Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study

Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study

Daiju Ueda, Toshimasa Matsumoto, Shoichi Ehara, Akira Yamamoto, Shannon L Walston, Asahiro Ito, Taro Shimono, Masatsugu Shiba, Tohru Takeshita, Daiju Fukuda, Yukio Miki

Summary

Background Chest radiography is a common and widely available examination. Although cardiovascular structures—such as cardiac shadows and vessels—are visible on chest radiographs, the ability of these radiographs to estimate cardiac function and valvular disease is poorly understood. Using datasets from multiple institutions, we aimed to develop and validate a deep-learning model to simultaneously detect valvular disease and cardiac functions from chest radiographs.

Methods In this model development and validation study, we trained, validated, and externally tested a deep learning-based model to classify left ventricular ejection fraction, tricuspid regurgitant velocity, mitral regurgitation, aortic stenosis, aortic regurgitation, mitral stenosis, tricuspid regurgitation, pulmonary regurgitation, and inferior vena cava dilation from chest radiographs. The chest radiographs and associated echocardiograms were collected from four institutions between April 1, 2013, and Dec 31, 2021: we used data from three sites (Osaka Metropolitan University Hospital, Osaka, Japan; Habikino Medical Center, Habikino, Japan; and Morimoto Hospital, Osaka, Japan) for training, validation, and internal testing, and data from one site (Kashiwara Municipal Hospital, Kashiwara, Japan) for external testing. We evaluated the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and accuracy.

Findings We included 22 551 radiographs associated with 22 551 echocardiograms obtained from 16 946 patients. The external test dataset featured 3311 radiographs from 2617 patients with a mean age of 72 years [SD 15], of whom 49·8% were male and 50·2% were female. The AUCs, accuracy, sensitivity, and specificity for this dataset were 0·92 (95% CI 0·90–0·95), 86% (85–87), 82% (75–87), and 86% (85–88) for classifying the left ventricular ejection fraction at a 40% cutoff, 0·85 (0·83–0·87), 75% (73–76), 83% (80–87), and 73% (71–75) for classifying the tricuspid regurgitant velocity at a 2·8 m/s cutoff, 0·89 (0·86–0·92), 85% (84–86), 82% (76–87), and 85% (84–86) for classifying mitral regurgitation at the none-mild versus moderate-severe cutoff, 0·83 (0·78–0·88), 73% (71–74), 79% (69–87), and 72% (71–74) for classifying aortic stenosis, 0·83 (0·79–0·87), 68% (67–70), 88% (81–92), and 67% (66–69) for classifying aortic regurgitation, 0·86 (0·67–1·00), 90% (89–91), 83% (36–100), and 90% (89–91) for classifying mitral stenosis, 0·92 (0·89–0·94), 83% (82–85), 87% (83–91), and 83% (82–84) for classifying tricuspid regurgitation, 0·86 (0·82–0·90), 69% (68–71), 91% (84–95), and 68% (67–70) for classifying pulmonary regurgitation, and 0·85 (0·81–0·89), 86% (85–88), 73% (65–81), and 87% (86–88) for classifying inferior vena cava dilation.

Interpretation The deep learning-based model can accurately classify cardiac functions and valvular heart diseases using information from digital chest radiographs. This model can classify values typically obtained from echocardiography in a fraction of the time, with low system requirements and the potential to be continuously available in areas where echocardiography specialists are scarce or absent.

Funding None.

Lancet Digit Health 2023; 5: e525–33

Published Online July 6, 2023 https://doi.org/10.1016/S2589-7500(23)00107-3

Department of Diagnostic and Interventional Radiology (D Ueda MD PhD, T Matsumoto PhD, A Yamamoto MD PhD, S L Walston MS, T Shimono MD PhD, Prof Y Miki MD PhD), **Department of Intensive Care Medicine** (S Ehara MD PhD), **Department of Cardiovascular Medicine** (A Ito MD PhD, Prof D Fukuda MD PhD), **and Department of Biofunctional Analysis** (M Shiba MD PhD), **Graduate School of Medicine, Osaka Metropolitan University, Osaka, Japan; Smart Life Science Lab, Center for Health Science Innovation, Osaka Metropolitan University, Osaka, Japan** (D Ueda, T Matsumoto, M Shiba); **Department of Radiology, Osaka Habikino Medical Center, Habikino, Japan** (T Takeshita MD PhD)

Correspondence to: Dr Daiju Ueda, Department of Diagnostic and Interventional Radiology, Graduate School of Medicine, Osaka Metropolitan University, Osaka 545-8585, Japan **ai.labo.ocu@gmail.com**

Introduction

Chest radiography remains the cornerstone of radiological imaging more than 100 years after it was first reported.[1] As the most frequently conducted radiological test in the world, this technique is central to the screening, diagnosis, and management of various lung diseases.[2–4] However, the relationship between chest radiographs and cardiac function remains poorly understood. Even the cardiothoracic ratio, which is a common indicator of cardiac enlargement, has a weak association with left ventricular ejection fraction.[5–8]

Transthoracic echocardiography is the most frequently used imaging technique for cardiac function and diseases, providing data on left ventricular ejection fraction and tricuspid regurgitant velocity for the diagnosis and monitoring of patients with mitral

Research in context

Evidence before this study

We searched PubMed, MEDLINE, and the Web of Science from database inception to Jan 1, 2023, using the keywords "deep learning", "convolutional neural network", "valvular heart disease", "valvular disease", "mitral regurgitation", "aortic stenosis", "left ventricular ejection fraction", "pulmonary hypertension", "chest x-ray", "chest radiography", and "chest radiograph". No multi-institutional research to estimate cardiac functions and valvular heart diseases from chest radiography was found. Two studies based on deep learning could detect only individual valvular diseases and used small, single-centre datasets, which are prone to model overfitting.

Added value of this study

We show that radiological images contain information that is clear to an artificial intelligence model but difficult for humans to identify. We use this capability of the model to classify information from chest radiographs, such as left ventricular ejection fraction and the presence of valvular heart disease, much faster than it is typically obtained from echocardiograms. The model has low system requirements and the potential to be continuously available in areas where echocardiography specialists are scarce or absent.

Implications of all the available evidence

Information typically obtained from electrocardiography, a dynamic examination, can be extrapolated from chest radiographs, a static examination. Further performance improvements and advancements in ancillary studies could enable quicker and more cost-effective diagnosis, monitoring, and treatment for cardiac function and valvular disease.

regurgitation, aortic stenosis, tricuspid regurgitation, and pulmonary regurgitation. Guidelines recommend that patients with these conditions are monitored regularly by echocardiography.[9–11] The incidence of these cardiac diseases is expected to increase in the future;[12–16] however, echocardiography requires specialised skills, and there is a shortage of technicians.[17,18]

See Online for appendix

Under these circumstances, chest radiography—which is accessible, reproducible, and quick—could complement echocardiography. Deep learning, a method in artificial intelligence,[19,20] has been applied to chest radiographs to estimate echocardiography results.[21,22] Unlike conventional machine-learning methods, deep learning extracts features from training data and does not require the manual definition of features in advance.[19] A deep-learning model is therefore advantageous for the classification and quantification of objects with complex or unknown features. In this study, we used datasets from multiple institutions to develop and evaluate a deep-learning model that can simultaneously detect valvular disease and cardiac functions from chest radiographs.

Methods

Study design and participants

In this model development and validation study, we trained, validated, and externally tested a deep learning-based model to estimate echocardiography results from digital chest radiographs. Chest radiographs, taken in the standing position with a posteroanterior view, were retrospectively collected from patients within 14 days of a two-dimensional echocardiographic examination at four institutions: Osaka Metropolitan University Hospital (dataset A) between May 1, 2019, and Dec 31, 2021; Habikino Medical Center (dataset B) between April 1, 2013, and Dec 31, 2020; Morimoto Hospital (dataset C) between Feb 1, 2018, and Dec 31, 2021; and Kashiwara Municipal Hospital (dataset D) between April 1, 2017, and Dec 31, 2021. An overview of the facility type and capacity of these institutions is available in the appendix (p 2). If a patient had more than one echocardiogram during the data collection period, all were included; if two or more radiographs were available, the radiograph taken closest to the day of the echocardiography examination was used. Patients with a postoperative valve and those with missing echocardiography results were excluded. We did not include sample size calculations but prepared a large external test dataset.

This study complies with the Declaration of Helsinki. The study protocol was reviewed and approved by the Ethical Committee of Osaka Metropolitan University Graduate School of Medicine, Osaka, Japan. The need for informed consent was waived because the images had been acquired during routine clinical practice. This manuscript was prepared in accordance with the Standards for Reporting Diagnostic Accuracy guidelines.[23]

Ground truth labelling

Chest radiographs were labelled using the echocardiography reports. We selected mitral regurgitation, aortic stenosis, aortic regurgitation, mitral stenosis, tricuspid regurgitation, and pulmonary regurgitation as examples of valvular heart disease. The severity of disease was classified as none, mild, moderate, or severe according to American Society of Echocardiography recommendations.[24,25] Two severity cutoffs were set for all valvular diseases: none versus mild-severe and none-mild versus moderate-severe. According to guidelines, we set two cutoffs for the left ventricular ejection fraction, one at 40% and one at 50%,[9] and two cutoffs for the tricuspid regurgitant velocity, one at 2·8 m/s and one at 3·4 m/s.[10] The cutoff for inferior vena cava dilation was set to 21 mm.[10]

Data partition

All labelled chest radiographs from datasets A, B, and C were randomly divided, on a patient basis and in a 9:1 ratio, into training and internal test datasets for each institution. The training datasets were used to train and validate the deep-learning model using five-fold cross-validation. The internal test datasets consisted of patients who were not part of the training dataset but were from the same institutional datasets. The external test dataset, dataset D, was collected from a separate institution. Detailed descriptions of the training, internal test, and external test datasets are shown in the appendix (pp 2, 4).

Model development

We developed a multilabel deep-learning model using EfficientNet[26] as a feature extractor. Multilabel model learning is more accurate than single-label learning because the model is more versatile as a result of extracting features from various labels.[27] 17 labels were chosen as classifiers: two cutoffs (none-mild *vs* moderate-severe and none *vs* mild-severe) for each of the six valvular heart diseases (mitral regurgitation, aortic stenosis, aortic regurgitation, mitral stenosis, tricuspid regurgitation, and pulmonary regurgitation), two cutoffs each for left ventricular ejection fraction (40% and 50%) and tricuspid regurgitant velocity (2·8 m/s and 3·4 m/s), and one cutoff for inferior vena cava dilation (21 mm). These classifiers comprised a fully connected layer connected to a single feature extractor. Each classifier was connected to a SoftMax activation function and followed by the cross-entropy loss function. All loss values were summed and used to measure model performance. The deep-learning model was trained on the basis of ImageNet pretrained parameters and tuned with the training dataset using five-fold cross-validation. All training radiographs were augmented using TrivialAugment.[28] Every development process was done using the PyTorch framework.[29] Detailed processes of the development and parameter search for the deep-learning model are shown in the appendix (p 2), in addition to the machine environment (p 2), class imbalance handling methods (p 2), and an outline of the model (p 5). The source code is available online.[30]

Model test

Using the best-performing model and the same thresholds as those for the validation dataset, we assessed the diagnostic performance of the deep-learning model on both the internal test and external test datasets. We set nine labels as the primary classifiers of the model and the remaining eight as the supplemental classifiers. The nine primary classifiers were as follows: a cutoff of none-mild versus moderate-severe for each of the six valvular heart diseases, a cutoff of 40% for left ventricular ejection fraction, 2·8 m/s for tricuspid regurgitant velocity, and 21 mm for interior vena cava dilation. The primary classifiers for valvular disease were set because disease of moderate or greater severity is recommended for follow-up by a specialist,[10,31] and those for left ventricular ejection fraction and tricuspid regurgitant velocity were chosen because these values are defined as important cutoff values for cardiac dysfunction according to the respective guidelines.[9,10]

To show the region of interest for each classifier as it discriminated each output in the external test dataset, saliency maps for the nine primary classifiers were created using the 10% of the true positive radiographs with the highest model output. The images were added together and divided by the total number of included images to create an averaged image per variable; these nine averaged images were also then averaged to generate an overall average image. To make each saliency map, Grad-CAM++[32] was applied by extracting the gradient information flowing into the last convolutional layer of the trained deep-learning model. The gradient indicates the importance and relevance of each feature map from the last convolutional layer. A detailed explanation of the saliency map generation model is shown in the appendix (p 6).

Statistical analysis

To evaluate the model, we assessed the sensitivity, specificity, accuracy, positive and negative predictive values, and area under the receiver operating characteristic curve (AUC) for each of the classifiers with both the internal and external test datasets. To evaluate the overall performance of the model per dataset, we evaluated the mean and SD of the AUC results for the primary classifiers. Using the external dataset, we evaluated AUCs by sex and age and generated confusion matrices for each primary classifier. 95% CIs for the AUC were calculated with a bootstrapping method. All analyses were done using R version 4.0.0, and statistical inferences were conducted with a two-sided significance level of 5% using the Clopper-Pearson method.

Role of the funding source

There was no funding source for this study.

Results

For the **source code** see https://github.com/xp-echo/Nervus

22 551 radiographs associated with 22 551 echocardiograms of 16 946 patients from four institutions were included in this study. The eligibility flowchart for each dataset is shown in the appendix (p 7). The training datasets, from three institutions, included 17 293 radiographs of 12 897 patients (7140 male and 5757 female) with a mean age of 69 years (SD 14, range 10–103) and the internal test datasets, from the same three institutions, included 1947 radiographs of 1432 patients (809 male and 623 female) with a mean age of 69 years (14, 14–99). The external test dataset, from one separate institution, included 3311 radiographs from 2617 patients (1304 male and 1313 female) with a mean age of 72 years (15, 13–102). Detailed demographics for each dataset are shown in

	Training dataset			Internal test dataset			External test dataset
	Dataset A	Dataset B	Dataset C	Dataset A	Dataset B	Dataset C	Dataset D
Total number of radiographs	8183	5894	3216	928	668	351	3311
Total number of echocardiography examinations	8183	5894	3216	928	668	351	3311
Total number of patients	6174	4202	2521	686	466	280	2617
Male	3370	2376	1394	391	258	160	1304
Female	2804	1826	1127	295	208	120	1313
Age, years	67 (15)	70 (14)	72 (14)	66 (15)	70 (13)	72 (14)	72 (15)
Period between examinations, days	3 (6)	4 (7)	4 (6)	3 (6)	4 (7)	3 (5)	3 (6)
Left ventricular ejection fraction							
≥50%	6891	5061	2409	794	566	275	2989
40–49%	484	418	333	49	36	24	154
<40%	808	415	474	85	66	52	168
Tricuspid regurgitant velocity							
<2·8 m/s	7305	4376	2696	853	493	297	2811
2·8–3·3 m/s	698	1051	425	60	111	40	417
≥3·4 m/s	180	467	95	15	64	14	83
Mitral regurgitation							
None	4706	3819	2182	543	437	246	2454
Mild	2353	1759	909	263	189	90	660
Moderate	798	269	89	87	34	12	146
Severe	326	47	36	35	8	3	51
Aortic stenosis							
None	7450	4029	3002	853	456	327	3144
Mild	147	1760	101	12	197	13	76
Moderate	251	78	87	26	10	10	50
Severe	335	27	26	37	5	1	41
Aortic regurgitation							
None	6098	4414	2537	690	508	278	2298
Mild	1663	1343	625	191	145	68	859
Moderate	339	133	49	40	11	5	142
Severe	83	4	5	7	4	0	12
Mitral stenosis							
None	8081	5825	3174	917	656	351	3295
Mild	32	52	37	2	11	0	10
Moderate	16	16	4	0	1	0	5
Severe	54	1	1	9	0	0	1
Tricuspid regurgitation							
None	6006	2247	2061	687	263	230	2385
Mild	1418	2991	1046	154	328	104	668
Moderate	632	581	80	68	69	15	218
Severe	127	75	29	19	8	2	40
Pulmonary regurgitation							
None	7115	3946	2533	813	462	274	2802
Mild	990	1912	678	105	203	73	381
Moderate	73	35	5	10	3	4	128
Severe	5	1	0	0	0	0	0

Data are n or mean (SD).

Table 1: **Dataset demographics**

table 1 and in the appendix (pp 12–13). Detailed vendor information for each dataset is shown in the appendix (p 14).

The mean AUC for the primary classifiers was 0·87 (SD 0·03) for external test dataset D, 0·89 (0·04) for internal test dataset A, 0·90 (0·04) for internal test

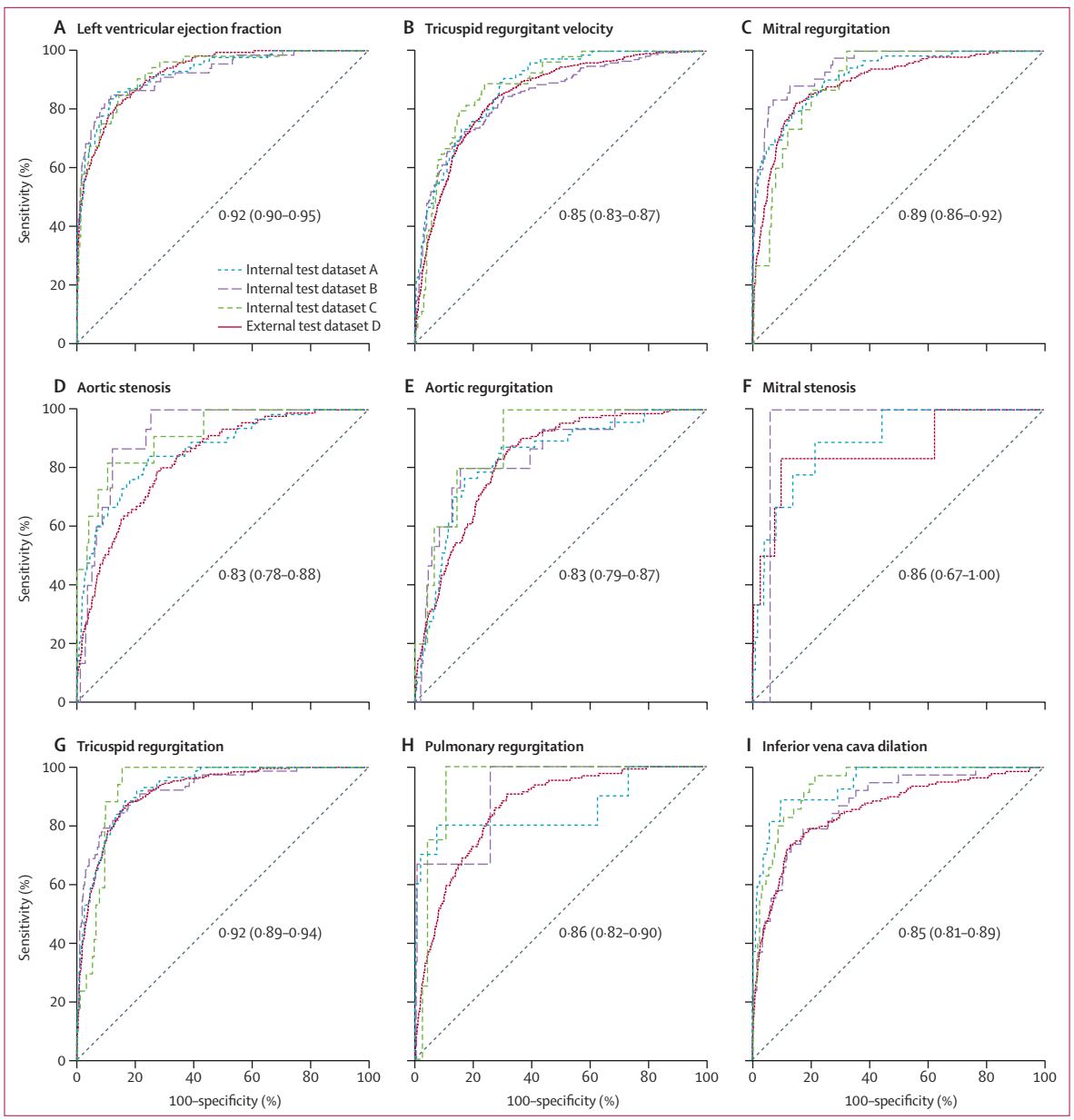

Figure 1: **Receiver operating characteristic curves for the internal and external test datasets**
Receiver operating characteristic curves are shown for the classification of left ventricular ejection fraction less than 40% (A), tricuspid regurgitant velocity greater than 2·8 m/s (B), mitral regurgitation (none-mild *vs* moderate-severe; C), aortic stenosis (none-mild *vs* moderate-severe; D), aortic regurgitation (none-mild *vs* moderate-severe; E), mitral stenosis (none-mild *vs* moderate-severe; F), tricuspid regurgitation (none-mild *vs* moderate-severe; G), pulmonary regurgitation (none-mild *vs* moderate-severe; H), and inferior vena cava dilation (I). The diagonal line signifies a random guess, equivalent to a model with no discriminative ability between positive and negative outcomes. The values on each graph are the areas under the receiver operating characteristic curve (95% CI) for the external dataset.

dataset B, and 0·92 (0·03) for internal test dataset C. For the external test dataset D, the AUC for the classification of left ventricular ejection fraction was 0·92 (95% CI 0·90–0·95) at the 40% cutoff and for the classification of tricuspid regurgitation velocity was 0·85 (0·83–0·87) at the 2·8 m/s cutoff. The AUCs for the classifiers of valvular diseases at the none-mild versus moderate-severe cutoff were 0·89 (0·86–0·92) for mitral regurgitation, 0·83 (0·78–0·88) for aortic stenosis, 0·83 (0·79–0·87) for aortic regurgitation, 0·86 (0·67–1·00) for mitral stenosis, 0·92 (0·89–0·94) for tricuspid regurgitation, and 0·86 (0·82–0·90) for pulmonary regurgitation. The AUC for interior vena cava dilation was 0·85 (0·81–0·89).

The two classifiers with AUCs greater than 0·90 in the external dataset were for left ventricular ejection fraction

	Internal test dataset			External test dataset
	Dataset A	Dataset B	Dataset C	Dataset D
Overall mean AUC (SD)	0·89 (0·04)	0·90 (0·04)	0·92 (0·03)	0·87 (0·03)
Left ventricular ejection fraction	0·92 (0·88–0·96)	0·92 (0·88–0·97)	0·93 (0·88–0·98)	0·92 (0·90–0·95)
Tricuspid regurgitant velocity	0·88 (0·83–0·93)	0·85 (0·81–0·89)	0·88 (0·82–0·94)	0·85 (0·83–0·87)
Valvular heart disease				
Mitral regurgitation	0·92 (0·88–0·95)	0·95 (0·90–0·99)	0·89 (0·79–1·00)	0·89 (0·86–0·92)
Aortic stenosis	0·86 (0·80–0·92)	0·91 (0·82–1·00)	0·91 (0·80–1·00)	0·83 (0·78–0·88)
Aortic regurgitation	0·83 (0·76–0·91)	0·85 (0·72–0·97)	0·89 (0·70–1·00)	0·83 (0·79–0·87)
Mitral stenosis	0·89 (0·75–1·00)	0·94 (0·61–1·00)	..	0·86 (0·67–1·00)
Tricuspid regurgitation	0·93 (0·89–0·97)	0·92 (0·88–0·96)	0·93 (0·85–1·00)	0·92 (0·89–0·94)
Pulmonary regurgitation	0·85 (0·70–1·00)	0·91 (0·69–1·00)	0·95 (0·79–1·00)	0·86 (0·82–0·90)
Inferior vena cava dilation	0·94 (0·88–1·00)	0·88 (0·80–0·95)	0·94 (0·88–0·99)	0·85 (0·81–0·89)

Data are mean (95% CI) unless otherwise stated. AUC=area under the receiver operating characteristic curve.

Table 2: **AUC for the internal and external test datasets**

	Accuracy, %	Sensitivity, %	Specificity, %	Positive predictive value, %	Negative predictive value, %
Left ventricular ejection fraction	86% (85–87)	82% (75–87)	86% (85–88)	24% (21–28)	99% (98–99)
Tricuspid regurgitant velocity	75% (73–76)	83% (80–87)	73% (71–75)	36% (33–38)	96% (95–97)
Valvular heart disease					
Mitral regurgitation	85% (84–86)	82% (76–87)	85% (84–86)	26% (22–29)	99% (98–99)
Aortic stenosis	73% (71–74)	79% (69–87)	72% (71–74)	8% (6–9)	99% (99–100)
Aortic regurgitation	68% (67–70)	88% (81–92)	67% (66–69)	12% (10–14)	99% (99–99)
Mitral stenosis	90% (89–91)	83% (36–100)	90% (89–91)	2% (0–3)	100% (100–100)
Tricuspid regurgitation	83% (82–85)	87% (83–91)	83% (82–84)	30% (27–34)	99% (98–99)
Pulmonary regurgitation	69% (68–71)	91% (84–95)	68% (67–70)	10% (9–12)	99% (99–100)
Inferior vena cava dilation	86% (85–88)	73% (65–81)	87% (86–88)	20% (16–24)	99% (98–99)

Data are % (95% CI).

Table 3: **Model performance for the external dataset**

at a 40% cutoff (AUC 0·92, accuracy 86%, sensitivity 82%, and specificity 86%) and tricuspid regurgitation (AUC 0·92, accuracy 83%, sensitivity 87%, and specificity 83%). Most results, including mean AUC, from the external test dataset are slightly lower than those from the internal test datasets. Detailed AUCs for the internal and external test datasets are shown in figure 1 and table 2 and precision-recall AUCs can be found in the appendix (p 15). AUCs for the inpatient and outpatient subgroups (pp 3, 16) and chest radiography first or echocardiography first subgroups (pp 3, 17) are given in the appendix, along with receiver operating characteristic curves (p 8) and AUCs (p 18) for the alternative cutoffs for each disease. Sensitivity, specificity, accuracy, positive predictive value, and negative predictive value for external test dataset D are shown in table 3; corresponding values for internal test datasets A (p 19), B (p 20), and C (p 21) can be found in the appendix. AUCs by sex (p 22) and age (p 23) for external test dataset D are also shown in the appendix, in addition to confusion matrices for each primary classifier for the external dataset (p 9) and label correlations among the ground truth labels (p 10).

We generated saliency maps for radiographs in the external test dataset. Representative maps for four classifiers (left ventricular ejection fraction less than 40%, tricuspid regurgitant velocity greater than 2·8 m/s, mitral regurgitation none-mild *vs* moderate-severe disease, and aortic stenosis none-mild *vs* moderate-severe disease) and an overall averaged saliency map of primary classifiers are shown in figure 2; all other saliency maps are provided in the appendix (p 11). All maps showed saliency focused on the cardiac shadow; the tricuspid regurgitant velocity and mitral regurgitation classifier maps also showed saliency over the hilar and lung regions.

Discussion

We developed and evaluated a deep-learning model for estimating the classification of transthoracic echocardiograms from chest radiographs. The mean AUC for the overall model performance was 0·87 in the external test dataset.

To the best of our knowledge, this study is the first to create and validate a deep learning-based classification model for cardiac functions and valvular heart disease using chest radiographs from multiple institutions. Deep-learning models that are trained and tested on a single dataset can be prone to overfitting, such that the final model works well only for images in the trained dataset.[19,33,34] To prevent overfitting and to confirm robustness and generalisability, the test set should ideally be collected from different facilities.[34,35] In two previous studies[21,22] of deep-learning models that detect single cardiac valvular diseases, only single-centre datasets were used. Our model included a larger cohort from several institutions, was more robust, and achieved higher AUCs than either of these studies. Our model could also classify basic echocardiographic results, including left ventricular ejection fraction, tricuspid regurgitant velocity, and several valvular diseases.

The overall saliency map showed strongest saliency over the cardiac shadow and slight saliency over the

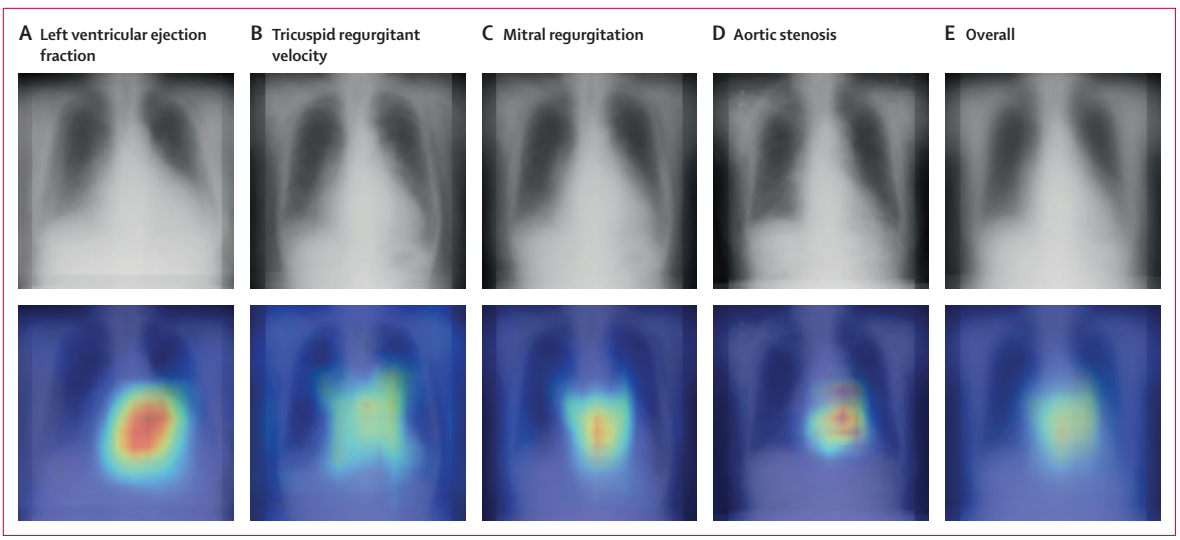

Figure 2: **Representative saliency maps for the external test dataset**
The top images show the averaged chest radiographs from the test dataset, generated by averaging the top 10% of true positive images—ie, those identified by the model as most likely to be showing features of disease. The number of images averaged for each classifier varies by the number of true positives identified by the model. The bottom images show the average of the saliency maps associated with the chest radiographs overlaid onto the averaged chest radiographs. Images are shown for the following classifiers: left ventricular ejection fraction less than 40% (A), tricuspid regurgitant velocity greater than 2·8 m/s (B), mitral regurgitation (none-mild *vs* moderate-severe; C), aortic stenosis (none-mild *vs* moderate-severe; D), and the average images of all primary classifiers (E). All saliency maps mainly highlight the cardiac shadow area. For images relevant to other classifiers, see appendix (p 11).

hilum. These results suggest that chest radiographs have intrinsic features that can help to classify cardiac functions and valvular diseases. The saliency map for the left ventricular ejection fraction classifier was elliptical in shape and covered the cardiac shadow; the long axis of this ellipse corresponded to the short axis of the cardiac shadow. This short axis is different from the horizontal distances measured by the cardiothoracic ratio,[5] suggesting that it might be a more important measure than the cardiothoracic ratio for estimating the left ventricular ejection fraction.

This model has some advantages over echocardiography-based evaluation. First, very little time is needed to take a radiograph and apply the model to it, which is advantageous for patients who are unable to lie still for the time required for echocardiography. Our model instantly classifies all variables that indicate cardiac function.[27] In addition, the small size of the input image (512×512 pixel) means that the system requirements to run the model are low; any computer used in daily clinical practice should be able to implement the model and rapidly output the results. Second, the model can be used by physicians at any time, making it useful in areas where echocardiography specialists are not available[36,37] or in emergency situations in hospitals during the night when an echocardiography technician is not working. In such situations, a provisional evaluation of the patient's cardiac function using chest radiography could be useful until an echocardiogram can be recorded. Two studies reported physicians' performance in classifying left ventricular ejection fraction from radiological images that show indications of disease.[6,8] In one study, cardiomegaly was reported as the most useful indicator of low left ventricular ejection fraction, but the sensitivity was low at 51% (95% CI 43–60) and the specificity was 79% (71–85).[6] Another study also found cardiomegaly to be the most sensitive indicator of heart failure, but again, the sensitivity (64%) and specificity (71%) were low.[8] Our model showed higher performance for the estimation of left ventricular ejection fraction than these two studies. After its initial implementation into medical systems, our model could be used without any specialised skills—unlike echocardiography, which requires specialised skills that are in short supply.[17,18] Finally, there are numerous clinical situations in which a chest radiograph has already been taken and stored in medical records[3] and could provide information on cardiac function when necessary, without the need for additional testing. Models can constitute a valid diagnostic method when used appropriately in the clinic; however, because the positive and negative predictive values depend on the prevalence of disease in the cohort in which the test is used,[34,38] it would be advisable to select patients with other indications of heart disease or to adjust the model thresholds according to the prevalence of disease in the cohort under study.

This study has several limitations. Because the data were collected retrospectively from multiple institutions, further validation with prospectively acquired test datasets from cohorts with various disease prevalences is

desirable. Additionally, we used echocardiography results as the ground truth. However, establishing left ventricular ejection fraction by indices obtained from cardiovascular MRI might be more accurate. Selection bias could be present in our model because, among patients who had echocardiographic data, we selected only those who also had chest radiographs. All chest radiographs were taken in the standing posteroanterior view, so tuning might be necessary to apply this model to patients who cannot stand. Our model is for classification, not for estimating the precise value of the echocardiographic results.

We developed a deep-learning model to classify echocardiography results using the features inherent in chest radiographs. The number of patients with cardiac disease will continue to increase worldwide.[12-15] Our artificial intelligence model, when used in appropriate situations, could have a complementary role to transthoracic echocardiography for cardiac assessments in the future.

Contributors

All authors contributed to the study conception and design. DU, SE, AY, AI, and TT collected, verified, and analysed the data and had access to the raw data. TM developed the model. DU wrote the first draft of the manuscript, which was edited by SLW and TM. All processes were supervised by AY, TS, MS, DF, and YM. All authors had access to all data and read and approved the final manuscript. DU had final responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

The study protocol and metadata are available from DU. Chest radiographs are not available at this time, as they have been withheld by the hospitals that participated in the trials to protect participant privacy.

Acknowledgments

We thank Morimoto Hospital (Osaka, Japan), Kashiwara Municipal Hospital (Kashiwara, Japan), and Habikino Medical Center (Habikino, Japan) for providing the data for this study.

References

1 Kevles BH, Kelsey CA. Naked to the bone: medical imaging in the twentieth century. *Phys Today* 1997; **50:** 56.

2 United Nations Scientific Committee on the Effects of Atomic Radiation. Sources and effects of ionizing radiation. 2000. https://www.unscear.org/unscear/uploads/documents/publications/UNSCEAR_2000_Annex-D.pdf (accessed Jan 1, 2023).

3 Mettler FA Jr, Bhargavan M, Faulkner K, et al. Radiologic and nuclear medicine studies in the United States and worldwide: frequency, radiation dose, and comparison with other radiation sources—1950–2007. *Radiology* 2009; **253:** 520–31.

4 Raoof S, Feigin D, Sung A, Raoof S, Irugulpati L, Rosenow EC 3rd. Interpretation of plain chest roentgenogram. *Chest* 2012; **141:** 545–58.

5 Danzer CS. The cardiothoracic ratio: an index of cardiac enlargement. *Am J Med Sci* 1919; **157:** 513.

6 Badgett RG, Mulrow CD, Otto PM, Ramírez G. How well can the chest radiograph diagnose left ventricular dysfunction? *J Gen Intern Med* 1996; **11:** 625–34.

7 Philbin EF, Garg R, Danisa K, et al. The relationship between cardiothoracic ratio and left ventricular ejection fraction in congestive heart failure. *Arch Intern Med* 1998; **158:** 501–06.

8 Mueller-Lenke N, Rudez J, Staub D, et al. Use of chest radiography in the emergency diagnosis of acute congestive heart failure. *Heart* 2006; **92:** 695–96.

9 Yancy CW, Jessup M, Bozkurt B, 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. *Circulation* 2013; **128:** e240–327.

10 Galiè N, Humbert M, Vachiery J-L, 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* 2016; **37:** 67–119.

11 Otto CM, Nishimura RA, Bonow RO, et al. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* 2021; **143:** e72–227.

12 Nkomo VT, Gardin JM, Skelton TN, Gottdiener JS, Scott CG, Enriquez-Sarano M. Burden of valvular heart diseases: a population-based study. *Lancet* 2006; **368:** 1005–11.

13 Savarese G, Lund LH. Global public health burden of heart failure. *Card Fail Rev* 2017; **3:** 7–11.

14 Chen J, Li W, Xiang M. Burden of valvular heart disease, 1990–2017: results from the Global Burden of Disease Study 2017. *J Glob Health* 2020; **10:** 020404.

15 Roger VL. Epidemiology of heart failure: a contemporary perspective. *Circ Res* 2021; **128:** 1421–34.

16 Mohebi R, Chen C, Ibrahim NE, et al. Cardiovascular disease projections in the United States based on the 2020 census estimates. *J Am Coll Cardiol* 2022; **80:** 565–78.

17 Wood PW, Choy JB, Nanda NC, Becher H. Left ventricular ejection fraction and volumes: it depends on the imaging method. *Echocardiography* 2014; **31:** 87–100.

18 De Geer L, Oscarsson A, Engvall J. Variability in echocardiographic measurements of left ventricular function in septic shock patients. *Cardiovasc Ultrasound* 2015; **13:** 19.

19 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521:** 436–44.

20 Hinton G. Deep learning—a technology with the potential to transform health care. *JAMA* 2018; **320:** 1101–02.

21 Ueda D, Ehara S, Yamamoto A, et al. Development and validation of artificial intelligence-based diagnosis method of mitral regurgitation from chest radiographs. *Radiol Artif Intell* 2022; **4:** e210221.

22 Ueda D, Yamamoto A, Ehara S, et al. Artificial intelligence-based detection of aortic stenosis from chest radiographs. *Eur Heart J Digit Health* 2021; **3:** 20–28.

23 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; **351:** h5527.

24 Zoghbi WA, Enriquez-Sarano M, Foster E, et al. Recommendations for evaluation of the severity of native valvular regurgitation with two-dimensional and Doppler echocardiography. *J Am Soc Echocardiogr* 2003; **16:** 777–802.

25 Baumgartner H, Hung J, Bermejo J, et al. Echocardiographic assessment of valve stenosis: EAE/ASE recommendations for clinical practice. *J Am Soc Echocardiogr* 2009; **22:** 1–23.

26 Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In: Kamalika C, Ruslan S, eds. Proceedings of the 36th international conference on machine learning. *PMLR* 2019; **97:** 6105–14.

27 Zhang M, Zhou Z. A review on multi-label learning algorithms. *IEEE Trans Knowl Data Eng* 2014; **26:** 1819–37.

28 Müller SG, Hutter F. TrivialAugment: tuning-free yet state-of-the-art data augmentation. 2021 IEEE/CVF international conference on computer vision. 754–62.

29 Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* 2019; **32:** 8026–37.

30 Matsumoto T, Walston SL, Miki Y, Ueda D. Nervus: a comprehensive deep learning classification, regression, and prognostication tool for both medical image and clinical data analysis. *arXiv* 2022; published online Dec 12. https://doi.org/10.48550/arXiv.2212.11113 (preprint).

31 Otto CM, Nishimura RA, Bonow RO, et al. 2020 ACC/AHA Guideline for the management of patients with valvular heart disease: executive summary: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* 2021; **143:** e35–71.

32 Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. 2018 IEEE winter conference on applications of computer vision. 839–47.

33 Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. *AJR Am J Roentgenol* 2017; **208:** 754–60.

34 Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology* 2018; **286:** 800–09.

35 Bluemke DA, Moy L, Bredella MA, et al. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers—from the *Radiology* editorial board. *Radiology* 2020; **294:** 487–89.

36 Mollura DJ, Culp MP, Pollack E, et al. Artificial intelligence in low- and middle-income countries: innovating global health radiology. *Radiology* 2020; **297:** 513–20.

37 Frija G, Blažić I, Frush DP, et al. How to improve access to medical imaging in low- and middle-income countries? *eClinicalMedicine* 2021; **38:** 101034.

38 Salmi LR, Coureau G, Bailhache M, Mathoulin-Pélissier S. To screen or not to screen: reconciling individual and population perspectives on screening. *Mayo Clin Proc* 2016; **91:** 1594–605.