJ, Breuninger K, Lehrke S et al (2015) Assessment of myocardial deformation with cardiac magnetic resonance strain imaging improves risk stratification in patients with dilated cardiomyopathy. Eur Heart J Cardiovasc Imaging 16(3):307–315. https://doi.org/10.1093/ehjci/jeu181
42. Litjens G, Ciompi F, Wolterink JM et al (2019) State-of-the-art deep learning in cardiovascular image analysis. JACC Cardiovasc Imaging 12(8 Pt 1):1549–1565. https://doi.org/10.1016/j.jcmg.2019.06.009
43. Chartrand G, Cheng PM, Vorontsov E et al (2017) Deep learning: a primer for radiologists. Radiographics 37(7):2113–2131. https://doi.org/10.1148/rg.2017170077
44. Shu S, Hong Z, Peng Q et al (2021) A machine-learning-based method to predict adverse events in patients with dilated cardiomyopathy and severely reduced ejection fractions. Br J Radiol 94(1127):20210259. https://doi.org/10.1259/bjr.20210259
45. Dawes TJW, de Marvao A, Shi W et al (2017) Machine learning of three-dimensional right ventricular motion enables outcome prediction in pulmonary hypertension: a cardiac MR imaging study. Radiology 283(2):381–390. https://doi.org/10.1148/radiol.2016161315
46. Lee DS, Gona P, Vasan RS et al (2009) Relation of disease pathogenesis and risk factors to heart failure with preserved or reduced ejection fraction: insights from the framingham heart study of the national heart, lung, and blood institute. Circulation 119(24):3070–3077. https://doi.org/10.1161/CIRCULATIONAHA.108.815944
47. Hogg K, Swedberg K, McMurray J (2004) Heart failure with preserved left ventricular systolic function; epidemiology, clinical characteristics, and prognosis. J Am Coll Cardiol 43(3):317–327. https://doi.org/10.1016/j.jacc.2003.07.046