npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-02011-4

SPEED-TR: a self-distilled and pre-trained transformer model for enhanced ECG detection of tricuspid regurgitation

Xiaolin Diao[1,5], Wei Xu[2,3,5], Huaibing Cheng[3,5], Ya Zhou[1,5], Yang Liu[1], Yanni Huo[1], Jianli Lu[1], Jinghan Huang[3], Jia He[3], Fang Liu[3], Zhihui Cao[1], Xue Zhang[1], Wei Zhao[4] ✉ & Xiaohan Fan[2,3] ✉

Tricuspid regurgitation (TR) remains underdiagnosed due to the lack of effective screening tools. We developed a self-distilled and pre-trained transformer model for detecting TR (SPEED-TR) from electrocardiography. The model was trained using 466,149 electrocardiogram-echocardiogram pairs from 291,673 patients and validated in one internal (63,925 patients) and two external cohorts (44,951 and 21,300 patients). SPEED-TR accurately detected moderate-to-severe TR in the hold-out set (AUROC 0.945, NPV 0.983; specificity 0.973) and maintained stable performance in multi-center testing sets (AUROCs 0.939–0.943; NPVs 0.978–0.988). Three thresholds enabled SPEED-TR severity grading: none (0–0.008), mild (0.008–0.255), moderate (0.255–0.755), and severe (0.755–1), achieving accuracies of 0.749 (hold-out), 0.730 (internal), 0.775 and 0.726 (external), with overall accuracy of 0.744. SPEED-TR remained robust in patients with 1 to ≥3 risk factors and 1 to ≥2 valvular diseases. SPEED-TR demonstrated potential as a screening tool and may provide reference for TR severity assessment.

Tricuspid regurgitation (TR) represents one of the most common forms of valvular heart diseases (VHDs), with a tendency for prevalence to increase with advancing age[1]. Prior investigations have reported that the incidence of asymptomatic moderate-to-severe TR patients ranges from approximately 2.7% to 13.8%[2,3]. Severe TR may lead to increased right ventricular load, decreased cardiac function, and even heart failure[4]. These patients are typically diagnosed and treated only after presenting with significant symptoms of right heart failure[5,6]. However, due to severe systemic end-organ involvement, pharmacological and surgical interventions are poorly effective during this period, with all-cause mortality rates ranging from 7.4% to 69% in the following years[7–9]. The advancement of surgical techniques has led to the observation that early surgical intervention is associated with lower rates of comorbidity morbidity, operative mortality, and long-term mortality[10–12]. Additionally, with the rapid development of transcatheter valve intervention in recent years, transcatheter tricuspid valve intervention (TTVI) has also been developed with great momentum. The optimal timing of TTVI is still inconclusive, but some studies pointed out that TR should be treated at early stage[13]. Thus, timely detection and intervention are likely to be more beneficial for the long-term prognosis of patients with TR. It is imperative to develop a practical and economical screening tool for the early detection of TR.

The traditional diagnosis of TR primarily relies on echocardiography (ECHO), computed tomography (CT), magnetic resonance imaging (MRI), and central venous pressure, which are costly and require professional technicians, contrast agents, and invasive procedures[14]. In contrast, electrocardiography (ECG) offers several advantages, including simple operation, real-time monitoring, wide availability, non-invasiveness, and low cost, making it an ideal tool for large-scale screening of high-risk populations. However, the ECG features may not be specific in VHDs, such as the ECG may show modifications in the P and QRS waveforms, right ventricular hypertrophy, right bundle branch block (RBBB), or atrial fibrillation (AF) in patients with TR[15]. It is difficult for clinicians to diagnose TR by using ECGs, due to the lack of specificity of ECG findings in TR. In recent decades, artificial intelligence (AI) technology is gradually being integrated into the diagnosis of heart diseases[16]. Among them, the application of AI in ECG analysis has received increasing attention[17]. Studies have demonstrated the

[1]Department of Information Center, Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [2]Department of Cardiology, Cardiac Arrhythmia Center, Fuwai Hospital, National Center for Cardiovascular Diseases, State Key Laboratory of Cardiovascular Disease, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [3]Function Test Center, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [4]Center for Health Statistics and Information, National Health Commission People's Republic of China, Beijing, China. [5]These authors contributed equally: Xiaolin Diao, Wei Xu, Huaibing Cheng, Ya Zhou.
✉e-mail: zhaowei@nhc.gov.cn; fanxiaohan@fuwaihospital.org

efficacy of deep learning models of ECGs in the diagnosis of AF[18], left ventricular systolic dysfunction[19], aortic stenosis (AS)[20], and mitral regurgitation (MR)[21]. Furthermore, AI technology may assist in identifying certain subtle features on ECGs that are challenging for humans to detect[22]. However, few studies have been conducted to construct AI-based ECG models for diagnosing TR.

The objective of this study was to evaluate the efficacy of a deep learning model for the detection of TR using a standard 12-lead ECG. We developed Self-Distilled and Pre-trained Transformer model for Enhanced ECG Detection of Tricuspid Regurgitation (SPEED-TR), a Transformer-based model that integrates self-distillation and self-supervised pre-training to enhance recognition of subtle ECG features and TR detection. The model was trained on a large ECG-ECHO paired dataset from Fuwai (FW) Hospital, the national center for cardiovascular diseases of China. The model was evaluated by using multi-center datasets, one internal (data from FW hospital) and two external testing sets [data from Yunnan Fuwai (YF) Hospital and Shenzhen Fuwai (SF) Hospital]. Model performance across multiple dimensions was assessed, including under different TR degrees, or specific subgroups with different TR risk factors or other VHDs.

Results

Population characteristics

The baseline characteristics of the model development datasets (training, validation, and hold-out testing sets) were presented in Table 1. Overall, patients with TR in the development datasets were older, had a higher proportion of female patients, and were more likely to have other VHDs. In terms of ECG characteristics, TR patients had faster heart rates, longer PR intervals and corrected QT intervals (QTc), wider QRS intervals, and higher rates of AF/atrial flutter (AFL), ventricular premature complexes, right bundle branch block (RBBB), intraventricular conduction delay, right or left axis deviation, and low voltage in limb leads. Regarding ECHO measurements, individuals with TR exhibited reduced left ventricular ejection fraction (LVEF) compared with patients without TR.

The baseline characteristics of multi-center evaluation datasets were shown in Table 2. The average age (56.5 ± 13.7, 56.2 ± 14.1, and 50.8 ± 24.6) of patients in three centers were similar to that of model development datasets (55.4 ± 14.2, 54.2 ± 14.2, and 54.4 ± 14.2). Some of the ECG and ECHO parameters were significantly different between patients with and without TR due to the large sample size.

Model performance across multi-center datasets

Figure 1 showed the receiver operating characteristic curves and precision-recall curves of the SPEED-TR model across all model development datasets and multi-center evaluation datasets. Table 3 summarized the evaluation metrics of the model in the hold-out testing set, FW testing set and the two external testing sets. In the hold-out testing set (Fig. 1C), the SPEED-TR achieved an area under the receiver operating characteristic curve (AUROC) of 0.945 [95% confidence interval (CI), 0.939–0.951] and an area under the precision-recall curve (AUPRC) of 0.519 (95% CI, 0.487–0.552). In addition, the model performed excellently with specificity of 0.973, sensitivity of 0.568, positive predictive value (PPV) of 0.448 and negative predictive value (NPV) of 0.983 (Table 3). Further analyses results demonstrated the effectiveness of various model components (Supplementary Table 1). Comparison of the performance among logistic regression (LR) models and SPEED-TR was reported in Supplementary Table 2, presenting a statistically significant difference ($P < 0.05$). The results showed that the SPEED-TR model outperformed the traditional LR models constructed based on baseline characteristics (LR Model 1) or risk factors (LR Model 2).

The performance of the SPEED-TR was also evaluated across multi-center validation testing datasets: FW internal testing set, and two external testing sets (YF and SF testing sets). In the FW testing set (Fig. 1D), the SPEED-TR achieved an AUROC of 0.939 (95% CI, 0.935–0.943) and an AUPRC of 0.550 (95% CI, 0.531–0.571), with specificity of 0.977, sensitivity of 0.543. In the YF and SF testing sets (Fig. 1E, F), the SPEED-TR model achieved AUROCs of 0.943 and 0.937, AUPRCs of 0.496 and 0.372, respectively. The SPEED-TR yielded specificities of 0.976 and 0.971, respectively. The PPVs varied across the three testing sets, with the highest value of 0.524 in the FW testing set and the lowest of 0.339 in the SF testing set. The NPVs were consistently high across all the testing sets (0.978 FW, 0.986 YF, 0.988 SF). The high NPV indicated the strong ability of the SPEED-TR model to correctly identify negative cases. Overall, the model achieved an accuracy ranging from 0.957 to 0.963, demonstrating a high ability of distinguishing between TR (moderate or severe) and non-TR (none or mild) across all testing sets. Calibration curves demonstrated good probability estimation, with brier scores varying from 0.021 to 0.025 across all the datasets (Supplementary Fig. 1).

TR degree Grading Based on SPEED-TR

We evaluated the ability of the SPEED-TR model for grading TR degree (none, mild, moderate, severe) based on the trained binary classification model for identifying TR. The prevalence of TR in different severity grades (none, mild, moderate, and severe) across all datasets is provided in Supplementary Table 3. Figure 2 illustrated the predicted probability distribution of the SPEED-TR model at different severity groups of TR on the hold-out testing set, FW, YF, SF testing set, and the overall multi-center testing sets (Fig. 2A–E). We used the boxplots and scatterplots in each subplot to demonstrate the predicted probabilities of the model for different TR degree cases. The prediction probability of the model was a decimal number between 0 and 1, where a value closer to 0 meant the model was more likely to judge the case as negative (with none or mild TR) and closer to 1 meant the model was more likely to judge the case as positive (with moderate to severe TR). The results demonstrated that the SPEED-TR consistently assigned higher predicted probabilities to moderate and severe TR cases across all testing sets. In cases with severe TR subgroups among all testing sets, the SPEED-TR exhibited the highest median predicted probabilities within 0.60–0.80.

As shown in Fig. 2F, three different probability thresholds were identified by maximizing the F1 scores on the validation set, using different ECHO TR positive outcome definitions (i.e., defining at least mild, at least moderate, or at least severe as positive). These thresholds were then used to classify TR degree severity as: none (0 to 0.008), mild (0.008 to 0.255), moderate (0.255 to 0.755), and severe (0.755 to 1). The accuracy of model's grading TR degree was also evaluated in the hold-out testing set, FW, YF, SF testing set, and the overall multi-center testing sets, achieving 0.749, 0.730, 0.775, 0.726 and 0.744, respectively.

Subgroup analyses

The SPEED-TR model achieved AUROC over 0.764 in the hold-out testing set (Supplementary Fig. 2A) and over 0.776 across all subgroups with different risk factors in the overall multi-center testing sets [0.837 MR, 0.898 LVEF < 50%, 0.925 female, 0.924 age≥60 years, 0.790 AF/AFL, and 0.776 pulmonary hypertension (PH)] (Fig. 3A). A modest reduction in AUROC was observed in patients with PH (0.776) and AF/AFL (0.790). Additionally, the model's performance maintained good discriminatory ability even in individuals combined with different numbers of risk factors, with the AUROC of 0.923 (zero), 0.903 (one), 0.883 (two), 0.784 (three) and 0.759 (over three risk factors), respectively (Fig. 3A). The performance of SPEED-TR model was also evaluated in subgroups based on other VHDs, including AS, aortic regurgitation (AR), mitral stenosis (MS), and MR in the hold-out testing set (Supplementary Fig. 2B) and the overall population of multi-center testing datasets (Fig. 3B, Supplementary Tables 6 and 7). The AUROC of SPEED-TR model were consistently stable in both hold-out testing set and the overall multi-center testing sets. Among patients with a single type of VHD, the AUROC was 0.929 for AS, 0.891 for AR, 0.837 for MR, and 0.783 for MS, with a reduction observed in the MS subgroup. When individuals were classified into subgroups by combining with zero, one, two, or over two VHDs, the AUROC of the model remained over 0.815, respectively, suggesting the model's ability to maintain performance with increasing valvular disease burden. The AUROCs in subgroups with different risk factors or VHDs in the hold-out testing set were consistent with that in the overall multi-center testing sets. Detailed results for other

Table 1 | Baseline characteristics of the model development datasets

Variables	Training Set				Validation Set				Hold-out Testing Set			
	Non-TR	TR	Total	*P* value	Non-TR	TR	Total Cases	*P* value	Non-TR	TR	Total	*P* value
Patients	224309	9030	233339	/	28057	1110	29167	/	28104	1063	29167	/
ECG-ECHO pairs	357722	14754	372476	/	28057	1110	29167	/	28104	1063	29167	/
Age at ECG, years	55.2 ± 14.1	58.3 ± 16.0	55.4 ± 14.2	<0.001	54.1 ± 14.1	56.8 ± 16.1	54.2 ± 14.2	<0.001	54.3 ± 14.1	56.9 ± 15.5	54.4 ± 14.2	<0.001
Age≥60 years (%)	152026 (42.50)	7780 (52.73)	159806 (42.90)	<0.001	11026 (39.30)	538 (48.47)	11564 (39.65)	<0.001	11129 (39.60)	525 (49.39)	11654 (39.96)	<0.001
Female at ECG (%)	140093 (39.16)	8419 (57.06)	148512 (39.87)	<0.001	11961 (42.63)	630 (56.76)	12591(43.17)	<0.001	12134 (43.18)	630 (59.27)	12764 (43.76)	<0.001
ECG characteristics												
Heart Rate, bpm	74.0 ± 14.9	84.6 ± 22.0	74.4 ± 15.4	<0.001	73.6 ± 14.6	87.1 ± 22.7	74.1 ± 15.2	<0.001	73.4 ± 14.2	85.7 ± 22.6	73.9 ± 14.8	<0.001
PR interval, ms	162.5 ± 28.0	174.5 ± 40.9	162.8 ± 28.5	<0.001	160.0 ± 25.6	169.3 ± 38.4	160.2 ± 26.0	<0.001	160.0 ± 25.8	171.9 ± 37.3	160.3 ± 26.2	<0.001
QTc, ms	433.5 ± 32.0	453.3 ± 39.5	434.2 ± 32.6	<0.001	428.7 ± 29.5	452.1 ± 41.1	429.6 ± 30.3	<0.001	428.5 ± 29.2	453.7 ± 38.5	429.4 ± 30.0	<0.001
QRS interval, ms	98.1 ± 18.7	105.3 ± 25.5	98.4 ± 19.1	<0.001	95.3 ± 16.8	103.4 ± 24.9	95.6 ± 17.3	<0.001	95.2 ± 17.0	104.1 ± 23.8	95.5 ± 17.4	<0.001
Sinus rhythm (%)	222570 (62.22)	4466 (30.27)	227036 (60.95)	<0.001	16815 (59.93)	261 (23.51)	17076 (58.55)	<0.001	16910 (60.17)	276 (25.96)	17186 (58.92)	<0.001
AF/AFL (%)	22737 (6.36)	7206 (48.84)	29943 (8.04)	<0.001	1655 (5.9)	572 (51.53)	2227 (7.64)	<0.001	1537 (5.47)	530 (49.86)	2067 (7.09)	<0.001
Ventricular premature complex(es) (%)	15833 (4.43)	1110 (7.52)	16943 (4.55)	<0.001	1322 (4.71)	77(6.94)	1399 (4.8)	<0.001	1339 (4.76)	74 (6.96)	1413 (4.84)	0.001
Atrial premature complex(es) (%)	9754 (2.73)	503 (3.41)	10257 (2.75)	<0.001	684 (2.44)	29(2.61)	713 (2.44)	0.787	782 (2.78)	38 (3.57)	820 (2.81)	0.15
LBBB (%)	5862 (1.64)	225 (1.53)	6087 (1.63)	0.301	308 (1.1)	19(1.71)	327 (1.12)	0.078	323 (1.15)	18 (1.69)	341 (1.17)	0.14
Complete RBBB (%)	13539 (3.78)	1834 (12.43)	15373 (4.13)	<0.001	876 (3.12)	137(12.34)	1013 (3.47)	<0.001	878 (3.12)	129 (12.14)	1007 (3.45)	<0.001
Incomplete RBBB (%)	7197 (2.01)	1281 (8.68)	8478 (2.28)	<0.001	472 (1.68)	113(10.18)	585 (2.01)	<0.001	482 (1.72)	105 (9.88)	587 (2.01)	<0.001
Intraventricular conduction delay (%)	9130 (2.55)	747 (5.06)	9877 (2.65)	<0.001	451 (1.61)	29(2.61)	480 (1.65)	0.014	426 (1.52)	52 (4.89)	478 (1.64)	<0.001
Left-axis deviation (%)	11573 (3.24)	618 (4.19)	12191 (3.27)	<0.001	879 (3.13)	38(3.42)	917 (3.14)	0.648	913 (3.25)	38 (3.57)	951 (3.26)	0.617
Right-axis deviation (%)	10217 (2.86)	2116 (14.34)	12333 (3.31)	<0.001	802 (2.86)	168(15.14)	970 (3.33)	<0.001	822 (2.92)	157 (14.77)	979 (3.36)	<0.001
Low voltage in limb leads (%)	4921 (1.38)	662 (4.49)	5583 (1.5)	<0.001	294 (1.05)	38(3.42)	332 (1.14)	<0.001	296 (1.05)	31 (2.92)	327 (1.12)	<0.001
Echo measurements												
LVEF, %	60.8 ± 8.9	56.6 ± 13.1	60.7 ± 9.1	<0.001	62.1 ± 8.1	57.4 ± 12.1	62.0 ± 8.4	<0.001	62.2 ± 8.0	57.3 ± 12.2	62.0 ± 8.3	<0.001
LVEF < 50% (%)	34977 (9.78)	2980 (20.20)	37957 (10.19)	<0.001	2049 (7.30)	187 (16.85)	2236 (7.67)	<0.001	2009 (7.15)	197 (18.53)	2206 (7.56)	<0.001
AS (moderate-to-severe) (%)	4377 (1.22)	403 (2.73)	4780 (1.28)	<0.001	370 (1.32)	42 (3.78)	412 (1.41)	<0.001	386 (1.37)	29 (2.73)	415 (1.42)	<0.001
AR (moderate-to-severe) (%)	9516 (2.66)	945 (6.41)	10461 (2.81)	<0.001	810 (2.89)	69 (6.22)	879 (3.01)	<0.001	803 (2.86)	62 (5.83)	865 (2.97)	<0.001
MS (moderate-to-severe) (%)	3175 (0.89)	1350 (9.15)	4525 (1.21)	<0.001	290 (1.03)	121 (10.9)	411 (1.41)	<0.001	300 (1.07)	111 (10.44)	411 (1.41)	<0.001
MR (moderate-to-severe) (%)	15228 (4.26)	4500 (30.5)	19728 (5.3)	<0.001	1120 (3.99)	358 (32.25)	1478 (5.07)	<0.001	1093 (3.89)	328 (30.86)	1421 (4.87)	<0.001
PH (%)	13929 (3.89)	8130 (55.10)	22059 (5.92)	<0.001	1062 (3.79)	617 (55.59)	1679 (5.76)	<0.001	1108 (3.94)	597 (56.16)	1705 (5.85)	<0.001
History of Surgery at patientss												
Aortic valve surgery (%)	9150 (4.08)	512 (5.67)	9662 (4.14)	<0.001	788 (2.81)	54 (4.86)	842 (2.89)	<0.001	799 (2.84)	45 (4.23)	844 (2.89)	<0.001
Mitral valve surgery (%)	6103 (2.72)	992 (10.99)	7095 (3.04)	<0.001	614 (2.19)	104 (9.37)	718 (2.46)	<0.001	532 (1.89)	122 (11.48)	654 (2.24)	<0.001

AF atrial fibrillation, *AFL* atrial flutter, *AR* aortic regurgitation, *AS* aortic stenosis, *ECG* electrocardiogram, *ECHO* echocardiography, *LBBB* left bundle branch block, *LVEF* left ventricular ejection fraction, *MR* mitral valve regurgitation, *MS* mitral valve stenosis, *PH* pulmonary hypertension, *QTc* QT interval correction, *RBBB* right bundle branch block, *TR* tricuspid regurgitation.

Table 2 | Baseline characteristics of multi-center evaluation datasets

Variables	FW Testing Set				YF Testing Set				SF Testing Set			
	Non-TR	TR	Total	*P* value	Non-TR	TR	Total Cases	P value	Non-TR	TR	Total	P value
Patients	61057	2868	63925	/	43555	1396	44951	/	20756	544	21300	/
ECG-ECHO pairs	61057	2868	63925	/	43555	1396	44951	/	20756	544	21300	/
Age at ECG, years	56.4 ± 13.6	58.4 ± 15.6	56.5 ± 13.7	<0.001	56.0 ± 14.0	62.0 ± 15.3	56.2 ± 14.1	<0.001	50.6 ± 24.7	59.3 ± 15.2	50.8 ± 24.6	<0.001
Age≥60 years (%)	27699 (45.37)	1538 (53.63)	29237 (45.74)	<0.001	18059 (41.46)	834 (59.74)	18893 (42.03)	<0.001	6220 (29.97)	283 (52.02)	6503 (30.53)	<0.001
Female at ECG (%)	22898 (37.5)	1631 (56.87)	24529 (38.37)	<0.001	18879 (43.35)	786 (56.3)	19665 (43.75)	<0.001	9627 (46.38)	352 (64.71)	9979 (46.85)	<0.001
ECG characteristics												
Heart Rate, bpm	73.7 ± 14.3	86.9 ± 23.0	74.3 ± 15.1	<0.001	71.9 ± 13.4	84.1 ± 22.7	72.3 ± 14.0	<0.001	74.3 ± 14.1	82.9 ± 20.3	74.5 ± 14.4	<0.001
PR interval, ms	163.6 ± 26.6	171.7 ± 39.7	163.8 ± 27.1	<0.001	162.2 ± 25.0	166.4 ± 41.2	162.3 ± 25.4	<0.001	159.0 ± 24.5	177.2 ± 47.8	159.3 ± 25.1	<0.001
QTc, ms	430.3 ± 28.0	452.9 ± 47.9	431.2 ± 29.4	<0.001	420.8 ± 27.8	434.6 ± 29.1	421.3 ± 27.9	<0.001	415.4 ± 25.2	438.7 ± 31.9	415.8 ± 25.5	<0.001
QRS duration, ms	98.3 ± 17.2	104.8 ± 24.2	98.6 ± 17.6	<0.001	94.7 ± 15.6	101.1 ± 21.9	94.9 ± 15.9	<0.001	100.6 ± 14.2	104.3 ± 20.0	100.7 ± 14.4	0.008
Sinus rhythm (%)	47380 (77.6)	1111 (38.74)	48491 (75.86)	<0.001	36807 (84.51)	643 (46.06)	37450 (83.31)	<0.001	16444 (79.23)	213 (39.15)	16657 (78.2)	<0.001
AF/AFL (%)	3112 (5.1)	1426 (49.72)	4538 (7.1)	<0.001	1197 (2.75)	583 (41.76)	1780 (3.96)	<0.001	1029 (4.96)	277 (50.92)	1306 (6.13)	<0.001
Ventricular premature complex(es) (%)	2836 (4.64)	247 (8.61)	3083 (4.82)	<0.001	1430 (3.28)	51 (3.65)	1481 (3.29)	0.492	815 (3.93)	52 (9.56)	867 (4.07)	<0.001
Atrial premature complex(es) (%)	1472 (2.41)	96 (3.35)	1568 (2.45)	0.002	924 (2.12)	73 (5.23)	997 (2.22)	<0.001	388 (1.87)	15 (2.76)	403 (1.89)	0.18
LBBB (%)	713 (1.17)	48 (1.67)	761 (1.19)	0.019	392 (0.9)	19 (1.36)	411 (0.91)	0.101	157 (0.76)	4 (0.74)	161 (0.76)	0.955
Complete RBBB (%)	2305 (3.78)	410 (14.3)	2715 (4.25)	<0.001	1664 (3.82)	149 (10.67)	1813 (4.03)	<0.001	642 (3.09)	56 (10.29)	698 (3.28)	<0.001
Incomplete RBBB (%)	1582 (2.59)	380 (13.25)	1962 (3.07)	<0.001	355 (0.82)	71 (5.09)	426 (0.95)	<0.001	267 (1.29)	16 (2.94)	283 (1.33)	0.002
Intraventricular conduction delay (%)	1486 (2.43)	115 (4.01)	1601 (2.5)	<0.001	1509 (3.46)	71 (5.09)	1580 (3.51)	0.002	173 (0.83)	4 (0.74)	177 (0.83)	0.803
Left-axis deviation (%)	2279 (3.73)	140 (4.88)	2419 (3.78)	0.002	2005 (4.6)	72 (5.16)	2077 (4.62)	0.365	1250 (6.02)	37 (6.8)	1287 (6.04)	0.508
Right-axis deviation (%)	1825 (2.99)	523 (18.24)	2348 (3.67)	<0.001	2137 (4.91)	304 (21.78)	2441 (5.43)	<0.001	884 (4.26)	105 (19.3)	989 (4.64)	<0.001
Low voltage in limb leads (%)	1239 (2.03)	145 (5.06)	1384 (2.17)	<0.001	766 (1.76)	103 (7.38)	869 (1.93)	<0.001	513 (2.47)	37 (6.8)	550 (2.58)	<0.001
Echo measurements												
LVEF, %	62.3 ± 8.5	58.2 ± 13.0	62.1 ± 8.8	<0.001	62.9 ± 8.8	54.8 ± 14.2	62.7 ± 9.1	<0.001	62.0 ± 6.4	58.7 ± 9.8	61.9 ± 6.6	<0.001
LVEF < 50% (%)	4725 (7.74)	531 (18.51)	5256 (8.22)	<0.001	3716 (8.53)	377 (27.01)	4093 (9.11)	<0.001	926 (4.46)	69 (12.68)	995 (4.67)	<0.001
AS (moderate-to-severe) (%)	875 (1.43)	73 (2.55)	948 (1.48)	<0.001	356 (0.82)	29 (2.08)	385 (0.86)	<0.001	148 (0.71)	14 (2.57)	162 (0.76)	<0.001
AR (moderate-to-severe) (%)	1857 (3.04)	164 (5.72)	2021 (3.16)	<0.001	937 (2.15)	71 (5.09)	1008 (2.24)	<0.001	281 (1.35)	32 (5.88)	313 (1.47)	<0.001
MS (moderate-to-severe) (%)	442 (0.72)	254 (8.86)	696 (1.09)	<0.001	175 (0.4)	96 (6.88)	271 (0.6)	<0.001	186 (0.9)	49 (9.01)	235 (1.1)	<0.001
MR (moderate-to-severe) (%)	2563 (4.2)	934 (32.57)	3497 (5.47)	<0.001	995 (2.28)	428 (30.66)	1423 (3.17)	<0.001	379 (1.83)	120 (22.06)	499 (2.34)	<0.001
PH (%)	2316 (3.79)	1615 (56.31)	3931 (6.15)	<0.001	2516 (5.78)	932 (66.76)	3448 (6.67)	<0.001	485 (2.34)	277 (50.92)	762 (3.58)	<0.001
History of Surgery at patients												
Aortic valve surgery (%)	1757 (2.88)	136 (4.74)	1893 (2.96)	<0.001	632 (1.45)	36 (2.58)	668 (1.49)	<0.001	632 (3.04)	61 (11.21)	693 (3.25)	<0.001
Mitral valve surgery (%)	955 (1.56)	236 (8.23)	1191 (1.86)	<0.001	288 (0.66)	56 (4.01)	344 (0.77)	<0.001	821 (3.96)	132 (24.26)	953 (4.47)	<0.001

AF atrial fibrillation, *AFL* atrial flutter, *AR* aortic regurgitation, *AS* aortic stenosis, *ECG* electrocardiogram, *ECHO* echocardiography, *FW* Fuwai Hospital, *LBBB* left bundle branch block, *LVEF* left ventricular ejection fraction, *MR* mitral valve regurgitation, *MS* mitral valve stenosis, *PH* pulmonary hypertension, *QTc* QT interval correction, *RBBB* right bundle branch block, *SF* Shenzhen Fuwai Hospital, *TR* tricuspid regurgitation, *YF* Yunnan Fuwai Hospital.

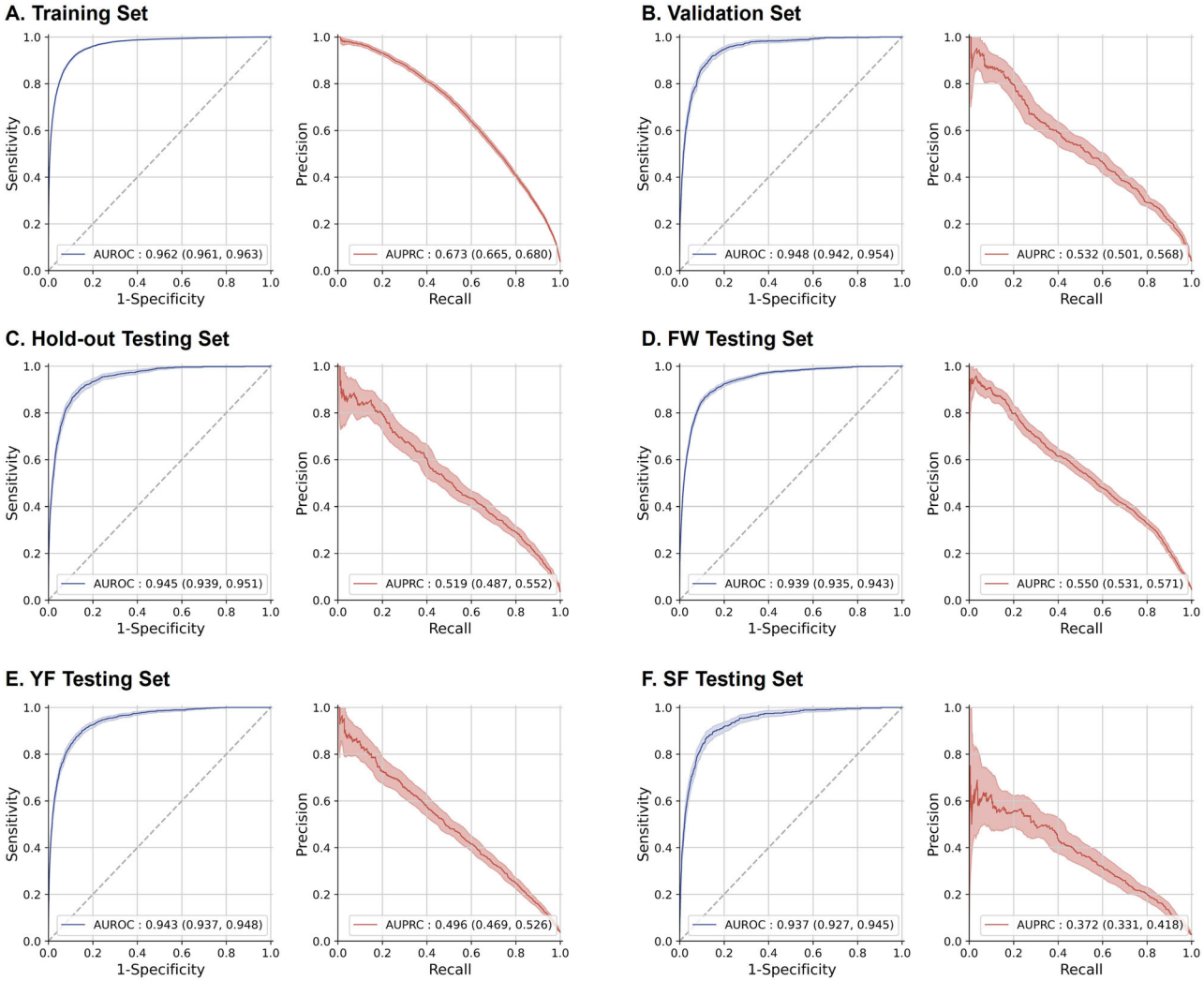

Fig. 1 | ROC curves and PRC curves for model development datasets and multi-center evaluation datasets. This figure presents the ROC curves and PRC curves with 95% CIs in the **A** training set, **B** validation set, **C** hold-out testing set, **D** FW testing set, **E** YF testing set and **F** SF testing set, respectively. AUROC area under the receiver operating characteristic curve, AUPRC area under the precision-recall curve, CI confidence interval, FW Fuwai Hospital, PRC the precision-recall curve, ROC the receiver operating characteristic curve, SF Shenzhen Fuwai Hospital, YF Yunnan Fuwai Hospital.

Table 3 | Model performance across hold-out testing set and multi-center testing sets

Testing set	Outcome rates	AUROC (95%CI)	AUPRC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	F1 (95%CI)	Accuracy (95%CI)
Hold-out testing set	3.64%	0.945 (0.939, 0.951)	0.519 (0.487, 0.552)	0.568 (0.538, 0.600)	0.973 (0.972, 0.975)	0.448 (0.421, 0.474)	0.983 (0.982, 0.985)	0.501 (0.475, 0.525)	0.959 (0.956, 0.961)
FW testing set	4.49%	0.939 (0.935, 0.943)	0.550 (0.531, 0.571)	0.543 (0.526, 0.560)	0.977 (0.976, 0.978)	0.524 (0.507, 0.544)	0.978 (0.977, 0.980)	0.533 (0.518, 0.550)	0.957 (0.956, 0.959)
YF testing set	3.11%	0.943 (0.937, 0.948)	0.496 (0.469, 0.526)	0.574 (0.549, 0.599)	0.976 (0.975, 0.977)	0.433 (0.411, 0.458)	0.986 (0.985, 0.987)	0.494 (0.475, 0.516)	0.963 (0.962, 0.965)
SF testing set	2.55%	0.937 (0.927, 0.945)	0.372 (0.331, 0.418)	0.568 (0.526, 0.611)	0.971 (0.969, 0.973)	0.339 (0.306, 0.372)	0.988 (0.987, 0.990)	0.424 (0.391, 0.457)	0.961 (0.958, 0.963)

AUPRC area under the precision-recall curve, *AUROC* area under the receiver operating characteristic curve, *CI* confidence interval, *FW* Fuwai Hospital, *NPV* negative predictive value, *PPV* positive predictive value, *SF* Shenzhen Fuwai Hospital, *YF* Yunnan Fuwai Hospital.

performance metrics for subgroups of risk factors and VHDs, including AUPRC, sensitivity, specificity, PPV, NPV, F1 score, and accuracy, are presented in Supplementary Tables 4–7.

Besides, we conducted additional subgroup analyses to evaluate model performance in different ECG subtypes, including: patients with both RBBB and right axis deviation (RAD), patients with only RBBB or only RAD, and patients without either RBBB or RAD. The results showed consistent model performance across all these subgroups in both the hold-out testing set and the multi-center external testing set (Supplementary Table 8, Supplementary Fig. 3).

Discussion

This study developed an AI model (the SPEED-TR) for screening TR using a single standard 12-lead ECG. The model performance demonstrated stable

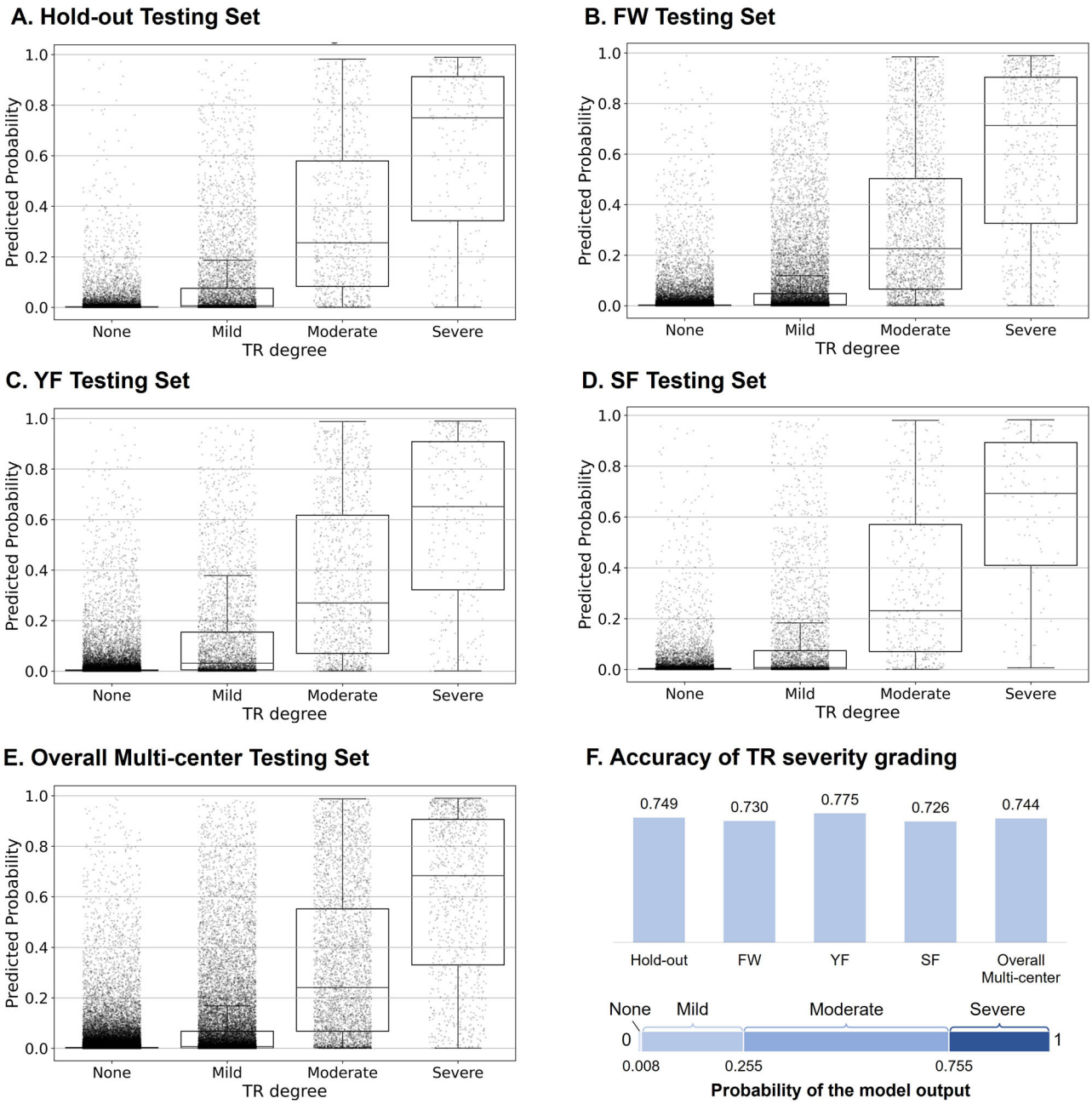

Fig. 2 | Distribution of prediction probabilities of SPEED-TR in different subgroups of TR severity and accuracy of SPEED-TR in grading TR severity across different datasets. The figure illustrates the distribution of the predicted probabilities of the SPEED-TR model on different subgroups of varying TR severity via scatter plots and box plots, in **A** hold-out testing set, **B** FW testing set, **C** YF testing set, **D** SF testing set and **E** overall multi-center testing set, respectively; and **F** the accuracy of SPEED-TR in grading TR severity across the five testing sets, based on the predicted probability for TR detection with the thresholds of 0.008 for mild, 0.255 for moderate and 0.755 for severe TR. FW Fuwai Hospital, SF Shenzhen Fuwai Hospital, SPEED-TR the self-distilled and pre-trained Transformer model for detecting tricuspid regurgitation, TR tricuspid regurgitation, YF Yunnan Fuwai Hospital.

and robust ability for detecting moderate to severe TR in the hold-out testing set (AUROC 0.945) and multi-center internal (AUROC 0.939) and external validation testing datasets (AUROC 0.943 and 0.937, respectively). The model also showed good performance across diverse populations with different TR risk factors or VHDs, alone or combined with various numbers of risk factors or VHDs. Furthermore, the model output probability values appeared to be associated with TR severity grading, indicating a possible role in distinguishing different degrees of TR, which warrants further validation. These findings demonstrated that the SPEED-TR model may offer a cost-effective screening tool for initially detecting TR in resource-limited settings.

Advanced AI techniques demonstrated strong performance in fields such as natural language processing and computer vision[23–25]. The majority of preceding studies on AI-based ECG diagnostic model have concentrated on aortic and mitral valve lesions, with a limited number of studies addressing TR diagnostic models[15,26]. Elias et al. developed and validated a Convolutional Neural Network (CNN) model using an ECG-ECHO dataset from 77,163 patients to identify left-sided VHDs, including AS, AR, and MR, which achieved AUROC values ranging from 0.77 to 0.88, a PPV of 20% and a NPV of 97.6%[21], but TR was not included in this study. A meta-analysis comprising 10 studies demonstrates the considerable value of ECG-based AI screening for VHD, with highly accurate models, excellent

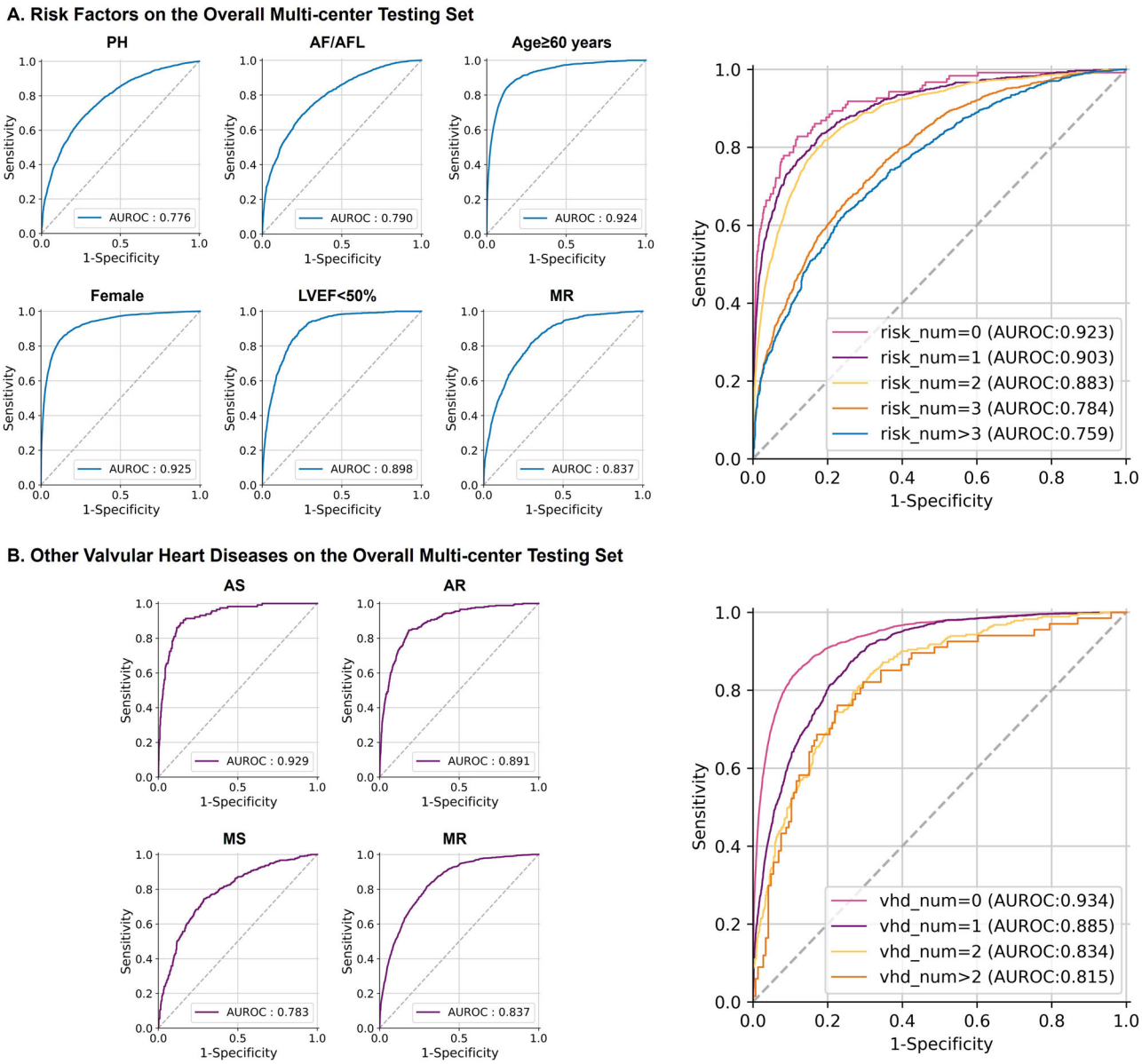

Fig. 3 | ROC curves of SPEED-TR performance in different subgroups on the overall multi-center testing set. A ROC curves of SPEED-TR in patients with different TR risk factors, including PH, AF/AFL, age ≥ 60 years, female, LVEF < 50% and MR, and patients characterized by differing numbers (0, 1, 2, 3 and over 3) of these risk factors; and **B** ROC curves of SPEED-TR across patients presenting with other typical VHDs, including AS, AR, MS or MR, and with 0, 1, 2 or over 2 types of these VHDs on the overall multi-center testing set. AF atrial fibrillation, AFL atrial flutter, AR aortic regurgitation, AS aortic stenosis, AUROC area under the receiver operating characteristic curve, LVEF left ventricular ejection fraction, MR mitral valve regurgitation, MS mitral valve stenosis, PH pulmonary hypertension, ROC the receiver operating characteristic curve, SPEED-TR the self-distilled and pre-trained Transformer model for detecting tricuspid regurgitation, TR tricuspid regurgitation, VHD valvular heart diseases.

sensitivity, specificity, and NPV, but the PPV was low (13%) which was speculated that may be related to the intrinsic variability of ECG recordings[27]. Nevertheless, among this meta-analysis, only one study by Ulloa-Cerna et al. has focused on TR and developed a CNN model with a PPV of merely 0.16 when predicting TR in isolation[26]. The study by Lin et al. also demonstrated that, despite adjusting for age and gender, the model's AUROC for diagnosing TR was only 0.841, with a PPV of 0.392, and an NPV of 0.942[15]. These findings further underscore the challenges associated with the detecting TR by using ECG AI model.

In contrast to conventional CNN-based supervised learning approaches, our study explores the application of Transformer-based self-supervised learning combined with a self-distillation strategy for the detection of TR. The proposed method leveraged the self-attention mechanism of the Transformer architecture, enabling more efficient capture of global information. This approach allows for the extraction of intrinsic ECG patterns from large unlabeled datasets through self-supervised learning. Furthermore, the incorporation of self-distillation enhanced the model's performance, boosting its potential. By using this self-distillation approach, our model achieved high and stable AUROC in the hold-out testing set and maintained consistent performance across multi-center internal and external testing datasets. The largest sample size of ECG-ECHO pairs data might be one important factor for the robust performance of our model. Therefore, the SPEED-TR model outperformed those developed in previous studies in terms of AUROC, specificity, PPV, and NPV[15,26].

Our results of subgroup analyses indicated that the SPEED-TR model maintained high discriminatory ability across diverse population, including those with one or more of risk factors for TR (e.g., reduced LVEF, female, age ≥ 60 years, AF/AFL, PH), as well as those with one or more of other type

of VHDs (e.g., AS, AR, MS and MR). The SPEED-TR model achieved AUROC of over 0.8 in those with one of other risk factors or VHDs except for in patients with PH, or AF/AFL, or MS. Despite the challenges associated with interpreting ECG features in patients with over 2 risk factors or VHDs, the SPEED-TR performed still well with AUROC between 0.759 and 0.883. In addition, the SPEED-TR model demonstrated consistently high NPV of above 0.97 in all testing sets. Consistently high NPV of over 0.832 was found across most patients with one of risk factors or one of VHDs except for those with PH (0.795) or with over 3 risk factors (0.712).

Although AUROC remains high, both PPV and AUPRC were relatively low, which is likely attributable to the low prevalence of TR cases in the dataset. Unlike AUROC, PPV and AUPRC are more sensitive to class imbalance and tend to decrease when the prevalence of positive cases is low[26]. We selected the classification threshold by maximizing the F1 score on the validation set to balance sensitivity and PPV for TR detection. The threshold selection process is important because it directly influences how well the model adapts to the clinical context, ensuring it optimizes performance for real-world applications. It is important to note that the model is threshold-independent and can be adjusted to accommodate different clinical priorities. For example, lowering the threshold may increase sensitivity to reduce missed cases in screening settings, while raising it may improve PPV to limit unnecessary follow-up in resource-constrained environments. These trade-offs are reflected in the precision–recall curve, which can assist clinicians in selecting thresholds appropriate to their specific context.

The SPEED-TR model also demonstrated the potential to estimate TR the grade of severity from ECG by using three different thresholds to identify the possible TR degrees (none, mild, moderate, and severe) with accuracy of over 72% across multiple datasets. Notably, the model exhibited high discrimination for identifying patients with no TR or severe TR, suggesting its utility in effectively ruling out TR in low-risk individuals and prompting further echocardiographic evaluation in those with a high predicted probability. However, the performance was relatively lower in differentiating mild from moderate TR, likely due to overlapping clinical features. The SPEED-TR model may be useful in large-scale screening programs or resource-limited settings to identify patients who are more likely to benefit from echocardiographic examination, while potentially reducing unnecessary imaging in those with low predicted risk. Future studies incorporating multimodal data are warranted to further optimize its clinical applicability, particularly in the assessment of borderline TR severity. Several limitations need to be mentioned. Firstly, the diagnosis of TR in this study was determined based on ECHO reports, which may be influenced by variations in ultrasound interpretation across different sonographers. However, huge ECHO-ECG paired data and multi-center datasets may balance the confounding of the variations among different sonographers. Second, the generalizability of the SPEED-TR model to other racial populations might be limited since only Chinese population were included into our study. Future studies involving diverse, international cohorts are warranted to further validate and enhance its generalizability. Third, patients with pacemaker implants or prior tricuspid valve surgery were excluded from this study. So it is unable to evaluate the performance of SPEED-TR in this part of population. Additionally, due to the absence of routine etiological classification in echocardiographic reports, the SPEED-TR could not be evaluated separately for atrial, ventricular, or mixed TR subtypes. As recent studies highlight important differences among these subtypes, future research is warranted to explore subtype-specific model performance[28].

In conclusion, this study developed the SPEED-TR model for screening TR by using a standard 12-lead ECG with robust performance across all diverse evaluation datasets. Subgroup analyses further validated the consistently good model performance in patients with or without VHDs and/or TR risk factors. The observed association between the model output probabilities and TR severity suggests that the model may provide reference value for TR grading. The SPEED-TR model may serve as a non-invasive and scalable screening tool to rapidly rule out TR or identify patients at higher risk who warrant further echocardiographic evaluation.

Methods

Study population

Figure 4 shows the schematic of study design and data inclusion and exclusion. Data from FW Hospital between January 2015 and December 2022 were consecutively collected and screened for model development. Multi-center data was used for performance evaluation, including internal testing set from FW Hospital between January 2023 and June 2023, and external testing sets from YF Hospital and SF Hospital, between January 2022 and December 2023. This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Fuwai Hospital (Approval No. 2023-1945), with local approval obtained from Shenzhen Fuwai Hospital and Yunnan Fuwai Hospital. The institutional review board allowed us to waive the requirement for obtaining informed consent to the study because the data are acquired for routine patient care and all data used for this study were acquired for clinical purposes and were handled anonymously.

In the model development stage (Fig. 4A), two datasets were used: the pre-training set for model pre-training, and the ECG-ECHO paired set for model post-training. Adult patients, who had at least one standard 10 s 12-lead ECG collected under supine position and with complete digital data and report of ECG from FW Hospital between January 2015 and December 2022, were initially considered, leading to a starting pool of 516,884 patients with 861,324 ECGs. For the ECG-ECHO paired set, patients without ECHO reports were excluded. 350,767 patients with 674,306 available ECGs and the paired ECHO reports closest to each ECG of these patients were gathered for further analysis. Exclusion criteria included: 1) ECG-ECHO pairs with their interval of more than 30 days; 2) ECG-ECHO pairs of patients underwent heart transplantation, implantation of defibrillator or pacemaker, left ventricular assist device implantation, tricuspid valve plasty or replacement; 3) ECG-ECHO pairs, without clear diagnosis or classification of valvular stenosis or regurgitation in ECHO reports. After exclusions, there were 466,149 ECG-ECHO paired data of 291,673 patients were included. We subsequently split them into 233,339 patients for training, 29,167 patients for validation and 29,167 patients for hold-out testing at a ratio of 8:1:1. Finally, for patients with multiple ECG-ECHO pairs, all 372,476 ECG-ECHO pairs were included in the training set, while only the earliest pairs were retained for the validation set (29,167 ECG-ECHO pairs) and the hold-out testing set (29,167 ECG-ECHO pairs). For the pre-training set, we excluded the ECGs of patients in the aforementioned validation and testing sets from the starting pool and finally included 666,020 ECGs of 458,550 patients.

For multi-center evaluation, we collected one internal testing set (FW testing set), and two external testing sets, YF and SF testing sets (Fig. 4B). Adult patients with at least one available ECG and paired ECHO report were included according to the same inclusion and exclusion criteria. Consistent with the validation and hold-out testing set, only the earliest ECG-ECHO pair data of each patient was used. Finally, the FW testing set included 63,925 ECG-ECHO pairs from January 2023 to June 2023. The YF testing set comprised 44,951 ECG-ECHO pairs from YF Hospital, and the SF testing set included 21,300 ECG-ECHO pairs from SF Hospital between January 2022 and December 2023. The overall multi-center sets included 130176 ECG-ECHO pairs from 130176 patients.

Outcome definition

The outcome TR in this study was confirmed based on ECHO reports by cardiac sonographers. ECHOs were performed using Epic7C Color Doppler Ultrasound Instrument, S5-1, S8-3 Probes, Frequency Rate 2.5 ~ 7.5 MHz by experienced cardiac sonographers following standardized protocols[29,30]. The echocardiographic evaluation included measurement of LVEF, left ventricular end-diastolic diameter (LVEDD), right ventricular diameter (RVD), and other echo parameters. TR degree was evaluated from multiple echocardiographic views, in accordance with the 2017 American Society of Echocardiography recommendations for valvular regurgitation[25]. Quantitative assessment using the vena contracta width was performed to assist in grading TR severity, which was classified as none, mild, moderate, or severe according to echocardiographic reports. We defined TR as positive for

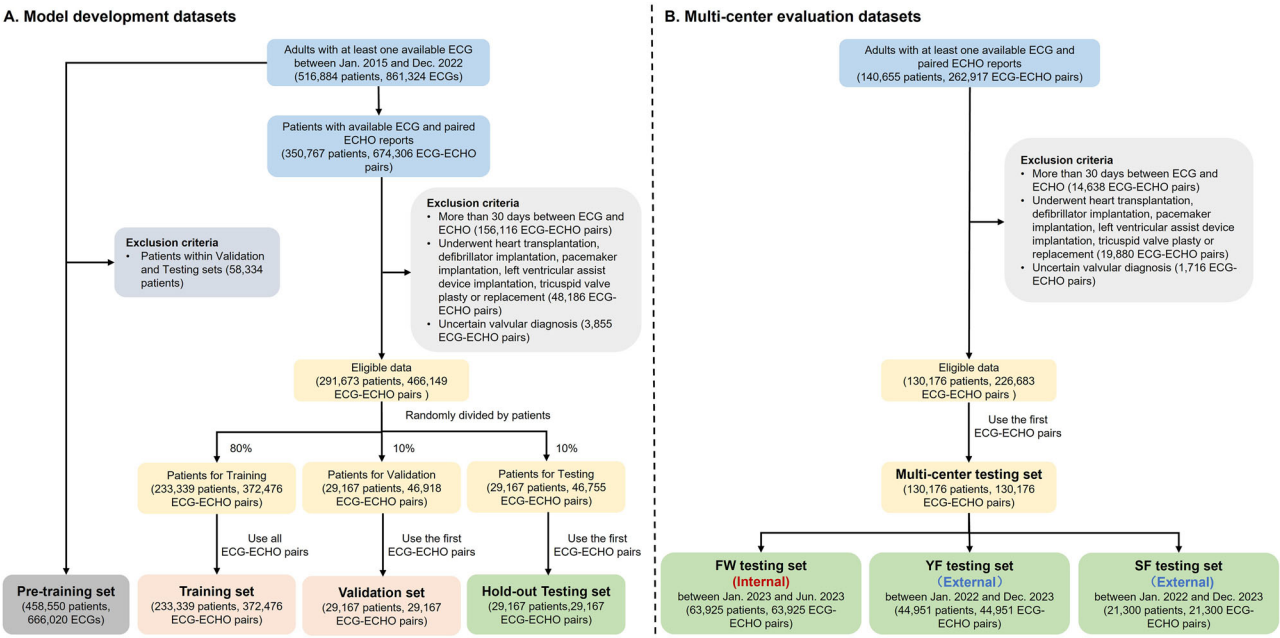

Fig. 4 | Flowchart illustrating patient inclusion and dataset distribution. The flowchart depicts data inclusion and exclusion criteria, as well as the division methods for different datasets. **A** The flowchart for model development datasets, including pre-training, training, validation, and hold-out testing sets. **B** The flowchart for multi-center evaluation datasets, including FW, YF and SF testing sets, these three datasets constitute the multi-center testing set. Dec., December; ECG, electrocardiogram; ECHO echocardiography, FW Fuwai Hospital, Jan. January, SF Shenzhen Fuwai Hospital, YF Yunnan Fuwai Hospital.

moderate or severe regurgitation. Negative TR was identified as no regurgitation, or mild regurgitation.

Clinical characteristics collection and definition

Demographics and clinical data, and relevant medical history were collected for all patients from electronic medical record systems. All digital ECGs in this study were acquired at 500 Hz by GE-Marquette ECG machine (Marquette, Milwaukee, Wisconsin) or Fukuda Denshi machine (Fukuda Denshi Co., Ltd, Tokyo, Japan). The ECG characteristics and ECHO measurements were reviewed by physicians from the ECG reports and ECHO reports. The VHDs considered in this study included AS, AR, MS, MR, and were identified as positive if moderate or severe was confirmed according to ECHO reports. PH was defined as systolic pulmonary artery pressure (sPAP) > 40 mmHg on echocardiography, or mean pulmonary artery pressure (mPAP) ≥ 25 mmHg in selected conditions where sPAP estimation may be unreliable (e.g., congenital heart disease with shunts, severe tricuspid regurgitation, right ventricular outflow obstruction, or significant pulmonary valve stenosis)[31–33]. Furthermore, the definition of aortic valve surgery in this study includes aortic valvuloplasty, aortic valve replacement, transcatheter aortic valve implantation. Mitral valve surgery includes mitral valvuloplasty, mitral valve replacement, or transcatheter mitral valve repair.

Model development

We developed a Transformer-based deep learning model, SPEED-TR, to enhance the detection of TR from standard 12-lead ECG signals. The model was based on Masked Transformer for ECG Classification-Tiny (MTECG-T) architecture[34], which has demonstrated effectiveness in ECG interpretation tasks. In this study, SPEED-TR was specifically adapted for TR detection, incorporating an encoder with 12 Transformer blocks, a single Transformer block decoder, and a binary linear classifier for TR detection. The encoder-decoder structure follows the original MTECG-T framework, allowing for effective feature extraction and sequence modeling. The model processed ECG signals by splitting them into a sequence of non-overlapping segments along the time dimension, preserving structural information related to TR.

To enhance model performance, we adopted a two-stage training process (Fig. 5): a self-supervised pre-training stage on the pre-training set comprising 666,020 unlabeled ECGs, followed by a post-training stage incorporating self-distillation using the training set with 372,476 ECGs labeled for TR.

During pre-training, we employed the self-supervised learning method of MTECG-T to reduce informational redundancy and enhance the extraction of intrinsic ECG patterns. The model is structured as an encoder-decoder network, tasked with reconstructing the per-segment normalization of 25% randomly masked ECG segments based on the remaining 75%, optimized using mean square error loss. All hyperparameter settings in this phase follow the configuration established in MTECG-T. By utilizing this self-supervised pre-training approach, the model was able to learn intrinsic ECG representations without relying on diagnostic labels. This approach aligns with well-established practices in computer vision and natural language processing, where large-scale unlabeled data is used to enhance performance on downstream tasks[24,35]. We observed that this pre-training method led to significant improvements across all evaluation metrics for TR detection (see Supplementary Table 1).

In post-training, we employed a self-distillation strategy to enhance performance[23]. Self-distillation is a technique where the model improves by learning from its own predictions, with a teacher-student setup using identical architectures. Specifically, we converted the pre-trained encoder into a TR detection model by removing the decoder and appending a global pooling layer followed by a binary linear classifier. Both the teacher and student models were initialized with identical weights. The teacher model was trained over 50 epochs using binary cross-entropy loss, with hyperparameters consistent with those established in MTECG-T. The checkpoint with the highest AUPRC on validation set is retained as the final teacher model. The student model was then trained over 50 epochs using a hybrid loss function, consisting of a weighted combination of binary cross-entropy (weight: 0.3) and Kullback-Leibler (KL) divergence between teacher and student logits (weight: 0.7), with a temperature parameter set to 2[36]. The student model checkpoint achieving the highest validation AUPRC was selected as the final SPEED-TR model. It is worth mentioning that there is an overlap of ECG signals between the training sets of the pre-training and

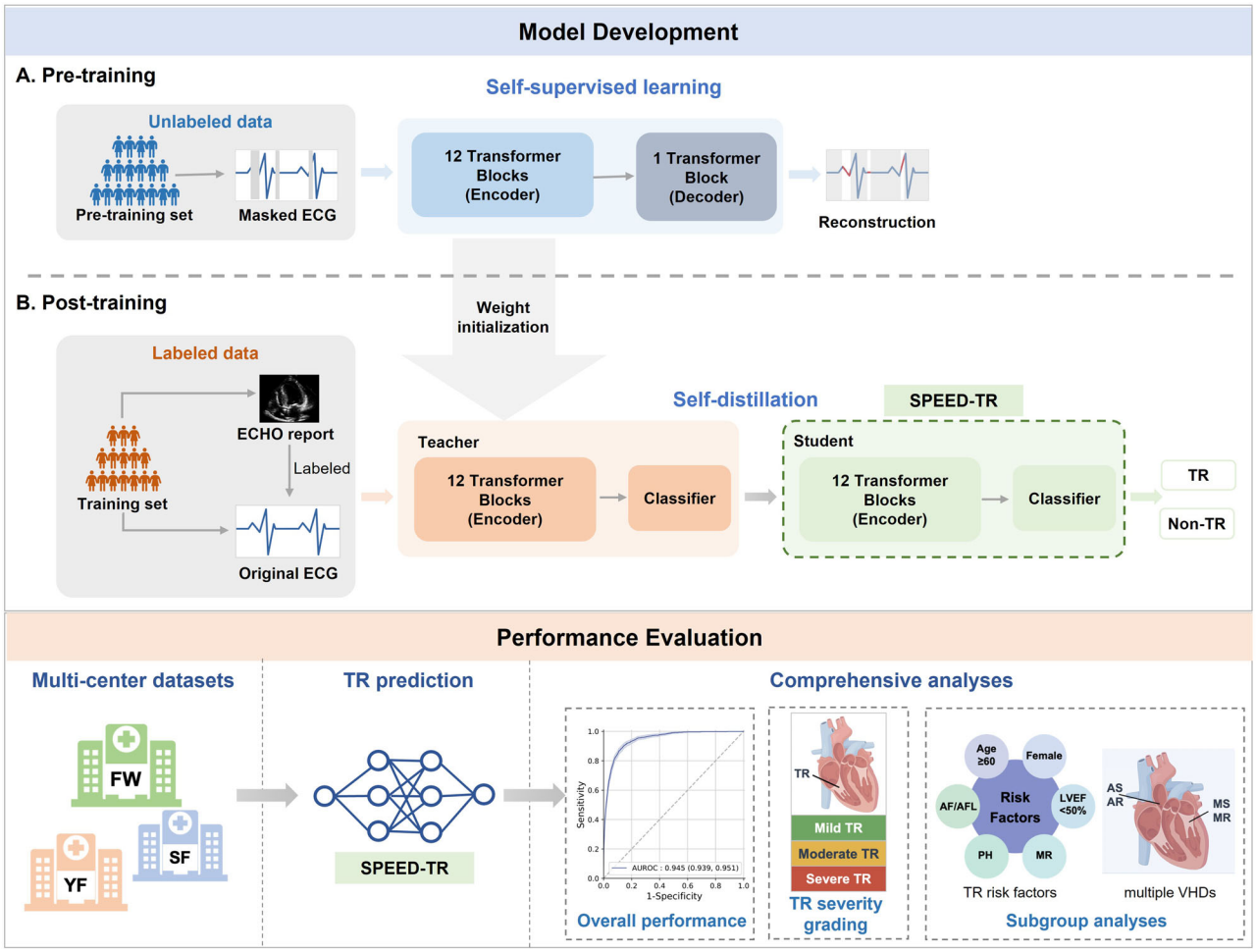

Fig. 5 | Flowchart for model development and evaluation. Model development comprised a pre-training stage on the pre-training set, followed by a post-training stage using the training set. During pre-training, a self-supervised learning method was employed to extract of intrinsic ECG patterns. The model adopted an encoder-decoder network, tasked with reconstructing the masked ECG segments. In post-training stage, a self-distillation strategy with a teacher-student setup using identical architectures was applied to enhance model performance. Specifically, the pre-trained encoder was transformed into a TR detection model, serving as the architecture for both the teacher and student models, with the pretrained weights providing initialization. After the self-distillation, the student model achieving the highest validation AUPRC was selected as the final SPEED-TR model. Performance evaluation was conducted on the hold-out testing set and FW testing set from Fuwai Hospital, YF testing set from Yunnan Fuwai Hospital and SF testing set from Shenzhen Fuwai Hospital. Comprehensive analysis results of SPEED-TR for TR recognition were presented, including not only overall performance, but also includes its capability in grading TR severity, along with subgroup analyses in patients with different TR risk factors and other typical VHDs. AF, atrial fibrillation; AFL, atrial flutter; AR, aortic regurgitation; AS aortic stenosis, AUPRC area under the precision-recall curve, ECG electrocardiogram, ECHO echocardiography, FW Fuwai Hospital, LVEF left ventricular ejection fraction, MR mitral valve regurgitation, MS mitral valve stenosis, PH pulmonary hypertension, SF Shenzhen Fuwai Hospital, SPEED-TR the self-distilled and pre-trained Transformer model for detecting tricuspid regurgitation, TR tricuspid regurgitation, VHD valvular heart diseases, YF Yunnan Fuwai Hospital.

post-training phases, which increases the overall training data size and aligns with common practices in transfer learning[24]. The testing set is strictly held out throughout all stages of the training process, ensuring no data leakage between the training and evaluation phases.

Performance evaluation

Model performance was evaluated on the hold-out testing set, FW testing set, YF testing set and SF testing set (Fig. 5). For comprehensive analysis, we computed various metrics, including the area under the AUROC, AUPRC, sensitivity, specificity, PPV, NPV, accuracy, and F1 score. 95% CI for these metrics were provided using bootstrapping. Calibration curves and brier scores were also presented for evaluating the calibration performance. Besides, the effectiveness of various model components was evaluated, and LR models were trained as 17 variables from baseline characteristics (LR Model 1) or risk factors of TR according to previous literature including MR (moderate and above), reduced ejection fraction (LVEF < 50%), female, elderly (≥60 years), atrial fibrillation or flutter (AF/AFL), and PH[37] (LR Model 2) for comparison with the SPEED-TR model. Delong's test was applied for statistical comparison of LR models and SPEED-TR. Besides, we evaluated the ability of model for grading TR severity (none, mild, moderate, severe) based on the trained binary classification model for identifying TR in the overall multi-center sets. We analyzed the distribution of model-derived prediction probabilities across different severity of TR by the boxplots and scatterplots. The probability thresholds were determined via maximizing the F1 scores of the validation set on different TR degree positive outcome definition (none, mild, moderate, severe) and the grading accuracy was evaluated in all four testing datasets.

Subgroup analyses and statistical analysis

In the overall multi-center sets, we assessed the performance of the SPEED-TR model in patients with specific TR risk factors according to previous literature, including MR, reduced ejection fraction (LVEF < 50%), female, elderly (≥60 years), AF/AFL, and PH[37]. AF/AFL was confirmed by ECG reports. Specifically, we computed the evaluation metrics on subsets with

and without each TR risk factor, and further compared the performance across patient groups characterized by differing numbers of these risk factors, for a comparative analysis of the model's capacity to identify TR in populations of varying complexity. Moreover, subgroup analysis was performed in subsets of patients presenting with other typical VHDs, including AS, AR, MS, MR, as well as the population with combination of two or more VHDs. Additional subgroup analysis to evaluate model performance in the important ECG subtypes of RBBB and RAD was also conducted.

In descriptive analyses, continuous variables were reported as mean ± standard deviation or median with interquartile range, depending on data normality by two-sample t-test or Mann-Whitney U test. Categorical variables were summarized as frequencies and percentages, and compared using the Chi-squared test. Statistical analyses were conducted using Python (version 3.8).

Data availability

The datasets generated and analyzed during the current study are not publicly available due to patient privacy and ethical restrictions but are available from the corresponding author on reasonable request.

Code availability

The final model architecture and the self-distillation code utilized in this study are publicly available on GitHub at: https://github.com/fwaiaccount/MTECG-SPEED-TR/.

Received: 15 June 2025; Accepted: 17 September 2025;
Published online: 12 November 2025

References

1. Hahn, R. T. Tricuspid Regurgitation. *N. Engl. J. Med.* **388**, 1876–1891, https://doi.org/10.1056/NEJMra2216709 (2023).
2. d'Arcy, J. L. et al. Large-scale community echocardiographic screening reveals a major burden of undiagnosed valvular heart disease in older people: the OxVALVE Population Cohort Study. *Eur. Heart J.* **37**, 3515–3522, https://doi.org/10.1093/eurheartj/ehw229 (2016).
3. Tsampasian, V. et al. Prevalence of asymptomatic valvular heart disease in the elderly population: a community-based echocardiographic study. *Eur. Heart J. Cardiovasc. Imaging* https://doi.org/10.1093/ehjci/jeae127 (2024).
4. Topilsky, Y. et al. Clinical presentation and outcome of tricuspid regurgitation in patients with systolic dysfunction. *Eur. Heart J.* **39**, 3584–3592, https://doi.org/10.1093/eurheartj/ehy434 (2018).
5. Adamo, M. et al. Epidemiology, pathophysiology, diagnosis and management of chronic right-sided heart failure and tricuspid regurgitation. A clinical consensus statement of the Heart Failure Association (HFA) and the European Association of Percutaneous Cardiovascular Interventions (EAPCI) of the ESC. *Eur. J. Heart Fail* **26**, 18–33, https://doi.org/10.1002/ejhf.3106 (2024).
6. Hahn, R. T. & Zamorano, J. L. The need for a new tricuspid regurgitation grading scheme. *Eur. Heart J. Cardiovasc. Imaging* **18**, 1342–1343, https://doi.org/10.1093/ehjci/jex139 (2017).
7. Obayashi, Y. et al. Tricuspid regurgitation in elderly patients with acute heart failure: insights from the KCHF registry. *ESC Heart Fail* **10**, 1948–1960, https://doi.org/10.1002/ehf2.14348 (2023).
8. Chorin, E. et al. Tricuspid regurgitation and long-term clinical outcomes. *Eur. Heart J. Cardiovasc. Imaging* **21**, 157–165, https://doi.org/10.1093/ehjci/jez216 (2020).
9. Samim, D. et al. Natural history and mid-term prognosis of severe tricuspid regurgitation: A cohort study. *Front Cardiovasc. Med.* **9**, 1026230. https://doi.org/10.3389/fcvm.2022.1026230 (2022).
10. Wang, T. K. M. et al. Early surgery is associated with improved long-term survival compared to class I indication for isolated severe tricuspid regurgitation. *J. Thorac. Cardiovasc. Surg.* **166**, 91–100, https://doi.org/10.1016/j.jtcvs.2021.07.036 (2023).
11. Hua, K. et al. Early surgery can improve the outcomes of patients with severe tricuspid regurgitation undergoing tricuspid replacement. *Cardiovasc. Diagn. Ther.* **11**, 1058–1066, https://doi.org/10.21037/cdt-21-311 (2021).
12. Sala, A. et al. Isolated tricuspid valve surgery: first outcomes report according to a novel clinical and functional staging of tricuspid regurgitation. *Eur. J. Cardiothorac. Surg.* **60**, 1124–1130, https://doi.org/10.1093/ejcts/ezab228 (2021).
13. Welle, G. A. et al. New approaches to assessment and management of tricuspid regurgitation beforE INTERVEntion. *Jacc. Cardiovasc. Intervent.* **17**, 837–858, https://doi.org/10.1016/j.jcin.2024.02.034 (2024).
14. Hahn, R. T. et al. Tricuspid regurgitation: recent advances in understanding pathophysiology, severity grading and outcome. *Eur. Heart J. Cardiovasc. Imaging* **23**, 913–929, https://doi.org/10.1093/ehjci/jeac009 (2022).
15. Lin, Y. T. et al. Comprehensive clinical application analysis of artificial intelligence-enabled electrocardiograms for screening multiple valvular heart diseases. *Aging* **16**, 8717–8731, https://doi.org/10.18632/aging.205835 (2024).
16. Sengupta, P. P., Kluin, J., Lee, S. P., Oh, J. K. & Smits, A. The future of valvular heart disease assessment and therapy. *Lancet* **403**, 1590–1602, https://doi.org/10.1016/s0140-6736(23)02754-x (2024).
17. Ose, B. et al. Artificial intelligence interpretation of the electrocardiogram: a state-of-the-art review. *Curr. Cardiol. Rep.* **26**, 561–580, https://doi.org/10.1007/s11886-024-02062-1 (2024).
18. Khurshid, S. et al. ECG-based deep learning and clinical risk factors to predict atrial fibrillation. *Circulation* **145**, 122–133, https://doi.org/10.1161/circulationaha.121.057480 (2022).
19. Attia, Z. I. et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat. Med.* **25**, 70–74, https://doi.org/10.1038/s41591-018-0240-2 (2019).
20. Cohen-Shelly, M. et al. Electrocardiogram screening for aortic valve stenosis using artificial intelligence. *Eur. Heart J.* **42**, 2885–2896, https://doi.org/10.1093/eurheartj/ehab153 (2021).
21. Elias, P. et al. Deep learning electrocardiographic analysis for detection of left-sided valvular heart disease. *J. Am. Coll. Cardiol.* **80**, 613–626, https://doi.org/10.1016/j.jacc.2022.05.029 (2022).
22. Cabitza, F. et al. Rams, hounds and white boxes: Investigating human-AI collaboration protocols in medical diagnosis. *Artif. Intell. Med.* **138**, 102506. https://doi.org/10.1016/j.artmed.2023.102506 (2023).
23. Pham, M. et al. Revisiting self-distillation. arXiv cs.LG, arXiv:2206.08491, https://doi.org/10.48550/arXiv.2206.08491 (2022).
24. He, K. et al. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16000–16009.
25. Devlin, J. et al. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguist.: Hum. Lang. Technol.* 1, 4171–4186. https://doi.org/10.18653/v1/N19-1423 (2019).
26. Ulloa-Cerna, A. E. et al. rECHOmmend: an ECG-based machine learning approach for identifying patients at increased risk of undiagnosed structural heart disease detectable by echocardiography. *Circulation* **146**, 36–47, https://doi.org/10.1161/circulationaha.121.057869 (2022).
27. Singh, S. et al. Meta-analysis of the performance of AI-driven ECG interpretation in the diagnosis of valvular heart diseases. *Am. J. Cardiol.* **213**, 126–131, https://doi.org/10.1016/j.amjcard.2023.12.015 (2024).
28. Muraru, D. et al. Atrial secondary tricuspid regurgitation: pathophysiology, definition, diagnosis, and treatment. *Eur. heart J.* **45**, 895–911, https://doi.org/10.1093/eurheartj/ehae088 (2024).

29. Lancellotti, P. et al. Recommendations for the echocardiographic assessment of native valvular regurgitation: an executive summary from the European Association of Cardiovascular Imaging. *Eur. heart J. Cardiovasc. Imaging* **14**, 611–644, https://doi.org/10.1093/ehjci/jet105 (2013).
30. Baumgartner, H. et al. 2017 ESC/EACTS guidelines for the management of valvular heart disease. *Eur. Heart J.* **38**, 2739–2791, https://doi.org/10.1093/eurheartj/ehx391 (2017).
31. McLaughlin, V. V. et al. ACCF/AHA 2009 expert consensus document on pulmonary hypertension a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents and the American Heart Association developed in collaboration with the American College of Chest Physicians; American Thoracic Society, Inc.; and the Pulmonary Hypertension Association. *J. Am. Coll. Cardiol.* **53**, 1573–1619, https://doi.org/10.1016/j.jacc.2009.01.004 (2009).
32. Augustine, D. X. et al. Echocardiographic assessment of pulmonary hypertension: a guideline protocol from the British Society of Echocardiography. *Echo Res. Pract.* **5**, G11–g24, https://doi.org/10.1530/erp-17-0071 (2018).
33. Galiè, 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): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur. Heart J.* **37**, 67–119, https://doi.org/10.1093/eurheartj/ehv317 (2016).
34. Zhou, Y. et al. Enhancing automatic multilabel diagnosis of electrocardiogram signals: A masked transformer approach. *Comput. Biol. Med*. 196 (Pt A), 110674, https://doi.org/10.1016/j.compbiomed.2025.110674 (2025).
35. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*. 4171–4186.
36. Hinton, G. et al. Distilling the knowledge in a neural network. arXiv arXiv:1503.02531 [stat.ML], https://doi.org/10.48550/arXiv.1503.02531 (2015).
37. Benfari, G. et al. Excess mortality associated with functional tricuspid regurgitation complicating heart failure with reduced ejection fraction. *Circulation* **140**, 196–206, https://doi.org/10.1161/circulationaha.118.038946 (2019).

Acknowledgements

This study was funded by the National High Level Hospital Clinical Research Funding (2022-GSP-GG-14). W.X. gratefully acknowledges support from the Postdoctoral Fellowship Program of the China Postdoctoral Science Foundation (Grant No. GZC20240151).

Author contributions

X.F. and W.Z. contributed to the conception or design of the work. X.D. contributed to the acquisition, analysis, or interpretation of data for the work, and drafting parts of the manuscript. W.X. and H.C. contributed to the interpretation of data and drafted the manuscript. Y.Z. led the algorithm design and source code development, performed data analysis, contributed to drafting parts of the manuscript, and prepared part of the tables. Y.L., Y.H., J.L., J. Huang., J.He, F.L., Z.C. and X.Z. contributed to the acquisition of data. Y.L., Y.H., J.L., and Z.C. conducted the experiments, analyzed the data and contributed to the interpretation and presentation of the results. All authors reviewed and revised the manuscript, approved the final version, and agreed to be accountable for all aspects of the work to ensure its integrity and accuracy.

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/s41746-025-02011-4.

Correspondence and requests for materials should be addressed to Wei Zhao or Xiaohan Fan.