European Heart Journal - Cardiovascular Imaging (2024) **25**, 937–946
https://doi.org/10.1093/ehjci/jeae037

ORIGINAL PAPER

A deep patient-similarity learning framework for the assessment of diastolic dysfunction in elderly patients

Rohan Shah [1], Marton Tokodi [1,2], Ankush Jamthikar [1], Sabha Bhatti [1], Ehimare Akhabue [1], Grace Casaclang-Verzosa [1], Naveena Yanamala[1], and Partho P. Sengupta [1]*

[1]Division of Cardiovascular Diseases and Hypertension, Robert Wood Johnson University Hospital (RWJUH), Rutgers Robert Wood Johnson Medical School (RWJMS), 1 Robert Wood Johnson Place, New Brunswick, NJ 08901, USA; and [2]Heart and Vascular Center, Semmelweis University, Budapest, Hungary

Received 11 June 2023; revised 27 January 2024; accepted 1 February 2024; online publish-ahead-of-print 5 February 2024

Aims

Age-related changes in cardiac structure and function are well recognized and make the clinical determination of abnormal left ventricular (LV) diastolic dysfunction (LVDD) particularly challenging in the elderly. We investigated whether a deep neural network (DeepNN) model of LVDD, previously validated in a younger cohort, can be implemented in an older population to predict incident heart failure (HF).

Methods and results

A previously developed DeepNN was tested on 5596 older participants (66–90 years; 57% female; 20% Black) from the Atherosclerosis Risk in Communities Study. The association of DeepNN predictions with HF or all-cause death for the American College of Cardiology Foundation/American Heart Association Stage A/B ($n = 4054$) and Stage C/D ($n = 1542$) subgroups was assessed. The DeepNN-predicted high-risk compared with the low-risk phenogroup demonstrated an increased incidence of HF and death for both Stage A/B and Stage C/D (log-rank $P < 0.0001$ for all). In multi-variable analyses, the high-risk phenogroup remained an independent predictor of HF and death in both Stages A/B {adjusted hazard ratio [95% confidence interval (CI)] 6.52 [4.20–10.13] and 2.21 [1.68–2.91], both $P < 0.0001$} and Stage C/D [6.51 (4.06–10.44) and 1.03 (1.00–1.06), both $P < 0.0001$], respectively. In addition, DeepNN showed incremental value over the 2016 American Society of Echocardiography/European Association of Cardiovascular Imaging (ASE/EACVI) guidelines [net re-classification index, 0.5 (CI 0.4–0.6), $P < 0.001$; *C*-statistic improvement, DeepNN (0.76) vs. ASE/EACVI (0.70), $P < 0.001$] overall and maintained across stage groups.

Conclusion

Despite training with a younger cohort, a deep patient-similarity–based learning framework for assessing LVDD provides a robust prediction of all-cause death and incident HF for older patients.

* Corresponding author. E-mail: partho.sengupta@rutgers.edu
Study work was performed at Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, USA.

Structured Graphical Abstract

Key Question

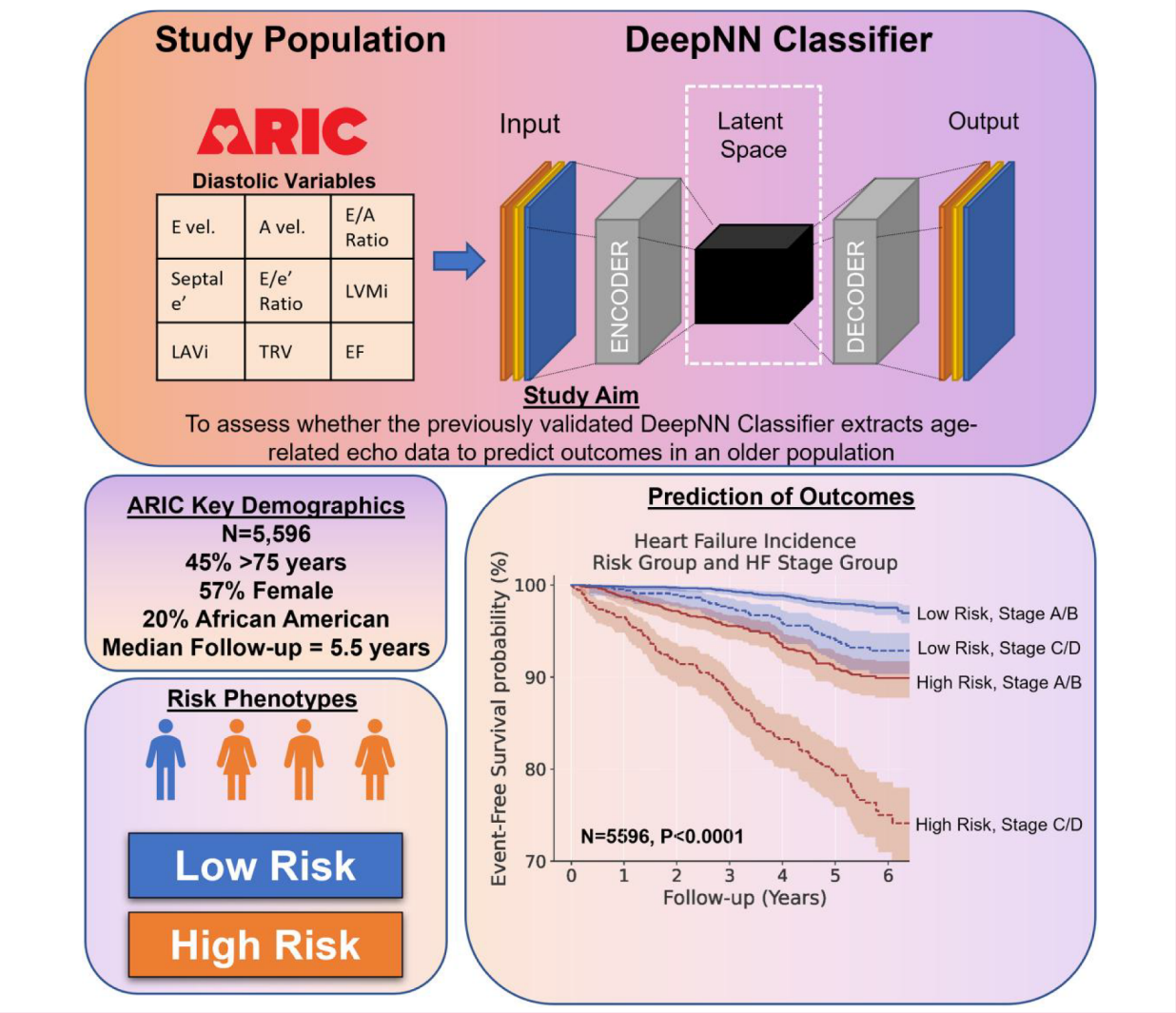
Can a deep neural network (DeepNN) model of left ventricular diastolic dysfunction (LVDD), previously validated in a younger cohort, be implemented for the prediction of all-cause death and heart failure (HF) incidence in an older adult population?

Key Finding

In an analysis of 5596 participants enrolled in the echocardiographic substudy of the Atherosclerosis Risk in the Communities (ARIC) Study, the DeepNN phenogroups independently predicted all-cause death and incident HF with incremental benefit over current guidelines.

Take-home Message

Despite being trained with a younger cohort, the DeepNN accounts for the age-related changes in diastolic function and can predict adverse events. The study's results suggest readiness to implement DeepNN in future HF clinical trials.



Echocardiographic parameters from the ARIC Study included ejection fraction (EF), left ventricular mass index (LVMi), early diastolic transmitral flow velocity (*E*), late diastolic transmitral flow velocity (*A*), *E*/*A* ratio, septal early diastolic relaxation velocity (e′), *E*/e′ ratio, left atrial volume index (LAVi), and tricuspid regurgitation peak velocity (TRV). These parameters were inputted into the classifier that extracts age-related echocardiographic data (and therefore able to predict age based on echo parameters alone) to batch predict the high-risk and low-risk phenogroups, and time-to-event outcomes were predicted.

Keywords left ventricular diastolic dysfunction • elderly • risk stratification • machine learning • phenogrouping

Introduction

Heart failure (HF) with preserved ejection fraction (HFpEF) affects greater than 64 million adults worldwide, with an increase of 50% predicted by 2035 in aging populations.[1] Echocardiography (echo) is the primary modality of risk stratification and prediction of incident HF, morbidity, and mortality. However, the increased prevalence of age-associated changes in left ventricular (LV) diastolic function makes risk stratification and grading of LV diastolic dysfunction (LVDD) especially difficult in older patients.[2–4] Moreover, HF is a heterogeneous disease, and differentiating the relative contribution of LVDD and other extra-cardiac co-morbidities in the pathophysiology is challenging.[5,6]

Age-adjusted thresholds for defining abnormalities in LVDD are not incorporated in the existing guidelines.[7] Moreover, univariable cut-offs in a linear decision tree may limit the assessment of age-dependent, highly non-linear relationships between LVDD variables. Additionally, not all LVDD variables may be measurable, which results in indeterminate or inaccurate grading.[8–10]

There is mounting evidence that deep neural networks (DeepNNs) can improve risk prediction by incorporating age-related changes directly as latent traits from the analysis of the imaging data.[11–13] We previously validated a DeepNN (https://wvu-model.herokuapp.com/) that integrates echo variables to determine whether an unknown patient is similar or dissimilar to patients at high risk for developing future adverse cardiac events.[14] We investigated this DeepNN in an older population with varying demographics using the multi-centric Atherosclerosis Risk in the Communities (ARIC) echo substudy.[15] We hypothesized that the deep learning echo models of LVDD would predict the future risk of incident HF even in the older population.

Methods

Study population

We investigated the application of a DeepNN framework for assessing LVDD in the individual-level data of the ARIC Study, a prospective epidemiologic cohort study started in 1987 investigating the aetiology of atherosclerosis and its clinical sequelae. The study initially enrolled 15 792 participants aged 45–64 from four US communities.[16,17] In the present study, we analysed all 5596 participants aged 76 ± 5 years (45% above the age of 75, 57% females, and 19.7% Black) who underwent a previously described echo substudy during the fifth visit between 2011 and 2013 with follow-up data until 2017 available from the National Heart Lung and Blood Institute's (NHLBI) BioLINCC database (*Figure 1*).[16] All available patients in the echo substudy, including those with coronary artery disease, atrial fibrillation, HFpEF, HF with reduced ejection fraction (HFrEF), right HF, and pulmonary hypertension (HTN), were included.

Outcomes of interest

The primary outcome of interest for the present study was incident HF, with all-cause death as secondary. The cohorts were followed from the time of their echo at the fifth visit to their last documented follow-up in the NHLBI BioLINCC data set. Incident HF events were identified by the first hospitalization event with HF as standardized by the coordinating study institutions and documented within the BioLINCC data set. HF was further classified into HFpEF and HFrEF based on the LV ejection fraction (LVEF) measured by echo. HFpEF was defined as ≥ 50% and HFrEF as LVEF < 50%.[16,18]

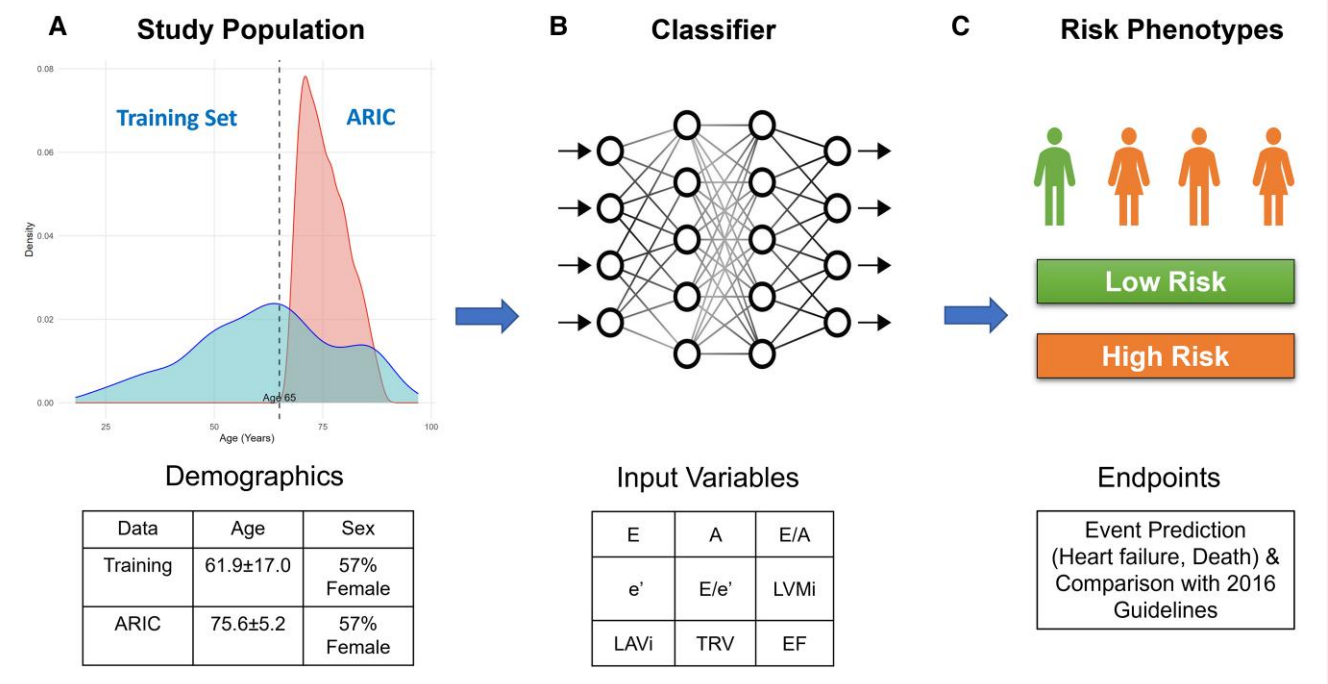


Figure 1 Study outline. (*A*) The age distribution of the training and the ARIC echocardiographic substudy used in the present external validation study. (*B*) Echocardiographic parameters used in DeepNN training included ejection fraction (EF), LV mass index (LVMi), early diastolic transmitral flow velocity (*E*), late diastolic transmitral flow velocity (*A*), *E*/*A* ratio, septal early diastolic relaxation velocity (e'), *E*/e' ratio, left atrial volume index (LAVi), and tricuspid regurgitation peak velocity (TRV). These parameters were inputted into the classifier and the 2016 ASE guidelines to batch predict the high-risk and low-risk phenogroups and ASE grades in the ARIC. (*C*) Model-predicted phenogroups and ASE grades were then compared.

Development of the deep patient-similarity learning framework

The deep patient-similarity learning framework makes its predictions (high- and low-risk) using a DeepNN model derived from a patient-similarity network that predicted the presence of LVDD by defining clusters of participants labelled high-risk and low-risk from the derivation cohort as previously published.[19] To keep it comparable with the 2016 American Society of Echocardiography/European Association of Cardiovascular Imaging (ASE/EACVI) guidelines, the model was developed using the nine echocardiographic variables used in the guidelines (*Table 1*). The model and its incremental value over 2016 ASE/EACVI LVDD guideline recommendations have been previously validated and outlined in Supplementary data online, *Section S1*.[14]

Model evaluation in older adults

We divided the ARIC cohort into Stage A/B or Stage C/D HF, based on the American College of Cardiology Foundation/American Heart Association (ACCF/AHA) guidelines to explore the significance of LVDD phenogrouping. We compared the model-predicted age with the actual chronological age to show a correlation and the ability of the DeepNN model to extract latent age-related data. We explored whether DeepNN-derived cut-offs could improve the 2016 ASE/EACVI guidelines' performance in outcomes prediction.

Statistical analysis

We explain in detail the model input features, the inclusion of ARIC data with missing values (see Supplementary data online, *Table S1*), model evaluation, outcomes comparisons, predicted and chronological age correlation, the incremental benefit of the DeepNN model over the ASE/EACVI LVDD guideline, and predicted cut-offs in Supplementary data online, *Section S3*. This manuscript used the ACC Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation reporting checklist to describe the study as noted in Supplementary data online, *Table S2*.[15] The following software solutions were used for analyses: Ayasdi workbench and SDK v7.9 (Ayasdi, Inc, Menlo Park, CA, USA, since acquired by SymphonyAI Sensa, Palo Alto, CA, USA) for the similarity network and phenogroup label generation, Stata v14.2 (StataCorp, College Station, TX, USA), and R v3.6.3 (R Foundation for Statistical Computing), and MedCalc® Statistical Software v20.027 (MedCalc Software Ltd, Ostend, Belgium). Statistical significance was tested with a two-sided *P*-value of <0.05.

Results

Study populations

The overall ARIC population's demographic, echo, and past medical history data are shown in *Table 1*. The ARIC population has a significantly higher average age than the DeepNN training cohort (ARIC: 76 ± 5 vs. DeepNN: 62 ± 17 years, $P < 0.0001$). Furthermore, ARIC had a higher percentage of participants aged over 65 and 75 compared with DeepNN's training data (see Supplementary data online, *Figure S1* and *Table S3*).

A total of 4054 (72.4%) and 1542 (27.6%) were at ACCF/AHA Stage A/B and Stage C/D, respectively. Of the 3635 (65.0%) participants predicted to be in the low-risk phenogroup, 2927 (80.5%) had Stage A/B, and 708 (19.5%) had Stage C/D HF. Of the 1961 (35.0%) participants in the high-risk phenogroup, 1127 (57.5%) had Stage A/B, and 834 (42.5%) had Stage C/D HF. A total of 359 participants (6.6%) met the HF hospitalization outcome and 811 (14.5%) participants met the all-cause death outcome (see Supplementary data online, *Table S4*).

In the Stage A/B group, the high-risk phenogroup included participants who were older, included more African Americans, and had a higher body mass index (BMI), body surface area (BSA), and blood pressure. The high-risk phenogroup also had higher baseline troponin T and N-terminal pro b-type natriuretic peptide (NT-ProBNP) levels. This group also had a higher rate of co-morbidities such as HTN, diabetes mellitus (DM), dyslipidaemia, anaemia, and chronic kidney disease (CKD). There were significant differences in the nine echo parameters used as input for the DeepNN model and in meeting the primary and secondary outcomes between the two the low- and high-risk phenogroups. Average peak longitudinal and average peak circumferential strain were statistically significant between the two phenogroups, with the low-risk group exhibiting higher strain magnitude than the high-risk group (*Table 1* and Supplementary data online, *Table S4*).

In the Stage C/D group, the high-risk phenogroup included the oldest subgroup of participants, more males, elevated BMI, BSA, and pulse pressure. Like the Stage A/B group, the high-risk phenogroup also had higher troponin T and NT-ProBNP levels. There were significant differences between eight of the nine parameters used for model training, with *E/A* ratio as the only variable with no difference between the two risk phenogroups ($P = 0.06$). There was also a larger percentage of those who met the primary and secondary outcomes in this HF group. Moreover, average peak longitudinal and average peak circumferential strain also remained statistically significant between the two phenogroups, with low-risk group showing higher strain magnitude than the high-risk group (*Table 1* and Supplementary data online, *Table S4*).

Model associations with outcomes predictions in older adults

In Kaplan–Meier analysis, the DeepNN model predicted a higher probability of incident HF and death in the high-risk phenogroup for both endpoints for both stage groups (all log-rank test $P < 0.0001$) (*Figure 2*).

Diastolic parameters used as input to the DeepNN correlated with age (see Supplementary data online, *Table S5*), and the addition of age, gender, or race to the DeepNN high-risk phenogroup provided no improvement in HF event prediction in multi-variable Cox analysis (*C*-statistic: both 0.76, $P \leq 0.0001$) (*Table 2*) and minimal improvement in predicting all-cause death (*C*-statistic: 0.64 vs. 0.68, both $P < 0.0001$) and the composite endpoint (*C*-statistic: 0.67 vs. 0.70, both < 0.0001).

In Cox univariable analysis across both endpoints and stage groups, high-risk phenogrouping, age, DM, anaemia, HTN, chronic obstructive pulmonary disease, and CKD remained significant covariates, with BMI remaining significant primarily in the Stage A/B cohorts. The high-risk phenogroup predicted an increased hazard ratio (HR) [confidence interval (CI)] for incident HF [11.84 (8.76–15.98), $P < 0.0001$], death [3.64 (3.04–4.36), $P < 0.0001$], and the composite endpoint [4.76 (1.04–5.60)] (*Table 3*; Supplementary data online, *Tables S6 and S7*). In multivariable analysis adjusted for underlying co-morbidities, the high-risk group continued to predict higher HRs for incident HF [7.92 (5.76–10.89), $P < 0.0001$], death [2.51 (2.06–3.05), $P < 0.0001$], and the composite endpoint [3.30 (2.76–3.93), $P < 0.0001$]. The high-risk phenogroup's predictive capability was upheld in both stage groups for each outcome (*Table 3*; Supplementary data online, *Tables S6 and S7*).

Re-classification of diastolic function grading

The 2016 ASE/EACVI guidelines identified 3499 (62%) normal patients, 432 (8%) with Grade 1 LVDD, 305 (6%) with high grade (2 or 3) LVDD, and 1360 (24%) with indeterminate grade LVDD (*Figure 3*). The cumulative risk for the composite outcome was higher in patients with high grade vs. Grade 1 and indeterminate LVDD (event rates: 26% vs. 13% vs. 10%). Among participants identified with high-grade LVDD, 277 (91%) were concordantly classified in the high-risk phenogroup. However, only 3009 (86%) and 69 (16%) of the normal participants and participants with Grade 1 LVDD, respectively, were concordantly classified in the low-risk phenogroup ($P < 0.0001$). Moreover, 829 (61%) indeterminate participants were classified as high

Table 1 Comparison of cross-sectional demographic, clinical data, and echocardiographic diastolic parameters between high- and low-risk phenogroups per ARIC HF stage cohorts and re-classification table of 2016 ASE/EACVI diastolic dysfunction grading guidelines to DeepNN classifier phenogroups

		ACC/AHA Stage A/B			ACC/AHA Stage C/D		
	Overall	Low risk	High risk	*P*-value	Low risk	High risk	*P*-value
Demographic and clinical information							
Age	75.6 (5.2)	74.5 (4.8)	76.5 (5.2)	<0.0001	76.1 (5.3)	77.6 (5.3)	<0.0001
Gender (female)	57%	63.2%	61.1%	0.21	45.1%	39.2%	0.0204
Race (Black)	19.7%	19.4%	24.6%	0.0003	15.4%	17.4%	0.29
BMI	28.5 (6.0)	28.1 (5.6)	29.4 (6.3)	<0.0001	27.9 (6.2)	29.3 (6.6)	<0.0001
BSA	1.90 (0.24)	1.87 (0.24)	1.91 (0.24)	<0.0001	1.91 (0.23)	1.96 (0.25)	0.0001
Systolic Blood Pressure, mm Hg	129.5 (19.5)	128 (17)	134 (22)	<0.0001	128 (18)	130 (23)	0.20
Pulse Pressure, mm Hg	63.6 (15.0)	62 (14)	67 (16)	<0.0001	63 (15)	66 (17)	0.0009
Lab values							
Troponin T, ng/mL	0.014 (0.016)	0.011 (0.008)	0.015 (0.013)	<0.0001	0.015 (0.012)	0.023 (0.033)	<0.0001
NT-ProBNP, pg/mL	308 (933)	139 (138)	286 (518)	<0.0001	302 (417)	940 (2186)	<0.0001
C-reactive protein, mg/L	4.06 (7.43)	3.8 (6.5)	4.0 (6.2)	0.21	4.3 (9.5)	5.0 (9.5)	0.17
Past medical history (at Visit 5)							
Hypertension	3872 (69.2%)	1802 (61.6%)	859 (76.2%)	<0.0001	521 (73.6%)	690 (82.7%)	<0.0001
Diabetes mellitus	1599 (28.7%)	687 (23.6%)	337 (30.0%)	<0.0001	197 (28.0%)	378 (45.5%)	<0.0001
Dyslipidaemia	2749 (50.7%)	1696 (57.9%)	562 (49.9%)	<0.0001	266 (37.6%)	262 (31.4%)	0.0112
Current smoking	317 (5.7%)	166 (5.7%)	65 (5.8%)	0.91	44 (6.2%)	42 (5.0%)	0.32
Chronic obstructive pulmonary disease	305 (5.6%)	137 (4.8%)	39 (3.6%)	0.15	53 (7.6%)	76 (9.4%)	0.054
Anaemia	1366 (25.0%)	561 (19.6%)	280 (25.7%)	<0.0001	182 (26.2%)	343 (42.1%)	<0.0001
Chronic kidney disease	1141 (20.5%)	455 (15.6%)	238 (21.2%)	<0.0001	174 (24.8%)	274 (33.0%)	0.0004
Atrial fibrillation	494 (8.8%)	—	—	—	194 (27.4%)	300 (36.0%)	0.0003
Coronary artery disease	823 (14.7%)	—	—	—	358 (50.6%)	465 (55.8%)	0.0418
Heart failure	290 (5.2%)	—	—	—	60 (8.5%)	230 (27.6%)	<0.0001
Echocardiographic parameters (used for model development)							
Ejection fraction, %	65.0 (6.9)	67.1 (4.8)	63.0 (15.9)	<0.0001	66.5 (5.3)	59.0 (9.9)	<0.0001
E-wave velocity, cm/s	0.68 (0.19)	0.65 (0.16)	0.68 (0.20)	0.0001	0.67 (0.18)	0.77 (0.26)	<0.0001
A-wave velocity, cm/s	0.80 (0.20)	0.78 (0.17)	0.86 (0.22)	<0.0001	0.76 (0.17)	0.84 (0.25)	<0.0001
E/*A* ratio	0.86 (0.30)	0.86 (0.24)	0.81 (0.28)	<0.0001	0.89 (0.33)	0.93 (0.44)	0.057
Septal e′, cm/s	5.68 (1.48)	6.11 (1.44)	4.65 (0.99)	<0.0001	6.30 (1.56)	4.99 (1.31)	<0.0001
E/e′ ratio	10.28 (4.06)	9.14 (2.66)	12.48 (4.45)	<0.0001	8.83 (2.72)	12.57 (5.91)	<0.0001
Left atrial volume index, mL/m^2	26.32 (9.33)	22.94 (6.05)	29.31 (8.48)	<0.0001	25.33 (7.14)	35.17 (13.62)	<0.0001
Tricuspid regurgitation velocity, cm/s	2.40 (0.30)	2.34 (0.25)	2.46 (0.32)	<0.0001	2.36 (0.26)	2.52 (0.36)	<0.0001
Left ventricular mass index, g/m^2	80.24 (21.12)	71.1 (13.0)	91.7 (20.0)	<0.0001	74.4 (14.2)	101.8 (26.9)	<0.0001
Strain measurements (not used for model development)							
Average peak longitudinal strain, %*	−17.9 (2.6)	−18.6 (2.1)	−17.2 (2.5)	<0.0001	−18.0 (2.3)	−15.9 (3.6)	<0.0001
Average peak circumferential strain, %*	−27.7 (3.9)	−28.4 (3.3)	−27.5 (4.3)	<0.0001	−27.8 (3.6)	−25.6 (4.9)	<0.0001
2016 ASE/EACVI LVDD grading							
Normal	3499	2152 (88%)	291 (12%)		493 (78%)	137 (22%)	
Indeterminate	1360	378 (50%)	383 (50%)		99 (22%)	355 (78%)	
Grade 1	432	53 (22%)	191 (78%)		8 (6%)	134 (94%)	
Grade 2	236	18 (15%)	106 (85%)		6 (5%)	106 (95%)	
Grade 3	31	0 (0%)	6 (100%)		1 (4%)	24 (96%)	

*Not used in model training or development.
Depicted as mean (standard deviation), count (percentage), or percentage of whole.

risk, with an adjusted HF event rate of 14% vs. 10% or 13% in the indeterminate and Grade 1 groups, respectively. The model provides clinical value as it accounts for age-related changes in this population and can assign an informative risk profile for those that were typically deemed 'low risk' (normal and Grade 1) or when there was no risk profile associated at all (indeterminate).

The DeepNN model had an increased Harrell's *C*-statistic in Cox proportional hazard modelling for the prediction of incident HF [DeepNN (0.76) vs. ASE/EACVI (0.70), $P \leq 0.001$], all-cause death [DeepNN (0.63) vs. ASE/EACVI (0.60), $P = 0.001$], and the composite endpoint [DeepNN (0.66) vs. ASE/EACVI (0.62), $P \leq 0.001$] (*Table 4*). When combined, the prediction benefits persist [combined model (HF incidence: 0.76; all-cause death: 0.63; composite endpoint: 0.67) vs. ASE/EACVI alone (as above); all $P < 0.001$]. This prediction performance is persistent within the stage groups as well. However, when combined, the *C*-statistic increase with the combined model was significant (0.60, $P = 0.002$). The addition of DeepNN phenogrouping to the 2016 ASE/EACVI guidelines also improved the re-classification [continuous net re-classification index (NRI) 0.50, 95% CI 0.39–0.62, $P < 0.0001$] for HF-free survival at 5–6 years after echo. The improved re-classification was maintained in both HF stage groups (*Table 5*).

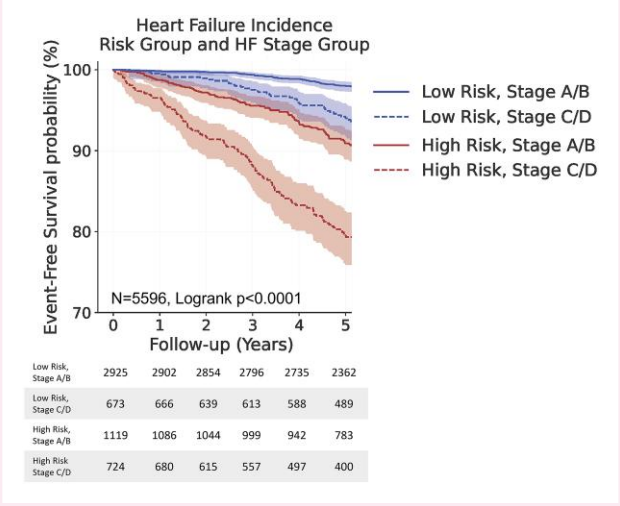


Figure 2 The risk of developing HF. This Kaplan–Meier analysis in the overall population by DeepNN phenogrouping (low- and high-risk) and HF stage groups (A/B and C/D) indicates that if a subject is classified as high risk, then they have a significantly higher risk for future events. Individual comparison between high- and low-risk phenogroups for each stage group log-rank $P < 0.0001$.

DeepNN-adjusted cut-off values

With the DeepNN model's improvement in classifying risk and prediction of future outcomes, we explored the possibility of adjusted cut-offs based on the DeepNN phenogrouping using the model's Youden index for each variable (see Supplementary data online, *Table S8*). Each DeepNN-derived adjusted cut-off is decreased compared with the 2016 ASE/EACVI guidelines and in more accordance with other proposed age-related cut-offs[7] and with published explorations in normal cardiac aging[20] (see Supplementary data online, *Figure S2*). However, there is a large amount of overlap between the different sets of cut-offs, and there was minimal difference in the change in HR when assessed using Cox proportional hazard modelling after adjustment of age, sex, and race. Only the adjusted septal e ′ velocity remained a unique independent predictor of incident HF to the DeepNN-derived cut-offs {adjusted cut-off HR 1.29 [95% CI 1.14–1.47], $P < 0.0001$ vs. ASE/EACVI 2016 cut-off HR 1.17 [95% CI 0.9 (0.98–1.40)], $P = 0.08$} (see Supplementary data online, *Table S9*). DeepNN-adjusted cut-offs, when used in the traditional linear decision tree format, however, showed no additive benefit over the conventional grading methods.

Table 2 Cox proportional hazards regression model with univariate and multi-variate analysis of HF incidence in the overall ARIC Study using DeepNN model

	Univariable analysis			Multi-variable analysis		
Covariates	***C*-index**	**HR**	***P*-value**	***C*-index**	**HR**	***P*-value**
Heart failure incidence						
High risk phenogroup	0.762 (0.735–0.789)	11.84 (8.76–15.98)	<0.0001	0.761 (0.735–0.787)	9.68 (7.13–13.16)	<0.0001
Age (years)	0.633 (0.604–0.662)	1.09 (1.07–1.11)	<0.0001		1.06 (1.04–1.08)	<0.0001
Gender	0.559 (0.533–0.585)	1.62 (1.31–1.99)	<0.0001		1.38 (1.12–1.70)	0.0024
Race	0.503 (0.482–0.524)	0.92 (0.71–1.21)	0.5583		—	—
All-cause death						
High-risk phenogroup	0.642 (0.622–0.663)	3.64 (3.04–4.36)	<0.0001	0.681 (0.662–0.700)	2.69 (2.24–3.24)	<0.0001
Age (years)	0.638 (0.618–0.658)	1.11 (1.09–1.12)	<0.0001		1.09 (1.07–1.10)	<0.0001
Gender	0.553 (0.535–0.570)	1.51 (1.32–1.74)	<0.0001		1.36 (1.18–1.56)	<0.0001
Race	0.507 (0.493–0.522)	1.05 (0.88–1.24)	0.5827		—	—
Composite endpoint						
High-risk phenogroup	0.672 (0.654–0.690)	4.76 (1.04–5.60)	<0.0001	0.704 (0.687–0.720)	3.63 (3.07–4.29)	<0.0001
Age (years)	0.646 (0.628–0.663)	1.11 (1.09–1.12)	<0.0001		1.09 (1.07–1.10)	<0.0001
Gender	0.548 (0.532–0.563)	1.46 (1.29–1.65)	<0.0001		1.29 (1.14–1.46)	<0.0001
Race	0.501 (0.489–0.514)	1.01 (0.87–1.18)	0.8749		—	—

Table 3 Univariate and multi-variate Cox regression models for risk factors associated with probability of the high-risk phenogroup for the prediction of the HF incidence in the overall, AHA Stage A/B, and Stage C/D population at ARIC Visit 5

	Univariable		Multi-variable	
Covariate	**HR (CI)**	***P***	**HR (CI)**	***P***
Overall ($n = 5596$)				
ML probability of high-risk phenogroup	11.84 (8.76–15.98)	<0.0001	7.92 (5.76–10.89)	<0.0001
Age (years)	1.09 (1.07–1.11)	<0.0001	1.05 (1.03–1.08)	<0.0001
Male	1.615 (1.31–1.99)	<0.0001	1.31 (1.05–1.64)	0.015
Black	0.92 (0.71–1.21)	0.56	—	—
Current smoking	1.45 (0.98–2.13)	0.06	—	—
Hypertension	2.12 (1.62–2.77)	<0.0001	1.45 (1.08–1.95)	0.0122
BMI (kg/m^2)	1.02 (1.00–10.4)	0.0169	1.00 (0.98–1.02)	0.76
Diabetes mellitus	1.91 (1.55–2.36)	<0.0001	1.34 (1.06–1.68)	0.0132
Chronic obstructive pulmonary disease	2.55 (2.07–3.15)	<0.0001	2.24 (1.61–3.11)	<0.0001
Anaemia	2.15 (1.73–2.68)	<0.0001	1.57 (1.24–1.97)	0.0001
Chronic kidney disease	0.47 (0.36–0.62)	<0.0001	1.44 (1.14–1.82)	0.0025
AHA/ACC Stage A/B ($n = 4054$)				
ML probability of high-risk phenogroup	8.63 (5.72–13.01)	<0.0001	6.52 (4.20–10.13)	<0.0001
Age (years)	1.10 (1.07–1.14)	<0.0001	1.07 (1.04–1.11)	<0.0001
Male	1.25 (0.92–1.70)	0.16	—	—
Black	1.16 (0.81–1.70)	0.42	—	—
Current smoking	1.6 (0.92–2.75)	0.1	—	—
Hypertension	1.83 (1.28–2.63)	0.001	1.31 (0.88–1.95)	0.19
BMI (kg/m^2)	1.04 (1.01–1.06)	0.0032	1.02 (0.99–1.05)	0.17
Diabetes mellitus	1.82 (1.33–2.49)	0.0003	1.35 (0.95–1.90)	0.09
Chronic obstructive pulmonary disease	2.46 (1.45–4.19)	0.0009	2.42 (1.40–4.19)	0.0017
Anaemia	2.52 (1.84–3.44)	<0.0001	1.92 (1.37–2.70)	0.0001
Chronic kidney disease	2.059 (1.47–2.88)	<0.0001	1.26 (0.88–0.18)	0.21
AHA/ACC Stage C/D ($n = 1542$)				
ML probability of high-risk phenogroup	8.89 (5.60–14.13)	<0.0001	6.51 (4.06–10.44)	<0.0001
Age (years)	1.05 (1.03–1.08)	0.0002	1.03 (1.00–1.06)	<0.0498
Male	1.30 (0.97–1.75)	0.0749	—	—
Black	0.87 (0.58–1.30)	0.589	—	—
Current smoking	1.29 (0.75–2.22)	0.3765	—	—
Hypertension	1.75 (1.16–2.62)	0.0041	1.40 (0.91–2.16)	0.12
BMI (kg/m^2)	1.00 (0.97–1.02)	0.8234	—	—
Diabetes mellitus	1.57 (1.18–2.09)	0.0021	1.15 (0.85–1.55)	0.3716
Chronic obstructive pulmonary disease	2.06 (1.37–3.11)	0.0014	1.83 (1.21–2.75)	0.0041
Anaemia	1.95 (1.47–2.60)	<0.0001	1.35 (1.00–1.82)	0.0475
Chronic kidney disease	1.70 (1.27–2.28)	0.0006	1.40 (1.03–1.90)	0.0318

'—' indicates the non-significant parameters in the univariate analyses that were omitted from the multi-variate analysis

Discussion

The primary aim of the study was to validate a previously developed and published machine learning (ML)-based model to predict the 5-year risk of incident HF in the older adult population using the ARIC echo substudy where over 45% of patients were above 75 years of age. Moreover, the diversity of the cohorts (57% females and 19.7% Black) enrolled from four communities in the USA is important for understanding model generalizability. First, the study demonstrates the ML model's added predictive value over the ASE/EACVI guidelines, especially for Stage A/B patients, evidenced by the improved Harrel's *C*-statistic in Cox modelling and NRI. Second, a key strength of the DeepNN classifier is its ability to classify and re-classify even in those with indeterminate grades or pre-clinical conditions like Grade 1 LVDD or Stage A/B HF. A total of 61% of the indeterminate grade and 84% of Grade 1 participants were re-classified as high risk per the ML model. Finally, the phenogroups remained

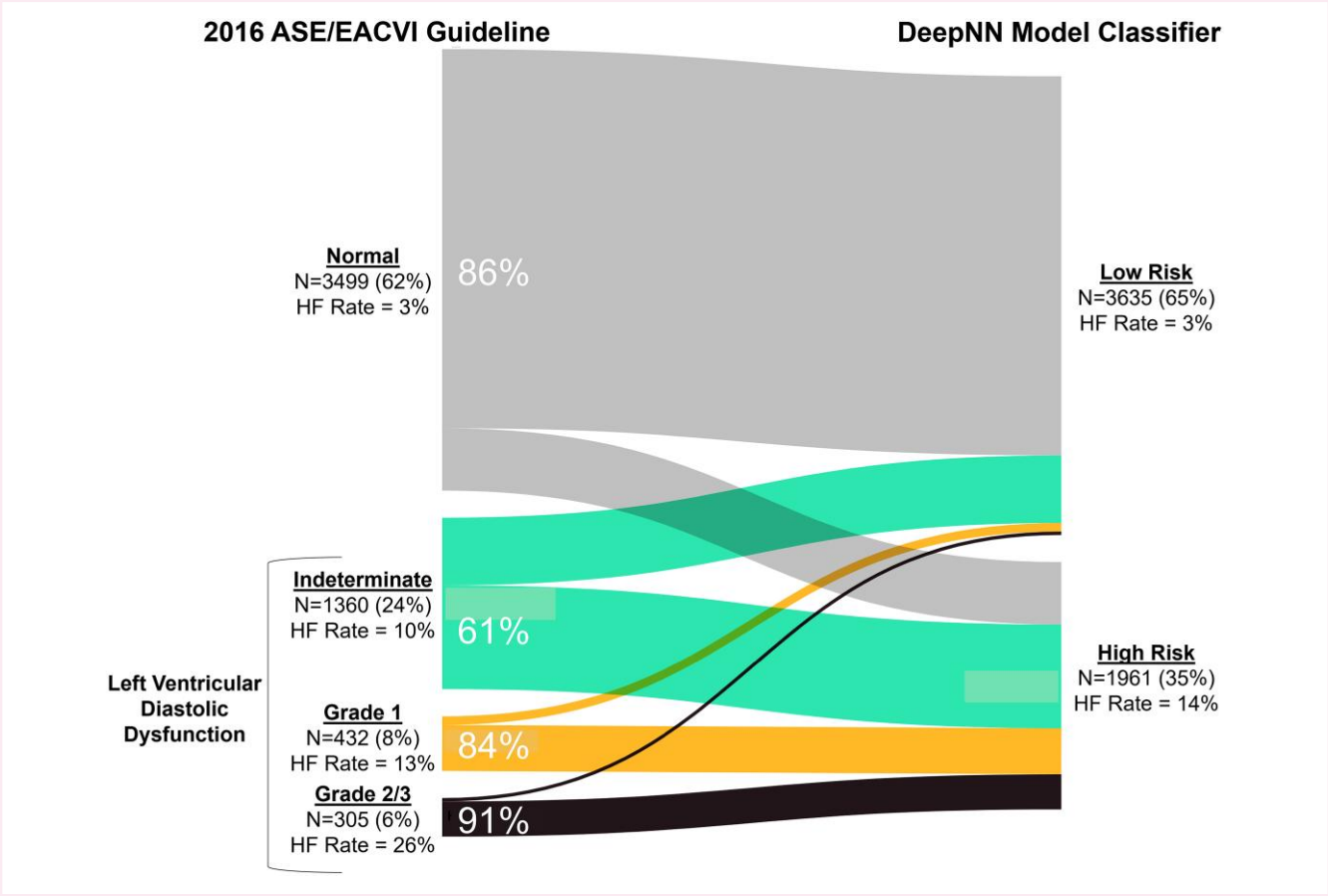


Figure 3 Alluvial plot demonstrating the re-classification of diastolic dysfunction clinical grades to DeepNN model classifier phenogroups. Most of the patients were deemed normal by the 2016 ASE guidelines. However, 14% of those patients were re-classified as high risk. Strikingly, 61% of indeterminate patients and 84% of Grade 1 patients were re-classified as high risk in this age group. Therefore, the biggest benefit of the classifier was in its ability to re-classify patients who may seem indeterminate or low grade using guideline-based grading but are high risk with a higher risk of incident HF.

Table 4 Comparison of Harrell's *C*-statistic in Cox proportional hazards modelling of the ASE/EACVI 2016 LVDD guideline classifications in comparison to the DeepNN for HF incidence, all-cause death, and composite endpoints in the overall, Stage A/B, and Stage C/D subgroups

Model	HF incidence	All-cause death	Composite endpoint
Overall			
ASE/EACVI	0.70	0.60	0.62
DeepNN	0.76 (<0.001)	0.63 (0.001)	0.66 (<0.001)
Combined	0.76 (<0.001)	0.63 (<0.001)	0.67 (<0.001)
Stage A/B			
ASE/EACVI	0.66	0.58	0.60
DeepNN	0.72 (0.0004)	0.60 (0.03)	0.63 (0.0006)
Combined	0.72 (0.0002)	0.60 (0.002)	0.64 (<0.0001)
Stage C/D			
ASE/EACVI	0.69	0.60	0.63
DeepNN	0.74 (0.001)	0.61 (0.50)	0.65 (0.047)
Combined	0.74 (<0.001)	0.61 (0.02)	0.65 (<0.001)

Presented as model *C*-statistic (*P*-value when compared to *C*-statistic of ASE model alone).

Table 5 Continuous NRI and IDI for the addition of DeepNN phenogrouping to 2016 ASE/EACVI grading for incident HF

Continuous net re-classification index Reference: 2016 ASE/EACVI grade Addition: DeepNN Phenogrouping			
Cohort	**NRI**	**95% CI**	***P*-value**
Overall	0.504	0.392–0.617	<0.0001
Stage A/B	0.422	0.258–0.587	<0.0001
Stage C/D	0.404	0.248–0.560	<0.0001
Integrated discrimination index Reference: 2016 ASE/EACVI grade Addition: DeepNN phenogrouping			
Cohort	**IDI**	**95% CI**	***P*-value**
Overall	−0.0094	−0.0182 to 0	0.04
Stage A/B	−0.0034	−0.0114 to −0.0047	0.4
Stage C/D	−0.0271	−0.0418 to −0.0125	0.0003

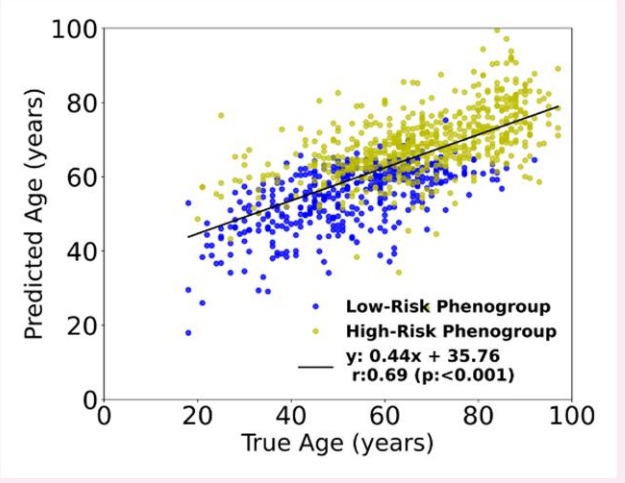


Figure 4 Correlation between model-predicted age vs. chronological age. Pearson correlation coefficient ($r = 0.69$) between the model-predicted and chronological ages.

independent predictors of HF incidence and all-cause death risk in Stage A/B and Stage C/D cardiomyopathy in the 5- to 8-year follow-up window after the echo. Notably, the high-risk phenogroup also predicted elevated NT-pro-BNP and troponin values and higher peak strain magnitudes further confirming the consistency of the model for providing individualized risk prediction.

The use of ML in HF risk prediction

Previous artificial intelligence (AI)/ML studies employed clustering algorithms, including K-means and hierarchical clustering, to identify novel diastolic dysfunction phenotypes.[21–25] However, our similarity-based approach is based upon a mathematical approach (topological data analysis) that extracts meaningful information by studying the shape and structure of the data. The algorithms are more robust to noise and outliers, can detect and represent the data structure at multiple scales, can handle missing data, and do not require a lot of tuning or prior knowledge to assess intricate data patterns.[19] This capability reduces the reliance on individual parameters and minimizes sensitivity to noise. We used a cluster-then-predict approach, in which a DeepNN model then predicts the cluster incorporated ASE/EACVI 2016 echocardiographic variables. The DeepNN model outperformed guideline-based classification while predicting HF hospitalization.

DeepNNs are labelled as 'black boxes' because of their opaque prediction processes. Nevertheless, these data-driven algorithms enhance classification by capturing non-linear relationships, rendering ML models more adaptable than conventional linear decision trees from traditional statistics. This flexibility enables risk prediction even with missing data, as demonstrated by the DeepNN model's re-classification of the indeterminate group.

Association of biological age with neural network activations

The relationship between diastolic function and biological age is well established.[26] There has been growing interest recently in using ML techniques to understand cardiovascular aging.[11–13,27] For example, a recent investigation using the UK Biobank database explored cardiovascular aging biomarkers derived from vascular function, cardiac motion, and myocardial fibrosis imaging.[27] It found that a fibrosis imaging marker, a key contributor to diastolic dysfunction, predicted accelerated aging. Additionally, 3D imaging analysis revealed uneven heart remodelling during aging, impacting LV contractility and relaxation patterns. Given this known relationship between biological age, LV remodelling, contractility, and diastolic function parameters, one can speculate that the DeepNN model used in the study may include some latent age-related information. To this end, in another technical study (see Supplementary data online, *Section* S2),[28] we dissected and isolated hidden nodes whose activation revealed that some of these nodes had learned to automatically regress age from echocardiography data (*Figure 4*). However, a bias is observed. This bias is consistent with previous studies that have explained the ability of the deep learning model to predict the biological age rather than the chronological age.[12,13,29–31] A model relying on functional traits that are both disease dependent and age dependent, however, may share the same latent features and have a confounding effect on model predictions. For example, age-related bias, as observed in this occasion, may lead to inherent weakness in distinguishing between a younger person with impaired function and an older person with preserved function. This bias could also be related to the characteristics of the training data set in which patients had a pre-existing indication for performing echocardiograms and thus may be sicker. Future use of a larger data set and new research approaches will be necessary to validate these concepts.

Clinical implications and future directions

The DeepNN model is superior to a linear decision tree models used previously and may improve age-related risk profiling in a multidimensional disease process such as LVDD.

Accurate risk stratification of the older population and therapeutic interventions may potentially diminish progression to HF, resulting in decreased hospitalizations and associated healthcare costs while maintaining the quality of life and exercise capacity. The DeepNN model has undergone external validation in various large study data sets and is prognostic of the risk of LVDD, increased LA pressures, worsening exercise performance, biomarker profiles, and incident HF. The current study reinforces the readiness of a publicly available model for identifying high-risk LVDD patients and warrants further prospective and longitudinal investigation and implementation for patient selection in clinical trials.

The current study has several limitations. First, the age of the training cohort of the DeepNN model was younger than the ARIC validation cohort, yet the DeepNN model still demonstrated a robust predictive performance. Further training of the model in an elderly cohort could further improve the predictive performance. Second, model performance can likely be improved by using a higher granularity of risk determination (i.e. more than two risk groups). Third, there is mounting evidence of the utility of strain or other biomarkers that reflect inflammation, cardiac remodelling, vascular changes, and physical performance markers (including exercise echocardiography and imaging) in determining cardiac functional capacity, and future investigations should explore the integration of these parameters. Finally, the strength of DeepNNs in the prediction of cardiovascular events in older patients opens up new opportunities in using such models for developing nomograms of cardiovascular aging in health and disease.

Supplementary data

Supplementary data are available at *European Heart Journal - Cardiovascular Imaging* online.

Funding

This work was supported by a National Science Foundation grant (grant #2125872 to P.P.S.).

Conflict of interest: P.P.S. is a consultant for RCE Technologies, Echo IQ. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Data availability

No new data were generated or analysed in support of this research. The data underlying this article are available in the National Heart, Lung, and Blood Institute Biologic Specimen and Data Repository Information Coordinating Center (NHLBI BioLINCC) at https://biolincc.nhlbi.nih.gov/studies/aric/.

References

1. Lloyd-Jones DM, Larson MG, Leip EP, Beiser A, D'Agostino RB, Kannel WB *et al.* Lifetime risk for developing congestive heart failure: the Framingham Heart Study. *Circulation* 2002;**106**:3068–72.
2. Forman DE, Fleg JL, Wenger NK. Cardiovascular disease in the elderly. In: Zipes DP, Libby P, Bonow RO, Mann DL, Tomaselli GF, Braunwald E, eds. *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*. 11th ed. Philadelphia, PA: Elsevier; 2019. p1735–66.
3. Kitzman DW, Scholz DG, Hagen PT, Ilstrup DM, Edwards WD. Age-related changes in normal human hearts during the first 10 decades of life. Part II (maturity): a quantitative anatomic study of 765 specimens from subjects 20 to 99 years old. *Mayo Clin Proc* 1988;**63**:137–46.
4. Cheng S, Xanthakis V, Sullivan LM, Lieb W, Massaro J, Aragam J *et al.* Correlates of echocardiographic indices of cardiac remodeling over the adult life course: longitudinal observations from the Framingham Heart Study. *Circulation* 2010;**122**:570–8.
5. Shah SJ, Kitzman DW, Borlaug BA, van Heerebeek L, Zile MR, Kass DA *et al.* Phenotype-specific treatment of heart failure with preserved ejection fraction: a multiorgan roadmap. *Circulation* 2016;**134**:73–90.
6. Zile MR, Baicu CF, Gaasch WH. Diastolic heart failure–abnormalities in active relaxation and passive stiffness of the left ventricle. *N Engl J Med* 2004;**350**:1953–9.
7. Shah AM, Claggett B, Kitzman D, Biering-Sorensen T, Jensen JS, Cheng S *et al.* Contemporary assessment of left ventricular diastolic function in older adults: the atherosclerosis risk in communities study. *Circulation* 2017;**135**:426–39.
8. Nagueh SF, Smiseth OA, Appleton CP, Byrd BF 3rd, 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. *Eur Heart J Cardiovasc Imaging* 2016;**17**:1321–60.
9. Oh JK, Miranda WR, Bird JG, Kane GC, Nagueh SF. The 2016 diastolic function guideline: is it already time to revisit or revise them? *JACC Cardiovasc Imaging* 2020;**13**:327–35.
10. Nikorowitsch J, Bei der Kellen R, Kirchhof P, Magnussen C, Jagodzinski A, Schnabel RB *et al.* Applying the ESC 2016, H2 FPEF, and HFA-PEFF diagnostic algorithms for heart failure with preserved ejection fraction to the general population. *ESC Heart Fail* 2021;**8**:3603–12.
11. Attia ZI, Friedman PA, Noseworthy PA, Lopez-Jimenez F, Ladewig DJ, Satam G *et al.* Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circulation: Arrhythmia and Electrophysiology* 2019;**12**:e007284.
12. Yang C-Y, Pan Y-J, Chou Y, Yang C-J, Kao C-C, Huang K-C *et al.* Using deep neural networks for predicting age and sex in healthy adult chest radiographs. *J Clin Med* 2021;**10**:4431.
13. Ganau A, Orrù M, Floris M, Saba PS, Loi F, Sanna GD *et al.* Echocardiographic heart ageing patterns predict cardiovascular and non-cardiovascular events and reflect biological age: the SardiNIA study. *Eur J Prev Cardiol* 2023. https://doi.org/10.1093/eurjpc/zwad254
14. 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;**14**:1887–900.
15. Sengupta PP, Shrestha S, Berthon B, Messas E, Donal E, Tison GH *et al.* Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME): a checklist: reviewed by the American College of Cardiology Healthcare innovation council. *JACC Cardiovasc Imaging* 2020;**13**:2017–35.
16. Shah AM, Cheng S, Skali H, Wu J, Mangion JR, Kitzman D *et al.* Rationale and design of a multicenter echocardiographic study to assess the relationship between cardiac structure and function and heart failure risk in a biracial cohort of community-dwelling elderly persons: the Atherosclerosis Risk in Communities study. *Circ Cardiovasc Imaging* 2014;**7**:173–81.
17. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) study: design and objectives. *Am J Epidemiol* 1989;**129**:687–702.
18. Reimer Jensen AM, Zierath R, Claggett B, Skali H, Solomon SD, Matsushita K *et al.* Association of left ventricular systolic function with incident heart failure in late life. *JAMA Cardiol* 2021;**6**:509–20.
19. 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.
20. Bello H, Norton GR, Peterson VR, Mmopi KN, Mthembu N, Libhaber CD *et al.* Hemodynamic determinants of age versus left ventricular diastolic function relations across the full adult age range. *Hypertension* 2020;**75**:1574–83.
21. Nouraei H, Nouraei H, Rabkin SW. Comparison of unsupervised machine learning approaches for cluster analysis to define subgroups of heart failure with preserved ejection fraction with different outcomes. *Bioengineering (Basel)* 2022;**9**:175.
22. 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.
23. 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.
24. Lancaster MC, Salem Omar AM, Narula S, Kulkarni H, Narula J, Sengupta PP. Phenotypic clustering of left ventricular diastolic function parameters: patterns and prognostic relevance. *JACC Cardiovasc Imaging* 2019;**12**:1149–61.
25. Chao CJ, Kato N, Scott CG, Lopez-Jimenez F, Lin G, Kane GC *et al.* Unsupervised machine learning for assessment of left ventricular diastolic function and risk stratification. *J Am Soc Echocardiogr* 2022;**35**:1214–1225.e8.
26. Munagala VK, Jacobsen SJ, Mahoney DW, Rodeheffer RJ, Bailey KR, Redfield MM. Association of newer diastolic function parameters with age in healthy subjects: a population-based study. *J Am Soc Echocardiogr* 2003;**16**:1049–56.
27. Shah M, de AIMH, Lu C, Schiratti PR, Zheng SL, Clement A *et al.* Environmental and genetic predictors of human cardiovascular ageing. *Nat Commun* 2023;**14**:4941.
28. Jamthikar AD, Shah R, Tokodi M, Sengupta PP, Yanamala N. Dissecting the latent representation of age inside a deep neural network's predictions of diastolic dysfunction using echocardiographic variables. *Biomedical Signal Processing and Controll* 2024;**92**. https://doi.org/10.1016/j.bspc.2024.106013
29. Libiseller-Egger J, Phelan JE, Attia ZI, Benavente ED, Campino S, Friedman PA *et al.* Deep learning-derived cardiovascular age shares a genetic basis with other cardiac phenotypes. *Sci Rep* 2022;**12**:22625.
30. Hwang I, Yeon EK, Lee JY, Yoo RE, Kang KM, Yun TJ *et al.* Prediction of brain age from routine T2-weighted spin-echo brain magnetic resonance images with a deep convolutional neural network. *Neurobiol Aging* 2021;**105**:78–85.
31. Duffy G, Clarke SL, Christensen M, He B, Yuan N, Cheng S *et al.* Confounders mediate AI prediction of demographics in medical imaging. *NPJ Digit Med* 2022;**5**:188.