

MINI-FOCUS: AI, MACHINE LEARNING, AND ECHOCARDIOGRAPHY

Deep-Learning Models for the Echocardiographic Assessment of Diastolic Dysfunction



Ambarish Pandey, MD, MSCS,^{a,*} Nobuyuki Kagiya, MD, PhD,^{b,c,d,*} Naveena Yanamala, MS, PhD,^{b,*} Matthew W. Segar, MD, MS,^a Jung S. Cho, MD, PhD,^{b,e} Márton Tokodi, MD,^{b,f} Partho P. Sengupta, MD, DM^b

ABSTRACT

OBJECTIVES The authors explored a deep neural network (DeepNN) model that integrates multidimensional echocardiographic data to identify distinct patient subgroups with heart failure with preserved ejection fraction (HFpEF).

BACKGROUND The clinical algorithms for phenotyping the severity of diastolic dysfunction in HFpEF remain imprecise.

METHODS The authors developed a DeepNN model to predict high- and low-risk phenogroups in a derivation cohort (n = 1,242). Model performance was first validated in 2 external cohorts to identify elevated left ventricular filling pressure (n = 84) and assess its prognostic value (n = 219) in patients with varying degrees of systolic and diastolic dysfunction. In 3 National Heart, Lung, and Blood Institute-funded HFpEF trials, the clinical significance of the model was further validated by assessing the relationships of the phenogroups with adverse clinical outcomes (TOPCAT [Aldosterone Antagonist Therapy for Adults With Heart Failure and Preserved Systolic Function] trial, n = 518), cardiac biomarkers, and exercise parameters (NEAT-HFpEF [Nitrate's Effect on Activity Tolerance in Heart Failure With Preserved Ejection Fraction] and RELAX-HF [Evaluating the Effectiveness of Sildenafil at Improving Health Outcomes and Exercise Ability in People With Diastolic Heart Failure] pooled cohort, n = 346).

RESULTS The DeepNN model showed higher area under the receiver-operating characteristic curve than 2016 American Society of Echocardiography guideline grades for predicting elevated left ventricular filling pressure (0.88 vs. 0.67; p = 0.01). The high-risk (vs. low-risk) phenogroup showed higher rates of heart failure hospitalization and/or death, even after adjusting for global left ventricular and atrial longitudinal strain (hazard ratio [HR]: 3.96; 95% confidence interval [CI]: 1.24 to 12.67; p = 0.021). Similarly, in the TOPCAT cohort, the high-risk (vs. low-risk) phenogroup showed higher rates of heart failure hospitalization or cardiac death (HR: 1.92; 95% CI: 1.16 to 3.22; p = 0.01) and higher event-free survival with spironolactone therapy (HR: 0.65; 95% CI: 0.46 to 0.90; p = 0.01). In the pooled RELAX-HF/NEAT-HFpEF cohort, the high-risk (vs. low-risk) phenogroup had a higher burden of chronic myocardial injury (p < 0.001), neuro-hormonal activation (p < 0.001), and lower exercise capacity (p = 0.001).

CONCLUSIONS This publicly available DeepNN classifier can characterize the severity of diastolic dysfunction and identify a specific subgroup of patients with HFpEF who have elevated left ventricular filling pressures, biomarkers of myocardial injury and stress, and adverse events and those who are more likely to respond to spironolactone.

(J Am Coll Cardiol Img 2021;14:1887-1900) © 2021 by the American College of Cardiology Foundation.

From the ^aDivision of Cardiology, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, USA; ^bCenter for Clinical Innovation, Division of Cardiology, West Virginia University Heart and Vascular Institute, Morgantown, West Virginia, USA; ^cDepartment of Cardiovascular Biology and Medicine, Juntendo University, Tokyo, Japan; ^dDepartment of Digital Health and Telemedicine R & D, Juntendo University, Tokyo, Japan; ^eDivision of Cardiology, Daejeon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea; and the ^fHeart and Vascular Center, Seemelweis University, Budapest, Hungary. *Drs. Pandey, Kagiya, and Yanamala contributed equally to this work and are joint first authors.

Jagat Narula, MD, served as the Guest Editor for this paper.

ABBREVIATIONS AND ACRONYMS

ASE = American Society of Echocardiography

AUROC = area under the receiver-operating characteristic curve

CI = confidence interval

DeepNN = deep neural network

GLS = global longitudinal strain

HF = heart failure

HFpEF = heart failure with preserved ejection fraction

HR = hazard ratio

LV = left ventricular

LVDD = left ventricular diastolic dysfunction

MLHFQ = Minnesota Living With Heart Failure Questionnaire

NT-proBNP = N-terminal pro-brain natriuretic peptide

PALS = peak atrial longitudinal strain

TDA = topological data analysis

VO₂peak = peak exercise oxygen uptake

The development of heart failure with preserved ejection fraction (HFpEF) encompasses a complex interplay of the varying burden of left ventricular diastolic dysfunction (LVDD) and extracardiac comorbidities (1-3). Because of this heterogeneous presentation, assessing the severity of LVDD is useful for understanding the relative contribution of cardiac and extracardiac factors underlying HFpEF (3,4). Although invasive hemodynamic assessment using cardiac catheterization is currently the gold standard, echocardiography is the most commonly used noninvasive test for routine assessment of diastolic function. Multiple studies have demonstrated the prognostic value of diastolic function parameters for predicting adverse clinical outcomes in patients with HFpEF (5,6). However, the algorithmic approaches for integrating the echocardiographic parameters of LVDD in patients with heart failure (HF), particularly HFpEF, using noninvasive echocardiographic methods remain imprecise (7,8). The correlation between echocardiographic and invasive measures of left ventricular (LV) filling pressure and diastolic function is modest (9-11). Furthermore, up to one-

third of patients with clinical HFpEF are deemed to have normal diastolic function (6). Moreover, the current guideline-based identification of LVDD relies on decision trees that provide limited insights into the multivariate and highly nonlinear interactions that underlie the development of LVDD (6,12,13). Thus, there is a need for novel approaches that can provide multiparametric integration of different echocardiographic characteristics to better delineate the phenotypic presentation of HFpEF. In this study, we combined unsupervised and supervised machine learning approaches to develop and validate a deep-learning solution for phenotyping patients with HFpEF. First, we used a patient-patient similarity network developed using compressed unsupervised representation of echocardiographic data to identify meaningful patient subgroups of LVDD (high and low risk). Following this, a deep neural network (DeepNN) classifier was developed using class labels assigned on the basis of identification of the high-

and low-risk subgroups. We validated the DeepNN in 5 external cohorts: 1) 2 prospectively recruited cohorts to test its ability to predict the presence of elevated LV filling pressure measured during cardiac catheterization and prognostic value in predicting adverse cardiac events; and 2) 3 diverse, multi-institutional independent datasets available from HFpEF clinical trials to test associations with cardiac biomarkers, exercise parameters, prognostic value, and the response to therapeutic interventions.

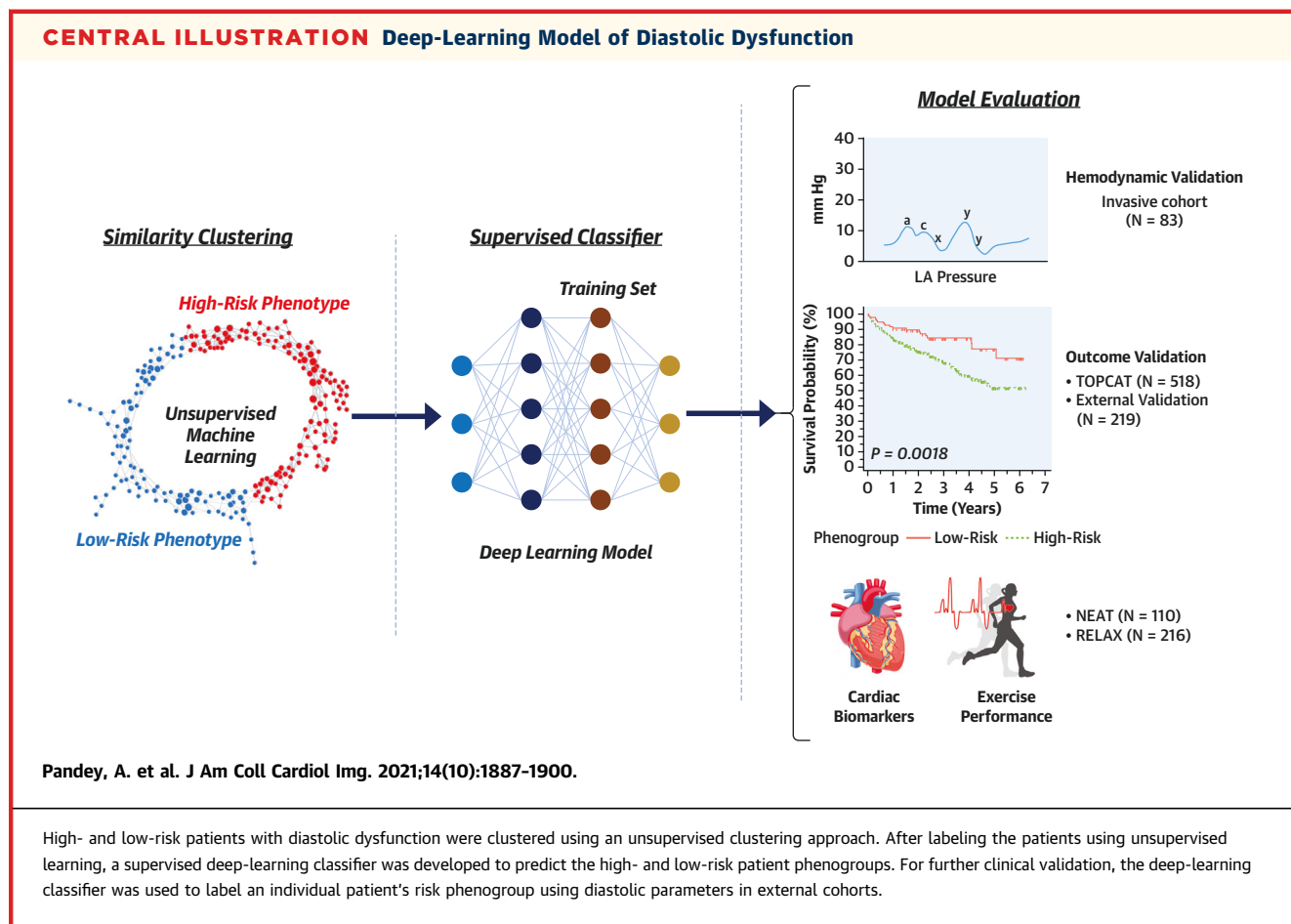
METHODS

DEVELOPMENT OF A DeepNN MODEL OF LVDD.

We previously described an unsupervised approach that integrates 9 echocardiographic variables into a similarity network using topological data analysis (TDA) in which patients (n = 1,242) with varying degrees of systolic and diastolic dysfunction are presented as a continuous loop (14). The details of the study cohort used for the model development are detailed in [Supplemental e-Methods-I](#). The developed unsupervised TDA network similarity model used routinely measured echocardiographic parameters, namely, ejection fraction, LV mass index, early diastolic transmitral flow velocity (E), late diastolic transmitral flow velocity (A), E/A ratio, early diastolic relaxation velocity (e'), E/e' ratio, left atrial volume index, and tricuspid regurgitation peak. We have previously published the method for defining the regions of a TDA loop with distinct prognostic outcomes (14). Briefly, once the TDA-based loop structure was created, the loop was subdivided into 4 groups of patients for statistical analysis. For this, for each patient with abnormal values according to the guidelines for the variables used in creating the model, the multidimensional Euclidean distance from the cutoff values were calculated. The distances were then used in the network to generate autogroups using agglomerative hierarchical clustering, which identified 40 clusters of nodes within the loop. The mean distance for each cluster was calculated, and the smallest distance was identified. Subsequently, 9 clusters adjacent to this cluster (the first 10 clusters) were selected to create the first region in the loop, and the remaining 30 clusters were divided into 3 regions containing 10 adjacent clusters of nodes each (14). The patients were assigned a low-risk (regions 1

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the [Author Center](#).

Manuscript received January 5, 2021; revised manuscript received March 3, 2021, accepted April 1, 2021.

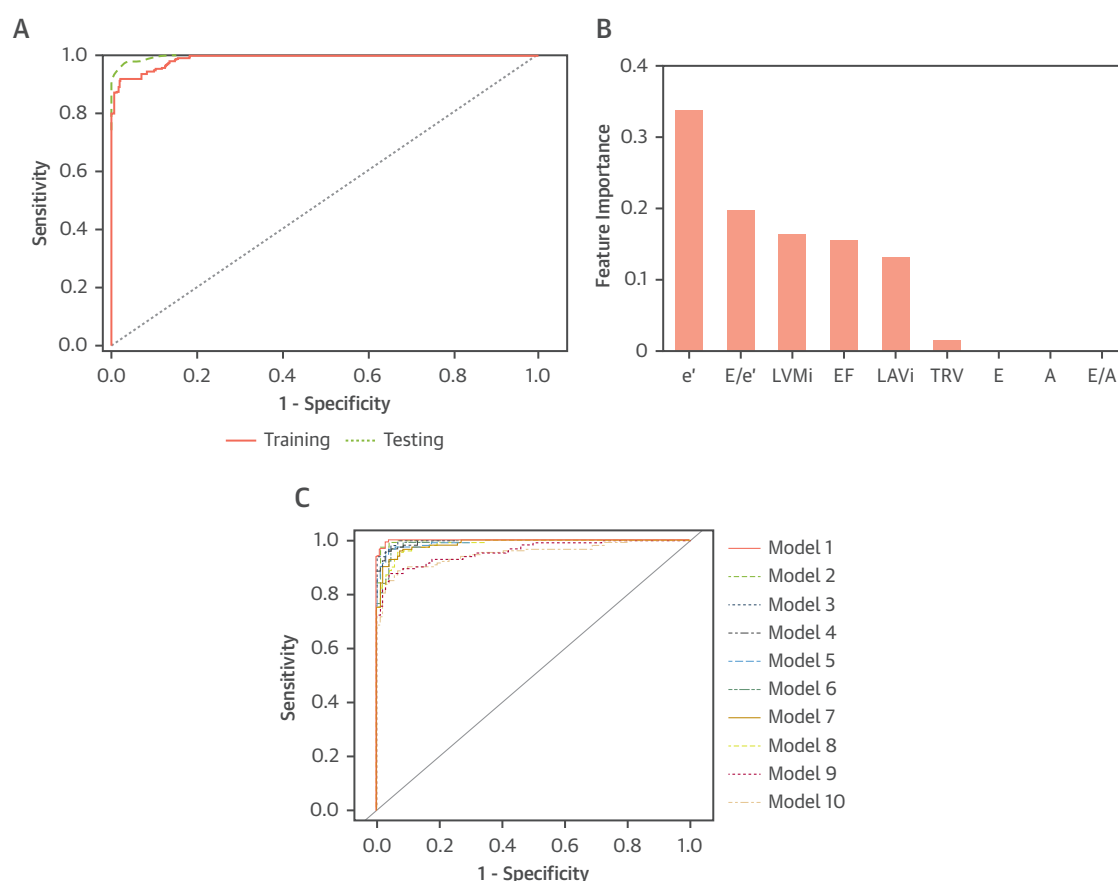


and 2) or high-risk (regions 3 and 4) label depending on their location on a TDA network loop structure (Central Illustration). After labeling the patients as high or low risk using the unsupervised TDA approach described previously (14), a cloud-based automated machine learning platform (OptiML, BigML, Corvallis, Oregon) was used to train multiple supervised machine learning algorithm Supplemental e-Table 1 and select the best performing classifier, a DeepNN for predicting the high- and low-risk phenogroups of LVDD. The technical specifications of the DeepNN classifier and the model development process are described in Supplemental e-Methods-II. The developed DeepNN classifier for LVDD is publicly accessible at <https://wvu-model.herokuapp.com>.

MODEL EVALUATION. Model performance and clinical relevance of the developed DeepNN model was evaluated as follows across different external validation cohorts: 1) model discrimination and interpretability using area under the curve and predicted probability correlations with invasively obtained

hemodynamic pressures to identify patients with elevated filling pressure; and 2) model generalizability and clinical relevance to evaluate the ability of the model to identify patients with low- and high-risk phenotypes for survival status and/or adverse cardiac events. The Institutional Review Board at West Virginia University approved the study protocol, and all study participants in the prospective validation studies provided written informed consent.

Step 1: external validation cohort for predicting invasive measures of LV filling pressure. The DeepNN model's external validity for LVDD was evaluated against an invasive assessment of LV filling pressure in a cohort of patients with echocardiographic and invasive hemodynamic studies (Supplemental e-Methods-III). Briefly, the hemodynamic validation cohort included 84 patients with varying degrees of systolic and diastolic dysfunction who underwent echocardiography and invasive hemodynamic assessment with right heart catheterization (n = 25) or left heart catheterization (n = 59). Sixty-three patients underwent both

FIGURE 1 Model Performance in the Development Cohort

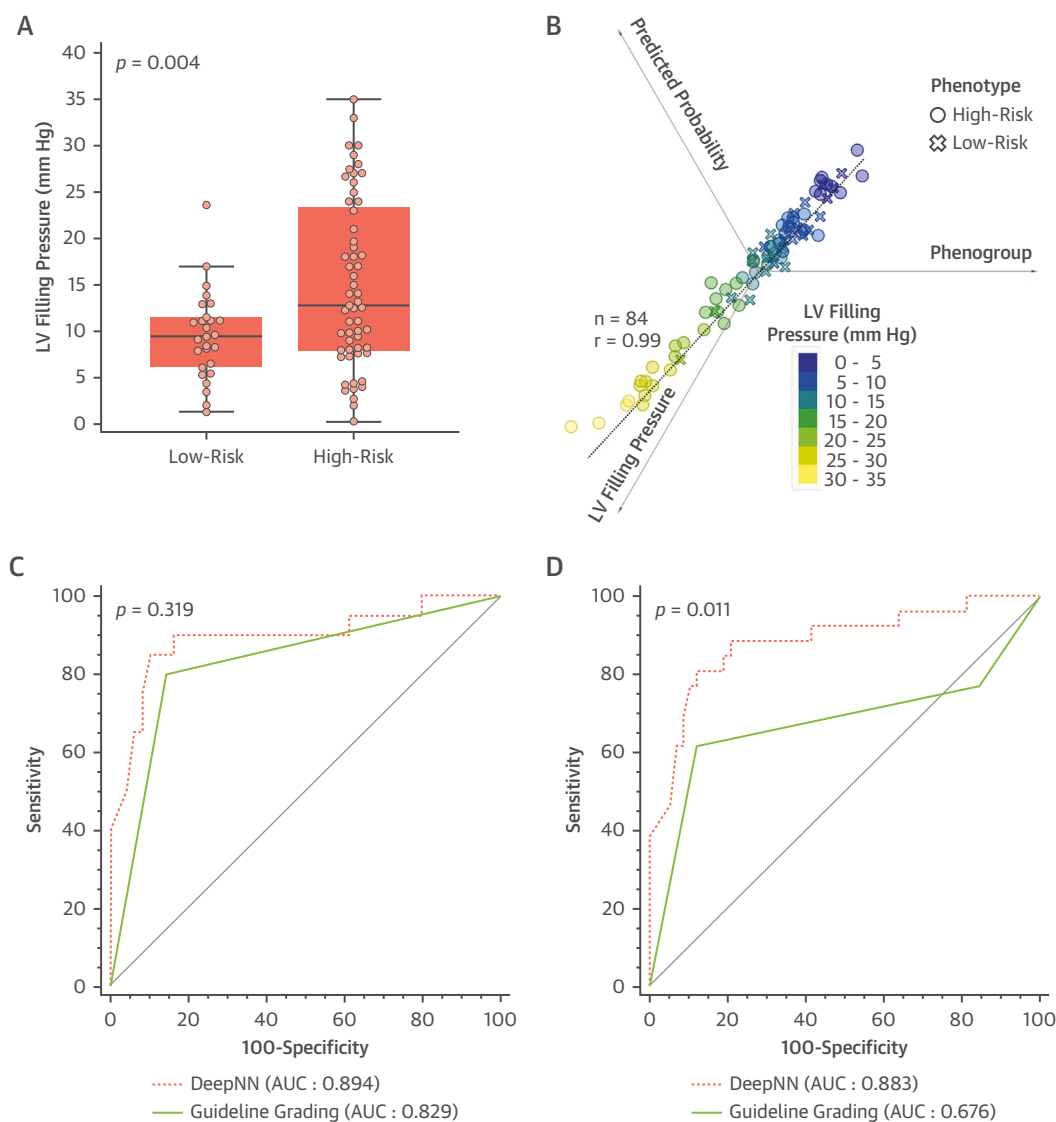
(A) The receiver-operating characteristic curves demonstrated the performance of the deep-learning model of diastolic dysfunction in the derivation and validation cohorts with **(B)** feature importance of the echocardiographic parameters and **(C)** model performance noted with serial removal of different echocardiographic variables. A = late diastolic transmitral flow velocity; E = early diastolic transmitral flow velocity; e' = early diastolic relaxation velocity at septal mitral annular position; EF = ejection fraction; LVMi = left ventricular mass index; TRV = tricuspid regurgitation velocity.

tests on the same day, and the remaining 21 underwent both tests within 72 h. We used a previously described protocol for invasive hemodynamic assessment in these cohorts (15,16). Elevated LV filling pressure was defined by pulmonary capillary wedge pressure, as measured on right heart catheterization, or LV pre-atrial contraction pressure, as assessed on left heart catheterization, >15 mm Hg (9,10,17). The DeepNN model was implemented in the hemodynamic external validation cohort to identify 2 phenogroups (high-risk vs. low-risk) of patients on the basis of the echocardiographic characteristics. The 2016 American Society of Echocardiography (ASE) guideline-based LVDD grade was also assessed in the study participants on the basis of the available echocardiographic parameters (18). To explore the relationship between predicted probabilities and the

distribution of LV filling pressure among the patients, the class prediction probabilities were first calibrated so that the probabilities were transformed to quantile or rank-based z-scores using the inverse of the cumulative distribution function. The inverse of the cumulative distribution function (or quantile function) returns values of x that would make $f(x)$ (e.g., logistic regression) return a predicted probability value of p .

Step 2: external validation cohort for predicting clinical outcomes. The DeepNN model's external validity for predicting clinical outcomes related to LVDD and its incremental value in relation to speckle-tracking echocardiography-derived biomarkers were prospectively evaluated in 219 consecutive subjects with varying degrees of systolic and diastolic dysfunction in whom echocardiography was performed between

FIGURE 2 Model Performance in the Invasive Hemodynamic Validation Cohort



Boxplot showing the distribution of left ventricular (LV) filling pressure by low- and high-risk phenogroups (**A**) and 3-dimensional scatterplot (**B**) showing the relationship between predicted probabilities and the distribution of LV filling pressure among the patients. The linear projections of calibrated class prediction probabilities (quantile or rank-based z-scores of deep neural network [DeepNN] class probabilities) and LV filling pressure (millimeters of mercury) were plotted against low- and high-risk phenogroups. The receiver-operating characteristic curves compare the performance of the 2016 American Society of Echocardiography guideline-based assessment of diastolic dysfunction versus the DeepNN model of diastolic dysfunction in predicting elevated filling pressures as assessed by invasive hemodynamic parameters in the subset of participants with classifiable diastolic function grade by the 2016 guideline (**C**) and in the overall cohort, including those with classifiable and indeterminate diastolic function grade by the 2016 guideline (**D**). AUC = area under the curve.

July 2017 and December 2018. In addition to detailed standard 2-dimensional and Doppler echocardiographic data with detailed diastolic function assessments, we measured global longitudinal strain (GLS) and peak atrial longitudinal strain (PALS) ([Supplemental e-Methods-IV](#)). Patients' electronic medical

records were reviewed for post-echocardiographic follow-up. Hospitalizations were classified on the basis of the International Classification of Diseases-Tenth Revision coding system. The primary end-points included a composite of all-cause death and hospitalization related to HF.

TABLE 1 Comparison of Baseline Characteristics of Study Participants in the External Clinical Outcome Validation Cohort Stratified by the Deep Neural Network Model-Based Phenogroups of Diastolic Dysfunction

	Low-Risk (n = 103)	High-Risk (n = 116)	p Value
Demographics			
Female	60 (58.25)	55 (47.41)	0.11
Age, yrs	49.90 ± 17.20	65.39 ± 13.69	<0.001 ^a
BMI, kg/m ²	29.14 ± 6.77	29.41 ± 6.82	0.76
Hypertension	55 (53.40)	97 (83.62)	<0.0001 ^a
Diabetes mellitus	18 (17.48)	48 (41.38)	<0.0001 ^a
CKD	7 (6.80)	20 (17.24)	0.02 ^a
History of CVA	10 (9.71)	19 (39.58)	0.15
COPD	18 (17.48)	26 (22.41)	0.36
Atrial fibrillation	3 (2.91)	33 (28.45)	<0.0001 ^a
NYHA functional class			<0.0001 ^a
I and II	96 (93.20)	61 (52.59)	
III and IV	7 (6.80)	55 (47.41)	
HF classification ^b			<0.0001 ^a
No HF	42 (40.78)	6 (5.17)	
Stage A	47 (45.63)	29 (25.00)	
Stage B	10 (9.71)	18 (15.52)	
Stage C	4 (3.88)	49 (42.24)	
Stage D	0 (0.00)	14 (12.07)	
LVH			<0.0001 ^a
No LVH	74 (71.84)	46 (39.66)	
Concentric remodeling	28 (27.18)	20 (17.24)	
Concentric hypertrophy	1 (0.97)	20 (17.24)	
Eccentric hypertrophy	0 (0.00)	30 (25.86)	
Clinical outcome			
MACCE	10 (9.71)	22 (18.97)	0.05
HF	2 (1.94)	18 (15.52)	
Death	12 (11.65)	37 (31.90)	0.0003 ^a
All-cause death	3 (2.91)	23 (19.83)	
Cardiac death	1 (0.97)	7 (6.03)	
Noncardiac death	1 (0.97)	5 (4.31)	
Unknown cause	1 (0.97)	11 (9.48)	
Hospitalization			
All-cause hospitalization	27 (26.21)	57 (49.14)	0.0005 ^a
Cardiac hospitalization	14 (13.59)	34 (29.31)	0.005 ^a

Continued on the next page

Step 3: external validation for predicting long-term outcomes in patients with HFpEF: TOPCAT echocardiography substudy. Following the validations performed in steps 1 and 2, which included mixed populations of patients with both reduced and preserved ejection fractions, we focused on understanding model performance specifically for patients with HFpEF. For this we used the data from the randomized, placebo-controlled TOPCAT (Aldosterone Antagonist Therapy for Adults With Heart Failure and Preserved Systolic Function) trial for clinical validation of the DeepNN model (Supplemental e-Methods-V). We used a publicly released and de-identified version of the trial database obtained from the National Heart, Lung, and Blood Institute's Biologic Specimen and

Data Repository Information Coordinating Center (BioLINCC). The outcome of interest for this analysis was the composite of aborted cardiac arrest, hospitalization for management of HF, or cardiovascular death (19). Similar to the hemodynamic validation cohort, the 2016 ASE guideline-based LVDD grade was also assessed among the study participants in the TOPCAT echocardiography cohort on the basis of the available echocardiographic parameters (18).

Step 4: external validation for assessing biomarkers and exercise performance in HFpEF: RELAX-HF and NEAT-HFpEF cohort. Data from the randomized, placebo-controlled RELAX-HF (Evaluating the Effectiveness of Sildenafil at Improving Health Outcomes and Exercise Ability in People With Diastolic Heart Failure) and NEAT-HFpEF (Nitrate's Effect on Activity Tolerance in Heart Failure With Preserved Ejection Fraction) trials were obtained from the National Heart, Lung, and Blood Institute's BioLINCC data repository and used for the external validation of the DeepNN model of LVDD against cardiac biomarkers and measures of exercise performance in patients with HFpEF. The details of the trial protocols, inclusion and exclusion criteria, and primary results of the RELAX-HF and NEAT-HFpEF trials have been reported previously and are summarized in Supplemental e-Methods-VI (20,21). The biomarkers of interest were baseline measures of troponin I and N-terminal pro-brain natriuretic peptide (NT-proBNP) levels. The exercise and quality-of-life parameters of interest were peak exercise oxygen uptake (VO_{2peak}) and Minnesota Living With Heart Failure Questionnaire (MLHFQ) score, respectively. Protocol details pertaining to exercise testing, quality-of-life assessment, and biomarker assessments have been reported previously (20).

STATISTICAL ANALYSIS. During model development, the study participants' baseline characteristics across the 2 phenogroups in the training cohort and the internal validation cohorts were compared using the chi-square test for categorical variables and 1-way analysis of variance for continuous variables. Furthermore, during internal validation, the DeepNN model's accuracy to predict phenogroups was assessed using the area under the receiver-operating characteristic curve (AUROC). To assess the contribution of different echocardiographic parameters toward the DeepNN model's performance, we sequentially removed diastolic parameters and evaluated the performance of the resulting models. Following the development of the DeepNN model of LVDD, we tested its performance in multiple external validation cohorts (invasive hemodynamic validation cohort, outcome validation cohort, and HFpEF trial

cohorts including the TOPCAT echocardiography, RELAX-HF, and NEAT-HFpEF cohorts) to identify the high- and low-risk phenogroups. In the invasive hemodynamic validation cohort, the classification performance of the DeepNN model and the 2016 ASE guideline-based algorithm to predict elevated LV filling pressure (>15 mm Hg) was assessed using the AUROC. In the clinical outcome validation cohort and the TOPCAT echocardiography cohort, the primary composite endpoint's risk across the phenogroups was assessed using time-to-event analysis with the log-rank test and adjusted Cox proportional hazards models. In the clinical outcome validation cohort, a multivariate Cox model was constructed for the prediction of primary clinical endpoints using phenogroup label, GLS, and PALS as covariates. The Cox models for TOPCAT echocardiography cohort were adjusted for age, sex, race, history of diabetes, country of enrollment, and enrollment stratum, consistent with the primary TOPCAT analysis. An additional Cox model was also constructed with adjustment for the MAGGIC (Meta-Analysis Global Group in Chronic Heart Failure) risk score to determine the association of the phenogroup with risk for the primary clinical endpoint independent of the well-established clinical risk score (22). We also evaluated the treatment effect of spironolactone (vs. placebo) for the primary composite endpoint within each phenogroup. Furthermore, the reclassification of participants with different LVDD grade determined by the 2016 ASE guideline-based algorithm using the DeepNN model was also assessed. In the RELAX-HF and NEAT-HFpEF cohorts, the association of the DeepNN model-based phenogroups with baseline measures of cardiac biomarkers (cardiac troponin I and NT-ProBNP) and exercise performance (VO₂peak). MLHFQ score was assessed using the following adjusted linear regression models constructed separately for each outcome: model 1, age, sex, and race; model 2, model 1 plus body mass index (except for VO₂peak model), diabetes, blood pressure, kidney function, hemoglobin, smoking status, atrial fibrillation, and New York Heart Association functional class. The data from the RELAX-HF and NEAT-HFpEF cohorts were pooled for the outcomes that were uniformly reported across the 2 cohorts (NT-proBNP and MLHFQ score). The following software programs were used for analyses: Ayasdi platform version 7.9 (Ayasdi, Palo Alto, California) for the TDA and phenogroup label generation, Stata version 14.2 (StataCorp, College Station, Texas), and R version 3.6.3 (R Foundation for Statistical Computing) for statistical analyses. Statistical significance was tested with a 2-sided p value <0.05.

TABLE 1 Continued

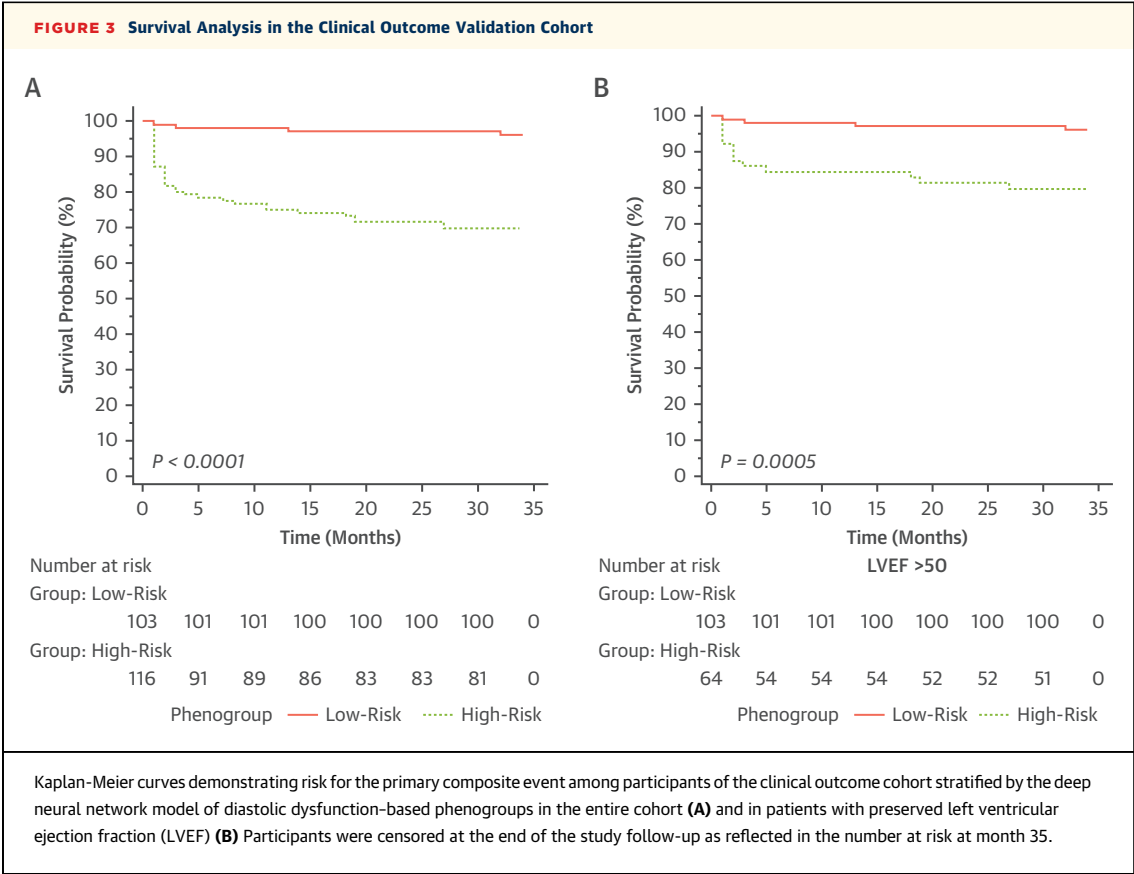
	Low-Risk (n = 103)	High-Risk (n = 116)	p Value
Echocardiography			
LVEDD, mm	44.20 ± 5.12	50.43 ± 9.05	<0.001 ^a
LVESV, ml	37.15 ± 11.01	70.59 ± 43.10	<0.001 ^a
LVEDV, ml	100.96 ± 24.58	132.67 ± 52.90	<0.001 ^a
LVSF, %	63.68 ± 16.83	63.03 ± 22.48	0.81
LVEF, %	63.64 ± 6.13	49.65 ± 14.34	<0.001 ^a
LV mass index, g/m ²	65.99 ± 16.83	106.10 ± 40.57	<0.001 ^a
E, cm/s	82.99 ± 16.38	85.23 ± 29.91	0.5
A, cm/s	69.77 ± 22.48	73.46 ± 26.77	0.28
E/A ratio	1.31 ± 0.56	1.27 ± 0.69	0.68
e', cm/s	9.10 ± 2.22	5.39 ± 1.89	<0.001 ^a
E/e' ratio	9.30 ± 2.41	17.62 ± 9.20	<0.001 ^a
LA volume index, ml/m ²	21.69 ± 7.92	37.08 ± 16.11	<0.001 ^a
Moderate or greater valve disease	2 (1.94)	22 (18.97)	0.00004 ^a
Speckle-tracking echocardiography			
GLS, %	-19.20 ± 3.41	-14.05 ± 5.82	<0.001 ^a
LV SRs, s ⁻¹	-0.90 ± 0.19	-0.63 ± 0.28	<0.001 ^a
LV SRe, s ⁻¹	0.77 ± 0.22	0.49 ± 0.23	<0.001 ^a
PALS, %	38.65 ± 10.99	23.43 ± 11.98	<0.001 ^a
PACS, %	17.56 ± 6.34	12.06 ± 7.04	<0.001 ^a
LA SRs, s ⁻¹	1.21 ± 0.34	0.75 ± 0.38	<0.001 ^a
LA SRe, s ⁻¹	-0.91 ± 0.38	-0.58 ± 0.36	<0.001 ^a

Values are n (%) or mean ± SD. ^ap < 0.05 comparing between the high-risk and low-risk groups using 1-way analysis of variance where mean is reported and using chi-square or Fisher exact test where frequency is reported. ^bClassification system of the American College of Cardiology and the American Heart Association.

A = late diastolic transmitral flow velocity; BMI = body mass index; CKD = chronic kidney disease; COPD = chronic obstructive lung disease; CVA = cerebrovascular accident; E = early diastolic transmitral flow velocity; e' = early diastolic relaxation velocity at septal mitral annular position; GLS = global longitudinal strain; HF = heart failure; LA = left atrial; LV = left ventricular; LVEDD = left ventricular end-diastolic dimension; LVEDV = left ventricular end-diastolic volume; LVEF = left ventricular ejection fraction; LVH = left ventricular hypertrophy; LVESD = left ventricular end-systolic dimension; LVESV = left ventricular end-systolic volume; MACCE = major adverse cardiac and cerebrovascular event(s); NYHA = New York Heart Association; PACS = peak atrial contraction strain; PALS = peak atrial longitudinal strain; SRe = peak early diastolic strain rate; SRs = peak systolic strain rate.

RESULTS

MODEL TRAINING AND INTERNAL VALIDATION OF THE DeepNN MODEL. In the training cohort (n = 990), the DeepNN model showed excellent accuracy by cross-validation in the training set (AUROC = 0.988, accuracy = 92.1%) (Figure 1). In the internal validation set (n = 252), the model continued to perform with high diagnostic accuracy (AUROC = 0.997, accuracy = 96.0%) for predicting the high- and low-risk patient phenogroups. The model continued to perform well even after performing feature elimination (other than e') to systematically exclude individual LV diastolic function parameters (AUROC = 0.945 to 0.997, accuracy = 92.1% to 96.4%). The model performance showed lower accuracy (72.3% to 74.7%) in the absence of e', although the AUROCs still suggested diagnostic value (0.945 to 0.955) (Figure 1, Supplemental e-Table 2). Tissue Doppler velocity (e') was the most important



echocardiographic parameter for predicting the DeepNN model-based phenogroups followed by E/e', LV mass, ejection fraction, and left atrial size (Figure 1B).

The study participants' baseline characteristics in the training and internal validation datasets stratified by the DeepNN model-based phenogroups (high risk: 52% training and 53.5% internal validation cohort participants) are shown in Supplemental e-Table 3. Participants in the high-risk phenogroup were older, had a higher burden of comorbidities, and had greater abnormalities in cardiac structure and function with higher E/e' ratio, LV mass, ejection fraction, and left atrial volume index.

INVASIVE HEMODYNAMIC VALIDATION. Baseline characteristics of study participants in the invasive hemodynamic validation cohort stratified by the DeepNN model-based phenogroups (60% high risk) are shown in Supplemental e-Table 4. The pattern of differences in patients' baseline clinical characteristics in the high-versus low-risk phenogroups in the invasive hemodynamic validation cohort was similar to that noted in the model development and internal validation cohorts discussed earlier. Compared with the low-risk phenogroup, participants in the high-risk phenogroup

showed higher LV filling pressure ($p = 0.004$) (Figure 2A). More important, within the high-risk phenogroup, the calibrated values of predicted probability (normalized quantile scores) showed a linear association ($r = 0.76$; $p < 0.0001$) with LV filling pressure, suggesting the ability of the trained DeepNN model in predicting elevated LV filling pressure. A correlation coefficient of 0.62 ($p < 0.001$) was observed across the entire invasive hemodynamic validation cohort. No significant correlation was observed between LV filling pressure and probabilities within the low-risk phenogroup. Furthermore, visualization of the linear projections of calibrated predicted probability and phenogroups against LV filling pressure showed a clear separation of patients with elevated LV filling pressure and a strong correlation ($r = 0.99$; $p < 0.001$) in the vector space by capturing the differences in probabilities within and across the low- and high-risk phenogroups (Figure 2B).

Among patients for whom the 2016 guideline was able to grade LVDD ($n = 69$) after excluding patients labeled as indeterminate, both the DeepNN model and the guideline had comparable discriminative performance in predicting elevated LV filling pressure, with comparable AUROCs (0.89 vs. 0.83,

respectively; $p = 0.32$) (Figure 2C). However, for the entire group, which also included patients labeled as indeterminate, the DeepNN model showed a significantly higher AUROC compared with the 2016 guideline (0.88 vs. 0.67; $p = 0.01$) (Figure 2D and Supplemental e-Table 5).

CLINICAL OUTCOME VALIDATION. The DeepNN model identified 53% ($n = 116$) participants as high risk and 47% ($n = 103$) as low risk. Compared with the low-risk phenogroup, participants in the high-risk phenogroup were older, had a higher burden of comorbidities, and had a higher incidence of death and HF hospitalization (Table 1). Among echocardiographic characteristics, participants in the high-risk phenogroup had a greater cardiac structure and functional impairment burden, with lower ejection fraction, higher LV end-diastolic volume and left atrial size, higher E/e' ratio, lower e' velocity, and impaired GLS and PALS (Table 1). The Kaplan-Meier survival curve demonstrated a significantly lower rate of the primary endpoint (all-cause mortality and/or HF hospitalization) in the low-risk phenogroup (log-rank $p < 0.0001$) both for the overall population and for patients with LV ejection fractions $>50\%$ (Figure 3). Even after incorporating GLS and PALS in a multivariate model, the high-risk phenogroup remained as a significant independent predictor of event-free survival (hazard ratio [HR]: 3.96; 95% confidence interval [CI]: 1.24 to 12.67; $p = 0.021$) (Supplemental e-Table 6).

CLINICAL OUTCOME VALIDATION (TOPCAT SUBSTUDY). Among participants in the TOPCAT echocardiography cohort, the DeepNN model was mapped on 518 participants after excluding those who did not have adequate echocardiographic information ($n = 136$). The clinical characteristics of the excluded cohort were not different from those of our final cohort (Supplemental e-Table 7). The DeepNN model identified 81.1% ($n = 420$) participants as high risk and 18.9% ($n = 98$) as low risk. Compared with the low-risk phenogroup, participants in the high-risk phenogroup were older, were more commonly African American, had higher blood pressure, and had lower hemoglobin levels (Table 2). Among echocardiographic characteristics, participants in the high-risk phenogroup had a greater cardiac structure and functional impairment burden, with lower ejection fraction, higher LV end-diastolic volume and left atrial size, higher E/e' ratio, and lower e' velocity (Table 2).

The primary composite outcome's cumulative incidence was higher in the high-risk than the low-risk phenogroup (34.0% vs. 17.3%; $p < 0.002$)

(Figure 4). In adjusted Cox models, high-risk phenogroup participants had a significantly increased risk for the primary composite endpoint (HR: 1.92; 95% CI: 1.16 to 3.22; $p = 0.01$). Consistent associations were also observed in the Cox model adjusted for the MAGGIC risk score (HR: 2.10; 95% CI: 1.20 to 3.47; $p = 0.0003$). In phenogroup-stratified analysis, spironolactone was significantly associated with a lower risk for the primary composite outcome in the high-risk phenogroup (HR for spironolactone vs. placebo: 0.65; 95% CI: 0.46 to 0.90; $p = 0.01$) but not in the low-risk phenogroup (HR for spironolactone vs. placebo: 1.13; 95% CI: 0.44 to 2.93; $p = 0.80$). There was no significant interaction between treatment arm and diastolic function phenogroup for the risk for the primary composite outcome on statistical testing.

RECLASSIFICATION OF DIASTOLIC FUNCTION GRADING IN TOPCAT ECHOCARDIOGRAPHY COHORT

The 2016 ASE guideline identified 33.8% participants as having grade I LVDD, 23.7% as having grade II or III LVDD, and 42.5% as having indeterminate-grade LVDD. The cumulative risk for the composite clinical outcome was higher in patients with grade II or III versus grade I LVDD and those with indeterminate grades of diastolic function (event rate 39% vs. 28% vs. 29%).

Among participants identified with high grade (II or III) LVDD by the 2016 guideline, almost all (94%) were concordantly classified in the high-risk phenogroup by the DeepNN model. In contrast, a substantial reclassification was noted for patients with low grade (grade I) LVDD or indeterminate diastolic function grade, with 73% and 80% of participants being discordantly reclassified into the high-risk phenogroup, respectively. Reclassification of the participants with grade I or indeterminate diastolic function grade by the DeepNN model was associated with restratification in the risk for the primary composite event, with a significantly higher risk noted among those reclassified into the high- versus low-risk phenogroup (Figure 5).

CARDIAC BIOMARKERS AND EXERCISE PERFORMANCE PATIENTS WITH HFpEF

In the pooled cohort of participants from the RELAX-HF and NEAT-HFpEF trials, the DeepNN model identified 74% of participants in the high-risk phenogroup. The baseline characteristics of the pooled cohort participants stratified by the DeepNN model-based phenogroup are shown in Supplemental e-Table 8. Consistent with observations in the TOPCAT trial cohort, participants in the high-risk phenogroup in the RELAX-HF/NEAT-HFpEF cohort were older, had higher blood pressure,

TABLE 2 Comparison of Baseline Characteristics of the DeepNN Model Based Phenogroups Diastolic in the TOPCAT Trial Echo-Sub Study Cohort

	Low-Risk (n = 98)	High-Risk (n = 420)	P-value
Clinical and demographic characteristics			
Age, y	69.1 ± 9.6	71.5 ± 10.0	0.04
Female	55 (56.1)	202 (48.1)	0.19
Race			0.02
White	82 (83.7)	296 (70.5)	
Black	14 (14.3)	95 (22.6)	
Other	2 (2.0)	29 (6.9)	
Hospitalization eligibility stratum	51 (52.0)	246 (58.6)	0.29
Country			0.36
US	72 (73.5)	315 (75.0)	
Canada	17 (17.3)	62 (14.8)	
Brazil	9 (9.2)	32 (7.6)	
Argentina	0 (0.0)	11 (2.6)	
BNP, pg/mL	320.9 ± 293.9	443.6 ± 499.6	0.13
NT-proBNP, pg/mL	854.8 ± 371.3	1812.8 ± 2240.5	0.10
Prior myocardial infarction	16 (16.3)	96 (22.9)	0.20
COPD	17 (17.3)	75 (17.9)	0.99
Hypertension	90 (91.8)	383 (91.4)	0.99
Diabetes category			0.10
No diabetes	59 (60.2)	203 (48.4)	
Insulin dependent diabetes	21 (21.4)	109 (26.0)	
Non-insulin dependent diabetes	18 (18.4)	107 (25.5)	
Current smoker	7 (7.1)	30 (7.2)	0.99
Ever smoker	50 (54.9)	211 (54.2)	0.99
NYHA functional class (≥III)	39 (39.8)	147 (35.3)	0.47
Heart rate, beats/min	70.9 ± 12.1	69.1 ± 11.6	0.17
Systolic blood pressure, mm Hg	121.8 ± 16.3	128.2 ± 15.8	<0.01
Body mass index, kg/m ²	34.9 ± 8.5	33.5 ± 7.9	0.12
Sodium, mg/dL	139.5 ± 3.4	139.5 ± 3.0	0.98
Potassium, mg/dL	4.2 ± 0.4	4.1 ± 0.4	0.35
Blood urea nitrogen, mg/dL	23.7 ± 12.7	25.2 ± 12.3	0.30
Glucose, mg/dL	122.1 ± 62.1	125.3 ± 62.9	0.65
Hemoglobin, g/dL	13.0 ± 1.6	12.6 ± 1.7	0.04
Aspartate Transaminase, U/L	26.0 ± 11.4	25.1 ± 12.3	0.51
Total bilirubin, mg/dL	0.6 ± 0.4	0.7 ± 0.5	0.34
Albumin, g/dL	3.9 ± 0.4	3.8 ± 0.4	0.7
QRS duration, ms	94.5 ± 22.7	108.2 ± 32.3	<0.01
Angiotensin-converting enzyme inhibitor	50 (51.0)	233 (55.5)	0.49
Angiotensin receptor blocker	21 (21.4)	128 (30.5)	0.10
Beta blocker	80 (81.6)	338 (80.5)	0.91
Diuretic	87 (88.8)	378 (90.0)	0.86
Aspirin	50 (51.0)	257 (61.2)	0.08
Statin	67 (68.4)	292 (69.5)	0.92
Warfarin	40 (40.8)	124 (29.5)	0.04

Continued on the next page

stress (Table 3). High-risk phenogroup membership was also significantly associated with lower exercise capacity (VO₂peak) and worse quality of life (MLHFQ score) independent of other potential confounders (Table 3).

DISCUSSION

Over the years, several echocardiographic algorithms have been proposed for the assessment of LVDD using a list of sequential heuristics that combine dichotomously defined abnormalities in diastolic echocardiographic parameters (18,23). However, recent studies demonstrated poor concordance among these algorithms, resulting in a prevalence of LVDD in the general population ranging from 2% to 35% (24). Although the 2016 ASE guideline on LVDD proposed a conceptually simplified approach (18), the algorithm has resulted in higher proportions of patients classified as normal or indeterminate (9,13,25,26). Moreover, lower sensitivity and negative predictive value have been reported for diagnosing HFpEF compared with gold-standard exercise invasive hemodynamic studies (26). The DeepNN model proposed in this study addresses some of these limitations. First, the pipeline integrates multiparametric echocardiographic assessment and classifies diastolic function for nearly all cases, including those considered indeterminate by the existing guidelines. Notably, the DeepNN model had higher discrimination than the 2016 ASE guidelines for assessing elevated LV filling pressures and successfully reclassified 70% to 80% of patients with HFpEF with low-grade (stage 1) or indeterminate LVDD into the high-risk phenogroup. Second, among patients with established HFpEF, the DeepNN model can risk-stratify patients with HFpEF such that those in the high-risk phenogroup had significantly elevated biomarkers to suggest chronic myocardial injury and stress, worse exercise capacity, and poor quality of life, independent of other confounders. Finally, the high-risk phenogroup DeepNN model of LVDD identified patients with clinical HFpEF who have worse long-term outcomes and may benefit from treatment with spironolactone (Central Illustration).

Several prior studies have used machine learning-based clustering approaches to identify distinct phenogroups of patients with HFpEF. Shah et al. (2) classified patients with HFpEF into 3 distinct subgroups, which differed in clinical, echocardiographic, and hemodynamic characteristics and risk for adverse outcomes on longitudinal follow-up. Similarly, Kao

and had a higher burden of comorbidities, including atrial fibrillation and renal dysfunction.

In adjusted analysis, participants in the high-risk phenogroup had significantly higher levels of high-sensitivity troponin and NT-proBNP, suggesting a greater burden of chronic myocardial injury and

et al. (27) identified 6 unique phenogroups of patients with HFpEF in a secondary analysis of the I-PRESERVE (Irbesartan in Heart Failure With Preserved Systolic Function) and CHARM Preserved (Candesartan Cilexetil in Heart Failure Assessment of Reduction in Mortality and Morbidity) trials using latent class analysis. Compared with the previous studies, the DeepNN approach presented in our study focused primarily on identifying patients with HFpEF who had a higher burden of cardiac abnormalities. Thus, the high-risk phenogroup had a higher prevalence of elevated E/e' ratio, lower e', and higher LV mass and higher levels of biomarkers of myocardial injury and stress, whereas the low-risk phenogroup had a higher burden of obesity and lesser degree of cardiac abnormalities. These findings suggest that the DeepNN model identified a biologically meaningful phenotype of cardiac impairment in HFpEF. Furthermore, the phenogroups identified by the DeepNN model also demonstrated significant differences in quality of life and exercise capacity and risk for adverse clinical outcomes.

Several cardioprotective pharmacotherapies have failed to improve clinical outcomes in large phase III clinical trials of patients with HFpEF (3). The pathophysiological and phenotypic heterogeneity in HFpEF has contributed substantially to the lack of effective therapies for this disease, highlighting the need for novel approaches to identify prognostically and therapeutically homogenous phenogroups (1-3,28). The DeepNN model evaluated in this study is an attempt to address this knowledge gap. The high-risk "cardiac impairment" phenotype identified by the DeepNN model was also more adaptable and benefited significantly from spironolactone. In contrast, the low-risk phenogroup did not have significant LVDD, and their risk for adverse events was not modified by spironolactone. Thus, the significant, independent association of DeepNN classifier-based phenogroups with different biologic measures of disease severity and risk for adverse clinical outcomes highlights the model's robustness, clinical validity, and external generalizability. With our DeepNN classifier being made publicly available, we further anticipate that more work can be done to confirm and discover the mechanistic underpinnings underlying the observed phenotypic differences. Moreover, the classifier may be useful to test the responsiveness of HFpEF phenogroups to new HF therapies in future clinical trials.

STUDY LIMITATIONS. First, the DeepNN classifier was developed primarily to use a set of echocardiographic

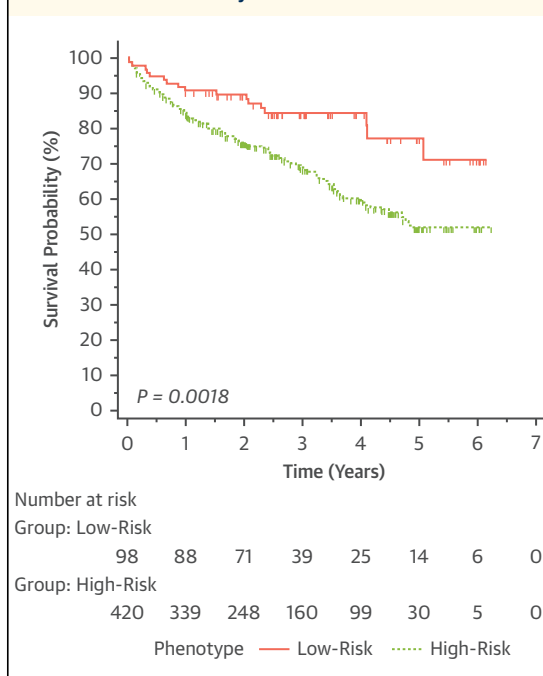
TABLE 2 Continued

	Low-Risk (n = 98)	High-Risk (n = 420)	P-value
Echocardiographic characteristics			
Ejection fraction, %	64.6 ± 5.5	58.7 ± 7.9	<0.01
LV diastolic diameter, cm	4.5 ± 0.5	4.9 ± 0.6	<0.01
E-wave velocity, m/s	0.9 ± 0.2	0.9 ± 0.3	0.19
A-wave velocity, m/s	0.7 ± 0.2	0.7 ± 0.3	0.99
E/a ratio	1.3 ± 0.6	1.4 ± 0.7	0.24
Peak TR velocity, cm/s	2.7 ± 0.4	2.8 ± 0.5	0.08
LA volume, cm ³	50.6 ± 18.2	63.7 ± 28.0	<0.01
LA volume indexed, cm ³ /m ²	23.8 ± 8.0	31.3 ± 13.5	<0.01
LV mass indexed, g/m ²	84.0 ± 20.3	112.6 ± 30.2	<0.01
E/e' septal	10.5 ± 3.3	18.2 ± 7.0	<0.01
E/e' lateral	9.1 ± 3.5	13.3 ± 6.2	<0.01
E' septal, cm/s	8.8 ± 2.5	5.6 ± 1.8	<0.01
E' lateral, cm/s	10.5 ± 3.5	7.9 ± 3.1	<0.01

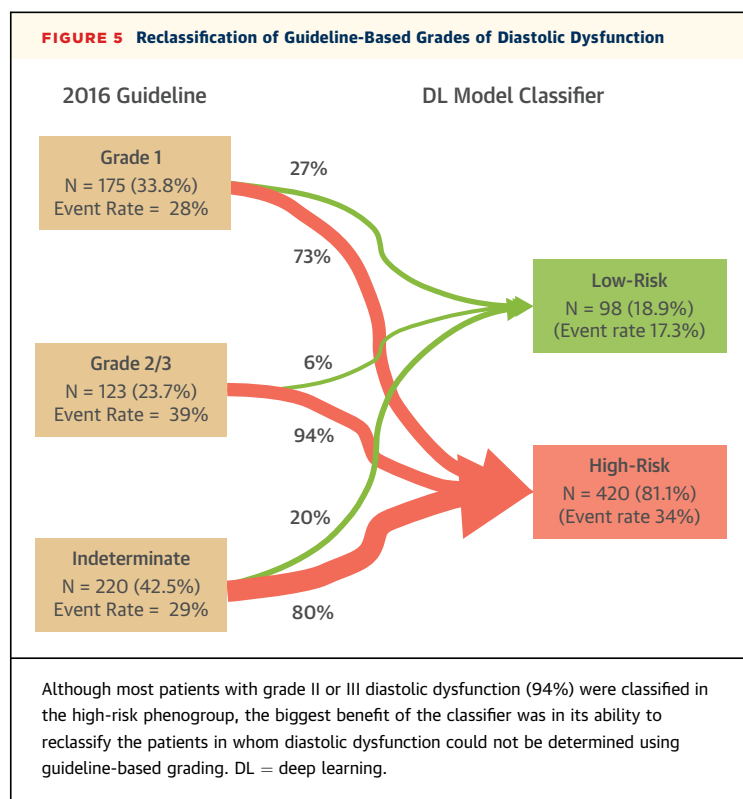
Values are n (%) or mean ± SD.

BNP = brain natriuretic peptide; NT-proBNP = N-terminal pro-brain natriuretic peptide; TOPCAT = Aldosterone Antagonist Therapy for Adults With Heart Failure and Preserved Systolic Function; TR = tricuspid regurgitation; other abbreviations as in Table 1.

FIGURE 4 Survival Analysis in the TOPCAT Trial



Kaplan-Meier curves demonstrating risk for the primary composite event among participants of the TOPCAT (Aldosterone Antagonist Therapy for Adults With Heart Failure and Preserved Systolic Function) echocardiography cohort stratified by the deep neural network model of diastolic dysfunction-based phenogroups. Participants were censored at the end of the study follow-up as reflected in the number at risk at year 7.



independent predictor of adverse outcomes even after addition of more novel echocardiographic parameters such as LV and atrial strain in phenotyping LVDD. This is consistent with a subgroup analysis from TOPCAT study, in which the association between left atrial strain and HF hospitalizations were not significant after adjustments for other echocardiographic systolic and diastolic parameters (29). Nevertheless, more in-depth considerations and potential integration of strain parameters with DeepNN classifiers need to be investigated in the future studies.

Second, the DeepNN classifier leverages only resting echocardiographic parameters to predict resting diastolic function phenotypes. The addition of new biomarkers that reflect inflammation, myocyte remodeling, extracellular matrix remodeling, and endothelial dysfunction and other variables such as ventriculovascular coupling could enrich the risk prediction and would need to be addressed in future studies. Moreover, a significant proportion of patients with HFpEF have LVDD unmasked with exercise only. Future studies with this model are needed to determine if the DeepNN classifier can predict exercise-related diastolic impairment.

CONCLUSIONS

Our DeepNN classifier thus offers a viable solution to overcome the limitations of the existing clinical standards for accurately characterizing the burden of LVDD in HFpEF. The use of this classifier to identify a high-risk subgroup of patients with HFpEF with underlying cardiac impairment will motivate investigators to refine strategies and therapies to

variables that are currently recommended for assessing LVDD and have well-established prognostic role for cardiovascular diseases. This was done to allow a more pragmatic and generalizable application of the DeepNN classifier to routinely collected echocardiographic data. Moreover, in the outcome validation cohort, the high-risk phenogroup remained an

TABLE 3 Adjusted Association of Deep Neural Network Model of Diastolic Dysfunction Phenogroups With Measures of Cardiac Injury, Stress, and Exercise Parameters in Patients With Heart Failure With Preserved Ejection Fraction

Outcome of Interest	Model 1		Model 2	
	Standardized Beta (95% CI)	p Value	Standardized Beta (95% CI)	p Value
Troponin I	0.37 (0.04 to 0.69)	0.001	0.42 (0.09 to 0.75)	0.01
NT-proBNP ^a	0.42 (0.18 to 0.66)	0.001	0.30 (0.07 to 0.54)	0.01
Peak VO ₂	−0.35 (−0.64 to −0.07)	0.02	−0.29 (−0.55 to −0.03)	0.03
Peak exercise workload	−0.35 (−0.63 to −0.07)	0.01	−0.25 (−0.52 to 0.01)	0.06
MLHFQ score	−0.31 (−0.56 to −0.06)	0.02	−0.32 (−0.56 to −0.08)	0.01

Separate multivariate-adjusted models were constructed for each cardiac biomarker and exercise parameter (dependent outcome variable of interest) with the phenogroup as the primary exposure variable (high risk vs. 1 [reference]) and adjustment for the following covariates: model 1, age, sex, and race; model 2: model 1 plus diabetes, systolic blood pressure, estimated glomerular filtration rate, hemoglobin, smoking status, atrial fibrillation, and New York Heart Association functional class. Additionally, model 2 was adjusted for body mass index for outcomes of troponin I, NT-proBNP, and MLHFQ score but not for peak VO₂, as peak VO₂ is already indexed for body weight. Models for NT-proBNP and MLHFQ score outcomes used pooled participants from the RELAX-HF (Evaluating the Effectiveness of Sildenafil at Improving Health Outcomes and Exercise Ability in People With Diastolic Heart Failure) and NEAT-HFpEF (Nitrate's Effect on Activity Tolerance in Heart Failure With Preserved Ejection Fraction) cohorts, as these parameters were consistently reported in both trials. Models for troponin I and exercise test parameters used only RELAX-HF cohort participants, as these parameters were not reported in the NEAT-HFpEF cohort. Beta (95% CI) represents the parameter estimate per 1 SD derived from the multivariate-adjusted linear regression models. The beta coefficient represents the change in outcome of interest per SD noted when the exposure variable was varied from low-risk to high-risk phenogroup keeping other covariates fixed. ^aLog transformed.

CI = confidence interval; MLHFQ = Minnesota Living With Heart Failure Questionnaire; NT-proBNP = N-terminal pro-brain natriuretic peptide; VO₂ = oxygen uptake.

manage HFpEF and discover the mechanistic underpinnings underlying the phenotypic differences and clinical outcomes.

ACKNOWLEDGMENTS The authors thank the participants, staff members, and investigators of the trials and study cohorts included in this study. The authors thank Heenaben B. Patel for her help in data collection.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

This study was supported partly by funds from the National Science Foundation (1920920) and an interinstitutional Smart Health Initiative. This study was judged as the winner of the 2020 Aurthur Weyman Young Investigator Award presented at the annual Scientific Sessions of the ASE. This project was also the winner of the National Heart, Lung, and Blood Institute Big-Data Challenge: Creating New Paradigms for Heart Failure Research Award. Dr. Kagiya was supported by a research grant from Hitachi Healthcare. Dr. Sengupta has served as a consultant to Ultrasonics and Kencor Health. Dr. Pandey has served on the advisory board of Roche Diagnostics; and is supported by research grants from the Texas Health Resources Clinical Scholarship, the Gilead Sciences Research Scholar Program, the National Institute on Aging GEMSSTAR Grant (1R03AG067960-01), and Applied Therapeutics. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr. Partho P. Sengupta, West Virginia University Heart and Vascular Institute, 1 Medical Center Drive, Morgantown, West Virginia 26506, USA. E-mail: partho.sengupta@wvumedicine.org. Twitter: [@appengupta1](https://twitter.com/appengupta1).

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: A DeepNN model of diastolic function can overcome the limitations of the existing guideline-based approach in identifying patients with elevated filling pressures and risk-stratify patients with greater impairment in cardiac structure, function, and exercise capacity who may benefit from therapies such as spironolactone

TRANSLATIONAL OUTLOOK: Future studies are needed to determine the mechanistic pathways underlying the usefulness of the DeepNN model for identifying patients with HFpEF who respond to specific cardioprotective therapies.

REFERENCES

1. Segar MW, Patel KV, Ayers C, et al. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail*. 2020;22:148-158.
2. Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. 2015;131:269-279.
3. Shah SJ, Kitzman DW, Borlaug BA, et al. Phenotype-specific treatment of heart failure with preserved ejection fraction: a multiorgan roadmap. *Circulation*. 2016;134:73-90.
4. 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;350:1953-1959.
5. Chetrit M, Cremer PC, Klein AL. Imaging of diastolic dysfunction in community-based epidemiological studies and randomized controlled trials of HFpEF. *J Am Coll Cardiol Img*. 2020;13:310-326.
6. Shah AM, Cikes M, Prasad N, et al. Echocardiographic features of patients with heart failure and preserved left ventricular ejection fraction. *J Am Coll Cardiol*. 2019;74:2858-2873.
7. Ho JE, Zern EK, Wooster L, et al. Differential clinical profiles, exercise responses, and outcomes associated with existing HFpEF definitions. *Circulation*. 2019;140:353-365.
8. Obokata M, Reddy YNV, Borlaug BA. Diastolic dysfunction and heart failure with preserved ejection fraction: understanding mechanisms by using noninvasive methods. *J Am Coll Cardiol Img*. 2020;13:245-257.
9. Andersen OS, Smiseth OA, Dokainish H, et al. Estimating left ventricular filling pressure by echocardiography. *J Am Coll Cardiol*. 2017;69:1937-1948.
10. Balaney B, Medvedofsky D, Mediratta A, 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.
11. Lancellotti P, Galderisi M, Edvardsen T, et al. Echo-Doppler estimation of left ventricular filling pressure: results of the multicentre EACVI Euro-Filling study. *Eur Heart J Cardiovasc Imaging*. 2017;18:961-968.
12. Flachskampf FA, Baron T. Echocardiographic algorithms for detecting elevated diastolic pressures: reasonable, not perfect. *J Am Coll Cardiol*. 2017;69:1949-1951.
13. Oh JK, Miranda WR, Bird JG, Kane GC, Nagueh SF. The 2016 diastolic function guideline: is it already time to revisit or revise them? *J Am Coll Cardiol Img*. 2020;13:327-335.
14. Tokodi M, Shrestha S, Bianco C, et al. Interpatient similarities in cardiac function: a platform for personalized cardiovascular medicine. *J Am Coll Cardiol Img*. 2020;13:1119-1132.
15. Benjamin MM, Bianco C, Caccamo M, et al. Non-invasive prediction of tissue Doppler-derived E/e' ratio using lung Doppler signals. *Eur Heart J Cardiovasc Imaging*. 2020;21:994-1004.
16. Kagiya N, Piccirilli M, Yanamala N, et al. Machine learning assessment of left ventricular diastolic function based on electrocardiographic features. *J Am Coll Cardiol*. 2020;76:930-941.
17. Rahimtoola SH, Loeb HS, Ehsani A, et al. Relationship of pulmonary artery to left ventricular diastolic pressures in acute myocardial infarction. *Circulation*. 1972;46:283-290.
18. Nagueh SF, Smiseth OA, Appleton CP, 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;17:1321-1360.
19. Pitt B, Pfeffer MA, Assmann SF, et al. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med*. 2014;370:1383-1392.
20. Redfield MM, Anstrom KJ, Levine JA, et al. Isosorbide mononitrate in heart failure with preserved ejection fraction. *N Engl J Med*. 2015;373:2314-2324.
21. Redfield MM, Chen HH, Borlaug BA, et al. Effect of phosphodiesterase-5 inhibition on exercise capacity and clinical status in heart failure with preserved ejection fraction: a randomized clinical trial. *JAMA*. 2013;309:1268-1277.
22. Meta-Analysis Global Group in Chronic Heart Failure. The survival of patients with heart failure with preserved or reduced left

ventricular ejection fraction: an individual patient data meta-analysis. *Eur Heart J*. 2012;33:1750-1757.

23. Nagueh SF, Appleton CP, Gillebert TC, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography. *Eur J Echocardiogr*. 2009;10:165-193.

24. Rasmussen-Torvik LJ, Colangelo LA, Lima JAC, et al. Prevalence and predictors of diastolic dysfunction according to different classification criteria: the Coronary Artery Risk Development in Young in Adults study. *Am J Epidemiol*. 2017;185:1221-1227.

25. Almeida JG, Fontes-Carvalho R, Sampaio F, et al. Impact of the 2016 ASE/EACVI recommen-

dations on the prevalence of diastolic dysfunction in the general population. *Eur Heart J Cardiovasc Imaging*. 2018;19:380-386.

26. Obokata M, Kane GC, Reddy YN, Olson TP, Melenovsky V, Borlaug BA. Role of diastolic stress testing in the evaluation for heart failure with preserved ejection fraction: a simultaneous invasive-echocardiographic study. *Circulation*. 2017;135:825-838.

27. Kao DP, Lewsey JD, Anand IS, et al. Characterization of subgroups of heart failure patients with preserved ejection fraction with possible implications for prognosis and treatment response. *Eur J Heart Fail*. 2015;17:925-935.

28. Sengupta PP, Kramer CM, Narula J, Dilsizian V. The potential of clinical phenotyping of heart

failure with imaging biomarkers for guiding therapies: a focused update. *J Am Coll Cardiol Img*. 2017;10:1056-1071.

29. Santos AB, Roca GQ, Claggett B, et al. Prognostic relevance of left atrial dysfunction in heart failure with preserved ejection fraction. *Circ Heart Fail*. 2016;9:e002763.

KEY WORDS deep learning, diastolic dysfunction, echocardiography, heart failure with preserved ejection fraction

APPENDIX For supplemental methods, tables, and references, please see the online version of this paper.