JAMA Cardiology | **Original Investigation**

Prediction of Pulmonary to Systemic Flow Ratio in Patients With Congenital Heart Disease Using Deep Learning–Based Analysis of Chest Radiographs

Shuhei Toba, MD; Yoshihide Mitani, MD, PhD; Noriko Yodoya, MD; Hiroyuki Ohashi, MD; Hirofumi Sawada, MD, PhD;
Hidetoshi Hayakawa, MD, PhD; Masahiro Hirayama, MD, PhD; Ayano Futsuki, MD; Naoki Yamamoto, MD; Hisato Ito, MD, PhD;
Takeshi Konuma, MD, PhD; Hideto Shimpo, MD, PhD; Motoshi Takao, MD, PhD

[+] **Supplemental content**

IMPORTANCE Chest radiography is a useful noninvasive modality to evaluate pulmonary blood flow status in patients with congenital heart disease. However, the predictive value of chest radiography is limited by the subjective and qualitive nature of the interpretation. Recently, deep learning has been used to analyze various images, but it has not been applied to analyzing chest radiographs in such patients.

OBJECTIVE To develop and validate a quantitative method to predict the pulmonary to systemic flow ratio from chest radiographs using deep learning.

DESIGN, SETTING, AND PARTICIPANTS This retrospective observational study included 1031 cardiac catheterizations performed for 657 patients from January 1, 2005, to April 30, 2019, at a tertiary center. Catheterizations without the Fick-derived pulmonary to systemic flow ratio or chest radiography performed within 1 month before catheterization were excluded. Seventy-eight patients (100 catheterizations) were randomly assigned for evaluation. A deep learning model that predicts the pulmonary to systemic flow ratio from chest radiographs was developed using the method of transfer learning.

MAIN OUTCOMES AND MEASURES Whether the model can predict the pulmonary to systemic flow ratio from chest radiographs was evaluated using the intraclass correlation coefficient and Bland-Altman analysis. The diagnostic concordance rate was compared with 3 certified pediatric cardiologists. The diagnostic performance for a high pulmonary to systemic flow ratio of 2.0 or more was evaluated using cross tabulation and a receiver operating characteristic curve.

RESULTS The study included 1031 catheterizations in 657 patients (522 males [51%]; median age, 3.4 years [interquartile range, 1.2-8.6 years]), in whom the mean (SD) Fick-derived pulmonary to systemic flow ratio was 1.43 (0.95). Diagnosis included congenital heart disease in 1008 catheterizations (98%). The intraclass correlation coefficient for the Fick-derived and deep learning–derived pulmonary to systemic flow ratio was 0.68, the log-transformed bias was 0.02, and the log-transformed precision was 0.12. The diagnostic concordance rate of the deep learning model was significantly higher than that of the experts (correctly classified 64 of 100 vs 49 of 100 chest radiographs; $P = .02$ [McNemar test]). For detecting a high pulmonary to systemic flow ratio, the sensitivity of the deep learning model was 0.47, the specificity was 0.95, and the area under the receiver operating curve was 0.88.

CONCLUSIONS AND RELEVANCE The present investigation demonstrated that deep learning–based analysis of chest radiographs predicted the pulmonary to systemic flow ratio in patients with congenital heart disease. These findings suggest that the deep learning–based approach may confer an objective and quantitative evaluation of chest radiographs in the congenital heart disease clinic.

Author Affiliations: Department of Thoracic and Cardiovascular Surgery, Mie University Graduate School of Medicine, Tsu, Mie, Japan (Toba, Futsuki, Yamamoto, Ito, Konuma, Shimpo, Takao); Department of Pediatrics, Mie University Graduate School of Medicine, Tsu, Mie, Japan (Mitani, Yodoya, Ohashi, Sawada, Hayakawa, Hirayama); Mie Prefectural General Medical Center, Yokkaichi, Mie, Japan (Shimpo).

Corresponding Authors: Shuhei Toba, MD, Department of Thoracic and Cardiovascular Surgery (s.toba.jp @gmail.com), and Yoshihide Mitani, MD, PhD, Department of Pediatrics (ymitani@clin.medic.mie-u.ac.jp), Mie University Graduate School of Medicine, 2-174 Edobashi, Tsu, Mie, Japan 514-8507.

A lthough the pulmonary to systemic flow ratio is an important hemodynamic parameter for clinical decision-making, including judgment of operative indication in patients with congenital heart disease, a precise evaluation of this parameter requires cardiac catheterization, which is still an invasive procedure and cannot be performed as daily practice. Previous studies have shown that less-invasive modalities, such as echocardiography and magnetic resonance imaging, are useful for assessing the pulmonary to systemic flow ratio.[1-3] However, echocardiography can be used only for patients with simple congenital heart disease, and magnetic resonance imaging can be used for a limited number of patients because of its magnetism and examination time. Chest radiography is another less-invasive method that is used to assess pulmonary blood flow status, but the evaluation is subjective and qualitative, and little is known about clinicians' diagnostic performance using chest radiography.

Deep learning, a branch of machine learning, is a method for automated image interpretation. Since deep learning became a rapidly evolving approach for computer vision, deep learning–based analysis has been recently performed in several medical settings using imaging modalities, such as diagnosis of lung cancer from chest computed tomography[4] or chest radiography,[5] diagnosis of tuberculosis[6] or pneumonia[7] from chest radiography, and assessment of the position of the endotracheal tube on a chest radiograph.[8] Previous studies have shown that a deep learning–based approach can be used to recognize diseases or findings objectively in various imaging modalities, and one of the studies showed the possibility that deep learning-based analysis may outperform clinicians.[7]

Given the potential capability of deep learning shown in the previous studies, we hypothesized that deep learning-based analysis can predict the pulmonary to systemic flow ratio from chest radiographs quantitatively in patients with congenital heart disease. In this study, we developed and validated a deep learning–based method to predict the pulmonary to systemic flow ratio from chest radiographs of patients with congenital heart disease.

Methods

Data Sets

We included 1689 consecutive cardiac catheterizations performed for 907 patients during the period from January 1, 2005, to April 30, 2019, in the Department of Pediatric Cardiology in Mie University Hospital, Mie, Japan. Catheterizations with no chest radiograph that was taken within 1 month before cardiac catheterization and catheterizations with no pulmonary to systemic flow ratio measurement were excluded. Among eligible patients, 78 patients (100 catheterizations) were randomly assigned for evaluation (evaluation group), and catheterizations performed for the other patients were used for training (training group). The pulmonary to systemic flow ratio was calculated by using the Fick method.[9] Radiographs were obtained using UD150L-40, Rad Speed Safire, and MUX-10 OHJ (Shimadzu Corp); Sirius 130HP (Hitachi Medical Corp); Beneo Fx 1T2P (Fujifilm Medical Co Ltd); and FCR XL-2, FCR

Key Points

Question Does deep learning–based analysis of chest radiographs predict the pulmonary to systemic flow ratio in patients with congenital heart disease?

Findings This retrospective observational study using 1031 cardiac catheterizations and chest radiographs showed that the pulmonary to systemic flow ratio predicted by a deep learning model was significantly correlated with the values calculated using the Fick method (intraclass correlation coefficient, 0.68). The diagnostic concordance rate of the model was significantly higher than that of experts (64 of 100 cases vs 49 of 100 cases).

Meaning These results may allow clinicians to quantify otherwise qualitative and subjective findings of pulmonary vascularity in chest radiographs.
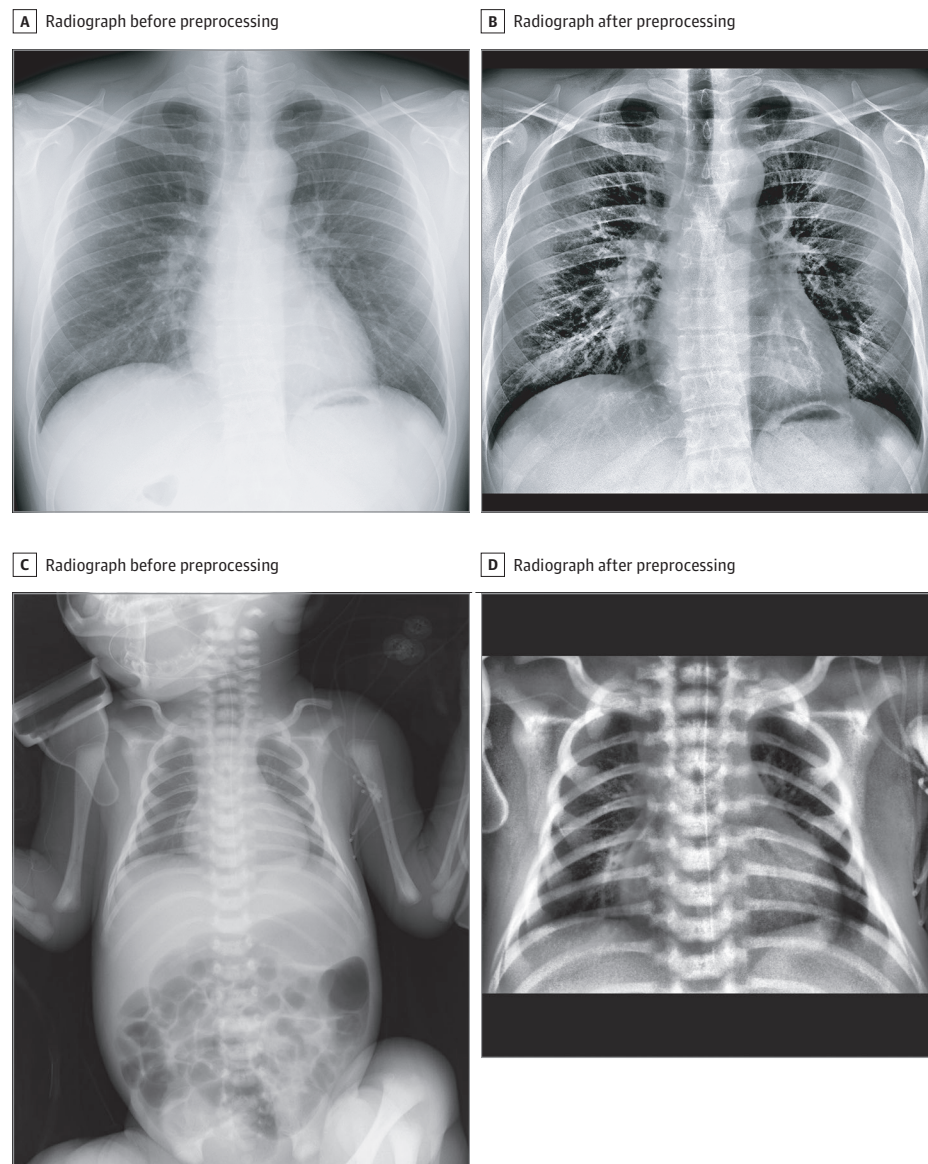
Verocity T, FCR Verocity U, Calneo HC SQ (SE), Calneo C 1417 Wireless, and Calneo Flex (Fujifilm Medical Co Ltd). As preparation of the data (**Figure 1**), we manually trimmed all radiographs to remove extrathoracic images and applied a contrast-limited adaptive histogram equalization. To square the images, we added black rectangles on the edge of each image. The images were converted into an 8-bit depth and 512 × 512-pixel matrix and Portable Networks Graphics format.

This study was approved by the Mie University Hospital Institutional Review Board. Waiver of informed consent was granted by the institutional review board (in accordance with Japanese government ethics guidelines for biomedical research, informed consent is waived if the study does not include personal information, has a retrospective design, and is carefully performed under the instruction of the ethics committee of the institute).

Transfer Learning

We developed a deep learning model using a method of transfer learning based on previous reports.[4-8] As a base model, we adopted pretrained Inception-v3, a 311-layer deep convolutional neural network model developed by Google,[10] which had been already trained with everyday color images from ImageNet (http://www.image-net.org/). Because the base model was designed for classifying 1000 categories of everyday objects, we replaced the final layer of the model with a global average pooling layer and a fully connected layer to make the model output a pulmonary to systemic flow ratio value, which is a continuous variable. We then retrained the model with the data sets obtained in our hospital to predict the pulmonary to systemic flow ratio from chest radiographs. Weights in the first 133 layers, which correspond to the first 5 inception blocks, were frozen, and weights in the other layers were retrained with our data. The following parameters were used for training: 100 epochs; the loss function, 2-way random and single-score intraclass correlation coefficient $ICC(2,1)$[11]; an optimizer, stochastic gradient descent; and a learning rate of 0.01. All images were randomly augmented by a rotation of less than 20°, a shift of less than 10%, and a horizontal flip. The accuracy of the prediction model was assessed using a 10-fold cross-validation. Ten models with a minimum loss of validation were obtained.

Figure 1. Preprocessing of the Input Images

A Radiograph before preprocessing

B Radiograph after preprocessing

C Radiograph before preprocessing

D Radiograph after preprocessing

Two examples of radiographs before preprocessing are shown, with corresponding images after preprocessing.

Evaluation and Data Analysis

We evaluated the obtained models using the evaluation group. The mean pulmonary to systemic flow ratio value predicted by the 10 models was used as the output of the whole retrained Inception-v3. Agreement between the Fick-derived pulmonary to systemic flow ratio and the deep learning-derived pulmonary to systemic flow ratio was analyzed by use of the root mean square deviation, ICC(2,1), and Bland-Altman analysis.[12] To compare the diagnostic performance of our model with the performance of clinicians at different levels of their experience, both the Fick-derived and deep learning-derived pulmonary to systemic flow ratios of the evaluation group were classified into the following 4 categories: low (ratio, <0.9), normal (ratio, ≥0.9 to <1.5), high (ratio, ≥1.5 to <2.0), and very high (ratio, ≥2.0), while 3 certified pediatric cardiologists (H.O., H. Sawada, and H.H.) and 3 pediatric cardi-

ology fellows (N.Y. and 2 others) classified the same chest radiographs into the 4 categories using deidentified original chest radiographs. When the answers differed among the 3 experts or fellows, the majority answer was adopted as their classification if there was a majority, otherwise the median of the answers was used. The concordance rate between the classification based on the Fick method and the classification based on deep learning was compared with that of the Fick-derived classification and the experts' or fellows' classification using the McNemar test with the Yates continuity correction. All *P* values were from 2-sided tests and results were deemed statistically significant at *P* < .05. We also assessed the capability of our model and the clinicians to diagnose a high pulmonary to systemic flow ratio of 2.0 or more using the cross tabulation and the area under the receiver operating characteristic curve (AUC). To examine how the experts interpret

Table. Characteristics of Patients

Characteristic	Training Group (n = 931)	Evaluation Group (n = 100)
Age, median (IQR), y	3.4 (1.2-7.9)	4.6 (1.4-11.4)
Weight, median (IQR), kg	12.6 (7.8-23.1)	15.4 (8.6-34.5)
Male sex, No. (%)	469 (50)	53 (53)
Pulmonary to systemic flow ratio, mean (SD)	1.42 (0.98)	1.43 (0.62)
PVR, mean (SD), Wood units · m^2	1.87 (1.62)[a]	1.81 (1.42)[a]
Mean PAP ≥25 mm Hg, No. (%)	143 (16)[a]	14 (14)[a]
PAWP ≥15 mm Hg, No. (%)	27 (3)[a]	3 (3)[a]
Disease, No. (%)		
Noncyanotic congenital heart disease	476 (51)	69 (69)
Single valve stenosis	23 (3)	4 (4)
Septal defects	345 (37)	46 (46)
No intervention	253 (27)	35 (35)
Palliated	48 (5)	4 (4)
Repaired	44 (5)	7 (7)
Coarctation of the aorta	10 (1)	2 (2)
Patent ductus arteriosus	66 (7)	12 (12)
No intervention	56 (6)	12 (12)
Repaired	10 (1)	0
Congenitally corrected transposition of the great arteries	2 (0.2)	0
Other	9 (1)	2 (2)
Cyanotic congenital heart disease	455 (49)	31 (31)
Dextro-transposition of the great arteries	37 (4)	2 (2)
No intervention	0	0
Palliated	4 (0.4)	0
Repaired	33 (4)	2 (2)
Single ventricle defects	214 (23)	12 (12)
No intervention	9 (1)	0
Palliated	124 (13)	6 (6)
Repaired	81 (9)	6 (6)
Tetralogy of Fallot	135 (15)	12 (12)
No intervention	26 (3)	5 (5)
Palliated	52 (6)	4 (4)
Repaired	57 (6)	3 (3)
Aortic disease	12 (1)	2 (2)
No intervention	0	0
Palliated	8 (0.9)	0
Repaired	4 (0.4)	2 (2)
Double-outlet right ventricle	33 (4)	1 (1)
No intervention	5 (0.5)	0
Palliated	17 (2)	0
Repaired	11 (1)	1 (1)
Pulmonary atresia with intact ventricular septum	22 (2)	2 (2)
No intervention	1 (0.1)	1 (1)
Palliated	13 (1)	1 (1)
Repaired	8 (0.9)	0
Other	2 (0.2)	0
No congenital heart disease	21 (2)	3 (3)

Abbreviations: IQR, interquartile range; PAP, pulmonary arterial pressure; PAWP, pulmonary arterial wedge pressure; PVR, pulmonary vascular resistance.

[a] PVR was calculated in 897 catheterizations in the training group and 99 catheterizations in the evaluation group, mean PAP was measured in 910 catheterizations in the training group and 99 catheterizations in the evaluation group, and PAWP was measured in 902 catheterizations in the training group and 99 catheterizations in the evaluation group.

pulmonary blood flow status on chest radiographs, they were asked to evaluate how much a finding based on chest radiographs was associated with the interpretation of pulmonary blood flow. Five findings (the size of the heart, the left second cardiac arch, the silhouette of the right lower pulmonary artery, the peripheral pulmonary vascularity, and the opacity of the lung fields) were evaluated on a scale of 4 (1, not important; 2, less important; 3, important; 4, very important). To visualize how our model predicted the pulmonary to systemic flow ratio from a chest radiograph in each case, we performed gradient-weighted class activation mapping (grad-CAM)[13] and guided backpropagation.[14] We also adopted an activation maximization technique[15] to visualize the deep neural network of our model by generating imaginary pictures that include characteristics of an increased or decreased pulmonary to systemic flow ratio.

Hardware and Software

Data preparation, training, and evaluation were performed on a Z800 workstation (HP) with a graphical processing unit of GeForce GTX 1080Ti (Nvidia) using Python 3.5 and its libraries, including Pydicom, PyPng, and scikit-image,[16] Keras 2.0.9 with TensorFlow[17] backend, and Keras-vis. Statistical analysis was performed using SPSS, version 25 (IBM Corp).

Results

A total of 1031 catheterizations performed for 657 patients were eligible for this study (eFigure 1 in the Supplement). Of 1031 catheterizations, 1008 (98%) were performed for patients with a diagnosis of congenital heart disease, and 217 (21%) were performed for patients younger than 12 months. The mean (SD) Fick-derived pulmonary to systemic flow ratio was 1.43 (0.95). The Fick-derived pulmonary to systemic flow ratio was 1.00 in 291 catheterizations (28%), and chest radiography was performed within 3 days before cardiac catheterization in 995 catheterizations (97%).

In the evaluation group, no surgical or catheter intervention that might have changed the pulmonary to systemic flow ratio was performed, and no circulatory drug treatment, including diuretics, vasodilators, or digoxin, was initiated between the chest radiography and the cardiac catheterization. Nineteen of 100 catheterizations (19%) in the evaluation group were performed for patients younger than 12 months, and all but 1 of these patients had chest radiography performed within 3 days before the catheterizations. The Fick-derived pulmonary to systemic flow ratio was 2.0 or more in 170 of 931 catheterizations (18%) in the training group and in 19 of 100 catheterizations (19%) in the evaluation group. The characteristics of the patients in the training and evaluation groups are summarized in the **Table**.

The scatterplot and the Bland-Altman plot between the Fick-derived and the deep learning–derived pulmonary to systemic flow ratio are shown in **Figure 2**. The root mean square deviation was 0.45, and the ICC(2,1) was 0.68 ($P < .001$). The percentage error was 0.62, and the relative error was less than 20% in 51 of 100 chest radiographs (51%). Bland-Altman analysis and

Figure 2. Performance of the Deep Learning Model
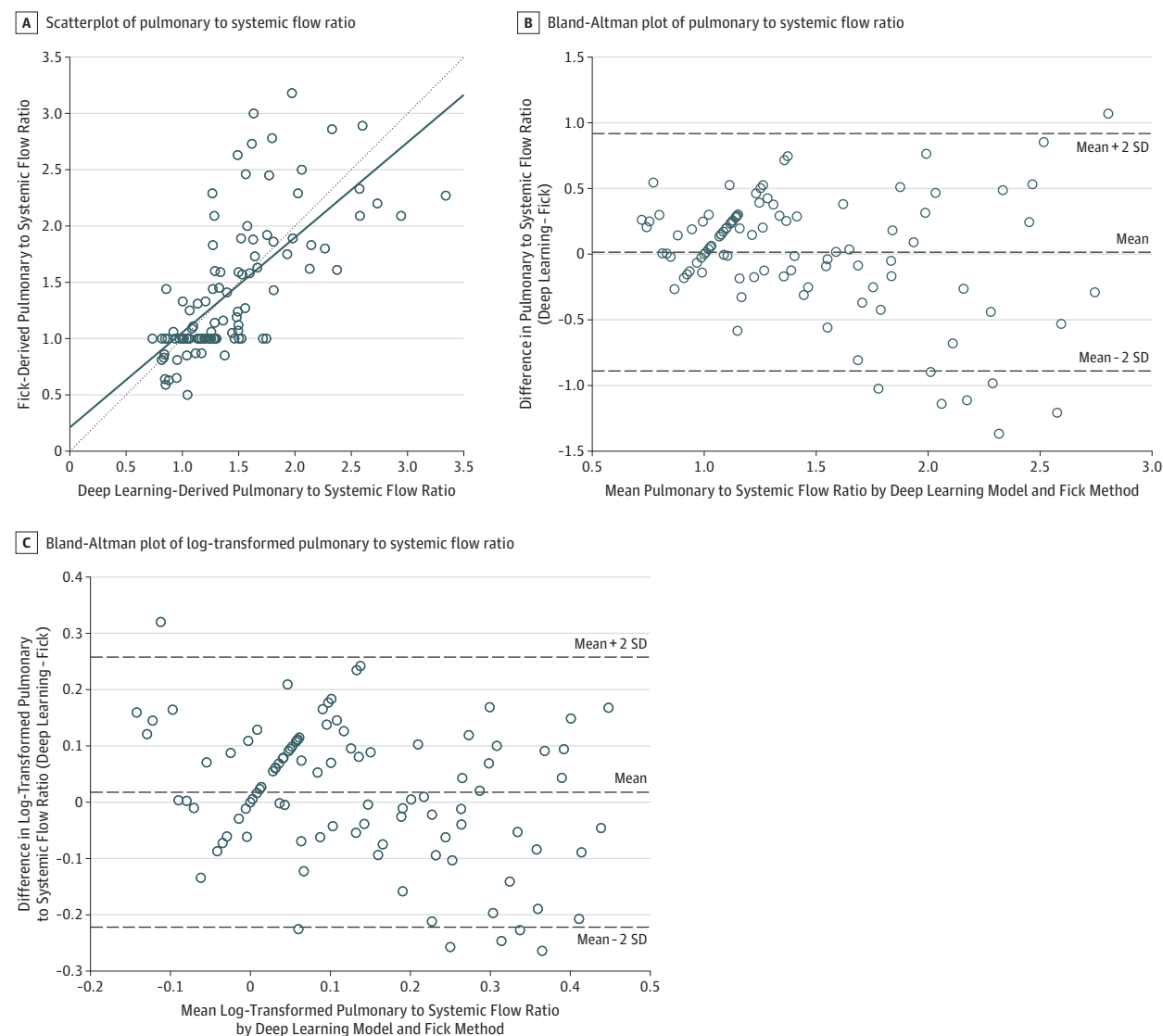
A, Scatterplot of Fick-derived pulmonary to systemic flow ratio and deep learning-derived pulmonary to systemic flow ratio in the evaluation group (100 catheterizations and chest radiographs). The solid line represents the linear regression fit. The dotted line represents the line of equality.
B, Bland-Altman plot comparing Fick-derived pulmonary to systemic flow ratio and deep learning-derived pulmonary to systemic flow ratio. Dashed lines indicate the bias (mean) and the limits of agreement (mean [2 SD]).
C, Bland-Altman plot comparing log-transformed Fick-derived pulmonary to systemic flow ratio and deep learning-derived pulmonary to systemic flow ratio. Dashed lines indicate the bias (mean) and the limits of agreement (mean [2 SD]). Each circle represents a single data point (of 100 sets) derived from the Fick-derived and deep learning-derived pulmonary to systemic flow ratio.

its plot (Figure 2B) showed that the bias was 0.01, the precision was 0.45, the upper limit of agreement (defined as mean [2 SD]) was 0.92, and the lower limit of agreement (defined as mean [2 SD]) was −0.89. We also performed the Bland-Altman analysis on the log-transformed data as recommended in a previous article[12] because the differences between the Fick-derived and deep learning-derived pulmonary to systemic flow ratios increased linearly with the mean pulmonary to systemic flow ratio. On the log scale, the bias was 0.02, the precision was 0.12, the upper limit of agreement was 0.26, and the lower limit of agreement was −0.22. After antilog transformation, the bias was 1.04, the upper limit of agreement was 1.81,

and the lower limit of agreement was 0.60. The Bland-Altman plot (Figure 2C) showed that the deep learning-derived pulmonary to systemic flow ratio was more different from the Fick-derived pulmonary to systemic flow ratio in cases with a pulmonary to systemic flow ratio of less than 0.79 (antilog of −0.1) and in cases with a pulmonary to systemic flow ratio of more than 1.58 (antilog of 0.2) and that the deep learning-derived pulmonary to systemic flow ratio tends to be lower in patients with an increased pulmonary to systemic flow ratio and higher in patients with a decreased pulmonary to systemic flow ratio.

The diagnostic concordance rate between deep learning-derived and Fick-derived classifications was 64 of 100 (64%),

Figure 3. Diagnostic Performance for a High Pulmonary
to Systemic Flow Ratio
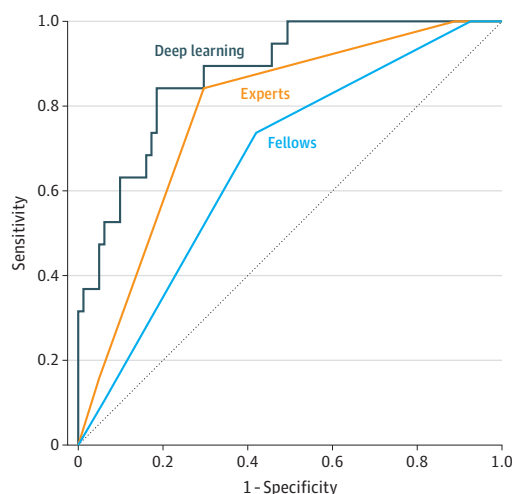
Receiver operating characteristic curves of the deep learning model, the
experts, and the fellows for detecting a high pulmonary to systemic flow ratio
of 2.0 or more. For the deep learning model, the accuracy was 0.86, the
positive predictive value was 0.69, the negative predictive value was 0.89,
sensitivity was 0.47, specificity was 0.95, and the area under the curve was
0.88. For the experts, the accuracy was 0.80, the positive predictive value was
0.43, the negative predictive value was 0.83, sensitivity was 0.16, specificity
was 0.95, and the area under the curve was 0.78. For the fellows, the accuracy
was 0.78, the positive predictive value was 0.29, the negative predictive
value was 0.82, sensitivity was 0.11, specificity was 0.94, and the area under
the curve was 0.67.

which was significantly higher than that of experts and fel-
lows (experts, 49 of 100 [49%]; P = .02; fellows, 40 of 100 [40%];
P = .001). The confusion matrix of our model and the clini-
cians are shown in eFigure 2 in the Supplement. Regarding the
detection of a high pulmonary to systemic flow ratio of 2.0 or
more, our model's accuracy was 0.86. Sensitivity was 0.47,
specificity was 0.95, and AUC was 0.88. For the experts, accu-
racy was 0.80, sensitivity was 0.16, specificity was 0.95, and AUC
was 0.78; for the fellows, accuracy was 0.78, sensitivity was 0.11,
specificity was 0.94, and AUC was 0.67. Receiver operating char-
acteristic curves and other statistics are shown in **Figure 3**.

The mean importance value of the findings of chest radio-
graphs evaluated by the experts were as follows: size of the
heart, 1.7; left second cardiac arch, 2.0; silhouette of the right
lower pulmonary artery, 2.3; peripheral pulmonary vascular-
ity, 3.7; and opacity of the lung fields, 3.3. Grad-CAM (**Figure 4**)
showed that our model recognized the area in the lung fields
and the area around the heart as valuable for predicting an
increased pulmonary to systemic flow ratio, while no particu-
lar focus pattern was found for predicting a decreased pulmo-
nary to systemic flow ratio. Guided backpropagation (Figure 4)
showed that our model recognized structures mainly in the
lung fields when predicting the pulmonary to systemic flow
ratio. Activation maximization (eFigure 3 in the Supplement)
showed that coarse nodules represented an increased pulmo-
nary to systemic flow ratio, while no well-described pattern
was found for a decreased pulmonary to systemic flow ratio.

In addition, to investigate whether our model could de-
tect changes in the pulmonary to systemic flow ratio in an
individual patient, we reviewed 34 repeated cardiac catheter-
izations performed for 12 patients. Changes in the deep learn-
ing-derived pulmonary to systemic flow ratio in each patient
were generally consistent with the changes in the Fick-
derived pulmonary to systemic flow ratio, with some range of
variance (eTable 1 in the Supplement). To investigate poten-
tial factors associated with the mismatch between deep
learning-derived and Fick-derived pulmonary to systemic flow
ratios, we reviewed 6 patients in whom the difference be-
tween the Fick-derived and deep learning-derived pulmo-
nary to systemic flow ratios was over the limits of agreement
(eTable 2 in the Supplement). Tracheomalacia and abnormal
position of the diaphragm were noted in such patients.

Discussion

Our study, which investigated whether deep learning-based
analysis of chest radiographs predicts the pulmonary to sys-
temic flow ratio in patients with congenital heart disease,
showed that our model could predict the pulmonary to sys-
temic flow ratio from chest radiographs and that its diagnos-
tic performance was higher than that of experts.

The methods in this study were developed using transfer
learning of a deep convolutional neural network. Deep
learning-based analysis of chest radiographs has been per-
formed in previous studies in which transfer learning was
shown to be an effective way to develop a deep learning
model for the interpretation of chest radiographs.[5-8] In trans-
fer learning, a model that has been pretrained on larger data
sets, such as images of everyday objects, is applied to another
task that becomes the focus. This approach is effective for
analyzing medical images because they have similar ele-
ments to everyday images, such as edges and blobs. We
adopted Inception-v3, which had been pretrained on
ImageNet, a data set of 1.2 million everyday color images, for
analysis of chest radiographs, considering the performance of
the model and our computational resource. Augmentation
of data sets has been shown to be an effective way to improve
the performance of deep learning.[18] In previous studies, aug-
mentation methods, including random rotation, contrast-
limited adaptive histogram equalization, horizontal flipping,
and random cropping, were used to improve the perfor-
mance of deep learning.[5,6,8] In this study, we applied hori-
zontal flipping, rotation, and shift. We applied contrast-
limited adaptive histogram equalization to all images
because they had been taken in various radiographic condi-
tions. As a loss function, we used ICC(2,1) because it is one of
the most important differentiable statistics for the evaluation
of agreement between 2 methods for measuring a quantita-
tive variable.[19]

We compared the deep learning-derived pulmonary to sys-
temic flow ratio with the Fick-derived pulmonary to systemic
flow ratio in the evaluation group. The ICC(2,1) showed that the
deep learning-derived pulmonary to systemic flow ratio was
significantly correlated with the Fick-derived pulmonary to

Figure 4. Results of Gradient-Weighted Class Activation Mapping (Grad-CAM) and Guided Backpropagation
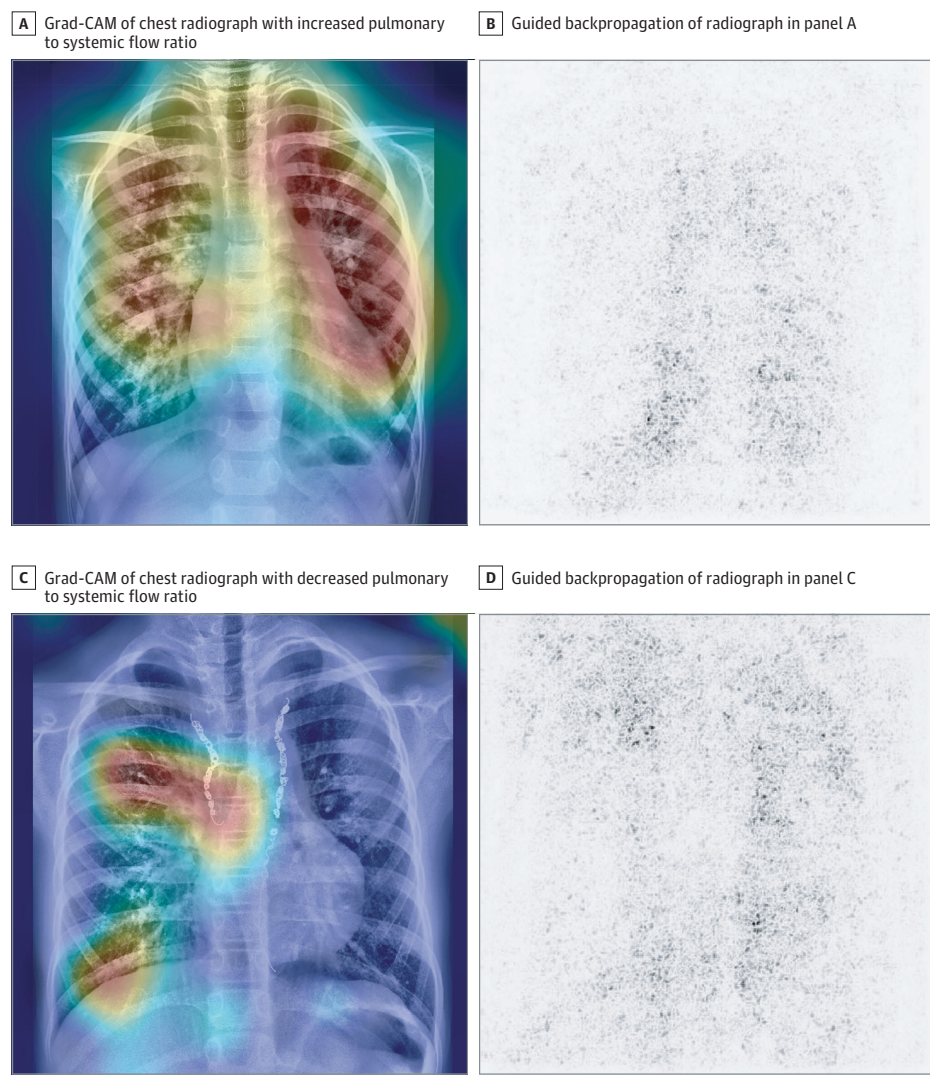
A Grad-CAM of chest radiograph with increased pulmonary to systemic flow ratio

B Guided backpropagation of radiograph in panel A

C Grad-CAM of chest radiograph with decreased pulmonary to systemic flow ratio

D Guided backpropagation of radiograph in panel C

A, Chest radiograph with increased pulmonary to systemic flow ratio (Fick-derived pulmonary to systemic flow ratio, 2.89; deep learning–derived pulmonary to systemic flow ratio, 2.60) was visualized using grad-CAM, which represents the area (yellow and red) that the deep learning model considered important for predicting an increased or decreased pulmonary to systemic flow ratio. B, Chest radiograph from panel A visualized with guided backpropagation, which emphasizes edges that were recognized by the model and shows how the model recognized chest radiographs. C, Chest radiograph with a decreased pulmonary to systemic flow ratio (Fick-derived pulmonary to systemic flow ratio, 0.86; deep learning–derived pulmonary to systemic flow ratio, 0.84) was visualized using grad-CAM. D, Chest radiograph from panel C visualized with guided backpropagation.

systemic flow ratio (Figure 2A). The ICC(2,1) of 0.68 represents good clinical significance according to a previous report.[20] Bland-Altman analysis showed that, after antilog transformation, the bias was 1.04, the upper limit of agreement was 1.81, and the lower limit of agreement was 0.60. Thus, the mean difference between the 2 methods was 4%, and the deep learning-derived pulmonary to systemic flow ratio may differ from the Fick-derived pulmonary to systemic flow ratio by 40% below to 81% above in 95% of cases. The Bland-Altman plot (Figure 2C) showed that the deep learning-derived pulmonary to systemic flow ratio was more different from the Fick-derived pulmonary to systemic flow ratio in cases with a pulmonary to systemic flow ratio of less than 0.79 (antilog of −0.1) and in cases with a pulmonary to systemic flow ratio of more than 1.58 (antilog of 0.2) and that the deep learning-derived pulmonary to systemic flow ratio tends to be lower in patients with an increased pulmonary to systemic flow ratio and higher in patients with a decreased pulmonary to systemic flow ratio. We considered that this might be caused by the distribution of the

training group. The cardiac catheterizations with a normal pulmonary to systemic flow ratio were dominant in our data set, which may have allowed our model to predict a value closer to 1.00. It may be effective to include more cases with increased or decreased pulmonary to systemic flow ratios for developing a better model. In an additional study, we confirmed that the deep learning-derived pulmonary to systemic flow ratio was generally consistent with the Fick-derived pulmonary to systemic flow ratio in repeated catheterizations. These findings suggest that our model may detect a temporal change in the pulmonary to systemic flow ratio in a patient, which should be verified in a further study.

In terms of classifying chest radiographs into 4 classes depending on pulmonary blood flow, our study showed that deep learning-based analysis outperformed the clinicians whose diagnostic performance was correlated with their level of experience. Although several studies reported a high level of performance for the deep learning model that is comparable to or even superior to that of clinicians, the mechanisms

involved in achieving such a performance are unclear.[7,21-23] In our study, because the data sets had not been labeled by clinicians but by another examination (catheterization), it is possible that our model was able to recognize features that are not consistently recognized by clinicians. In addition, the AUC of our model for detecting a high pulmonary to systemic flow ratio of 2.0 or more was 0.88, while that of the experts was 0.78. This finding suggests the potential capability of our model to help physicians to justify intervention. The clinicians in our study were not provided with information (including diagnosis or past medical history) other than the chest radiograph, which was not similar to the real-world clinical setting. This change may have allowed us to underestimate the clinicians' capability to evaluate pulmonary blood flow on chest radiographs. If patients are confined to a specific disease, clinicians as well as the deep learning model may be able to predict the pulmonary to systemic flow ratio better.

One of the problems of deep learning is its difficulty in showing the reasoning behind outputting a value.[24] To understand our model's recognition of chest radiographs, we adopted 3 visualization methods. Grad-CAM and guided backpropagation (Figure 4) showed that our model recognized and focused on the structures in the lung fields and the area around the heart when predicting the pulmonary to systemic flow ratio in each chest radiograph. Through the activation maximization technique (eFigure 3 in the Supplement), the model generated the imaginary pictures that were considered to maximize or minimize the output value. The results of activation maximization showed that findings indicating coarse nodule-like structures, which were possibly equivalent to the pulmonary vasculature, were associated with a high pulmonary to systemic flow ratio. According to a standard textbook of pediatric cardiology, findings that suggest increased pulmonary blood flow include enlarged pulmonary arteries that extend into the lateral third of the lung field, increased vascularity to the lung apices, and wider diameter of the right pulmonary artery visible in the right hilum compared with the diameter of the trachea, whereas findings that suggest decreased pulmonary blood flow include a small hilum, black lung field, and small or thin vessels.[25] Our findings focusing on how experts interpreted chest radiographs were consistent with those descriptions. The recognition of our deep learning model is thereby consistent with that of the experts and the textbooks, except our model focused on the area around the heart as well as the structures in the peripheral lung fields. In addition to understanding the characteristics of our model, because grad-CAM and guided backpropagation can be applied to each chest radiograph, it is possible that one can use these 2 visualization techniques in clinical settings as well.

Because of our model's capability to quantitatively predict the pulmonary to systemic flow ratio from chest radiographs and to outperform clinicians, the present proof-of-concept study suggests that there may be hidden information in routine imaging tests that deep learning can identify, adding clinical value.

Limitations

Several limitations of this study should be acknowledged. First, this is a single-institution study. The methods may therefore perform less well in other situations because of the lack of an external validation sample in the present study. Second, because the present evaluation group included patients with a variety of congenital heart diseases, including complex congenital heart diseases, that were indicated for diagnostic catheterization, the chest radiography findings may have been influenced by associated lesions. Third, because the results of the visualization methods represent only limited aspects of our model, the predicting process of our model may not have been completely revealed. Fourth, the performance of our model may not have been properly evaluated because of the different characteristics of the diagnostic methods. Chest radiography may reflect pulmonary blood flow over a longer period, whereas the measurement of cardiac catheterization is instantaneous and influenced by sedation, dehydration, or choice of mixed venous saturation. Fifth, the time difference between chest radiography and cardiac catheterization in the evaluation group may have influenced the performance of our model, especially for infants, owing to the temporal change in pulmonary vascular resistance. However, such effects may be minimal because the time difference was minimal (<3 days in 99% of all catheterizations and 95% of infantile catheterizations) and no surgical or catheter intervention or initiation of circulatory drugs was performed between the studies. Sixth, considering the sensitivity and the specificity for diagnosing a high pulmonary to systemic flow ratio, our model seems to be relatively specific but not very sensitive. However, in the future, such an evaluation of our model could be improved by increasing the training data. Seventh, the predictive value of our model might be limited for patients with several conditions, including tracheomalacia and abnormal position of the diaphragm, as shown in eTable 2 in the Supplement, which may be associated with the limited number of such patients in the training group. Eighth, similarly, the deep learning–based evaluation of patients with idiopathic pulmonary arterial hypertension or Eisenmenger syndrome will be limited in the present algorithm because neither of those conditions was included in the present study sample.

Conclusions

The present study showed that deep learning–based analysis of chest radiographs could predict the pulmonary to systemic flow ratio quantitatively and objectively, which may confer an opportunity to quantify otherwise qualitative and subjective findings of pulmonary vascularity in the clinical setting. Further studies are warranted to improve the performance of the model and to understand how the model predicts the pulmonary to systemic flow ratio.

REFERENCES

1. Sanders SP, Yeager S, Williams RG. Measurement of systemic and pulmonary blood flow and QP/QS ratio using Doppler and two-dimensional echocardiography. *Am J Cardiol*. 1983;51(6):952-956. doi:10.1016/S0002-9149(83)80172-6

2. Kitabatake A, Inoue M, Asao M, et al. Noninvasive evaluation of the ratio of pulmonary to systemic flow in atrial septal defect by duplex Doppler echocardiography. *Circulation*. 1984;69(1): 73-79. doi:10.1161/01.CIR.69.1.73

3. Beerbaum P, Körperich H, Barth P, Esdorn H, Gieseke J, Meyer H. Noninvasive quantification of left-to-right shunt in pediatric patients: phase-contrast cine magnetic resonance imaging

compared with invasive oximetry. *Circulation*. 2001; 103(20):2476-2482. doi:10.1161/01.CIR.103.20.2476

4. Hua KL, Hsu CH, Hidayati SC, Cheng WH, Chen YJ. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *Onco Targets Ther*. 2015;8:2015-2022.

5. Ausawalaithong W, Marukatat S, Thirach A, Wilaiprasitporn T. Automatic lung cancer prediction from chest x-ray images using deep learning approach. https://arxiv.org/abs/1808.10858. Published August 31, 2018. Accessed November 29, 2018.

6. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*. 2017;284(2):574-582. doi:10.1148/radiol.2017162326

7. Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning. https://arxiv.org/abs/1711. 05225. Updated December 25, 2017. Accessed November 29, 2018.

8. Lakhani P. Deep convolutional neural networks for endotracheal tube position and X-ray image classification: challenges and opportunities. *J Digit Imaging*. 2017;30(4):460-468. doi:10.1007/ s10278-017-9980-7

9. Taggart NW, Cabalka AK. Cardiac catheterization and angiography. In: Allen HD, Shaddy RE, Penny DJ, Feltes TF, Cetta F, eds. *Moss and Adams' Heart Disease in Infants, Children, and Adolescents, Including the Fetus and Young Adult*. 9th ed. Philadelphia, PA: Wolters Kluwer Health; 2016: 444-449.

10. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. https://arxiv.org/abs/1512.00567. Updated December 11, 2015. Accessed November 10, 2018.

11. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86 (2):420-428. doi:10.1037/0033-2909.86.2.420

12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-310. doi:10.1016/S0140-6736(86)90837-8

13. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. https://arxiv.org/abs/1610.02391. Updated March 21, 2017. Accessed November 10, 2018.

14. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striving for simplicity: the all convolutional net. https://arxiv.org/abs/1412.6806.

Updated April 13, 2015. Accessed November 10, 2018.

15. Erhan D, Bengio Y, Courville A, Vincent P. Visualizing higher-layer features of a deep network: technical report. https://pdfs.semanticscholar.org/ 65d9/94fb778a8d9e0f632659fb33a082949a50d3. pdf. Published June 9, 2009. Accessed August 19, 2019.

16. van der Walt S, Schönberger JL, Nunez-Iglesias J, et al; scikit-image Contributors. scikit-image: Image processing in Python. *PeerJ*. 2014;2:e453. doi:10.7717/peerj.453

17. Abadi M, Barham P, Chen J, et al. Tensorflow: a system for large-scale machine learning. https://arxiv.org/abs/1605.08695. Updated May 31, 2016. Accessed November 10, 2018.

18. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Paper presented at: Advances in Neural Information Processing Systems 25; December, 2012; Lake Tahoe, Nevada. http://papers.nips.cc/ paper/4824-imagenet-classification-with-deep- convolutional-neural-networ. Accessed December 6, 2019.

19. Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. *Comput Biol Med*. 1989;19(1): 61-70. doi:10.1016/0010-4825(89)90036-X

20. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284-290. doi:10.1037/1040-3590. 6.4.284

21. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22): 2402-2410. doi:10.1001/jama.2016.17216

22. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542 (7639):115-118. doi:10.1038/nature21056

23. Wang P, Xiao X, Glissen Brown JR, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat Biomed Eng*. 2018;2(10):741-748. doi:10.1038/s41551-018-0301-3

24. Yosinski J, Clune J, Nguyen A, Fuchs T, Lipson H. Understanding neural networks through deep visualization. https://arxiv.org/abs/1506.06579. Published June 22, 2015. Accessed August 19, 2019.

25. Myung KP. *Pediatric Cardiology for Practitioners*. 5th ed. Philadelphia, PA: Mosby Elsevier; 2008.