



OPEN

Artificial intelligence improves detection and classification of pulmonary venous hypertension related to left ventricular diastolic dysfunction by chest radiography

Richard D. White^{1,2,3}✉, Mutlu Demirer¹, Ronnie A. Sebro¹, Isabel O. Cortopassi², Justin T. Stowell², Matthew R. McCann², Timothy Barry⁴, Christopher P. Appleton⁴, Scott A. Helgeson⁵ & Barbaros S. Erdal¹

Isolated-Left Ventricular Diastolic Dysfunction [LVDD] ranges (and may progress) from preclinical asymptomatic, symptomatic-LVDD, to LVDD-predominate Heart Failure [HF] presentations; if recognized early, LVDD progression might be preventable. Current early-HF screening remains limited, providing opportunities for insights from a standard Chest X-Ray [CXR]. While CXR assessment for “pulmonary congestion” supports suspected-HF evaluation in evidence-based guidelines, the potential for systematic Pulmonary Venous Hypertension [PVH]-Staging to contribute to initial detection and scaling of LVDD is unclear. This study compared CXR-based PVH-Staging to Doppler Echocardiography [DEcho]-based LVDD-Grading in the absence of systolic dysfunction. Questions included: (1) With PVH-Staging performed by cardiothoracic radiologists, what intra-/inter-reader variabilities remain? (2) Does PVH-Staging track LVDD-Grading? and (3) Can AI-assisted PVH prediction of LVDD-Grade match human performance? CXR examinations of 1,682 (including 750 asymptomatic/healthy) subjects, without: (1) Anatomical/physiological confounders of DEcho or CXR examinations (≤ 24 h apart), and (2) AI model-training confounders, were independently assigned 1 of 11 (9 PVH-related) Pulmonary Vasculature Patterns [PVPs] by 4 cardiothoracic radiologists and repeated for reliability evaluation. Expert-consensus Human Ground Truth [HGT] PVH PVPs were correlated with LVDD Grades (0 to 3–4), as were PVH-Rank predictions by a transformer-based AI model [“PVPI”]. Despite experience-dependent intra-/inter-reader reliability in PVP assignment, there was significant ($p < 0.001$) overall consistency. With increasing HGT PVH Stage, a significant ($p < 0.001$) trend towards increasing LVDD Grade was found; while PVH-Staging achieved confidence backing Grade 0/No LVDD, confident LVDD Grade recognition was not achieved until Grades 3–4/Restrictive Filling. However, a significantly ($p < 0.001$) stronger incrementally positive trend in PVPI PVH-Ranking with LVDD-Grading was demonstrated. Although validated, PVH-Staging for LVDD-Grading is limited by reader variabilities. AI-assisted PVH-Ranking may facilitate earlier and widespread objective CXR screening for LVDD which is ubiquitous in HF.

Dysfunction of the left ventricle during the typically longer diastolic filling phase of the cardiac cycle (from impaired muscle relaxation and/or chamber distention) is a pathophysiological manifestation of multiple conditions^{1,2}. Left Ventricular Diastolic Dysfunction [LVDD] is ubiquitous in Heart Failure [HF] regardless of the LV Ejection Fraction [LVEF] during the systolic contraction phase^{1–3}, and it plays a greater role than systolic dysfunction in determining baseline clinical capabilities and exercise-intolerance^{1,4}.

¹Division of Augmented Intelligence in Imaging, Mayo Clinic Florida, Jacksonville, FL, USA. ²Division of Cardiothoracic Imaging, Department of Radiology, Mayo Clinic Florida, Jacksonville, FL, USA. ³Department of Radiology, Mayo Clinic Florida, 4456 San Pablo Road, Jacksonville, FL 32224, USA. ⁴Department of Cardiovascular Medicine, Mayo Clinic Arizona, Phoenix, AZ, USA. ⁵Division of Pulmonary Medicine, Department of Medicine, Mayo Clinic Florida, Jacksonville, FL, USA. ✉email: white.richard1@mayo.edu

From both functional and clinical standpoints, isolated-LVDD can precede HF. Its severity ranges from: (1) Preclinical asymptomatic, but at-risk (Stage A⁵) or mildly dysfunctional (Stage B⁵, found in 20–30% of general population^{6–10}, to (2) Symptomatic-LVDD (Stage C⁵, especially exercise-related^{6,11}, to (3) LVDD-predominate HF (Stage D⁵, with morbidity and mortality resembling systolic HF^{6,12}). The potential for progression along this pathophysiological and prognostic spectrum is well-recognized^{6–8,10,11,13–15}.

Preclinical LVDD is associated with aging¹⁶ (especially in individuals ≥ 50 years old and/or of female sex^{7,14,17,18}). These initially subclinical changes (affecting almost 40% of elderly population¹⁹ and the prospects for symptom development (e.g., dyspnea on exertion)^{6,7} are amplified in persons with modifiable predisposing factors (e.g., hypertension⁷, obesity², diabetes⁷, renal failure²⁰).

At the other extreme, LVDD is an important component of HF with preserved LVEF $\geq 50\%$ [HFpEF]¹², an entity accounting for more than half of newly diagnosed HF cases. HFpEF diagnosis relies on evidence of: (1) HF-related signs and symptoms from high mean Left Atrial Pressure [mLAP] with “pulmonary congestion”^{6,21}; (2) high LV End-Diastolic Pressure [LVEDP] (at rest or only during exercise^{6,22} from invasive Right Heart Catheterization [RHCath]^{12,23} or non-invasive transthoracic Doppler Echocardiography [DEcho]^{24–26}; and/or (3) elevated serum B-type Natriuretic Peptide [BNP] levels^{12,21}. Due to LVDD, the following may result: 1. LA enlargement with dysfunction; 2. Pulmonary Venous Hypertension [PVH], potentially achieving isolated post-capillary Pulmonary Hypertension [PH], due to LA-PV continuity; and 3. Combined pre-/post-capillary PH [CpcPH], causing vascular remodeling and eventually failure of the Right Ventricle [RV]^{27,28}.

RHCath-based increased mean Pulmonary Capillary Wedge Pressure [mPCWP] (a surrogate for directly measured high LVEDP or mLAP^{1,4,29,30} indicates the cumulative hemodynamic burden of LVDD on both LA operating compliance^{12,31} and atrioventricular coupling^{23,32} causing mLAP elevation. Thus, while LVDD-related increases in LVEDP primarily reflect high LV preload (input volume or pressure) and reduced LV diastolic operating compliance (myocardial relaxation and inherent viscoelastic stretching)^{1,4,23,33}, increased mLAP and mPCWP signify its net hemodynamic impact directed retrograde into the pulmonary circulation, including the tendency for PVH^{4,24,34,35}. DEcho now dominates LVDD evaluations, with various parameters correlating closely with elevated mLAP predisposing to PVH²⁶; although a multi-parameter DEcho examination delineates LVDD severity, no single parameter reliably indicates elevated LVEDP or mLAP levels per individual^{24,26}.

Standard screening algorithms for the initial recognition of HF-related LVDD have relied on clinical history, physical findings, electrocardiographic changes, and BNP levels, but to a much lesser extent Chest X-Ray [CXR] results^{21,35}. Nevertheless, despite very limited objective supportive evidence, a CXR has been consistently included in initial diagnostic approaches to suspected-HF based on expert consensus, with appropriateness justified by perceived value in: (1) Identifying other causes (e.g., emphysema) of HF-like symptoms; and (2) Recognizing undefined “pulmonary congestion”^{21,35–38}.

CXR “pulmonary congestion” has been evaluated using systematic PVH-Staging (based on active superior redistribution or central distension of pulmonary vasculature, progressing to superimposed interstitial followed by alveolar pulmonary edema)^{39–41}. PVH-Staging has been validated against catheter-determined increased mPCWP (or infrequently LVEDP) within a range of pathophysiological conditions (varying LV-preload, LA-operating, and/or LV-operating conditions or with confounding factors) with inconsistent corroborative results and without definitively establishing a direct relationship with LVDD^{40,42–49}. Due to considerable CXR-interpretation variability, especially for characteristics of mild PVH^{50–53}, consensus readings have been proposed^{43,44,53}.

Thus, CXR-based PVH-Staging has yet to: (1) Be confirmed as directly indicating LVDD, the universal trait of HF^{1–3}, especially in isolated-LVDD where ventricular size is relatively normal⁵⁴ and unlikely a HF marker⁵⁵; or (2) Address the appearance of PVH with secondary PH^{27,56,57}. In addition, reported PVH-Staging results have not accounted for co-existent confounding causes of LVDD (e.g., conduction delay)^{55,58–61}, on one hand, or lung conditions (e.g., fibrosis) altering pulmonary vasculature⁶², on the other. Last, PVH-Staging has relied on subjective CXR evaluations leading to intra-/inter-reader variability without delineation of expectations for interpreter qualifications influencing diagnostic success^{43,63,64}.

Hence, the potential for CXR-based PVH-Staging to make a significant contribution to current diagnostic algorithms^{35,36} or scoring systems^{21,24,65} for the initial exclusion, versus detection with scaling, of LVDD has remained unanswered; ongoing limitations in screening approaches^{36,66} indicate persistent enhancement opportunities, possibly using PVH-Staging. It is particularly opportune, considering that a CXR remains: (1) The most widely accessible and frequently performed imaging examination worldwide; and (2) Inexpensive; these two attributes combined could facilitate early and widespread community-based application of PVH-Staging in suspected-HF, and possibly HF at-risk, cases. Also, despite regional differences in Cardiothoracic Radiology expertise, digital CXR formats and Artificial Intelligence [AI] processing now support more uniform and less human-dependent PVH-Staging^{67,68}. While CXR-dependent AI models may help predict an elevated BNP level in CXR-identified congestive HF⁶⁹ or help detect an elevated mPCWP value⁷⁰, AI prediction of LVDD absence, versus its presence and/or severity independent of co-existent pathophysiological confounders^{55,58–61}, has not been described.

Therefore, considering the worldwide epidemic of pre-HF and HF cases^{8,9,71}, especially those emphasizing LVDD^{6–8,10,11,13–15} that, if recognized early (along with associated risk factors), might have its progression prevented or reversed^{2,28,72}, optimized timely and widespread LVDD recognition is paramount. To that end, this study was designed to elucidate the potential for CXR-based PVH-Staging to identify levels of isolated-LVDD. Prevailing questions included: (1) When PVH-Staging is performed only by cardiothoracic radiologists, what intra-/inter-reader reliabilities remain and do they relate to reader experience?; (2) Does expert PVH-Staging confirm LVDD absence or, when detected, correctly track LVDD severity?; and (3) Can AI-assisted PVH predictions of LVDD degree at least match human performance, while eliminating human interpretation-reliability issues?

Results

Study population

The Study Population consisted of 1,682 subjects with or without suspected LVDD [Table 1] in the absence of possible anatomic or physiological imaging confounders or possible confounders of AI-model training (described in detail as exclusion criteria under **Methods**).

Suspected-LVDD subjects

The 846 subjects with characteristic symptoms (e.g., dyspnea on exertion) had undergone within the same 24-hour period: (1) a DEcho examination for LVDD which confirmed preserved LVEF $\geq 50\%$; and (2) a CXR examination [Table 1]. DEcho LVDD-Grading had categorized each subject as one of the following^{25,26,73}:

- Grade 0 (aka Normal Filling): $N = 250$.
- Grade 1 (aka Delayed Relaxation): $N = 253$.
- Grade 2 (aka Pseudo-Normal Filling): $N = 252$.
- Grade 3–4 (aka Restrictive Filling, either Grade 3/Reversible or Grade 4/Fixed): $N = 91$.

Unsuspected-LVDD subjects

A group of 750 asymptomatic and largely healthy subjects [Healthy] (without DEcho examinations), all free of LVDD predisposing factors (e.g., renal failure^{2,7,20} was added [Table 1]. In addition, 86 subjects with RHC-confirmed pre-capillary PH [Group 1 PH]^{74,75} were also included.

Human-based CXR PVH-staging vs. DEcho LVDD-grading

Reviewer-based assignment of PVP

Description

The digital CXR examination of each study subject was independently reviewed twice (separated in time) by four cardiothoracic-trained radiologists (varying in experience level) for the assignment of one of the following 11 possible Pulmonary Vascular Patterns [PVPs]:

- 1: Normal.
- 2–5: PVH-Stages 1, 2-Early, 2-Late, or 3 alone^{39,53,62,76}.
- 6–9: PVH-Stages (above) + CpcPH^{56,57,76}.
- 10: Group 1 PH^{74,75}.
- 11: Uncertain.

	Subjects (N)	Demographics		Concurrent DEcho findings			Previous confirming RHCath findings			
		Age Mean (YO) [Range]	Sex %F/%M	CXR-DEcho Interval ^o (Days)	No PH (N)	Isolated post-cap PH (N)	CpcPH (N)	RHCath-CXR Interval (Days)	mPCWP Mean (mmHg) [Range]	mPAP Mean (mmHg) [Range]
Healthy (No Suspected LVDD)	750	56 [20–87]	35%/65%	-	-	-	-	-	-	-
Grade 0 (Normal Filling)	250	55 [17–87]	48%/52%	≤ 1	98% (245)	2% (5)	0% (0)	-	-	-
Grade 1* (Delayed Relaxation)	253	75 [32–101]	61%/39%	≤ 1	87% (221)	11% (27)	2% (5)	-	-	-
Grade 2** (Pseudo-Normal Filling)	252	72 [21–99]	54%/46%	≤ 1	57% (143)	34% (87)	9% (22)	-	-	-
Grade 3–4*** (Restrictive Filling)	91	70 [30–94]	47%/53%	≤ 1	33% (30)	49% (45)	18% (16)	-	-	-
Group 1 PH	86	52 [21–78]	88%/12%	Mean 130	-	-	-	Mean 255	10 [1–15]	49 [26–82] < 49 (mean 40): $N = 52$ > 49 (mean 63): $N = 34$
Total	1,682									

Table 1. Study population. CpcPH: Combined pre-/post-capillary PH, CXR: Chest X-Ray, DEcho: Doppler Echocardiography, LVDD: Left Ventricular Diastolic Dysfunction, mPAP: mean Pulmonary Artery Pressure, mPCWP: mean Pulmonary Capillary Wedge Pressure, mmHg: millimeters of mercury, N: Number of subjects, %F: % Female, %M: % Male, post-cap: post-capillary, PH: Pulmonary Hypertension, RHCath: Right Heart Catheterization, YO: Years Old. ^o When > 2 suitable CXRs available, selection favored first those preceding, and then closest to, the DEcho. * vs. Grade 0: Significantly [$p < 0.001$] greater mean age, % female, Isolated post-cap PH %, and CpcPH %. ** vs. Grade 1: Significantly [$p < 0.001$] greater Isolated post-cap PH % and CpcPH %. *** vs. Grade 2: Significantly [$p < 0.001$] greater Isolated post-cap PH % and CpcPH %.

Human Ground Truth [HGT] PVP assignments (by consensus between the two most-experienced Expert Reviewers: Reviewer 3 and Reviewer 4) of PVH Stage were used in determining the relationship between PVH-Staging and LVDD-Grading.

Intra-reviewer reliabilities

Intra-Reviewer variability in assigning PVPs by Reviewers 1–4 ranged from 14.0 to 19.4%. Nevertheless, there was significant ($p < 0.001$) overall intra-Reviewer reliability by good agreement (Intra-Class Correlation [ICC] 0.78–0.81) with PVP-order ignored, and by strong agreement (Reviewers 1–3: Kendall rank correlation coefficient [Ktau] 0.55–0.67) or very strong agreement (Reviewer 4: Ktau 0.74) with PVP-order considered [Table 2].

For intra-Reviewer reliability related exclusively to a pulmonary vascular abnormality [Table 2], significant ($p < 0.001$) values, increasing with Reviewer experience, applied to Group 1 PH detection (versus Normal) and its differentiation from PVH Stage 1; respectively, the two and three most-experienced Reviewers demonstrated progressing good-excellent consistency (ICC 0.77–0.95). While significant ($p < 0.001$), poor-moderate agreement (ICC 0.29–0.56) applied to PVH Stage 1 detection (versus Normal) across Reviewers 1–4.

For intra-Reviewer reliability related to pulmonary edema, a positive experience-to-consistency trend for edema detection (versus Normal, PVH Stage 1, or Group 1 PH) was noted [Table 2]. While significant ($p < 0.001$), the agreements were moderate (ICC 0.55–0.72) across Reviewers 1–4.

Regarding CpcPH detection, only the most-experienced Reviewer achieved significant ($p < 0.001$) moderate consistency (ICC 0.64) in the setting of PVH Stage 2 [Table 2].

Inter-expert reviewer concordance

While 18.3% of final PVP assignments by the two Expert Reviewers (i.e., Reviewer 3 and Reviewer 4) were discordant, there was significant ($p < 0.001$) overall agreement between them at good (ICC 0.76) to strong (Ktau 0.62) levels [Table 2].

Regarding inter-Expert Reviewer concordance related exclusively to a pulmonary vascular abnormality in the absence of edema, significant ($p < 0.001$) agreement at good (approaching excellent) levels applied to the detection of Group 1 PH (vs. Normal) (ICC 0.88) and its differentiation from PVH Stage 1 (ICC 0.86). Though, poor agreement (ICC 0.09) applied to the detection of PVH Stage 1 (vs. Normal).

		Intra-reviewer [between initial and repeat assignments]				Inter-expert reviewer [Pre-HGT consensus]
		Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 3 vs. Reviewer 4
Overall Intra-Reviewer Reliability	PVP-Order Ignored ¹	0.78* (0.76, 0.80)	0.78* (0.76, 0.80)	0.78* (0.76, 0.80)	0.81* (0.79, 0.82)	-
	PVP-Order Considered ²	0.67*	0.66*	0.55*	0.74*	-
Overall Inter-Expert Reviewer Concordance	PVP-Order Ignored ¹	-	-	-	-	0.76* (0.74, 0.79)
	PVP-Order Considered ²					0.62*
Intra-reviewer reliability per basic PVP characteristic						
Exclusively pulmonary vascular abnormality (- Edema)						
	• Detect: PVH Stage 1 (+/- Group 2 PH) vs. Normal ¹	0.29* (0.25, 0.35)	0.41* (0.36, 0.45)	0.34* (0.29, 0.39)	0.56* (0.52, 0.59)	0.09* (0.08, 0.25)
	• Detect: Group 1 PH vs. Normal ¹	0.71* (0.68, 0.74)	0.71* (0.69, 0.74)	0.77* (0.75, 0.79)	0.79* (0.77, 0.81)	0.88* (0.87, 0.89)
	• Differentiate: PVH Stage 1 (+/- Group 2 PH) vs. Group 1 PH ¹	0.72* (0.54, 0.84)	0.79* (0.69, 0.86)	0.80* (0.73, 0.86)	0.95* (0.92, 0.97)	0.86* (0.75, 0.93)
Pulmonary edema (+/- Group 2 PH)						
	• Detect: Edema (Stage 2 or 3) vs. Normal, PVH Stage 1, or Group 1 PH ¹	0.55* (0.52, 0.59)	0.65* (0.62, 0.68)	0.67* (0.63, 0.69)	0.72* (0.70, 0.75)	0.52* (0.48, 0.57)
	• Stage: Matching rating of detected edema (Between Stage 2–3) ¹	0.67* (0.56, 0.76)	0.65* (0.54, 0.74)	0.60* (0.41, 0.75)	0.47* (0.34, 0.59)	0.78* (0.61, 0.87)
Group 2 PH						
	• Detect: Group 2 PH Superimposed of PVH Stage 1	0.00 (-0.37, 0.37)	0.19 (-0.10, 0.45)	0.00 (-0.21, 0.21)	0.37 (0.05, 0.62)	-
	• Detect: Group 2 PH Superimposed of PVH Stage 2	0.39* (0.20, 0.55)	0.27* (0.11, 0.43)	0.00 (-0.32, 0.32)	0.64* (0.53, 0.72)	0.00 (-0.89, 0.47)
	• Detect: Group 2 PH Superimposed of PVH Stage 3	-	0.00 (-0.12, 0.12)	0.00 (-1.4e-14, 5.1e-14)	0.00 (-0.45, 0.83)	0.00 (-0.21, 0.55)

Table 2. Intra-reviewer reliability and inter-expert reviewer concordance in PVP assignment. HGT: Human Ground Truth, PH: Pulmonary Hypertension, PVH: Pulmonary Venous Hypertension, PVP: Pulmonary Vascular Pattern. * $p < 0.001$. ²Ktau: 0.26–0.48.26.48 (Moderate), 0.49–0.70.49.70 (Strong), 0.71–1.00.71.00 (Very Strong). ¹ICC (confidence interval): < 0.50 (Poor), 0.50–0.74.50.74 (Moderate), 0.75–0.89.75.89 (Good), 0.90–1.00.90.00 (Excellent).

However, regarding inter-Expert Reviewer concordance related to pulmonary edema, a significant ($p < 0.001$) agreement at a good level (ICC 0.78) was demonstrated for rating edema, once detected.

Final PVP assignments

Reviewers 1–4 finally assigned PVPs to be abnormal in 14.6–17.9% of Study Population CXR examinations, while 40.5% (682/1,682) of subjects had pathophysiology confirmed by DEcho or RHCath [Table 3].

By consensus HGT assignment [Table 3], the prevalence of abnormal PVPs was 15.5%. Remarkably, HGT assignments of PVH Stage 1 (3.2%) were less common than those individually by Reviewers 1–4 (3.4–9.2%). On the other hand, HGT assignments of PVH Stage 2-Early without superimposed CpcPH (4.9%), were more common than individually by both Expert Reviewers (3: 2.0%, 4: 4.8%), while less common than by the other two Reviewers (1: 7.6%, 2: 5.1%). Similarly, HGT assignments of Group 1 PH (3.2%) were more common than those individually by Reviewers 1–4 (1.0–2.9.0.9%).

Correlating PVH-staging with LVDD-grading

Suspected absence of LVDD

When LVDD was not suspected (by lack of symptoms and predisposing factors) in Healthy subjects, HGT-assigned PVPs included: Normal 98.0%, PVH 1.7%, and Group 1 PH 0.3% [Fig. 1]. The designation of PVH applied to 13 subjects aged 46–83 (median 51) years old.

Increasing LVDD grade

With progression from DEcho LVDD Grade 0 to Grade 3–4, a significant ($p < 0.001$) positive trend towards increasing HGT-assigned PVH Stage 1 to PVH Stage 3 PVPs was found in symptomatic subjects with suspected LVDD [Fig. 1].

With confirmed absence of LVDD based on Grade 0/Normal Filling, HGT-assigned PVPs indicated Normal in 94.0% and PVH in 5.6% [Fig. 1]. Despite presumed normal mLAP levels, a significantly ($p < 0.050$) higher PVH prevalence was found in symptomatic subjects with Grade 0 than in unsuspected-LVDD Healthy subjects. Despite this apparent PVH background, the assessment of likelihood of excluding DEcho Grade 1 to Grade 3–4 produced an HGT -LR value of 0.18 (Reviewers 1–4: 0.14–0.38), approximating a strong level of confidence (i.e., < 0.10) in favor of PVH-Staging to rule-out LVDD at rest, even in the presence of characteristic symptoms [Table 4].

Although expected to be highly prevalent in Grade 1/Delayed Relaxation due to presumed absence of any mLAP elevation, HGT-assigned PVPs indicated Normal in 87.3% of affected subjects [Fig. 1]; varying PVH PVPs, especially PVH Stage 2 in 9.5%, were assigned in the remaining 12.7%. This relatively higher PVH prevalence in Grade 1, compared to Healthy and/or Grade 0 conditions, was statistically significant ($p < 0.050$). However, neither -LR nor +LR values provided strong support to PVH-Staging to rule-out or rule-in Grade 1, respectively [Table 4].

With further progression of LVDD to Grade 2/Pseudo-Normal Filling with its anticipated mildly elevated mLAP levels, the prevalence of a HGT-assigned Normal PVP continued to significantly ($p < 0.050$) decrease to 64.7% with a corresponding additional significant increase in prevalence of PVH PVPs to 34.5%. This rise in PVH prevalence included increasing representation of PVH Stage 2 in 26.6%, with superimposed CpcPH

PVP Assignment Options (1,682 CXR Examinations)	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Consensus HGT
Normal	82.5% (1,387/1,682)	82.1% (1,380/1,682)	85.4% (1,434/1,682)	82.3% (1,385/1,682)	84.5% (1,422/1,682)
PVH Stage 1	3.9% (66/1,682)	4.9% (82/1,682)	9.2% (155/1,682)	3.4% (57/1,682)	3.2% (53/1,682)
PVH Stage 1 + Group 2 PH	0.5% (8/1,682)	1.0% (17/1,682)	0.1% (2/1,682)	0.1% (2/1,682)	0.1% (2/1,682)
PVH Stage 2-Early	7.6% (128/1,682)	5.1% (85/1,682)	2.0% (34/1,682)	4.8% (81/1,682)	4.9% (82/1,682)
PVH Stage 2-Early + Group 2 PH	0.4% (7/1,682)	0.7% (12/1,682)	0.0% (0/1,682)	1.4% (24/1,682)	0.6% (10/1,682)
PVH Stage 2-Late	2.7% (46/1,682)	3.4% (57/1,682)	0.5% (9/1,682)	2.6% (43/1,682)	1.8% (30/1,682)
PVH Stage 2-Late + Group 2 PH	0.0% (0/1,682)	0.3% (5/1,682)	0.0% (0/1,682)	2.1% (35/1,682)	1.3% (22/1,682)
PVH Stage 3	1.4% (23/1,682)	0.4% (7/1,682)	0.7% (12/1,682)	0.0% (1/1,682)	0.3% (5/1,682)
PVH Stage 3 + Group 2 PH	0.0% (0/1,682)	0.0% (1/1,682)	0.0% (0/1,682)	0.4% (6/1,682)	0.1% (2/1,682)
Group 1 PH	1.0% (17/1,682)	2.1% (35/1,682)	2.1% (36/1,682)	2.9% (48/1,682)	3.2% (54/1,682)
Uncertain	0.0% (0/1,682)	0.0% (1/1,682)	0.0% (0/1,682)	0.0% (0/1,682)	0.0% (0/1,682)

Table 3. Final PVP assignments. HGT: Human Ground Truth, PH: Pulmonary Hypertension, PVH: Pulmonary Venous Hypertension, PVP: Pulmonary Vascular Pattern.

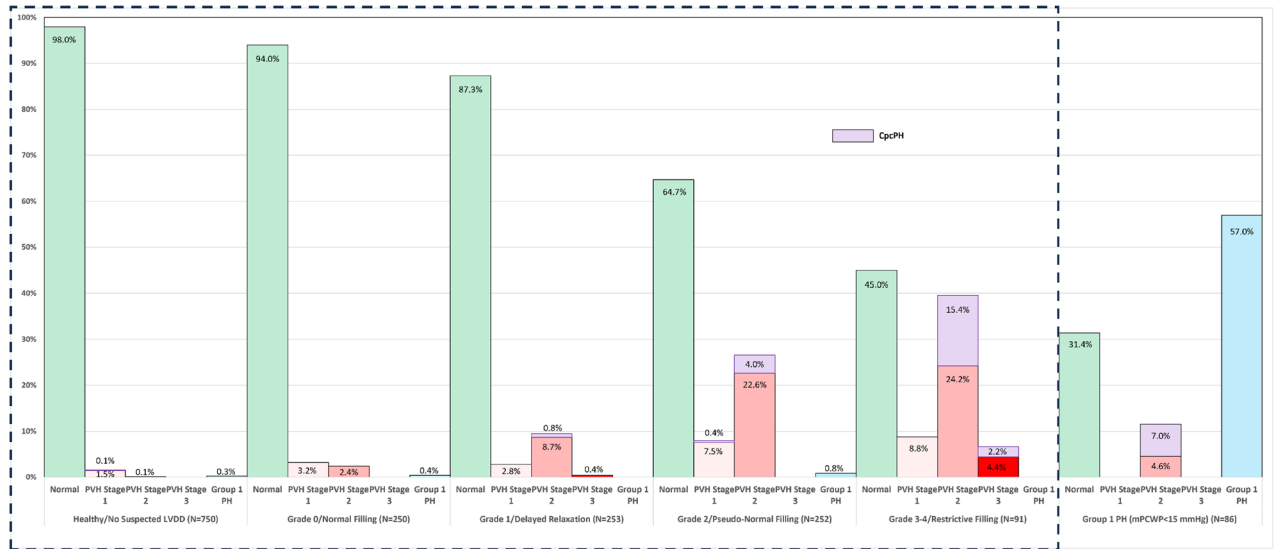


Fig. 1. HGT-Assigned CXR-Based PVH Stage vs. DEcho-Based LVDD Grade (or Confirmed Group. For LVDD severity ranging from no suspected LVDD in Healthy subjects through DEcho LVDD Grade 0 to Grade 3–4, the prevalence of an HGT-assigned PVH PVP [dash-outlined box], at Stage 1 [light pink], Stage 2 [medium pink], and/or Stage 3 [red] alone or with superimposed CpcPH [purple], significantly increased while the assignment of a Normal PVP [green] correspondingly declined. For previously confirmed Group 1 PH, the correct PVP assignment [blue] predominated. CXR: Chest X-Ray, CpcPH: Combined pre-/post-capillary Pulmonary Hypertension, DEcho: Doppler Echocardiography, HGT: Human Ground Truth, LVDD: Left Ventricular Diastolic Dysfunction, mmHg: millimeters of mercury, N: Number of subjects, PCWP: Pulmonary Capillary Wedge Pressure, PH: Pulmonary Hypertension, PVH: Pulmonary Venous Hypertension, PVP: Pulmonary Vasculature Pattern.

	DEcho LVDD Grade				Group 1 PH
	0	1	2	3–4	
	Normal Filling	Delayed Relaxation	Pseudo-Normal Filling	Restrictive Filling	
+LR					
Reviewer 1	1.47	0.45	2.29	23.10	423.00
Reviewer 2	1.39	0.51	3.10	18.48	167.23
Reviewer 3	1.55	0.66	2.52	64.62	162.12
Reviewer 4	1.37	0.95	1.82	10.08	Undef
Consensus HGT	1.42	0.52	2.38	55.45	160.67
-LR					
Reviewer 1	0.20	1.03	0.78	0.95	0.50
Reviewer 2	0.21	1.11	0.94	0.92	0.61
Reviewer 3	0.14	1.04	0.80	0.92	0.62
Reviewer 4	0.38	1.00	0.85	0.88	0.80
Consensus HGT	0.18	1.04	0.83	0.94	0.43

Table 4. Likelihoods of identifying DEcho LVDD grade per reviewer or HGT. HGT: Human Ground Truth, -LR: Negative Likelihood Ratio (values <0.10 indicate strong evidence for ruling-out diagnosis), +LR: Positive Likelihood Ratio (values >10.0 indicate strong evidence for ruling-in diagnosis), Undef: Undefined due of specificity of 1.

accounting for 4.0%. Again, neither -LR nor +LR values provided strong evidence for PVH-Staging in ruling-out or ruling-in Grade 2, respectively [Table 4].

Last, the progression to Grade 3–4/Restrictive Filling, with its anticipated moderately elevated mLAP levels, was associated with another significant ($p < 0.050$) increase in PVH prevalence in 55.0% of affected subjects. Once more, the rise in PVH was characterized by increasing representation of PVH Stage 2 in 39.6% (superimposed CpcPH accounting for 15.4%) and PVH Stage 3 in 6.6% (including 2.2% with CpcPH). The assessment of likelihood of detecting the presence of Grade 3–4 produced an HGT +LR value of 55.5 (high

individual Reviewer values of 10.1–64.6), indicating strong evidence (i.e., >10.0) in favor of PVH-Staging to rule-in Grade 3–4 at rest [Table 4].

HGT results in group 1 PH

Correct HGT PVP assignments of Group 1 PH were made in 57.0% (mPAP 26–77/mean 49 mmHg), versus assignments of Normal in 31.4% (mPAP 27–70/mean 45 mmHg), of subjects with previously RHC-confirmed Group 1 PH [Fig. 1]. In the remaining 11.6% of subjects, assignments of PVH Stage 2 were made (despite mPCWP 6–15/mean 10 mmHg), 4.6% represented by Stage 2 alone (mPAP 28–77/mean 60 mmHg) and 7.0% characterized by superimposed CpcPH (mPAP 46–82/mean 59 mmHg). Thus, a HGT assignment of a form of pre-capillary PH was made in 64% of Group 1 PH subjects. The high +LR values by both HGT (160.7) and individual-Reviewers (162.1–423.0) indicated very strong confidence in identifying Group 1 PH by CXR; based on -LR values (0.43–0.80), its exclusion confidence was found to be equivocal [Table 4].

AI-based CXR PVH-ranking vs. DEcho LVDD-grading

AI-based assignment of “PVP”

Description

Our cascading two-component “PVP Identifier” [PVPI] AI model consisted of the: (1) *Thoracic-Content Segmentator* (for automatic thoracic-cavity segmentation); and (2) *PVP Multi-Classifer* to differentiate between image-data characteristics based on a four-class physiological pattern, as follows:

- Normal.
- PVH Stage 1 (Vascular redistribution without edema +/- CpcPH).
- PVH “Stage 2+” (Vascular redistribution/congestion with predominantly interstitial edema +/- CpcPH).
- Group 1 PH.

The PVPI was used in processing the same 1,682 CXR examinations to predict probabilities for these four classes. PVPI-derived CXR-based PVH-rank predictions were correlated with LVDD-Grading.

AI performance compared to reviewer performance

PVPI activity maps [Fig. 2] demonstrated an increasingly redistributed balance of activity (green) from lung bases to upper-lung regions with worsening HGT-assigned PVH Stage; for each example, the highest PVPI PVH-Ranking prediction (PVH Predict) corresponded well with HGT PVH-Staging assignment. However, with both the pulmonary vasculature and cardiac silhouette reflecting relative inactivity (blue shades), changing physiological “States” rather than PVPs were apparently recognized by the PVPI.

For PVH detection, the PVPI achieved an overall accuracy of 0.89 and balanced accuracy of 0.86 by a 3-class output (i.e., Normal versus PVH Stages 1–2 + versus Group 1 PH), with near-equally strong normalized class accuracies for both Normal at 0.91 and PVH at 0.90 [Table 5].

On the other hand, for basic PVH-Ranking, the PVPI by a 4-class output (i.e., Normal versus PVH Stage 1 versus PVH Stage 2 + versus Group 1 PH) reached a lower overall accuracy of 0.79, and balanced accuracy of 0.72, with normalized class accuracies still strongest for Normal at 0.91, followed by PVH Stage 2 + at 0.84 [Table 5]. The intermediary PVH Stage 1 was predicted with a low normalized accuracy of 0.35, near-equivalent to PVH Stage 2 + misclassification at 0.38.

Correlating PVH-ranking with LVDD-grading

As with HGT PVH-Staging assignments, a significant ($p < 0.001$) very positive trend towards increasing PVH-Ranking by the PVPI with progression from Grade 0 to Grade 3–4 was found in symptomatic subjects [Fig. 3]; compared to HGT PVH-Staging the PVPI-related trend was significantly ($p < 0.001$) stronger both overall as well as at each LVDD Grade.

Suspected absence of LVDD

In the absence of suspected LVDD (i.e., in Healthy), the PVPI predicted Normal in 91.6% of subjects [Fig. 3] (vs. HGT 98.0%). Although still representing a very small proportion of this group, 4.8% (vs. HGT: 1.7%) had PVPI-recognized PVH.

Increasing LVDD grade

In contrast, with Grade 0/Normal Filling, the PVPI predicted Normal in only 30.4% (vs. HGT 98.2%) but PVH in 47.6% (including Stage 1: 29.6%, Stage 2+: 18%) [Fig. 3]. This relatively higher PVH prediction in suspected LVDD but Grade 0, compared to unsuspected LVDD (i.e., Healthy), was statistically significant ($p < 0.001$).

With Grade 1/Delayed Relaxation, PVPI prediction of PVH predominated at 70.7% overall (including Stage 2+: 39.1%, Stage 1: 31.6%), while Normal was predicted in 10.7% [Fig. 3]. Again, PVH prediction by the PVPI has significantly ($p < 0.001$) higher than that for Grade 0.

Advances to Grade 2/Pseudo-Normal Filling and then Grade 3–4/Restrictive Filling, were tracked by rapid and significant ($p < 0.001$) additional stepwise increases in PVH predictions at 79.4% and 83.5% overall (including Stage 2+: 55.2% and 68.1%), respectively [Fig. 3].

Trialing in confounded test group

To initially gauge PVPI robustness, it was trialed in a Confounded Test Group of 40 previously excluded suspected-LVDD subjects, each demonstrating on CXR an example of implanted cardiovascular material, while otherwise meeting the inclusion criteria of Study Population subjects.

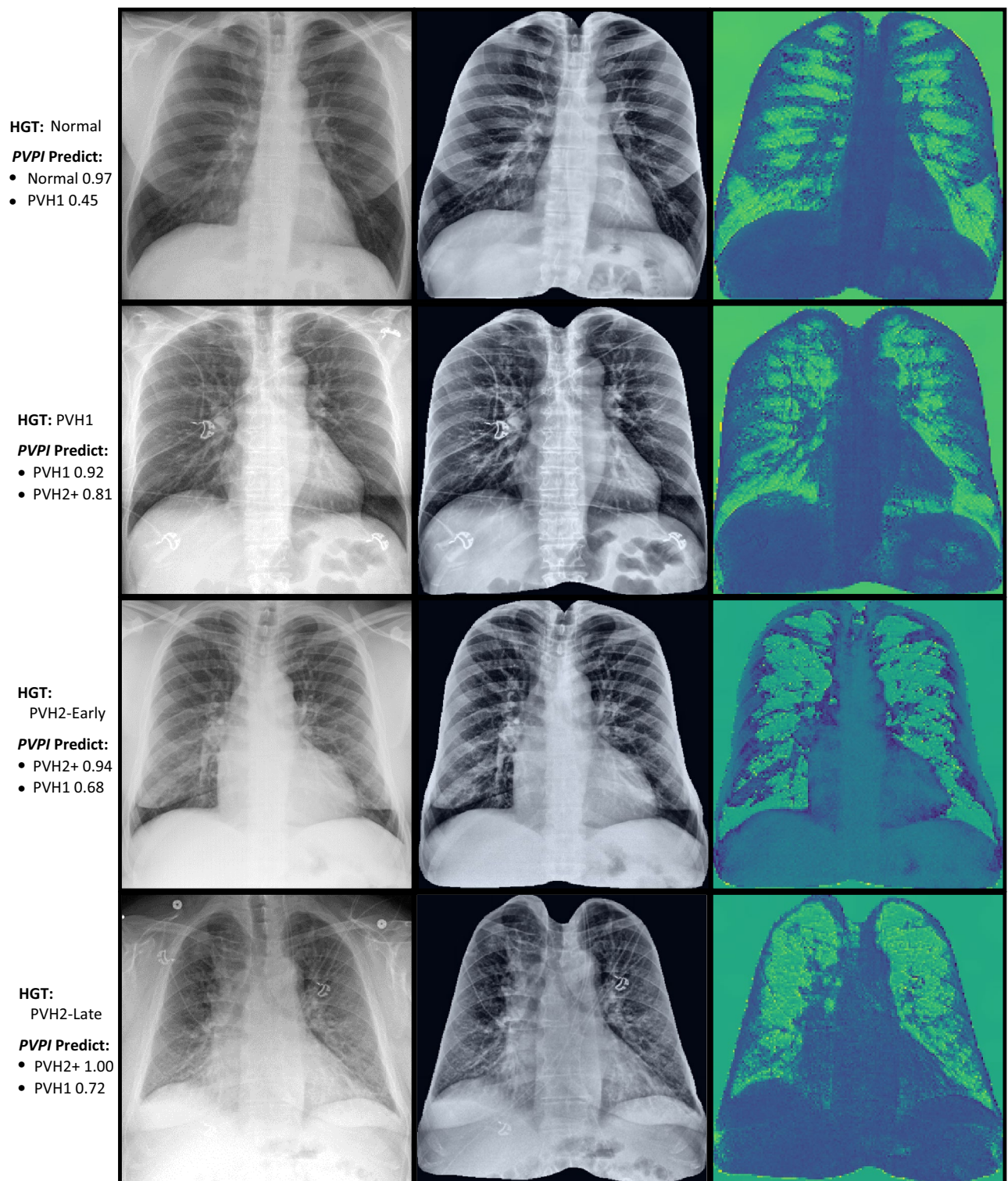


Fig. 2. PVPI activity maps.

By Expert Reviewer adjudication, initial auto-segmentations of Frontal-view CXRs produced by the *Thoracic-Content Segmentator* were deemed adequate in 38 of 40 subjects, resulting in an overall Dice coefficient approaching 1.00 (i.e., 0.9995). For the remaining two auto-segmentations (truncated left hemithorax apex at clavicle containing orthopedic plate-screw device; blunted left costophrenic angle by moderate-sized pleural effusion) in which minor manual modifications were thought to be potentially needed, the Dice coefficients were initially 0.99.

HGT Assignments	PVH Detection [3-Class Output Normalized Accuracies]			HGT Assignments	PVH-Severity Ranking [4-Class Output Normalized Accuracies]			
	Normal	PVH Stages 1 & 2+	Group 1 PH		Normal	PVH Stage 1	PVH Stage 2+	Group 1 PH
Normal	0.91 (138/151)	0.03 (5/151)	0.05 (8/151)	Normal	0.91 (138/151)	0.03 (5/151)	0.00 (0/151)	0.05 (8/151)
PVH Stages 1 & 2	0.05 (10/199)	0.90 (179/199)	0.05 (10/199)	PVH Stage 1	0.16 (9/55)	0.35 (19/55)	0.38 (21/55)	0.11 (6/55)
				PVH Stage 2	0.01 (1/144)	0.13 (18/144)	0.84 (121/144)	0.03 (4/144)
Group 1 PH	0.07 (4/54)	0.15 (8/54)	0.78 (42/54)	Group 1 PH	0.07 (4/54)	0.06 (3/54)	0.09 (5/54)	0.78 (42/54)
Overall Accuracy = 0.89 (359/404) Overall Balanced Accuracy = 0.86				Overall Accuracy = 0.79 (320/404) Overall Balanced Accuracy = 0.72				

Table 5. Performance of PVPI relative to HGT assignments of PVP. HGT: Human Ground Truth, PH: Pulmonary Hypertension, PVH: Pulmonary Venous Hypertension, PVP: Pulmonary Vasculature Pattern, PVPI: Pulmonary Vasculature Pattern Identifier.

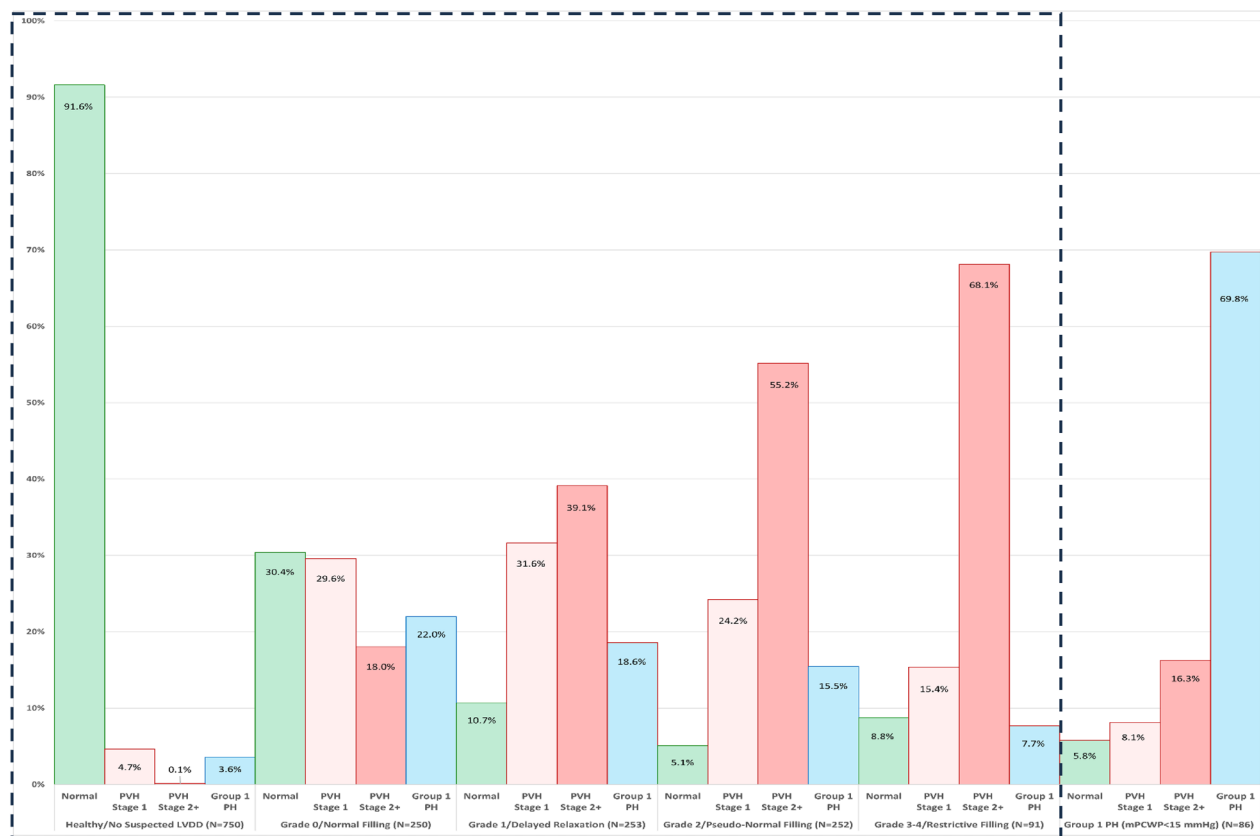


Fig. 3. PVPI CXR-Based PVH Rank vs. DEcho-Based LVDD Grade (or Confirmed Group 1 PH). For LVDD severity, ranging from no suspected LVDD in Healthy subjects through DEcho LVDD Grade 0 to Grade 3–4, the prevalence of an PVPI ranking of PVH level [dash-outlined box] as Stage 1 [light pink] or Stage 2 [medium pink] (but to a lesser degree incorrectly as Group 1 PH [blue]), quickly increased while the ranking as Normal [green] correspondingly declined rapidly. In the setting of previously confirmed Group 1 PH, the correct assignment predominated. CXR: Chest X-Ray, DEcho: Doppler Echocardiography, LVDD: Left Ventricular Diastolic Dysfunction, PH: Pulmonary Hypertension, PVH: Pulmonary Venous Hypertension, PVPI: Pulmonary Vasculature Pattern Identifier.

Evaluation of the subsequent 3-class and 4-class output performances by the *PVP Multi-Classifer* (using the unmodified auto-segmentation set) indicated a negative impact with overall accuracies declining to 0.73 and 0.58, and balanced accuracies to 0.68 and 0.58, respectively [Table 6]. The decrease was primarily attributable to recognition of a Normal “State” at 0.40 and PVH Stage 2+ “State” at 0.70; no association with the represented cardiovascular material types was identified.

Despite evidence of anatomical-confounder impairment of physiological-“State” differentiation, the *PVP Multi-Classifer* predictions of LVDD severity remained strong, with CXR PVH-Ranking of DEcho LVDD-Grading demonstrating a significantly positive ($p < 0.001$) relationship [Fig. 4]. Above Grade 1/Delayed Relaxation, only PVH “States” were detected, including a purely PVH Stage 2+ “State” for Grade 3–4/Restrictive Filling.

Discussion

This work provides overdue insights into the implications of CXR “pulmonary congestion” in suspected HF. We believe it is the first to describe the: (1) Potential for systematic CXR-based PVH-Staging to make evidence-based contributions to current diagnostic algorithms^{35,36} or scoring systems^{21,24,66} in the initial exclusion versus detection with scaling of LVDD in asymptomatic pre-HF^{6–10}, symptomatic pre/early HF^{6,11}, or HFpEF^{6,12} conditions; (2) Relationship between PVH-Staging (presumably indicating mLAP reflecting both LA-operating compliance^{12,31} and atrioventricular coupling^{22,32} and LVDD-Grading (indicating LVEDP reflecting both LV preload and LV-operating compliance^{1,4,23}); (3) Interpretation reliability of PVH-Staging by cardiothoracic radiologists with differing experience; and (4) Performance of completely AI-assisted CXR-based PVH-Ranking in predicting DEcho-determined LVDD Grade.

Despite the fact that, unlike prior studies^{39–49}, we restricted CXR interpretations to cardiothoracic radiologists, intra-/inter-interpret variability were demonstrated during Reviewer assignments of one of 11 possible PVPs, including 9 PVH-related. Nevertheless, there was significant overall intra-Reviewer reliability. In addition, for Group 1 PH detection and its differentiation from PVH Stage 1, as well for pulmonary edema detection, at least moderate intra-Reviewer reliability was achieved, with enhancement by Reviewer experience, as previously recognized by others⁶³. In addition, comparable inter-Expert Reviewer concordance for rating edema (i.e., early versus late interstitial) was also shown.

Like related studies^{50–53}, we found intra-Reviewer and inter-Expert Reviewer identification of pulmonary vascular redistribution without edema (i.e., PVH Stage 1) to be unreliable; only the most-experienced Reviewer achieved moderate consistency. Thus, the application of PVH-Staging, especially in early phases preceding identifiable edema, would be limited by such dependency on experience; prior proposals for consensus interpretations^{43,44,51} constitute impractical solutions.

Difficulties recognizing a PVH Stage 1 PVP characterized by redistribution (with PCWP or mLAP 13–17 mmHg), representing a narrow LV-preload transition between a Normal PVP (with PCWP or mLAP 4–12 mmHg) and “pulmonary congestion” plus interstitial edema (with PCWP or mLAP \geq 18–24 mmHg)^{39,53,62,76}, have been previously reported^{43,45,47,51,52,55,77}. Considering the diverse prior experience, the absence of a PVH Stage 1 PVP might only help exclude increased LV preload for HF pretest probability $< 9\%$, whereas it might help confirm HF with pretest probability $> 91\%$ ⁵³; the prospects for PVH Stage 1 “State” recognition in isolated-LVDD with AI assistance are uniquely addressed in this work.

Based on HGT-assigned PVH-related PVPs, significant correlation between CXR PVH-Staging and LVDD-Grading in the setting of preserved LVEF was found.

Remarkably, low levels of PVH were recognized even in asymptomatic unsuspected-LVDD (i.e., Healthy) subject. With their median age of 51 years old, detected PVH in half could have represented age-related (i.e., \geq 50-year-old) LVDD^{7,17,18} but no DEcho data was available to test this conjecture.

Nevertheless, significantly increased PVH prevalence in symptomatic DEcho-confirmed Grade 0/Normal Filling suggested influences by fluctuating hemodynamics from episodic LVDD or atrial fibrillation^{6,18,22,78}.

HGT Assignments	PVH Detection [3-Class Output Normalized Accuracies]			HGT Assignments	PVH-Severity Ranking [4-Class Output Normalized Accuracies]			
	Normal	PVH Stages 1 & 2+	Group 1 PH		Normal	PVH Stage 1	PVH Stage 2+	Group 1 PH
Normal	0.40 (4/10)	0.60 (6/10)	0.00 (0/10)	Normal	0.40 (4/10)	0.30 (3/10)	0.30 (3/10)	0.00 (0/10)
PVH Stages 1 & 2+	0.05 (1/20)	0.85 (17/20)	0.10 (2/20)	PVH Stage 1	0.00 (0/10)	0.40 (4/10)	0.40 (4/10)	0.20 (2/10)
				PVH Stage 2+	0.10 (1/10)	0.20 (2/10)	0.70 (7/10)	0.00 (0/10)
Group 1 PH	0.00 (0/10)	0.20 (2/10)	0.80 (8/10)	Group 1 PH	0.00 (0/10)	0.00 (0/10)	0.20 (2/10)	0.80 (8/10)
	Overall Accuracy = 0.73 (29/40) Overall Balanced Accuracy = 0.68				Overall Accuracy = 0.58 (23/40) Overall Balanced Accuracy = 0.58			

Table 6. Performance of *PVP multi-classifier* in confounded test group. HGT: Human Ground Truth, PH: Pulmonary Hypertension, PVH: Pulmonary Venous Hypertension, PVP: Pulmonary Vasculature Pattern, *PVPI*: Pulmonary Vasculature Pattern Identifier.

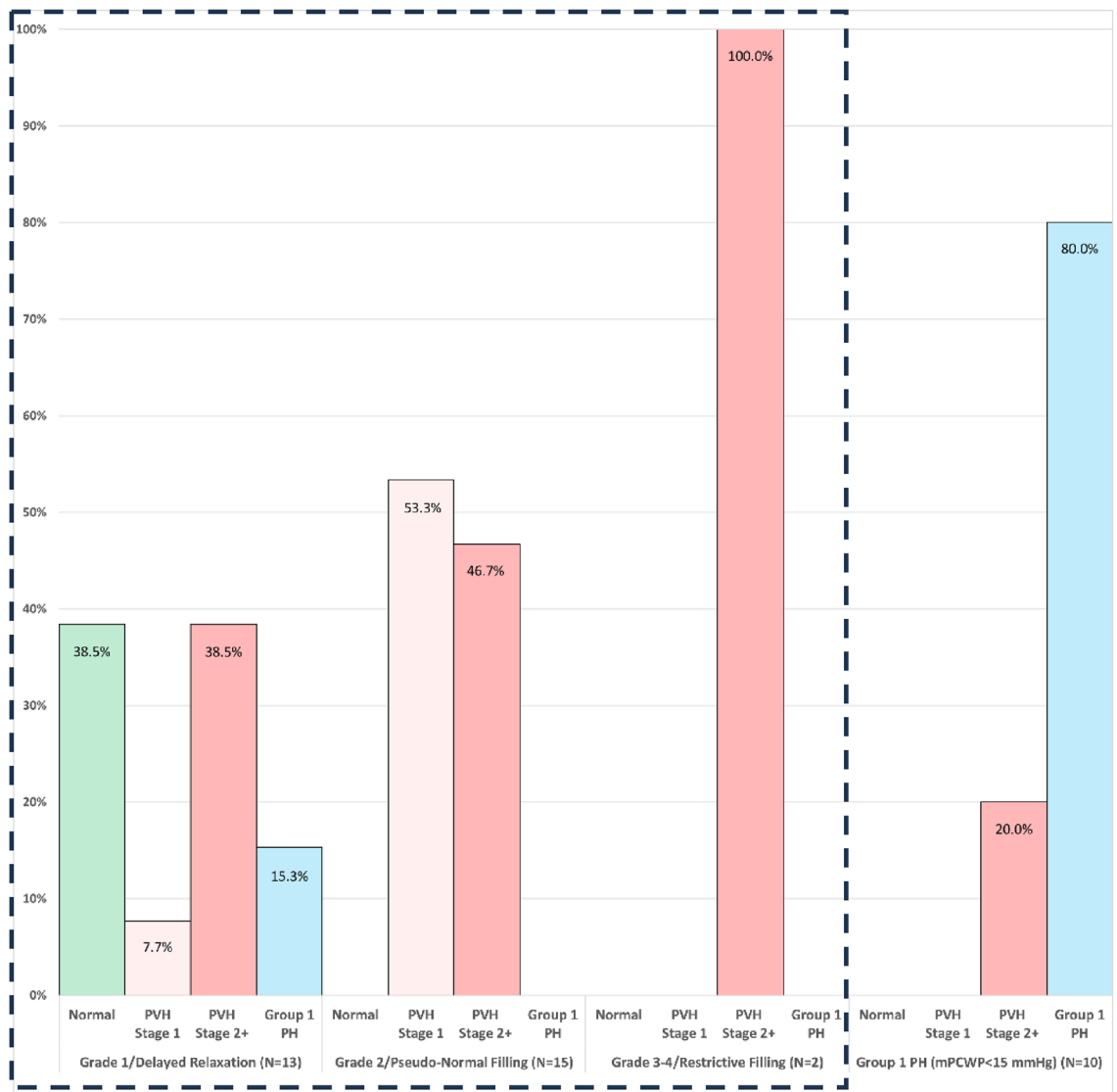


Fig. 4. PVPI CXR-Based PVH Rank vs. DEcho-Based LVDD Grade (or Confirmed Group 1 PH) in Confounded Test Group. For LVDD severity in the Confounded Test Group, ranging from DEcho LVDD Grade 0 to Grade 3–4, the prevalence of an PVPI ranking of PVH level [dash-outlined box] as Stage 1 [light pink] or Stage 2 [medium pink] (but to a lesser degree incorrectly as Group 1 PH [blue]), quickly increased while the ranking as Normal [green] correspondingly declined rapidly. In the setting of previously confirmed Group 1 PH, the correct assignment predominated. CXR: Chest X-Ray, DEcho: Doppler Echocardiography, LVDD: Left Ventricular Diastolic Dysfunction, PH: Pulmonary Hypertension, PVH: Pulmonary Venous Hypertension, PVPI: Pulmonary Vasculature Pattern Identifier.

While false-positive PVH recognition in Healthy and Grade 0 conditions is a possible explanation, generally low Reviewer sensitivities to PVH, especially to PVH Stage 1 (3.3% overall, 8.6% Stage 2) was observed in this study.

Despite presumably normal mLAP levels^{39,53,62,76}, significantly higher PVH assignments in Grade 1/Delayed Relaxation, implicating hemodynamic lability with intermittent mLAP elevations or LVDD exacerbations^{6,18,22,78} and more prolonged or relatively fixed pulmonary vascular manifestations (e.g., distention)⁷⁹. Thus, combined DEcho Grades 0–1 and CXR-indicated PVH might serve as a marker for unstable and/or worsening mild or early LVDD^{10,78,80,81}.

As expected from mild mLAP elevation, Grade 2/Pseudo-Normal Filling was associated with significantly higher PVH prevalence, reflecting background PVH accentuated by chronically fluctuating hemodynamics^{22,78,80}. Consequently, PVH on CXR might serve as an adjunct LVDD indicator for Grade 2 at rest, thereby obviating total dependency on stressing for its differentiation from Grade 0^{82,83}. Last, Grade 3–4/Restrictive Filling, with expected moderate mLAP elevation, demonstrated another significant increase in PVH prevalence involving slightly more than half of affected subjects.

However, while LR-based assessment of confidence in excluding DEcho-confirmed LVDD was approximated at Grade 0, adequate confidence in LVDD recognition was not achieved until Grade 3–4. Therefore, despite

significant direct correlation between PVH-Staging and LVDD-Grading, human-based PVH assignments did not definitely predict incremental increases in LVDD severity. Accordingly, we saw justification for investigating the potential of AI to support CXR-based PVH evaluation in facilitating LVDD assessment or complementing DEcho-based LVDD examinations.

Our cascading two-component *PVPI* model includes the *PVP Multi-Classifier* which was created to freely identify distinguishing anatomical and/or functional cardiopulmonary differences, hence our reference to a physiological “State”, rather than a PVP, of PVH. Considering that accompanying HGT-assigned increasing PVH PVPs, there were progressing upper-lung redistributions of AI-model inference activity (without correspondence to cardiac or PVP silhouettes), *PVP Multi-Classifier*-identified parameter(s) likely reflected: (1) Subtle or complex variations in radiodensity not humanly detectable⁸⁴; or (2) physiological indicators of lung blood flow and/or water content^{85,86}.

Nonetheless, compared to HGT-assigned PVH, the *PVPI* demonstrated significantly greater: (1) Sensitivity to PVH-related changes, starting in symptomatic proven-absent LVDD (Grade 0); and (2) Correlation between increases in ranked CXR PVH-“State” (especially PVH Stage 2+) and LVDD-Grading. Both should facilitate contributions by CXR-based PVH evaluation to initial LVDD recognition in suspected-HF individuals or those at-risk of HF, as well as help realize opportunities for complementing DEcho LVDD examinations, while eliminating issues in reliability of human-based CXR interpretation.

In Group 1 PH, *PVPI* performance exceeded that by HGT-assigned PVPs. Low-level PVH prevalences detected by both human-based and AI-based CXR evaluations are at least partly explained by: (1) An inclusion allowance of RHCath-confirmed mPCWP 13–14 mmHg, shared with PVH Stage 1^{39,53,62,74–76}; (2) Interdependence of RV systolic function and LV filling⁸⁷; and (3) The prolonged RHCath-CXR periods, allowing interval PVH development. However, unlike HGT assignments, the *PVPI* exhibited a tendency towards Group 1 PH detection in symptomatic subjects, especially those with Grade 0. While, with few exceptions, affected subjects received HGT-assigned Normal PVP, this *PVPI* predilection may have reflected recognition of Group 1 PH with normal baseline spatial lung-perfusion heterogeneity, differentiable after vasodilatation⁸⁸.

Limitations in this study are recognized. First, the relatively small numbers of subjects representing PVH Stages and LVDD Grades limited human-/AI-based evaluations, resulting from the purposeful exclusion of over 14,000 potential subjects to avoid possible: (1) anatomical/physiological confounders of DEcho-LVDD or CXR-PVP examinations^{55,58–62,89,90}; or (2) AI model-training confounders^{91–93}. Nevertheless, our stringent inclusion restrictions ensured the: (1) Needed assessment of a relationship between PVH-Staging and LVDD-Grading, and (2) Development of an AI model which focused only on that relationship. Consequently, our Subject Population uniquely reflected suspected HF with isolated-LVDD in the absence of systolic dysfunction. Thus, while a complete understanding of the robustness of our *PVPI* for AI-assisted objective CXR “pulmonary congestion” assessment awaits real-world deployment allowing for confounding factors, we performed initial trialing in the Confounded Test Group and showed that *PVPI* prediction remained strong.

In addition, semi-erect positioning during CXR examination was allowed since: (1) Basilar preferential lung flow is maintained even when supine⁹⁴; and (2) Prior reports demonstrated no significant negative impact on PVH assessment with supine positioning⁵⁰.

Last, the DEcho reports reviewed from the greater than 20-year span did not uniformly indicate use of all now-possible LVDD-Grading methodology^{25,26,73}. However, the fundamental components⁹⁵ were considered adequate for the needed basic LVDD-Grading.

In conclusion, work reveals that: (1) There is, in fact, a previously unverified significant direct relationship between CXR PVH-Staging and LVDD-Grading; (2) CXR PVH-Staging remains, however, clinically limited by experience-dependent intra-/inter-interpreter variabilities; (3) LVDD exclusion by PVH-Staging is adequate but its inclusion is not confidently reached until Restrictive Filling is present, thereby inhibiting its impact on recognition of milder or earlier phases of HF; although (4) Our AI-assisted CXR PVH-Ranking appears to exceed the sensitivity of human performance which may facilitate earlier and widespread pathophysiological assessments in persons with suspected HF or at-risk of HF, potentially further enhanced when blended with other proposed screening indicators (e.g., AI-enabled ECG⁹⁶ in a multi-modal generative AI model⁹⁷ advancing the diagnostic needs of cardiometabolic care⁹⁸).

Methods

All methods were performed in accordance with the relevant guidelines and regulations.

Selection of study population

With approval by the Mayo Clinic Institutional Review Board, enterprise-wide data-mining of consecutive patients (spanning: 02/22/2003–08/15/2023) was completed using its shared Electronic Medical Record [EMR] system (Epic Systems, Verona, WI) to retrospectively identify potential candidates for the Study Population. Due to the retrospective nature of the study, the Mayo Clinic Institutional Review Board waived the need for obtaining informed consent from patients.

Screened personal EMRs reflected the highly multi-racial/cultural population of individuals drawn (locally, regionally, nationally, or internationally) to our integrated healthcare enterprise which consists of a multi-state network with three geographically dispersed quaternary-referral medical centers (located in Upper-Midwestern, Southeastern, and Southwestern regions of the United States), as well as > 70 Midwestern satellite hospitals or ambulatory clinics.

Suspected-LVDD subject identification and characterization

By electronic EMR searching, potential study subjects with suspected isolated-LVDD were first identified by having undergone within the same 24-hour period: (1) a DEcho examination confirming preserved LVEF ≥

50% during a comprehensive LVDD evaluation^{24,25,73} for characteristic symptoms (e.g., dyspnea on exertion)^{2,6}; and (2) a CXR examination. Preliminary filtering also included early exclusion of individuals with concurrent Chronic Kidney Disease at either Stage-5 or hemodialysis-dependent Stage-4 levels to avoid potentially confounding influences of cardiopulmonary volume-overloading and/or rapid dialysis-related fluctuations in LV-preload volume^{89,90}. From this initial filtered search, 15,880 potential subjects remained.

Next, the individual EMR of each of these potential subjects was manually screened (Author 1) for the presence of possible anatomical or physiological confounders of either the DEcho evaluation for isolated-LVDD or the CXR assessment of PVPs [Table 7]^{55,58–62,89,90}; those potential subjects affected by any of such possible confounders were excluded from further consideration. Last, the CXR images of still-eligible potential subjects were visually reviewed (Author 1) for: (1) Supine positioning (reducing gravitational contribution to normal pulmonary blood distribution)⁹⁴; or (2) Possible AI model-training confounders [Table 8] leading to spurious associations between characteristic cardiovascular materials and a disease type or severity during model training^{91–93}; the presence of either type of condition caused those affected to also be excluded. Ultimately, this extensive filtering resulted in marked reductions in the number of eligible subjects to 846 study subjects [Table 1], negatively impacting the representation of those with more-severe LVDD Grades, in which exclusion criteria (e.g., atrial fibrillation, indwelling infusion catheters) were more often met.

LV systolic dysfunction
• LVEF < 50%
Alternative Causes of LVDD
• LV-Inflow Abnormality
o ≥ Mild Mitral Valve Stenosis (Diastolic mean Doppler gradient > 2 mmHg)
o Left Atrial Myxoma
o Active Atrial Fibrillation or Flutter
• Conduction Abnormality
o First Degree Atrio-Ventricular Block with PR Interval ≥ 300 msec
o First Degree Atrio-Ventricular Block with PR Interval 200–299 msec + Left Fascicular Block
o Left Bundle Branch Block
• Extrinsic Abnormality
o Constrictive Pathophysiology
o Tamponade
o > Mild Right Ventricular Dilatation (to avoid Interventricular Septal Flattening/Reversal)
Increased and/or Variable LV Pre-Load Conditions
• > Mild Mitral Valve Insufficiency
• > Mild Aortic Valve Insufficiency
• > Mild Atrial-Level Shunting
• Stage 5 CKD or Hemodialysis-dependent Stage 4 CKD
Pulmonary Vascular Disease Altering Pulmonary Vasculature Patterns
• Pulmonic Valve Stenosis
• Peripheral Pulmonary Artery Stenosis
• Chronic Thrombo-Embolic Disease
• Pulmonary Vein Stenosis
Lung Disease Altering or Obscuring Pulmonary Vasculature Patterns
• Emphysema
• Basilar Pulmonary Fibrosis
• Diffuse Interstitial or Alveolar Disease
• Pneumonia
• Non-Cardiogenic Edema
• Diffuse Pulmonary Masses
• Surgical Lung Resections or Transplantation
• Endobronchial Valve Volume-Reductions
Acute Cardiovascular/Cardiopulmonary Events
• Acute Myocardial Infarction
• Acute Pulmonary Embolus
• Pneumothorax

Table 7. Exclusion criteria - possible imaging anatomical or physiological confounders. CKD: Chronic Kidney Disease, mmHg: millimeters of mercury, msec: milliseconds.

Implanted cardiovascular materials possibly promoting spurious associations during model training
• Sternal Wires
• Mediastinal or Lung Surgical Items
• Cardiac Valve Prostheses
• Pacing or Defibrillating Components
• Intra-Cardiac/Vascular Catheters
• Ventricular Assist Devices
• Loop Recorders

Table 8. Exclusion criteria - possible AI model-training confounders.

	Marker	DEcho Criteria				
		Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
		Normal Filling	Delayed Relaxation	Pseudo-Normal Filling	Reversible Restrictive Filling	Fixed Restrictive Filling
Mitral Inflow E/A Ratio	LV Preload	≥ 0.8	≤ 0.8	> 0.8 to < 2.0 (Reversed on Valsalva)	> 2.0 (Reversed on Valsalva)	> 2.0 (No Reversal on Valsalva)
Peak TR velocity: (Normal ≤ 2.8 m/sec)	RV & MPA Systolic Pressure	Normal	Normal	Abnormally Increased	Abnormally Increased	Abnormally Increased
Associated level of LV diastolic pathophysiology						
LV Myocardial Relaxation		Normal	Mildly Impaired	Moderately Impaired	Severely Impaired	Severely Impaired
LV Myocardial Stiffness		Normal	Mildly Increased	Moderately Increased	Severely Increased	Severely Increased
LV Preload (i.e., mLAP)		Normal	Normal	Mildly Increased	Moderately Increased	Moderately Increased
LA Contractility		Normal	Normal	Normal	Decreased	Decreased

Table 9. Representative criteria for DEcho grading of LVDD. DEcho: Doppler Echocardiography, E/A : peak E-wave velocity/peak A-wave velocity, LA: Left Atrial, LV: Left Ventricular, LVDD: Left Ventricular Diastolic Dysfunction, m/sec: meters/second, mLAP: mean Left Atrial Pressure, MPA: Main Pulmonary Artery, RV: Right Ventricle, TR: Tricuspid Regurgitation.

DEcho-examination results evaluation

All comprehensive DEcho-based LVDD evaluations represented were selected from examinations performed enterprise-wide by its Intersocietal Accredited Commission (Ellicott City, MD)-accredited Echocardiology Laboratories to support clinical standard-of-care over the greater than 20-year span. During this period, varying generations of different DEcho system manufacturers and/or versions were used.

LVDD-Grading based on the final report of each DEcho examination (≤ 1 day from CXR), was confirmed or corrected (Authors 1,7,8) for each of the remaining 846 study subjects with suspected LVDD in the absence of systolic dysfunction. Using well-established criteria for Grading of LVDD, including standard E/A-ratio parameters^{24,25,73}, each subject was categorized as one of the following [Table 9]:

- Grade 0 (aka Normal Filling).
- Grade 1 (aka Delayed Relaxation).
- Grade 2 (aka Pseudo-Normal Filling).
- Grade 3–4 (aka Restrictive Filling): Grade 3 (Reversible E/A ratio with patient Valsalva) and Grade 4 (Fixed E/A ratio despite patient Valsalva) were combined due to independently small subject numbers.

In addition, on an individual subject basis, DEcho determination (often corroborated by RHCath) of the severity of resulting PVH confirmed the concurrence of one of the following [Table 10]^{26,74,99}: (1) No PVH; (2) Insignificant PVH; Significant PVH (i.e., isolated post-capillary PH); or 4. CpcPH.

Unsuspected-LVDD subject identification

After the exclusion of initially considered subjects based on the presence of LVDD predisposing clinical factors (e.g., history of myocardial infarction), or for possible anatomical or physiological confounders of imaging (DEcho or CXR) [Table 7]^{55,58–62,89,90} or possible confounders of AI model-training [Table 8]^{91–93}, the following groups of subjects without suspected LVDD were added.

Healthy subjects

To support the possibility that a CXR-based PVP depicting baseline normal LV diastolic function was differentiable from a PVP associated symptomatic suspected-LVDD but with DEcho Grade 0, a group of 750 asymptomatic and relatively Healthy subjects (no DEcho performed) was added (Author 1) [Table 1].

	Marker	DEcho			
		No PVH	Insignificant PVH	Isolate post-cap PH	CpcPH
Peak TRV: Normal ≤ 2.8 m/sec	Estimated sPAP (RV-RA gradient plus estimated RAP)	Normal	Normal	Elevated	Elevated
$(TRV/TVI_{RVOI} \times 10) + 0.16$	Pulmonary Vascular Resistance	-	-	Normal	Elevated
RHCath					
mPCWP		4–12 mmHg	13–15 mmHg	> 15 mmHg	> 15 mmHg
mPAP		≤ 20 mmHg	≤ 20 mmHg	≥ 20 mmHg	≥ 20 mmHg
Pulmonary Vascular Resistance		< 2 WU	< 2 WU	< 2 WU	≥ 2 WU

Table 10. DEcho and RHCath measures of PVH severity. CpcPH: Combined pre-/post-capillary Pulmonary Hypertension, DEcho: Doppler Echocardiography, m/sec: meters/second, mmHg: millimeters of mercury, mPAP: mean Pulmonary Artery Pressure, mPCWP: mean Pulmonary Capillary Wedge Pressure, PH: Pulmonary Hypertension, post-cap: post-capillary, PVH: Pulmonary Venous Hypertension, RAP: Right Atrial Pressure, RHCath: Right Heart Catheterization, sPAP: Systolic Pulmonary Artery Pressure, TRV: Tricuspid Regurgitation Velocity, TVI_{RVOI} : Time-Velocity Integral of Right Ventricular Outflow Tract, WU: Wood Units.

Group 1 PH subjects

In addition, to help guarantee that PVPs reflecting CpcPH were distinguishable from the Group 1 PH^{74,75} (the other form of cardiopulmonary-derived pre-capillary PH), a group of 86 subjects with previously RHCath-confirmed (frequently corroborated by interval DEcho) Group 1 PH was also included (Authors 1,9) [Table 1]; in these patients, simultaneous PCWP measurements proved the absence of isolated post-capillary PH^{26,74,99}.

Supporting human-based CXR PVH-staging

CXR examinations and CXR-image reviewing

All CXR examinations of the 1,682 subjects in the Study Population [Table 1] originally promoted clinical standard-of-care by direct-digital or computed radiography using varying generations of nine manufacturers of fixed and/or portable CXR systems operating enterprise-wide over the greater than 20-year span. Each of digital CXR examination consisted of an upright or semi-upright Frontal view (postero-anterior or antero-posterior), 91% (1,524/1,682) were accompanied by a Lateral view.

The total of 3,206 CXR images were downloaded from the enterprise deconstructed Picture Archiving and Communication System, consisting of the: (1) Radiology Information System (Radiant from Epic Systems, Verona, WI); (2) Vendor Neutral Archive system (Synapse from TeraMedica/Fujifilm Medical Systems USA, Inc., Wauwatosa, WI); and (3) Viewer (Visage Imaging from Pro Medicus Ltd, Richmond, Australia) to a secure shared-drive. The shared-drive supported locally developed: (1) Graphical User Interface [GUI]¹⁰⁰ which allowed modifications of the underlying commercial software (MeVisLab from MeVis Medical Solutions AG, Bremen, Germany) for either bulk CXR image-reviewing or image-segmentation prior to AI-model training; and (2) Zero-footprint viewer (“CAI Viewer”)¹⁰¹ functioning on a backend database manager for CXR-image reviews to establish individual-Reviewer or consensus-HGT PVP assignments by the two Expert Reviewers, on one hand, or AI-model inference display for adjudication, on the other.

PVP assigning and PVH-staging

Initially, the de-identified CXR examinations were presented in random order for independent assessments (while blinded to all information regarding corresponding clinical status or DEcho findings) by four cardiothoracic-trained radiologists (Authors 1,4–6) to establish individual-Reviewer PVP assignments, and eventually consensus-HGT PVP assignments by the two most-experienced “Expert Reviewers”. CXR Reviewers 1–4 had different subspecialty experiences related to: (1) Training in CXR-based PVH-Staging during Radiology residency and/or Cardiothoracic Radiology fellowship; as well as (2) Years of post-training Cardiothoracic Radiology practice. These Reviewers included:

- Reviewer 1 (Fellowship-level).
 - Fellowship: 0.5-year.
 - PVH-Staging training: Moderate residency/strong fellowship.
 - Practice years: None.
- Reviewer 2 (Early-Career):
 - Fellowship: 1-year.
 - PVH-Staging training: Mild residency/mild fellowship.
 - Practice years: 4.
- Reviewer 3/HGT Expert Reviewer (Mid-Career).
 - Fellowship: 1-year.
 - PVH-Staging training: Moderate residency/strong fellowship.

- Practice years: 14.
- Reviewer 4/HGT Expert Reviewer (Late-Career).
 - Fellowship: 2-year.
 - PVH-Staging training: Strong residency/strong fellowship.
 - Practice years: 37.

While blinded to associated clinical and DEcho information, each Reviewer independently assessed the CXR examinations to assign each with one of the following 11 possible PVPs, including:

- 1: Normal.
- 2–5: PVH-Stages 1, 2-Early, 2-Late, or 3 alone^{39,53,62,76} [Table 11] [Fig. 5].
- 6–9: PVH-Stages (above) + CpcPH (with superimposed main & central pulmonary artery dilatation)^{56,57,76} [Fig. 6].
- 10: Group 1 PH (Main & central pulmonary artery dilatation without PVH, lung disease, etc.)^{74,75} [Fig. 6].
- 11: Uncertain.

Evaluating intra-reviewer reliabilities and inter-expert reviewer concordance

After a four-week “washout” period (with interval re-randomization of examinations), independent CXR assessments were repeated by Reviewers 1–4; this supported the determination of intra-Reviewer reliability per subspecialty experience (training and practice). During a subsequent (≥ 2 weeks later) adjudication by each Reviewer of any personal-assignment inconsistencies between assessments (previous assignments provided for consideration, without restriction to their re-use), Reviewers 1–4 individually committed to their final PVP assignments, thereby facilitating the determination of inter-Reviewer reliability.

Determining final HGT PVP assignments

Last, to achieve HGT assignments of PVPs for the 1,682 CXR examinations, consensus between the Expert Reviewers was reached per examination by: (1) Initial concordance between final assignments; or (2) Subsequent final concordance during “face-to-face” review of discordant final assignments.

Correlating human-based CXR PVH-staging with DEcho LVDD-grading

HGT-based PVH-severity determinations by PVP assignments were used in determining the relationship between CXR-based PVH-Staging and DEcho-based LVDD-Grading.

Supporting AI-based CXR PVH-ranking Creation of AI model for CXR assessment

Technical infrastructure

AI-model development utilized secure on-site and remote Graphics Processing Unit [GPU] (Nvidia, Santa Clara, CA)-dependent systems. Data curations and initial model development relied on: (1) One dual-GPU workstation (2 RTX 8000) with 96 GB video memory/128 GB system memory/12 TBs disk storage/2 TB SSD drive for operating-system support (Windows 10); and (2) Two single-GPU workstations containing (RTX 8000) with 48 GB video memory/128 GB system memory/18 TBs disk storage/2 TB SSD drive for operating-system support (Windows 10). For base-model training, a node (4xA100 80GB GPUs) from a 32-GPU high-performance cluster was utilized. For subsequent image-classification tasks, involving training, validation, and testing, a DGX A100 System (8xA100 40GB GPUs) was used.

	Vascular Redistribution (aka “Cephalization”)	Central Vascular Congestion	Perihilar Peribronchial Cuffing/Haziness	Peripheral Interlobular Septal Thickening (i.e., Kerley B Lines)	Perihilar & Lower-Lung Alveolar Opacity (aka “Batwing” Pattern)	Expected mPCWP (mmHg)
Normal	-	-	-	-	-	4–12
PVH Stage 1	XXX	-	-	-	-	13–17
PVH Stage 2-Early	XX+	XX	XXX	-	-	18–22
PVH Stage 2-Late	XX	XX+	XX	XXX	-	22–24
PVH Stage 3	-	XXX	-	XX	XXX	≥ 25

Table 11. CXR PVH-staging characteristics. CXR: Chest X-Ray, mPCWP: mean Pulmonary Capillary Wedge Pressure, mmHg: millimeters of mercury, PVH: Pulmonary Venous Hypertension, XX: moderately characteristic, XX+: moderately-highly characteristic, XXX: highly characteristic.

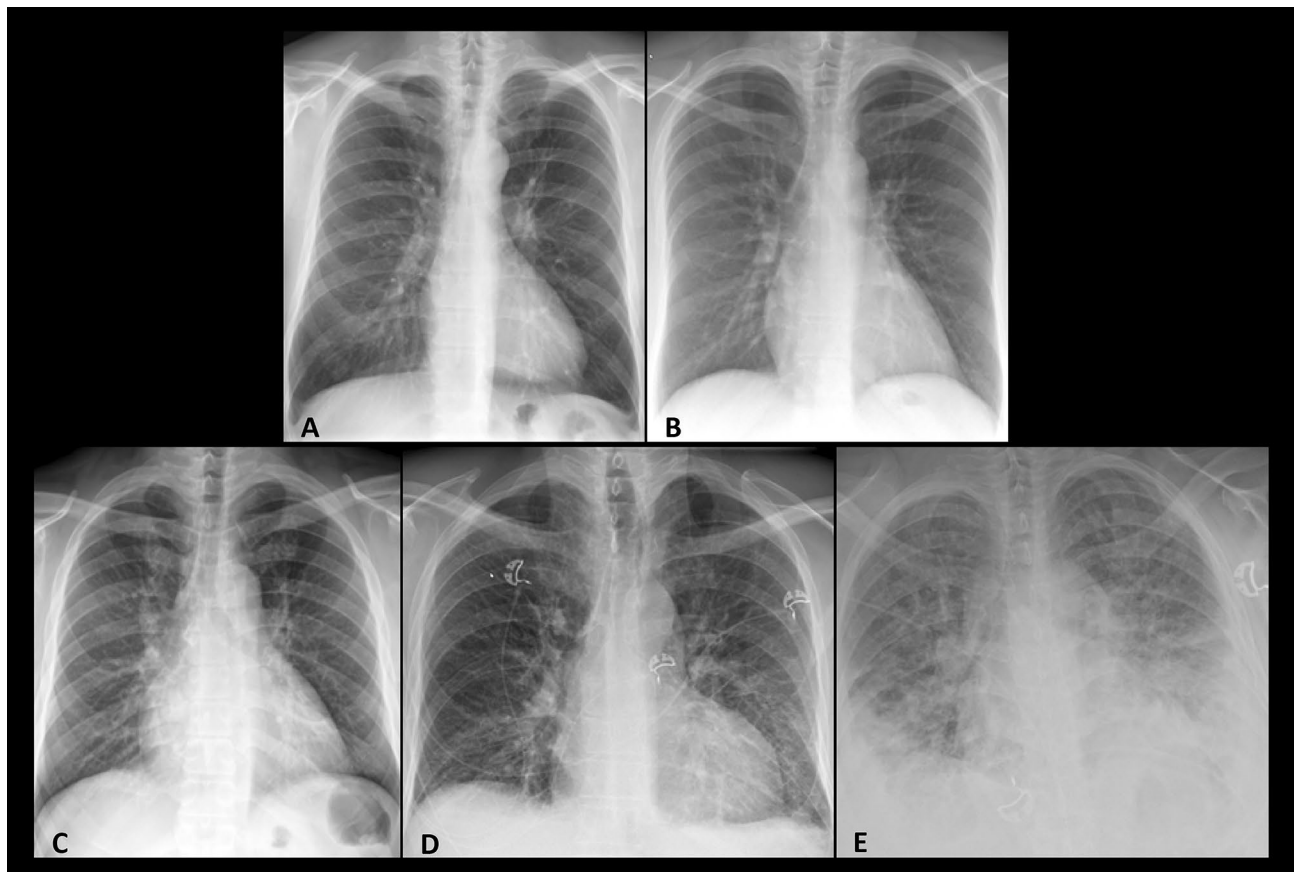


Fig. 5. Frontal Chest X-Ray images from 5 Study Population subjects represent the pulmonary vasculature patterns of Normal physiology and increasing Pulmonary Venous Hypertension [PVH]. They include [Table 11]: Normal: Mid-to-lower lung vascular predominance (A). PVH Stage 1: Vascular redistribution (aka “cephalization”) to upper lungs without pulmonary edema. (B) PVH Stage 2-Early: Vascular redistribution/ congestion with mild edema (perihilar peribronchial cuffing/haziness) (C). PVH Stage 2-Late: Central vascular congestion with moderate edema (perihilar haziness plus peripheral interlobular septal thickening (i.e., Kerley B Lines)) (D). PVH Stage 3: Central vascular congestion with severe edema (perihilar and lower-lung alveolar opacification (aka “batwing” pattern)) (E).

Data curation/annotation

Initially hypothesizing that increasing DEcho LVDD-Grade would be tracked closely by increasing PVH Stage on CXR, relatively equal-sized numbers of symptomatic subjects with Grades 0–4 were anticipated during formation of the Study Population. Accordingly, relatively equivalent numbers of study subjects with DEcho Grades 0–2 were compiled, expecting corresponding incremental increases in milder PVH Stages. However, despite Grade 3 (Reversible) and Grade 4 (Fixed) subjects being combined, the cumulative number of study subjects representing Restrictive Filling remained relatively small, attributable due to greater frequency of confounding conditions causing subject exclusion.

While maintaining clinical significance, subject bundling across PVPs was eventually needed for optimization of multi-classifier creation. This bundling was justified by the following:

- Unexpected imbalances in representations of the original 11 possible PVPs, including the segregation of those affected by CpcPH; nevertheless, the prevalence of CpcPH interpretations was recorded.
- Very few representations of PVH Stage 3 ($n=7$), insufficient for creation of a separate class for an already readily identifiable PVP (i.e., alveolar edema largely obscuring pulmonary vasculature); while included in multi-classifier training and validation, these subjects were omitted from model testing.

PVP identifier model components

Our cascading “PVP Identifier” [PVPI] incorporated the automatic consecutive application of the following two components (Authors 2,10).

“Thoracic-content segmentator” component

To avoid spurious associations related to extra-pulmonary characteristics (e.g., patient identification labels)^{91–93}, the 3,206 (1,682 Frontal + 1,524 Lateral) CXR images were manually segmented by an Expert Reviewer (Author

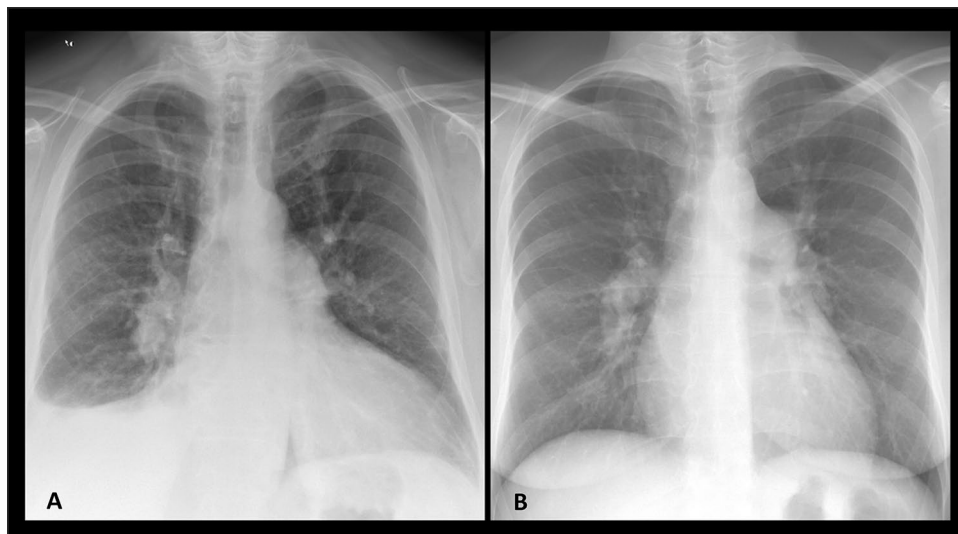


Fig. 6. Frontal Chest X-Ray images from 2 Study Population subjects represent the PVPs of both forms of cardiopulmonary-derived Pre-Capillary PH, including: CpcPH: Main and central pulmonary artery dilatation superimposed on PVH (Stage 2-Late in this case) (A). Group 1 PH: Main and central pulmonary artery dilatation without PVH, lung disease, etc. (B). CpcPH : Combined pre-/post-capillary Pulmonary hypertension PH: Pulmonary Hypertension, PVH: Pulmonary Venous Hypertension, PVP: Pulmonary Vasculature Pattern.

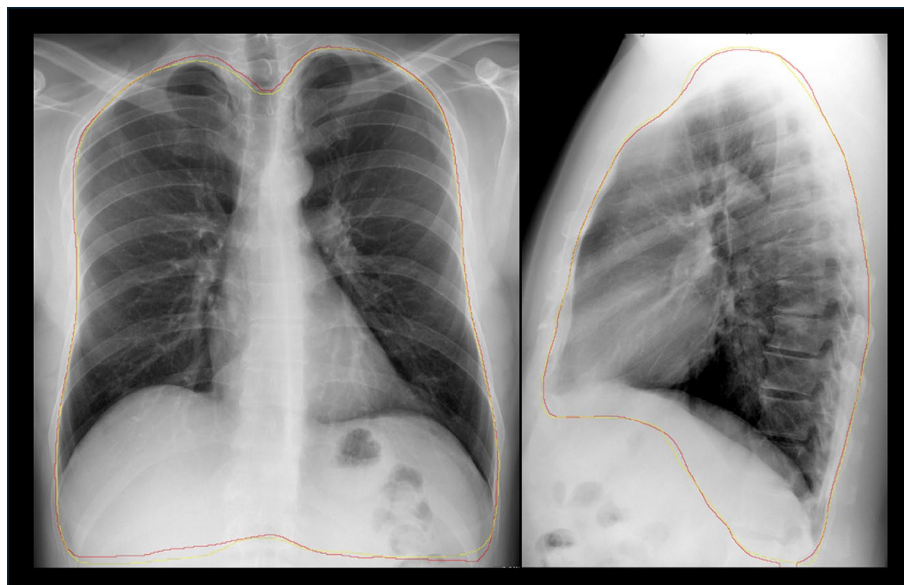


Fig. 7. Automatic thoracic-content segmentation on the CXR Frontal view (Left) and the Lateral view (Right). The original expert ground-truth segmentations (yellow) and the model inference segmentations produced by the *Thoracic-Content Segmentator* (red) corresponded very closely.

1) using a modification of the GUI¹⁰⁰. These segmentations followed the inner surface of the ribs and diaphragm to delineate the thoracic-cavity contents (e.g., lungs) from the apex to the posterolateral costophrenic sulci, using the L1-L2 disc space as a reference whenever they were obscured, such as by a pleural effusion¹⁰².

The resulting segmentations of the thoracic contents were used to create by supervised learning the *Thoracic-Content Segmentator* for automatic thoracic segmentation as the first *PVPI* component. The *Thoracic-Content Segmentator* was developed based on the DeepLabV3 semantic segmentation architecture¹⁰³, with ResNet-50 as a backbone¹⁰⁴. The final Dice coefficients^{105,106} achieved by the *Thoracic-Content Segmentator* were 0.98 for Frontal views and 0.97 for Lateral views [Fig. 7].

“PVP multi-classifier” component

The *PVP Multi-Classifier*, the second *PVPI* component, was developed using the DINO transformer architecture to: (1) Subdivide CXR-image data into patches of pixels; (2) Flatten patches into a sequence of visual vector tokens by linear transformation; and (3) Pass the token sequence into a transformer¹⁰⁷. During the model creation, a base model was initially developed using the publicly available MIMIC-CXR dataset comprised of 337,100 images¹⁰⁸, where a training loss of 1.7 was achieved with 200 epochs. The base model weights were then utilized to train an image-classification task^{107,109} for final *PVP Multi-Classifier* creation using *Thoracic-Content Segmentator* output from the 3,206 CXR images.

For creation of the *PVP Multi-Classifier*, a 5-fold stratified cross-validation approach (i.e., K-fold folded test set)¹¹⁰ was employed to rotate through training (60%), validation (20%), and testing (20%) subsets of input data after balancing of sample size per class via undersampling or oversampling. While supervised from the standpoint of PVP categorization, the *PVP Multi-Classifier* was developed in a relatively unsupervised fashion, free of any ground-truth expert delineation of the pulmonary vasculature itself. Despite the execution of multiple iterations based on use of non-bisected vs. bisected Frontal-view data and/or Lateral-view data, the non-bisected complete/full-resolution Frontal auto-segmentations alone provided the highest-yield input into the transformer.

PVP-“state” assigning and PVH-ranking

The *PVP Multi-Classifier* was designed to differentiate between image-data characteristics based on a more basic physiological classification, as follows:

- Normal.
- PVH Stage 1 (Vascular redistribution without edema +/- CpcPH).
- PVH “Stage 2+” (Vascular redistribution/congestion with predominantly interstitial edema +/- CpcPH).
- Group 1 PH.

The *PVPI* was ultimately used to process the Frontal views from same 1,682 CXR examinations to predict probabilities for these classes with either a 3-class output for PVH detection or a 4-class output for basic PVH-Ranking.

Correlating AI-based CXR PVH-ranking with DEcho LVDD-grading*Re-evaluation of unconfounded study population*

As with the HGT PVP assignments, *PVPI*-derived predictions of PVP “State” were used in determining the relationship between CXR-based PVH-Ranking and DEcho-based LVDD-Grading.

Trialing in confounded test group

The *PVPI* was purposely designed without representation of implanted cardiovascular materials commonly found in the setting of HF (e.g., surgical items) [Table 8]. Therefore, to preliminarily gauge *PVPI* robustness, it was trialed in a Confounded Test Group of previously excluded subjects, each demonstrating on CXR an example of implanted cardiovascular material while otherwise meeting the same Study Population inclusion criteria.

Based on HGT analysis of CXR examinations, the 40 subjects (23 males and 17 females) in the Confounded Test Group included 10 examples each of Normal, PVH Stage 1, PVH Stage 2+, and Group 1 PH PVPs. Simultaneously, 30 examples of DEcho Grade 1 ($N=13$), Grade 2 ($N=15$), or Grade 3–4 ($N=2$) LVDD, and 10 examples of RHCath-confirmed Group 1 PH, were represented.

For the assessment of overall *PVPI* performance in the Confounded Test Group, the following component evaluations were completed:

- *Thoracic-Content Segmentator*: The proportion of satisfactory Frontal-view auto-segmentations without versus with perceived-needed expert modification were determined. Overall Dice coefficients for all 40 segmentations, as well as for perceived deficient segmentations, were calculated.
- *PVP Multi-Classifier*: The accuracy of correct (i.e., highest probability) physiological assignment relative to HGT-assigned PVP was established. *PVP Multi-Classifier*-generated PVH predictions were again correlated with LVDD-Grading.

Evaluations, comparisons, and statistical analyses

No large language models were used in any portion of this research, including the preparation of this report. Statistical analyses were supervised internally (Author 3).

Using the ICC statistic with PVP-order ignored¹¹¹, both intra-Reviewer reliability (between first and second individual assignments) and inter-Expert Reviewer concordance (prior to HGT consensus) in the assignment of the original 11 possible PVPs by Reviewers 1–4 were evaluated. These included calculations of both overall agreement and agreement per PVP-characteristic; strength of agreement was indicated by the following ICC values: < 0.50 (Poor), 0.50–0.74 (Moderate), 0.75–0.89 (Good), or 0.90–1.00.90.00 (Excellent)¹¹¹.

The same evaluations of overall agreement were also performed with PVP-order considered using the *Ktau*¹¹². The *Ktau* expresses the following values for strength of agreement: 0.26–0.48 (Moderate), 0.49–0.70 (Strong) or 0.71–1.00.71.00 (Very Strong)¹¹³.

HGT assignments were used in evaluating the relationship between PVH-Staging and LVDD-Grading; for this purpose, *Ktau* methodology was again applied. In addition, to assess the impact of Reviewer-to-Reviewer PVH-Staging on this relationship, a Positive Likelihood Ratio [+ LR] and a Negative Predictive Value [-LR] were

calculated for Reviewers 1–4, as well as for HGT; an LR is the ratio of the probability of the specific test result in cases vs. the probability in cases without disease¹¹⁴, where LRs above 10 and below 0.10 are considered to provide strong evidence to rule in or rule out diagnoses, respectively¹¹⁵.

The overall performance of the *PVP Multi-Classifer* was expressed as both unbalanced and balanced average accuracies of the model^{116,117} with an output of 3 classes (i.e., Normal vs. combined PVH Stages 1 & 2 + vs. Group 1 PH) to emphasize PVH detection, as well as an output of 4 classes (i.e., Normal vs. PVH Stage 1 vs. PVH Stage 2 + vs. Group 1 PH) to concentrate on PVH-severity ranking. For both outputs, normalized accuracies^{117,118} were calculated to construct confusion matrices¹¹⁹.

PVP Multi-Classifer-derived PVH prediction (highest probability classification) and DEcho Grades were correlated using *Ktau* methodology. A comparison of this relationship with the corresponding relationship based on HGT assignments was then made.

A Chi-squared association test was used to confirm statistical significance [$p < 0.050$] of correspondence between described measures of CXR-based PVH and DEcho-based LVDD¹²⁰. On the other hand, a Fisher's exact test was used to gauge the statistical significance of differences between the proportions of categories in HGT-derived vs. *PVP Multi-Classifer*-derived measures of CXR-based PVH and DEcho-based LVDD¹²⁰.

Data availability

Data sets containing representative images and labels (re-sized and de-identified for downloading) used in this study are available in the supplementary zip file. While the full data sets are not publicly available due to institutional restrictions, researchers interested in accessing high-resolution images may contact the Center for Augmented Intelligence in Imaging of the Mayo Clinic Florida (mail to: erdal.barbaros@mayo.edu) for further information.

Received: 19 June 2025; Accepted: 25 September 2025

Published online: 31 October 2025

References

- Angeja, B. G. & Grossman, W. Evaluation and management of diastolic heart failure. *Circulation* **107**, 659–663 (2003).
- Aljaroudi, W. A., Thomas, J. D., Rodriguez, L. L. & Jaber, W. A. Prognostic value of diastolic dysfunction: state of the Art review. *Cardiol. Rev.* **22**, 79–90 (2014).
- Brucks, S. et al. Contribution of left ventricular diastolic dysfunction to heart failure regardless of ejection fraction. *Am. J. Cardiol.* **95**, 603–606 (2005).
- Fukuta, H. & Little, W. C. The cardiac cycle and the physiologic basis of left ventricular contraction, ejection, relaxation, and filling. *Heart Fail. Clin.* **4**, 1–11 (2008).
- Yancy, C. W. et al. 2013 ACCF/AHA guideline for the management of heart failure: A report of the American college of cardiology Foundation/American heart association task force on practice guidelines. *J. Am. Coll. Cardiol.* **62**, e147–239 (2013).
- Gaasch, W. H. & Zile, M. R. Left ventricular diastolic dysfunction and diastolic heart failure. *Annu. Rev. Med.* **55**, 373–394 (2004).
- Wan, S. H., Vogel, M. W. & Chen, H. H. Pre-clinical diastolic dysfunction. *J. Am. Coll. Cardiol.* **63**, 407–416 (2014).
- Naylor, M. et al. Left ventricular diastolic dysfunction in the community: impact of diagnostic criteria on the burden, correlates, and prognosis. *J. Am. Heart Assoc.* **7**, e008291 (2018).
- Redfield, M. M. et al. Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic. *JAMA* **289**, 194–202 (2003).
- Kosmala, W. & Marwick, T. H. Asymptomatic left ventricular diastolic dysfunction: predicting progression to symptomatic heart failure. *JACC Cardiovasc. Imaging.* **13**, 215–227 (2020).
- Zile, M. R. & Brutsaert, D. L. New concepts in diastolic dysfunction and diastolic heart failure: part I: diagnosis, prognosis, and measurements of diastolic function. *Circulation* **105**, 1387–1393 (2002).
- Borlaug, B. A., Sharma, K., Shah, S. J. & Ho, J. E. Heart failure with preserved ejection fraction: JACC scientific statement. *JACC* **81**, 1810–1834 (2023).
- Young, K. A., Scott, C. G., Rodeheffer, R. J. & Chen, H. H. Progression of preclinical heart failure: A description of stage A and B heart failure in a community population. *Circ. Cardiovasc. Qual. Outcomes.* **14**, 622–632 (2021).
- Young, K. A. et al. Association of impaired relaxation mitral inflow pattern (grade 1 diastolic function) with long-term noncardiovascular and cardiovascular mortality. *J. Am. Soc. Echocardiogr.* **38**, 367–377 (2025).
- Kane, G. C. et al. Progression of left ventricular diastolic dysfunction and risk of heart failure. *JAMA* **306**, 856–863 (2011).
- Bello, H. et al. Hemodynamic determinants of age versus left ventricular diastolic function relations across the full adult age range. *Hypertension* **75**, 1574–1583 (2020).
- Okura, H. et al. Age- and gender-specific changes in the left ventricular relaxation: A doppler echocardiographic study in healthy individuals. *Circ. Cardiovasc. Imaging.* **2**, 41–46 (2009).
- Klein, A. L. et al. Effects of age on left ventricular dimensions and filling dynamics in 117 normal persons. *Mayo Clin. Proc.* **69**, 212–224 (1994).
- Lam, C. S. et al. Cardiac dysfunction and noncardiac dysfunction as precursors of heart failure with reduced and preserved ejection fraction in the community. *Circulation* **124**, 24–30 (2011).
- Vogel, M. W., Slusser, J. P., Hodge, D. O. & Chen, H. H. The natural history of preclinical diastolic dysfunction: A population-based study. *Circ. Heart Fail.* **5**, 144–151 (2012).
- Pieske, B. et al. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm: A consensus recommendation from the heart failure association (HFA) of the European society of cardiology (ESC). *Eur. Heart J.* **40**, 3297–3317 (2019).
- Omote, K. et al. Central haemodynamic abnormalities and outcome in patients with unexplained dyspnoea. *Eur. J. Heart Fail.* **25**, 185–196 (2023).
- Reddy, Y. N. V., El-Sabbagh, A. & Nishimura, R. A. Comparing pulmonary arterial wedge pressure and left ventricular end diastolic pressure for assessment of left-sided filling pressures. *JAMA Cardiol.* **3**, 453–454 (2018).
- Reddy, Y. N. V., Carter, R. E., Obokata, M., Redfield, M. M. & Borlaug, B. A. A simple, evidence-based approach to help guide diagnosis of heart failure with preserved ejection fraction. *Circulation* **138**, 861–870 (2018).
- Smiseth, O. A. et al. Multimodality imaging in patients with heart failure and preserved ejection fraction: an expert consensus document of the European association of cardiovascular imaging. *Eur. Heart J. Cardiovasc. Imaging.* **23**, e34–e61 (2022).

26. Nagueh, S. F. et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American society of echocardiography and the European association of cardiovascular imaging. *J. Am. Soc. Echocardiogr.* **29**, 277–314 (2016).
27. Humbert, M. et al. 2022 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur. Heart J.* **43**, 3618–3731 (2022).
28. Obokata, M., Reddy, Y. N. V. & Borlaug, B. A. Diastolic dysfunction and heart failure with preserved ejection fraction: Understanding mechanisms by using noninvasive methods. *JACC Cardiovasc. Imaging.* **13**, 245–257 (2020).
29. Lundqvist, C. B., Olsson, S. B. & Varnauskas, E. Transseptal left heart catheterization: A review of 278 studies. *Clin. Cardiol.* **9**, 21–26 (1986).
30. Connolly, D. C., Kirklin, J. W. & Wood, E. H. The relationship between pulmonary artery wedge pressure and left atrial pressure in man. *Circ. Res.* **2**, 434–440 (1954).
31. Thomas, L., Marwick, T. H., Popescu, B. A., Donal, E. & Badano, L. P. Left atrial structure and function, and left ventricular diastolic dysfunction: JACC state-of-the-art review. *JACC* **73**, 1961–1977 (2019).
32. Terlizzi, V. D. et al. The atrioventricular coupling in heart failure: pathophysiological and therapeutic aspects. *Rev. Cardiovasc. Med.* **25**, 169–181 (2024).
33. Nishimura, R. A. & Carabello, B. A. Hemodynamics in the cardiac catheterization laboratory of the 21st century. *Circulation* **125**, 2138–2150 (2012).
34. Hemnes, A. R. et al. Features associated with discordance between pulmonary arterial wedge pressure and left ventricular end diastolic pressure in clinical practice: implications for pulmonary hypertension classification. *Chest* **154**, 1099–1107 (2018).
35. Heidenreich, P. A. et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: A report of the American college of Cardiology/American heart association joint committee on clinical practice guidelines. *Circulation* **145**, e895–e1032 (2022).
36. Bozkurt, B. et al. Universal definition and classification of heart failure: A report of the heart failure society of America, heart failure association of the European society of Cardiology, Japanese heart failure society and writing committee of the universal definition of heart failure: endorsed by the Canadian heart failure society, heart failure association of India, cardiac society of Australia and new Zealand, and Chinese heart failure association. *Eur. J. Heart Fail.* **23**, 352–380 (2021).
37. Patel, M. R. et al. 2013 ACCF/ACR/ASE/ASNC/SCCT/SCMR appropriate utilization of cardiovascular imaging in heart failure: A joint report of the American college of radiology appropriateness criteria committee and the American college of cardiology foundation appropriate use criteria task force. *JACC* **61**, 2207–2231 (2013).
38. McDonagh, T. A. et al. 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur. Heart J.* **42**, 3599–3726 (2021).
39. Elliott, L. P. Pulmonary vascularity on the plain chest radiograph. *Cardiol. Clin.* **1**, 545–564 (1983).
40. Milne, E. N. Physiological interpretation of the plain radiograph in mitral stenosis, including a review of criteria for the radiological Estimation of pulmonary arterial and venous pressures. *Br. J. Radiol.* **36**, 902–913 (1963).
41. Turner, A. F., Lau, F. Y. & Jacobson, G. A method for the Estimation of pulmonary venous and arterial pressures from the routine chest roentgenogram. *Am. J. Roentgenol. Radium Ther. Nucl. Med.* **116**, 97–106 (1972).
42. Kostuk, W., Barr, J. W., Simon, A. L. & Ross, J. Jr. Correlations between the chest film and hemodynamics in acute myocardial infarction. *Circulation* **48**, 624–632 (1973).
43. Baumstark, A. et al. Evaluating the radiographic assessment of pulmonary venous hypertension in chronic heart disease. *Am. J. Roentgenol.* **142**, 877–884 (1984).
44. Herman, P. G. et al. Limited correlation of left ventricular end-diastolic pressure with radiographic assessment of pulmonary hemodynamics. *Radiology* **174**, 721–724 (1990).
45. Chakko, S. et al. Clinical, radiographic, and hemodynamic correlations in chronic congestive heart failure: conflicting results May lead to inappropriate care. *Am. J. Med.* **90**, 353–359 (1991).
46. Sharma, S., Bhargava, A., Krishnakumar, R. & Rajani, M. Can pulmonary venous hypertension be graded by the chest radiograph? *Clin. Radiol.* **53**, 899–902 (1998).
47. Dash, H., Lipton, M. J., Chatterjee, K. & Parmley, W. W. Estimation of pulmonary artery wedge pressure from chest radiograph in patients with chronic congestive cardiomyopathy and ischaemic cardiomyopathy. *Br. Heart J.* **44**, 322–329 (1980).
48. Costanzo, W. E. & Fein, S. A. The role of the chest X-ray in the evaluation of chronic severe heart failure: things are not always as they appear. *Clin. Cardiol.* **11**, 486–488 (1988).
49. Jögi, J., Al-Mashat, M., Rådegran, G., Bajc, M. & Arheden, H. Diagnosing and grading heart failure with tomographic perfusion lung scintigraphy: validation with right heart catheterization. *ESC Heart Fail.* **5**, 902–910 (2018).
50. Studler, U. et al. Accuracy of chest radiographs in the emergency diagnosis of heart failure. *Eur. Radiol.* **18**, 1644–1652 (2008).
51. Nørgaard, H., Gjørup, T., Brems-Dalgaard, E., Hartelius, H. & Brun, B. Interobserver variation in the detection of pulmonary venous hypertension in chest radiographs. *Eur. J. Radiol.* **11**, 203–206 (1990).
52. Henriksson, L., Sundin, A., Smedby, O. & Albrektsson, P. Assessment of congestive heart failure in chest radiographs. Observer performance with two common film-screen systems. *Acta Radiol.* **31**, 469–471 (1990).
53. Badgett, R. G., Mulrow, C. D., Otto, P. M. & Ramirez, G. How well can the chest radiograph diagnose left ventricular dysfunction? *J. Gen. Intern. Med.* **11**, 625–634 (1996).
54. Borlaug, B. A. & Redfield, M. M. Diastolic and systolic heart failure are distinct phenotypes within the heart failure spectrum. *Circulation* **123**, 2006–2013 (2011).
55. Fonseca, C. et al. The value of the electrocardiogram and chest X-ray for confirming or refuting a suspected diagnosis of heart failure in the community. *Eur. J. Heart Fail.* **6**, 807–812 (2004).
56. Oudiz, R. J. Pulmonary hypertension associated with left-sided heart disease. *Clin. Chest Med.* **28**, 233–241 (2007).
57. Milne, E. N. Forgotten gold in diagnosing pulmonary hypertension: the plain chest radiograph. *Radiographics* **32**, 1085–1087 (2012).
58. Vucic, E., Chakhtoura, E. Y., Sohal, S. & Waxman, S. Pathophysiological concepts of constrictive pericarditis in cardiac imaging: back to basics. *Circ. Cardiovasc. Imaging.* **14**, e012136 (2021).
59. Grodecki, P. V. & Klein, A. L. Pitfalls in the echo-Doppler assessment of diastolic dysfunction. *Echocardiography* **10**, 213–234 (1993).
60. Xiao, H. B., Lee, C. H. & Gibson, D. G. Effect of left bundle branch block on diastolic function in dilated cardiomyopathy. *Br. Heart J.* **66**, 443–447 (1991).
61. Dickinson, M. G. et al. Atrial fibrillation modifies the association between pulmonary artery wedge pressure and left ventricular end-diastolic pressure. *Eur. J. Heart Fail.* **19**, 1483–1490 (2017).
62. Milne, E. N. Correlation of physiologic findings with chest roentgenology. *Radiol. Clin. North. Am.* **11**, 17–47 (1973).
63. Kennedy, S., Simon, B., Alter, H. J. & Cheung, P. Ability of physicians to diagnose congestive heart failure based on chest X-ray. *J. Emerg. Med.* **40**, 47–52 (2011).
64. Feldmann, E. J., Jain, V. R., Rakoff, S. & Haramati, L. B. Radiology residents' on-call interpretation of chest radiographs for congestive heart failure. *Acad. Radiol.* **14**, 1264–1270 (2007).
65. Przewlocka-Kosmala, M., Butler, J., Donal, E., Ponikowski, P. & Kosmala, W. Prognostic value of the MAGGIC Score, H₂FPEF Score, and HFA-PEFF algorithm in patients with exertional dyspnea and the incremental value of exercise echocardiography. *J. Am. Soc. Echocardiogr.* **35**, 966–975 (2022).

66. van de Bovenkamp, A. A. et al. Validation of the 2016 ASE/EACVI guideline for diastolic dysfunction in patients with unexplained dyspnea and a preserved left ventricular ejection fraction. *J. Am. Heart Assoc.* **10**, e021165 (2021).
67. Lavista Ferres, J. M., Fishman, E. K., Rowe, S. P. & Chu, L. C. Lugo-Fagundo, E. Artificial intelligence as a public service. *JACR* **20**, 919–921 (2023).
68. Farina, J. M. et al. Artificial intelligence-based prediction of cardiovascular diseases from chest radiography. *J. Imaging* **9**, 236–247 (2023).
69. Seah, J. C. Y., Tang, J. S. N., Kitchen, A., Gaillard, F. & Dixon, A. F. Chest radiographs in congestive heart failure: visualizing neural network learning. *Radiology* **290**, 514–522 (2019).
70. Hirata, Y. et al. Deep learning for detection of elevated pulmonary artery wedge pressure using standard chest x-ray. *Can. J. Cardiol.* **37**, 1198–1206 (2021).
71. Bozkurt, B. et al. HF STATS 2024: heart failure epidemiology and outcomes statistics - An updated 2024 report from the heart failure society of America. *J. Cardiac Fail.* **31**, 66–116 (2025).
72. Beladan, C. C., Botezatu, S. & Popescu, B. A. Reversible left ventricular diastolic dysfunction—Overview and clinical implications. *Echocardiography* **37**, 1957–1966 (2020).
73. Xu, B. & Klein, A. L. Utility of echocardiography in heart failure with preserved ejection fraction. *J. Card Fail.* **24**, 397–403 (2018).
74. Galie, N. et al. 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension: the joint task force for the diagnosis and treatment of pulmonary hypertension of the European society of cardiology (ESC) and the European respiratory society (ERS). *Eur. Respir. J.* **46**, 903–975 (2015).
75. Hassoun, P. M. Pulmonary arterial hypertension. *N Engl. J. Med.* **385**, 2361–2376 (2021).
76. Downey, R., White, R. D. & Krasuski, R. A. Chest (ed.) Chest radiography: What the cardiologist needs to know. Chapter 36, 403–420. In: Griffin, B.P., Rimmerman, C.M., E.J., *The Cleveland Clinic Cardiology Board Review*. (Philadelphia: Lippincott Williams & Wilkins), (2007).
77. Cardinale, L., Priola, A. M., Moretti, F. & Volpicelli, G. Effectiveness of chest radiography, lung ultrasound and thoracic computed tomography in the diagnosis of congestive heart failure. *World J. Radiol.* **6**, 230–237 (2014).
78. Weil, B. R., Techiryan, G., Suzuki, G., Konecny, F. & Canty, J. M. Jr. Adaptive reductions in left ventricular diastolic compliance protect the heart from stretch-induced stunning. *JACC Basic. Transl. Sci.* **4**, 527–541 (2019).
79. Fayyaz, A. U. et al. Global pulmonary vascular remodeling in pulmonary hypertension associated with heart failure and preserved or reduced ejection fraction. *Circulation* **137**, 1796–1810 (2018).
80. Rosenberg, M. A. & Manning, W. J. Diastolic dysfunction and risk of atrial fibrillation: A mechanistic appraisal. *Circulation* **126**, 2353–2362 (2012).
81. Gong, F. F., Campbell, D. J. & Prior, D. L. Noninvasive cardiac imaging and the prediction of heart failure progression in preclinical stage A/B subjects. *JACC Cardiovasc. Imaging* **10**, 1504–1519 (2017).
82. Xie, G. Y. & Smith, M. D. Pseudonormal or intermediate pattern? *JACC* **39**, 1796–1798 (2002).
83. Grewal, J., McCully, R. B., Kane, G. C., Lam, C. & Pellikka, P. A. Left ventricular function and exercise capacity. *JAMA* **301**, 286–294 (2009).
84. Chen, Z. et al. Exploring explainable AI features in the vocal biomarkers of lung disease. *Comput. Biol. Med.* **179**, 108844. <https://doi.org/10.1016/j.compbiomed.2024.108844> (2024).
85. Seemann, F. et al. Imaging gravity-induced lung water redistribution with automated inline processing at 0.55 T cardiovascular magnetic resonance. *J. Cardiovasc. Magn. Reson.* **24**, 35–47 (2022).
86. James, A. E. Jr., Cooper, M., White, R. I. & Wagner, H. N. Jr. Perfusion changes on lung scans in patients with congestive heart failure. *Radiology* **100**, 99–106 (1971).
87. Motoji, Y. et al. Interdependence of right ventricular systolic function and left ventricular filling and its association with outcome for patients with pulmonary hypertension. *Int. J. Cardiovasc. Imaging* **31**, 691–698 (2015).
88. Winkler, T. et al. Perfusion imaging heterogeneity during NO inhalation distinguishes pulmonary arterial hypertension (PAH) from healthy subjects and has potential as an imaging biomarker. *Respir. Res.* **23**, 325–340 (2022).
89. Lim, K., McGregor, G., Coggan, A. R., Lewis, G. D. & Moe, S. M. Cardiovascular functional changes in chronic kidney disease: integrative physiology, pathophysiology and applications of cardiopulmonary exercise testing. *Front. Physiol.* **11**, 572355, 1–14 (2020).
90. Loutradis, C., Sarafidis, P. A., Papadopoulos, C. E., Papagianni, A. & Zoccali, C. The ebb and flow of echocardiographic cardiac function parameters in relationship to Hemodialysis treatment in patients with ESRD. *J. Am. Soc. Nephrol.* **29**, 1372–1381 (2018).
91. Adeli, E. et al. Representation learning with statistical independence to mitigate bias. *IEEE Winter Conf Appl Comput Vis.* 2512–2522 (2021). (2021).
92. Zhao, Q., Adeli, E. & Pohl, K. M. Training confounder-free deep learning models for medical applications. *Nat. Commun.* **11**, 6010, 1–9 (2020).
93. DeGrave, A. J., Janizek, J. D. & Lee, S. I. AI for radiographic COVID-19 detection selects shortcuts over signal. *MedRxiv Preprint Published with.* <https://doi.org/10.1038/s42256-021-0018-7> (2020). PMID: 32995822; PMCID: PMC7523163.
94. Galvin, I., Drummond, G. B. & Nirmalan, M. Distribution of blood flow and ventilation in the lung: gravity is not the only factor. *Br. J. Anaesth.* **98**, 420–428 (2007).
95. Andersen, O. S. et al. Estimating left ventricular filling pressure by echocardiography. *JACC* **69**, 1937–1948 (2017).
96. Lee, E. et al. Artificial intelligence-enabled ECG for left ventricular diastolic function and filling pressure. *NPJ Digit. Med.* **7** <https://doi.org/10.1038/s41746-023-00993-7> (2024). PMID: 38182738; PMCID: PMC10770308.
97. Rao, V. M. et al. Multimodal generative AI for medical image interpretation. *Nature* **639**, 888–896 (2025).
98. Theodorakis, N., Nikolaou, M. & Krentz, A. cardiovascular-endocrine-metabolic medicine: proposing a new clinical sub-specialty amid the cardiometabolic pandemic. *Biomolecules* **15**, 373–399 (2025).
99. Abbas, A. E. et al. A simple method for noninvasive Estimation of pulmonary vascular resistance. *JACC* **41**, 1021–1027 (2003).
100. Demirer, M. et al. A user interface for optimizing radiologist engagement in image data curation for artificial intelligence. *Radiol. Artif. Intell.* **1** (e180095), 1–7 (2019).
101. White, R. D. et al. Pre-deployment assessment of an AI model to assist radiologists in chest X-ray detection and identification of lead-less implanted electronic devices for pre-MRI safety screening: realized implementation needs and proposed operational solutions. *J. Med. Imaging (Bellingham)* **9** (054504), 1–34 (2022).
102. Restrepo, C. S. et al. The diaphragmatic Crura and retrocrural space: normal imaging appearance, variants, and pathologic conditions. *Radiographics* **28**, 1289–1305 (2008).
103. Chen, L. C., Papandreou, G., Schroff, F. & Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv*. (2017). <https://arxiv.org/abs/1706.05587>
104. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. (2015). <https://arxiv.org/abs/1512.03385> (2015).
105. Zou, K. H. et al. Statistical validation of image segmentation quality based on a Spatial overlap index. *Acad. Radiol.* **11**, 178–189 (2004).
106. Wilson, S. M., Bautista, A., Yen, M., Lauderdale, S. & Eriksson, D. K. Validity and reliability of four Language mapping paradigms. *Neuroimage Clin.* **16**, 399–408 (2016).
107. Caron, M. et al. Emerging properties in self-supervised vision Transformers. *ArXiv* <https://doi.org/10.48550/ArXiv.2104.14294> (2021).

108. Johnson, A., Pollard, T., Mark, R., Berkowitz, S. & Horng, S. MIMIC-CXR database. *mimic.mit.* (2021). <https://mimic.mit.edu/docs/iv/modules/cxr/>
109. Fernando Pérez-García, F. et al. RAD-DINO: Exploring scalable medical image encoders beyond text supervision. *arXiv.* (2024). <https://doi.org/10.48550/arXiv.2401.10815>
110. Bradshaw, T. J., Huemann, Z., Hu, J. & Rahmim, A. A guide to cross-validation for artificial intelligence in medical imaging. *Radiol. Artif. Intell.* **5**, e220232 (2023).
111. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
112. Kendall, M. G. A new measure of rank correlation. *Biometrika* **30**, 81–93. <https://doi.org/10.2307/2332226> (1938). JSTOR.
113. Wicklin, R. Weak or strong? How to interpret a Spearman or Kendall correlation. *SAS.* (2023). <https://blogs.sas.com/content/iml/2023/04/05/interpret-spearman-kendall-corr.html>
114. Sackett, D. L., Straus, S., Richardson, W. S., Rosenberg, W. & Haynes, R. B. (eds) *Evidence-Based Medicine. How To Practise and Teach EBM* 2nd edn 67–93 (Churchill Livingstone), 2000).
115. Jaeschke, R., Guyatt, G., Lijmer, J. & Diagnostic tests., 121–140. In: Guyatt, G. & Rennie, D., eds. *Users' Guides to the Medical Literature.* (Chicago: AMA Press), (2002).
116. Brodersen, K. H., Ong, C. S., Stephan, K. E. & Buhmann, J. M. The balanced accuracy and its posterior distribution. *IEEE Xplore. 20th International Conference on Pattern Recognition.* (2010). <https://ieeexplore.ieee.org/abstract/document/5597285> (2010).
117. Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
118. Meyer-Baese, A. & Schmid, V. J. *Pattern Recognition and Signal Analysis in Medical Imaging* (Elsevier, 2014).
119. Erickson, B. J. & Kitamura, F. Magician's corner: 9. Performance metrics for machine learning models. *Radiol. Artif. Intell.* **3** (e200126), 1–7 (2021).
120. Kim, H. Y. Statistical notes for clinical researchers: Chi-squared test and fisher's exact test. *Restor. Dent. Endod.* **42**, 152–155 (2017).

Author contributions

All authors read and contributed to the creation of this manuscript. Individual contributions follow: RW: Project leader (e.g., idea development, study execution, results analysis, manuscript writing)/Primary image reviewer-MD: Data manager/Image-review interface developer/AI model processor/Technical text for manuscriptRS: Supervisor and performer of statistical analyses/Statistical text for manuscriptIC: Independent and Ground-Truth image reviewer/Manuscript reviewerJS: Independent early-career image reviewer/Manuscript reviewerMM: Independent trainee image reviewer/Manuscript reviewerTB: Set standards for Echo/LVDD subject ID/Confirmed Echo Results per subject/Echo text for manuscriptCA: Set standards for Echo/LVDD subject ID/Confirmed Echo Results per subject/Echo text for manuscriptSH: Set standards for Right Heart Cath/PH subject ID/Confirmed Right Heart Cath results per subject/PH text for manuscript BE: Primary Radiology-Informatics-AI Lead/Data mining/AI model training supervision/Technical text for manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-22026-x>.

Correspondence and requests for materials should be addressed to R.D.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025