Using Deep Learning to Detect Pediatric Congenital Heart Disease by Chest Radiography

Rui Chen
Guangdong Provincial People's Hospital
Xinyi Wu
Guangdong Provincial People's Hospital
Taylor Chung
UCSF Benioff Children's Hospital
Longwei Sun
Shenzhen Children's Hospital
Paul Schoenhagen
Cleveland Clinic
Zhicheng Du
Sun Yat-Sen University
Weitao Ye
Guangdong Provincial People's Hospital
Zekun Tan
Guangdong Provincial People's Hospital
Tingyu Li
Guangdong Provincial People's Hospital
Hongyu Wu
Guangdong Provincial People's Hospital
Weidong Qiao
Guangdong Provincial People's Hospital
Xiaomei Zhong
Guangdong Provincial People's Hospital
Shushui Wang
Guangdong Provincial People's Hospital
Xi Chen
Sichuan Provincial Hospital for Women and Children
Zhuliang Yu
South China University of Technology
Hui Liu (liuhuijiujiu@gmail.com)
Guangdong Provincial People's Hospital https://orcid.org/0000-0003-0148-0125

Research Article

Keywords: Congenital heart disease, Chest radiography, Diagnostic tool, Deep learning

DOI: https://doi.org/10.21203/rs.3.rs-491844/v1

Abstract

Purpose: To detect congenital heart disease (CHD) patients is a challenging task in low-income areas with limited resources. Chest radiography is usually available in these areas, but the diagnostic accuracy needs to be improved. The aim of the study was to establish a deep learning-based diagnostic tool for paediatric CHD patients using chest radiography.

Methods: Totally 11,105 chest radiographs from 11,105 paediatric patients in two centres were labeled, and a convolutional neural network (CNN) was trained for diagnosing CHD in the dataset from Hospital 1. To demonstrate the generalizability and the clinical usefulness of CNN, the accuracy in test set from two centres were both reported. We also trained another CNN to evaluate pulmonary blood flow (PBF) on chest radiographs of the patients. Moreover, the predictive results of both CNN were compared to the decisions of three experienced radiologists (range 10–35 years) without any patient information.

Results: The accuracy of CNN for detecting CHD was 81.4% (95% confidence interval, [78.3%–84.2%]) in Hospital 2, while the accuracy was 85.6% (95% confidence interval, [84.0%–87.1%]) in Hospital 1. The overall accuracy of CNN for evaluating PBF was 78.9% (95% confidence interval, [74.3%–83.0%]). Compared to three experienced radiologists, CNN showed statistically non-inferiority at a 5% margin for the accuracy ($P < 0.025$) in detecting CHD and evaluating PBF.

Conclusions: The deep learning-based diagnostic tool we developed could effectively detect CHD and evaluate for PBF on chest radiography. It can potentially aid the radiologists and cardiologists in the future.

Introduction

Congenital heart disease (CHD) is a significant cause of paediatric morbidity and mortality. The prevalence and incidence of CHD are reported to be approximately 3.7–4.3 % and 1 % per live births, respectively [1]. The pediatric CHD patients would benefit from timely detection, treatment and intervention. In highly developed countries in the world, the combination of fetal echocardiography, pulse oximetry and routine clinical assessment has been considered as an effective and suitable method to detect CHD [2; 3]. However, many children with simple but serious CHD in low-income, low-resource areas cannot be detected timely enough without those diagnostic methods and potentially create preventable suboptimal clinical outcomes [4; 5].

In areas with limited resources, chest radiography is a helpful and common method to detect paediatric CHD patients. The radiologists could evaluate the heart size, heart shape, and pulmonary blood flow (PBF) from the chest radiography. However, the accuracy of chest radiography has been reported from 30% to 78 % in different clinical settings in distinguishing children with CHD from normal children [6; 7]. Diagnostic accuracy varied even among experienced paediatric radiologists who were blinded to the patients’ history. Moreover, the majority of the local radiologists and cardiologists in these rural areas without the benefit of working experience in major tertiary medical centers may not be familiar with CHD, which may lead to the more serious misdiagnosis of those children with CHD.

Deep learning has the potential to help human experts with disease diagnosis and management, especially in areas with limited resources. In Diagnostic Radiology, convolutional neural network (CNN) was shown to automatically and effectively diagnose paediatric pneumonia on chest radiographs with high accuracy [8]; CNN was also used for estimating prognosis and Pulmonary to Systemic Flow Ratio in CHD patients [9; 10]. Therefore, we hypothesize CNN can help radiologists and cardiologists with the diagnosis and detection of CHD on chest radiographs of pediatric CHD patients. To verify this hypothesis, we utilized a large dataset of labeled chest radiographs in Hospital 1 to train the CNN, then prove the predictive performance of CNN in test set from Hospital 1 and Hospital 2. To provide further assistance to the radiologists and cardiologists, we also trained another CNN to classify PBF of the patients in chest radiography in the present study.

Methods

Study Population

A total of 11,105 paediatric patient under the age of 14 years from two centers in different cities from January 2014 to March 2020 were included in the study. The dataset from Hospital 1 consisted of 5165 patients with CHD and 5240 normal controls, while the dataset from Hospital 2 consisted of 350 patients with CHD and 350 normal controls. Final diagnoses were determined from clinical information, computed tomography, echocardiography or surgical reports. The CHD patients with increased PBF was defined as the ratio of pulmonary-to-systemic blood flow (Qp:Qs) of greater than 1.5 derived from cardiac catherization [11]. The patients were diagnosed to have decreased PBF by a combined evaluation of oxygen saturation and morphological features investigated with computed tomography angiography [12]. The normal controls were pediatric patients with other diseases, but without cardiopulmonary disease based on clinical history and physical examination. No patient information such as age, sex, clinical history, physical examination results, or other diagnostic examinations results were included with the chest radiographs.

Chest Radiography Acquisition

All chest radiographs were performed as part of routine clinical care, and each patient had only one chest radiograph. These chest radiographs were acquired by multiple different radiology technologists utilizing various radiographic equipment manufactured by different vendors (including GE Healthcare, Philips Healthcare, Siemens Healthineers, Shimadzu, and Fujifilm). All chest radiographs were retrieved from the PACS of the respective hospitals with the highest quality possible in original size without any specific window setting.

Deep Learning and Statistics

The deep learning system we developed consisted of two separate CNNs in the present study (Fig 1). The first CNN was used for determining whether the patient has CHD (label: positive or negative). The second CNN was used to evaluate the PBF of the patients (label: normal, increased, or decreased). Both CNN were trained limited only to the chest radiographs without any other information such as clinical history.

CNN1 for detecting CHD

A total of 11,105 chest radiographs in two centers were included in this deep learning experiment. The 10,405 chest radiographs in Hospital 1 were randomly separated into training set (CHD: 3165; normal control: 3240), validation set (CHD: 1000; normal control: 1000) and test set (CHD: 1000; normal control: 1000). To evaluate the ability of CNN to generalize across populations and screening settings, 700 chest radiographs (CHD: 350; normal control: 350) in Hospital 2 were recognized as the external test set.

We retrained a VGG-16 network as the final model, which was pretrained on approximately 1.28 million images from the ImageNet Large Scale Visual Recognition Challenge [13]. The training process was performed by stochastic gradient descent per step using an Adam Optimizer; the loss was calculated using cross-entropy. The activation function of the output layer was a sigmoid layer. To monitor the training procedure, the cross-entropy loss in validation set was recognized as the main metric of selecting the best model in the training procedure. The early stopping technique was used to avoid the over-fitting problem and improve the training efficiency.

The accuracy of the test sets in both Hospital 1 and Hospital 2 were utilized as the main metric to evaluate the predictive performance of CNN (Figs 2 and 3). Moreover, the final decisions of the CNN on the 2000 chest radiographs in test set in Hospital 1 and 700 chest radiographs in Hospital 2 were compared with the interpretations made by three experienced radiologists (a 35-year experienced radiologist, a 25-year experienced radiologist and a 10-year experienced radiologist). These chest radiographs were individually reviewed by experienced radiologists blinded to patient history, case composition and ratio. The confusion matrices of CNN and experienced radiologists were constructed.

CNN2 for evaluating PBF

In this part of the present study, a total of 3604 chest radiographs in Hospital 1 from November 2017 to March 2020 were included. The chest radiographs were randomly separated into training set (normal control: 1453; increased PBF: 856; decreased PBF: 575), validation set (normal control: 120; increased PBF: 120; decreased PBF: 120), and test set (normal control: 120; increased PBF: 120; decreased PBF: 120).

We fine-tuned the weights of the upper convolutional layers, and retrained the weights of the fully connected layers utilizing a pre-trained VGG-16 architecture [14]. The activation function of the output layer was performed by using a Softmax layer. The training process was performed by stochastic gradient descent per step using an Adam Optimizer; the loss was also calculated using cross-entropy; and early stopping technique was also performed.

The accuracy of the test set was recognized as the main metric in this multi-class classification problem. The output decisions of the CNN on a test set of 360 chest radiographs were also compared with the interpretations by the same three experienced radiologists as described previously. The confusion matrices of CNN and experienced radiologists were also constructed (Fig 4).

To demonstrate the non-inferiority in diagnostic performance of CNN with each experienced radiologist in test set, the non-inferiority test was performed, with a pre-defined margin of -5% [15]. A one-sided P value <0.025 was recognized as statistically significant.

The hardware and software used in our experiments were as follows: Desktop computer with Intel (Intel Corporation, Santa Clara, California, USA) i7 3.6Ghz CPU, 16G DDR3 RAM, Nvidia (Nvidia Corporation, Santa Clara, California, USA) Titan X, Ubuntu 16.04, Keras with TensorFlow backend. The ROC plots were generated by statistical software R version 3.4.1 (R Foundation, Vienna, Austria) and pROC package.

Results

Totally 11,105 patients under the age of 14 years from two centers were enrolled in this study. The distribution of chest radiographs from CHD patients in two centers was shown at table 1.

CNN1 for detecting CHD

The accuracy of CNN for detecting CHD was 81.4% (95% confidence interval, [78.3%–84.2%]) in test set in Hospital 2 (Fig 2), while the accuracy was 85.6% (95% confidence interval, [84.0%–87.1%]) in test set in Hospital 1 (Fig 3). The sensitivity in Hospital 2, sensitivity in Hospital 1, specificity in Hospital 2, specificity in Hospital 1 were 92.9%, 91.8%, 70.0%, 79.4%, respectively (Fig 2 and Fig 3).

In test set in Hospital 2, the accuracy of the three experienced radiologists was 81.7% (95% confidence interval, [78.7%–84.5%]) for the 35-year experienced radiologist, 80.3% (95% confidence interval, [77.1%–83.2%]) for the 25-year experienced radiologist, and 73.9% (95% confidence interval, [70.4%–77.1%]) for 10-year experienced radiologist, respectively. Compared to those three radiologists, CNN showed non-inferiority at a 5% margin for the accuracy (-0.3%, 95% confidence interval [-4.3%, 3.8%]; P = 0.017), (+1.1%, 95% confidence interval [-3.0%, 5.3%]; P = 0.002), (+7.6%, 95% confidence interval [3.2%, 11.9%]; P < 0.001), respectively.

In test set in Hospital 1, the accuracy of the three experienced radiologists was 82.4% (95% confidence interval, [80.6%–84.0%]) for the 35-year experienced radiologist, 79.8% (95% confidence interval, [78.0%–81.5%]) for the 25-year experienced radiologist, and 72.0% (95% confidence interval, [70.0%–74.0%]) for 10-year experienced radiologist, respectively. Compared to those three radiologists, CNN showed non-inferiority at a 5% margin for the accuracy (+3.2%, 95% CI [0.9%, 5.5%]; $P < 0.001$), (+5.8%, 95% CI [3.5%, 8.1%]; $P < 0.001$), (+13.6%, 95% CI [11.1%, 16.1%]; $P < 0.001$), respectively.

CNN2 for evaluating PBF

The overall accuracy of CNN for evaluating PBF was 78.9% (95% confidence interval, [74.3%–83.0%]) in test set, which exceeded the accuracy of the three experienced radiologists at 71.7% (95% confidence interval, [66.7%–76.3%]) for the 35-year experienced radiologist, 69.2% (95% confidence interval, [64.1%–73.9%]) for the 25-year experienced radiologist, and 56.1% (95% confidence interval, [50.8%–61.3%]) for the 10-year experienced radiologist. Compared with those three radiologists, CNN showed non-inferiority at a 5% margin for the accuracy (+7.2%, 95% CI [0.9%, 13.5%]; $P < 0.001$), (+9.7%, 95% CI [3.4%, 16.1%]; $P < 0.001$), (+22.7%, 95% CI [16.1%, 29.4%]; $P < 0.001$), respectively.

In detecting decreased PBF, the true positive rate of CNN2 was 70.8% (85/120), statistically higher than that achieved by the 35-year experienced radiologist 55.8% (67/120) (+15.0%, 95% confidence interval, [3.0%–27.0%], $P<0.001$), 25-year experienced radiologist 48.3% (58/120) (+22.5%, 95% confidence interval, [10.4%–34.6%], $P<0.001$), 10-year experienced radiologist 21.7% (26/120) (+49.2%, 95% confidence interval, [38.2%–60.1%], $P<0.001$), respectively.

Discussion

In the present study, we utilized a large dataset of paediatric chest radiographs and trained the CNN for detecting paediatric CHD. To provide further assistance for the radiologists, we also trained another CNN to evaluate PBF of the patients from the chest radiographs. Our results indicated that CNN performed well in the detection of CHD (accuracy, 81.4% (95% confidence interval, [78.3%–84.2%]) in the external test set), and the evaluation of PBF of paediatric patients (accuracy, 78.9% (95% confidence interval, [74.3%–83.0%])) on chest radiographs in test set.

The deep learning based diagnostic tool we have developed has the potential to be widely used in hospitals in small cities and rural areas with limited availability of advanced imaging modalities such as echocardiography, CT, and MRI, to help the radiologists and cardiologists with the diagnosis CHD from chest radiographs. Classically, chest radiographs had been established as a valuable tool in the evaluation of patients with heart murmur or the suspicion of CHD [16]. It had been taught that specific CHDs can have classic manifestations on chest radiography. Typical examples include a boot-shaped heart [11] for tetralogy of Fallot [17], “egg on a string” for D-transposition of the great arteries [18], and a snowman appearance for total anomalous pulmonary venous connection [17], etc. Furthermore, chest radiography is useful for the assessment of PBF in CHD patients; and PBF is one of the most important findings for differential diagnosis of CHD, especially when availability of cross-sectional imaging is limited. PBF represented the severity of the hemodynamic status in CHD patients, and it usually affected the treatment of CHD patients. According to the standard management guideline in CHD, non-hemodynamically significant patients are recommended for medical treatment, whereas interventional or surgical treatment is recommended in patients with significant hemodynamic change [19]. However, the role of chest radiography has been limited. Diagnostic accuracy varied with even among experienced paediatric radiologists who were blinded to the patients' history. Laya et al studied the accuracy of chest radiography in detecting pediatric CHD patients. The accuracy ranged from 71.7% to 82.4% for 5 experienced radiologists in distinguishing normal from CHD patients [7], which was similar with the results in the present study. Therefore, we could imagine that, the majority of the radiologists and cardiologists in rural areas without the benefit of working experience in major tertiary medical centers may lead to the more serious misdiagnosis of those children with CHD.

One of the advantages of CNN was to address the deficiencies of low accuracy, and aid the cardiologists or cardiologists. In the present study, CNN performed well in detecting paediatric CHD, the accuracy was 85.6% (95% confidence interval, [84.0%–87.1%]) in test set in Hospital 1. In addition, CNN showed statically non-inferiority at a 5% margin for the accuracy (+3.2%, 95% CI [0.9%, 5.5%]; P < 0.001), (+5.8%, 95% CI [3.5%, 8.1%]; P < 0.001), (+13.6%, 95% CI [11.1%, 16.1%]; P < 0.001), respectively. To prove the ability of CNN to generalize across populations and screening settings, 700 chest radiographs from Hospital 2 were recognized as the external test set. The accuracy of CNN for detecting CHD was 81.4% (95% CI, [78.3%–84.2%]) in Hospital 2. Also, CNN showed non-inferiority at a 5% margin for the accuracy (-0.3%, 95% CI [-4.3%, 3.8%]; P = 0.017), (+1.1%, 95% CI [-3.0%, 5.3%]; P = 0.002), (+7.6%, 95% CI [3.2%, 11.9%]; P < 0.001), respectively. Thus, our CNN provides similar interpretation of the chest radiographs compared with experienced radiologists in diagnosing CHD.

Recently, Toba et al trained a CNN to detect the CHD patients with significantly increased PBF (Qp:Qs >2.0), and the CNN correctly classified 64 of 100 chest radiographs. However, the accuracy or true positive rate of detecting patients with decreased PBF was not mentioned [10]. In Tumkosit et al, the true positive rate of chest radiography for experienced radiologists or cardiologists to detect decreased PBF was much lower than to detect increased PBF [11; 20]. Therefore, to detect CHD patients with decreased PBF should be regarded as an important goal and a difficult task. To address these problems, the CNN we have trained could effectively classify the patients with increased PBF, the patients with and decreased PBF, and normal control. In the present study, the true positive rate of detecting patients with decreased PBF of CNN was 70.8% , achieved by the 35-year experienced radiologist 55.8% (+15.0%, 95% confidence interval, [3.0%–27.0%], P<0.001), 25-year experienced radiologist 48.3% (+22.5%, 95% confidence interval, [10.4%–34.6%], P<0.001), 10-year experienced radiologist 21.7% (+49.2%, 95% confidence interval, [38.2%–60.1%], P<0.001), respectively. It turned out that CNN could provide a potential to address the low-true-positive-rate detection problems in decreased PBF, probably because CNN may have a stronger ability in image identification or classification.

Though the sample size of paediatric CHD patients was over 10,000, the present study was still a double-center study in detecting CHD and a single-center study in evaluating PBF. The patients in different groups were neither age-matched nor gender-matched. The CNN we had trained seemed to be relatively sensitive but not very specific. These problems could be addressed by a mult-centre studies in the future.

In evaluating PBF in CHD patients, we did not have hemodynamic data from cardiac catheterization or by MRI flow quantification in all CHD patients suspected with decreased PBF. The cardiac catheterization was the routine examination for the CHD patients suspected with pulmonary hypertension, but not the essential examination for all CHD patients. In addition, the CNN for evaluating PBF still needed to be improved. The reason may be the lack of the dataset of CHD patients with increased PBF and decreased PBF confirmed by cardiac catheterization.

Conclusions

In the present study, the two CNNs we have trained could effectively detect pediatric CHD patients and evaluate for PBF on chest radiography with similar accuracy as experienced radiologists. We postulate that the CNNs can assist the radiologists and cardiologists who are in small rural hospitals and have limited experience in the detection of presence of CHD on chest radiographs of paediatric patients.

Abbreviations

CHD, congenital heart disease

PBF, pulmonary blood flow

CNN, convolutional neural network

Declarations

Funding The study was supported by the National Natural Science Foundation of China (No. 81974262, to Hui Liu).

Conflicts of interest All authors declare no conflicts of interest.

Availability of data and material All data and materials support the published claims and are available from the corresponding author.

Code availability The code was available from the corresponding author.

Authors' contributions We are grateful to our authors for accomplishment of this manuscript. RC, XW, TC, PS, ZD, WY, ZY, HL substantially contributed to collection and statistical analysis of data, the writing of the manuscript; Rui Chen, Xinyi Wu, ZD, WY, XC, ZY, HL provided the statistical analysis method and revised the manuscript; ZY and HL contributed to the coordination of the study; RC, XW, LS, ZD, WY, ZT, TL, HW, WQ, XZ, SW, XC, ZY, HL contributed to the examination of patients, data collection; RC, XW, LS, ZD, WY, ZT, TL, HW, WQ, XZ, SW, XC, ZY, HL contributed to enrollment of patients, RC, XW, ZD, ZY, HL contributed to data analysis; RC, XW, TC, LS, PS, ZD, WY, ZT, TL, HW, ZY, HL contributed to linguistic revision of this manuscript; RC, XW, TC, PS, ZY, HL contributed to design of the study, substantially revision of the manuscript.

Ethics approval This study was approved by the Ethics Committee of Guangdong Provincial People's Hospital (Ethical code: 2017079H[R1]).

Consent to participate Written informed consent was obtained from all participants.

Consent for publication Written informed consent for publication was obtained from all participants.

References

1 Marelli AJ, Mackie AS, Ionescu-Ittu R, Rahme E, Pilote L (2007) Congenital Heart Disease in the General Population. Circulation 115:163-172. doi:10.1161/circulationaha.106.627224

2 Ewer AK, Middleton LJ, Furmston AT et al (2011) Pulse oximetry screening for congenital heart defects in newborn infants (PulseOx): a test accuracy study. The Lancet 378:785-794. doi:10.1016/s0140-6736(11)60753-8

3 Nikolaus A. Haas KRS (2017) Guidelines for the Management of Congenital Heart Diseases in Childhood and Adolescence. Cardiology in the Young 27:S1–S105. doi:10.1017/S1047951116001955

4 Ho TC, Ouyang H, Lu Y, Young AH, Chintala K, Detrano RC (2011) Postprocedural outcomes of rural children undergoing correction of congenital heart lesions in Yunnan Province, China. Pediatr Cardiol 32:811-814. doi:10.1007/s00246-011-9991-y

5 A S (2009) Congenital cardiac surgery in the less privileged regions of the world. Expert Rev Cardiovasc Ther 7:1621-1629.

6 Birkebak NH, Hansen LK, Elle B et al (1999) Chest Roentgenogram in the Evaluation of Heart Defects in Asymptomatic Infants and Children With a Cardiac Murmur: Reproducibility and Accuracy. Pediatrics 103:e15-e15. doi:10.1542/peds.103.2.e15

7 Laya BF, Goske MJ, Morrison S et al (2006) The accuracy of chest radiographs in the detection of congenital heart disease and in the diagnosis of specific congenital cardiac lesions. Pediatr Radiol 36:677-681. doi:10.1007/s00247-006-0133-2

8 Kermany DS, Goldbaum M, Cai W et al (2018) Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell 172:1122-1131 e1129. doi:10.1016/j.cell.2018.02.010

9 Gerhard-Paul Diller, Aleksander Kempny, Sonya V. Babu-Narayan et al (2019) Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10019 patients. European Heart Journal 40:1069-1077. doi:10.1093/eurheartj/ehy915

10 Toba S, Mitani Y, Yodoya N et al (2020) Prediction of Pulmonary to Systemic Flow Ratio in Patients With Congenital Heart Disease Using Deep Learning-Based Analysis of Chest Radiographs.

JAMA Cardiol. 10.1001/jamacardio.2019.5620doi:10.1001/jamacardio.2019.5620

11 Tumkosit M, Yingyong N, Mahayosnond A, Choo KS, Goo HW (2012) Accuracy of chest radiography for evaluating significantly abnormal pulmonary vascularity in children with congenital heart disease. Int J Cardiovasc Imaging 28 Suppl 1:69-75. doi:10.1007/s10554-012-0073-x

12 Shi G, Zhu Z, Chen J et al (2017) Total Anomalous Pulmonary Venous Connection: The Current Management Strategies in a Pediatric Cohort of 768 Patients. Circulation 135:48-58. doi:10.1161/CIRCULATIONAHA.116.023889

13 Russakovsky O, Deng J, Su H et al (2015) ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision 115:211-252. doi:10.1007/s11263-015-0816-y

14 Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? NIPS'14 Proceedings of the 27th International Conference on Neural Information Processing Systems 2:3320–3328.

15 Soyeon Ahn, Seong Ho Park, Kyoung Ho Lee (2013) How to demonstrate similarity by Using noninferiority and equivalence statistical testing in Radiology Research. Radiology 267:328–338. doi:10.1148/radiol.12120725

16 Swenson JM, Fischer DR, Miller SA, Boyle GJ, Ettedgui JA, Beerman LB (1997) Are chest radiographs and electrocardiograms still valuable in evaluating new pediatric patients with heart murmurs or chest pain? Pediatrics 99:1-3. doi:10.1542/peds.99.1.1

17 Fayad LM, Boxt LM (1999) Chest film diagnosis of congenital heart disease. Semin Roentgenol 34:228-248.

18 Patrick M. Bossuyt, Johannes B. Reitsma, David E. Bruns et al (2015) STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. Radiology 277:826-832.

19 Baumgartner H, Bonhoeffer P, De Groot NM et al (2010) ESC Guidelines for the management of grown-up congenital heart disease (new version 2010). Eur Heart J 31:2915-2957. doi:10.1093/eurheartj/ehq249

20 Arnois DC, Silverman FN, Turner ME (1959) The radiographic evaluation of pulmonary vasculature in children with congenital cardiovascular disease. Radiology 72:689-698. doi:10.1148/72.5.689

Tables

Table 1 Distribution of Chest Radiographs from CHD patients

	Number of Patients in Guangzhou (n = 5165)	Number of Patients in Shenzhen (n = 350)
Ventricular septal defect	1265 (24.5%)	28 (8.0%)
Atrial septal defect	678 (13.1%)	52 (14.9%)
Patent ductus arteriosus	639 (12.4%)	29 (8.3%)
Pulmonary stenosis	213 (4.1%)	20 (5.7%)
Coarctation of the aorta	121 (2.3%)	17 (4.9%)
Tetralogy of Fallot	183 (3.5%)	58 (16.6%)
Transposition of great vessels	96 (1.9%)	4 (1.1%)
Atrioventricular septal defect	55 (1.1%)	2 (0.5%)
Double outlet right ventricle	43 (0.8%)	4 (1.1%)
More than one diagnosis	1872 (36.2%)	136 (38.9%)

Figures

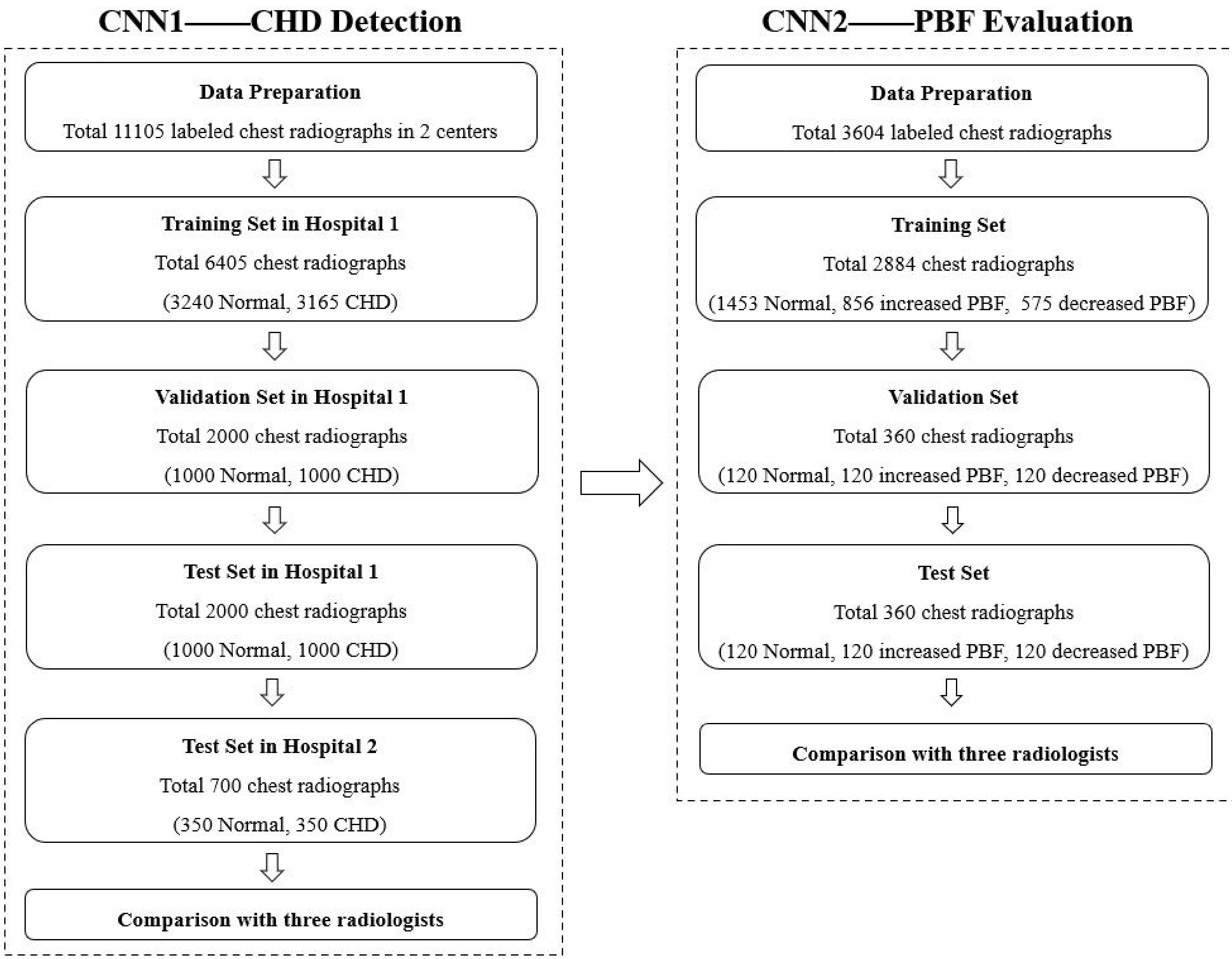

Figure 1

Workflow Diagram for the deep learning based diagnostic tools including 2 CNNs, one for CHD detection, and another for PBF evaluation. CNN, convolutional neural network; CHD, congenital heart disease; PBF, pulmonary blood flow;

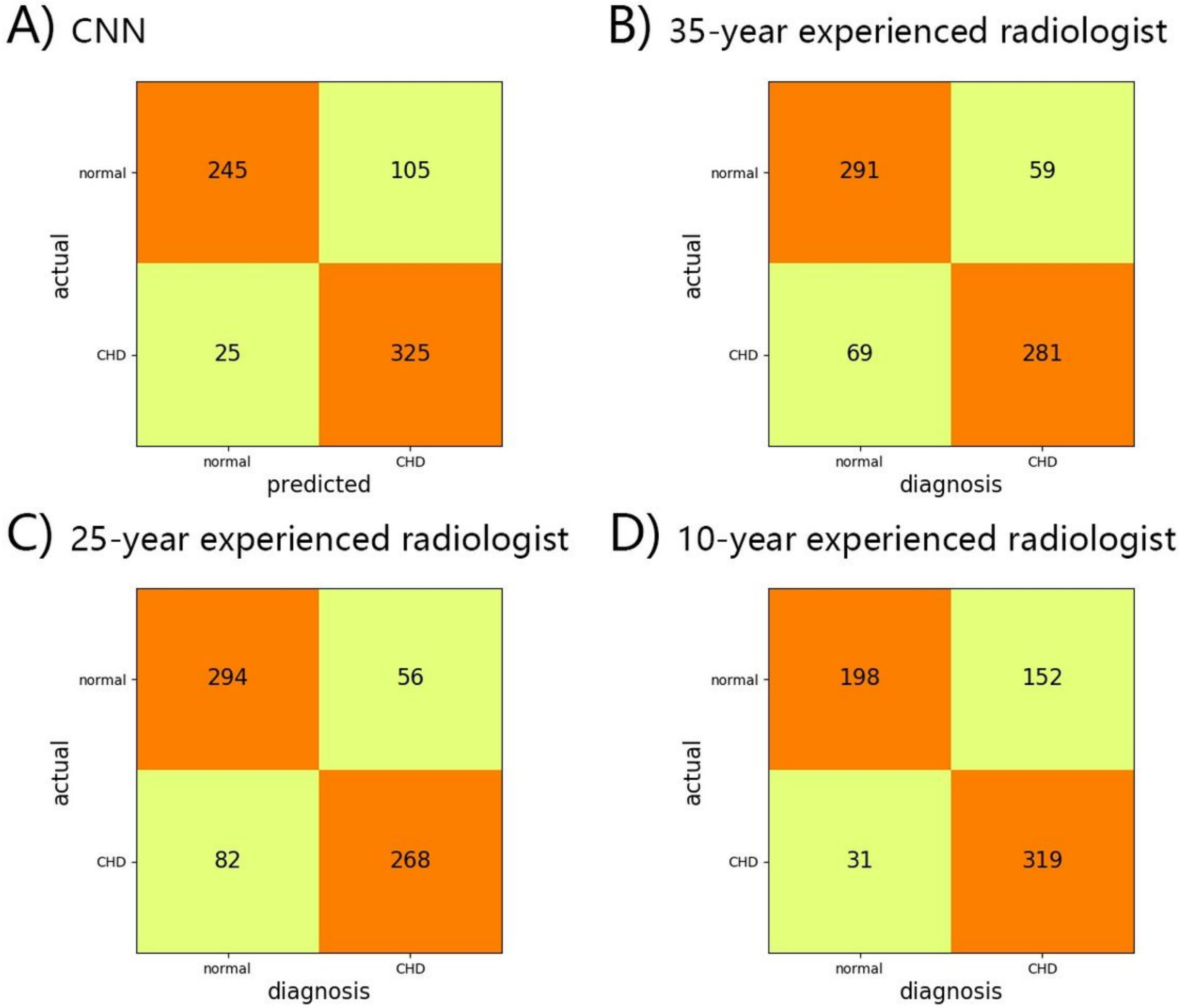

Figure 2

Confusion matrix of CNN and experienced radiologists in the detection of congenital heart disease in test set from Hospital 2. (A) the confusion matrix of CNN, accuracy: 81.4% (95% CI, [78.3%–84.2%]); (B) the confusion matrix of a 35-year experienced radiologist, accuracy: 81.7% (95% CI, [78.7%–84.5%]); (C) the confusion matrix of a 25-year experienced radiologist, accuracy: 80.3% (95% CI, [77.1%–83.2%]); (D) the confusion matrix of a 10-year experienced radiologist, accuracy: 73.9% (95% CI, [70.4%–77.1%]); Compared with those three radiologists, CNN showed non-inferiority at a 5% margin for the accuracy (-0.3%, 95% CI [-4.3%, 3.8%]; P = 0.017), (+1.1%, 95% CI [-3.0%, 5.3%]; P = 0.002), (+7.6%, 95% CI [3.2%, 11.9%]; P < 0.001), respectively. CHD, congenital heart disease; CNN, convolution neural network; CI, confidence interval

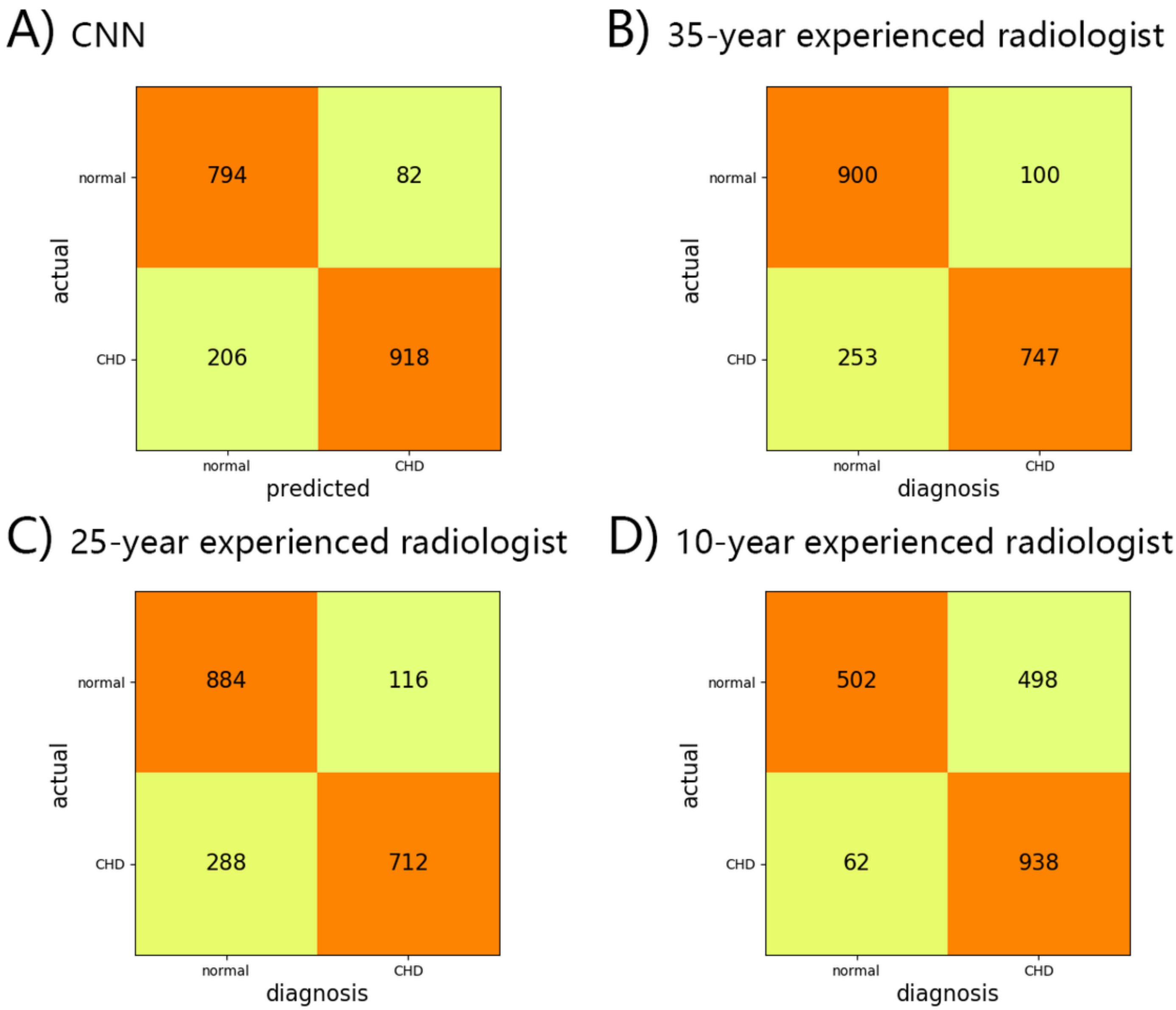

Figure 3

Confusion matrix of CNN and experienced radiologists in the detection of congenital heart disease in test set from Hospital 1. (A) the confusion matrix of CNN, accuracy: 85.6% (95% CI, [84.0%–87.1%]); (B) the confusion matrix of a 35-year experienced radiologist, accuracy: 82.4% (95% CI, [80.6%–84.0%]); (C) the confusion matrix of a 25-year experienced radiologist, accuracy: 79.8% (95% CI, [78.0%–81.5%]); (D) the confusion matrix of a 10-year experienced radiologist, accuracy: 72.0% (95% CI, [70.0%–74.0%]); Compared with those three radiologists, CNN showed non-inferiority at a 5% margin for the accuracy (+3.2%, 95% CI [0.9%, 5.5%]; $P < 0.001$), (+5.8%, 95% CI [3.5%, 8.1%]; $P < 0.001$), (+13.6%, 95% CI [11.1%, 16.1%]; $P < 0.001$), respectively. CHD, congenital heart disease; CNN, convolution neural network; CI, confidence interval

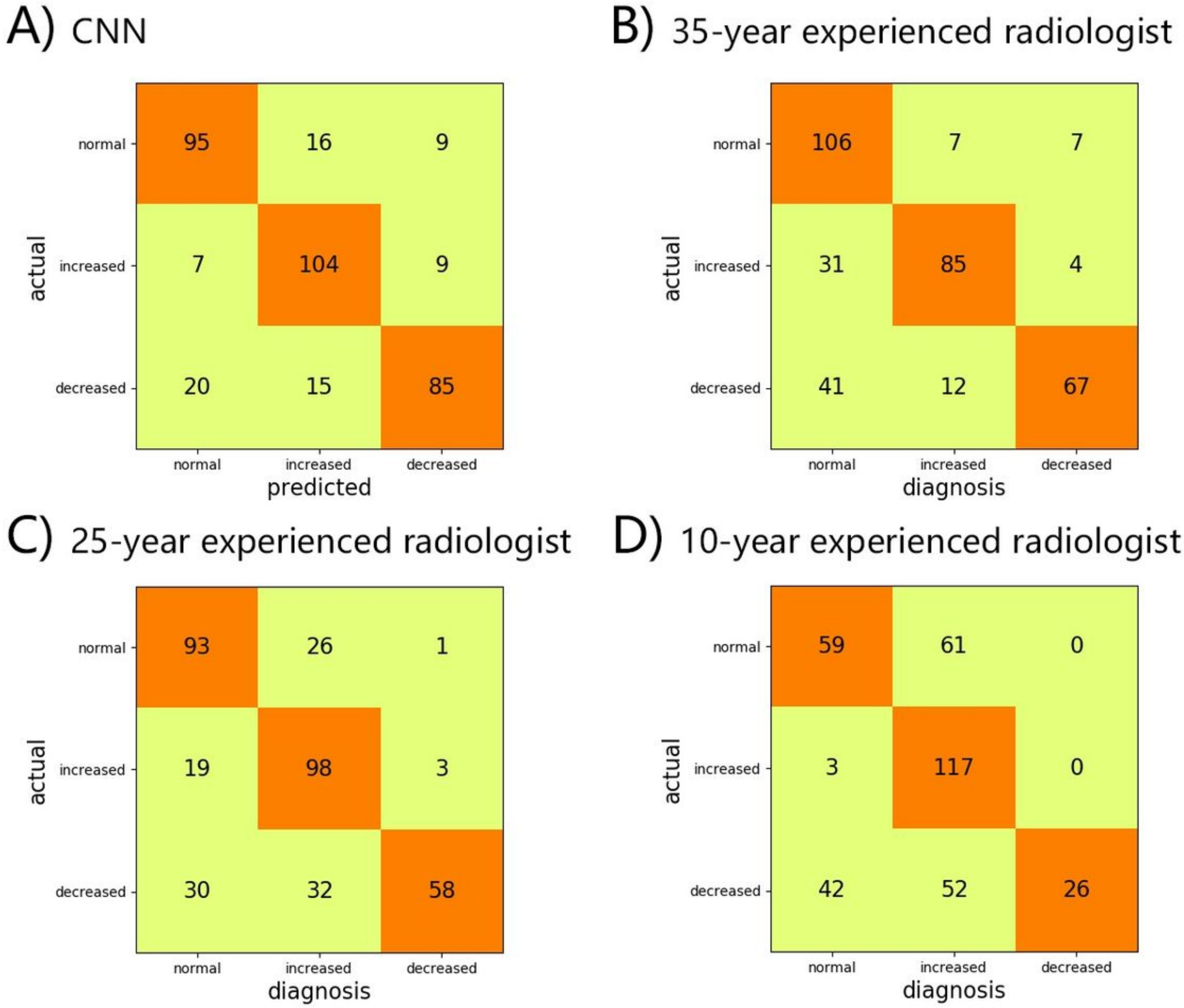

Figure 4

Confusion matrix of best model's classification of CNN in the test set for the evaluation of PBF. (A) the confusion matrix of CNN, accuracy, 78.9% (95% CI, [74.3%–83.0%]); (B) the confusion matrix of a 35-year experienced radiologist, accuracy, 71.7% (95% CI, [66.7%–76.3%]); (C) the confusion matrix of a 25-year experienced radiologist, accuracy, 69.2% (95% CI, [64.1%–73.9%]); (D) the confusion matrix of a 10-year experienced radiologist, accuracy, 56.1% (95% CI, [50.8%–61.3%]); Compared with those three radiologists, CNN showed non-inferiority at a 5% margin for the accuracy (+7.2%, 95% CI [0.9%, 13.5%]; P < 0.001), (+9.7%, 95% CI [3.4%, 16.1%]; P < 0.001), (+22.7%, 95% CI [16.1%, 29.4%]; P < 0.001), respectively. PBF, pulmonary blood flow; CI, confidence interval

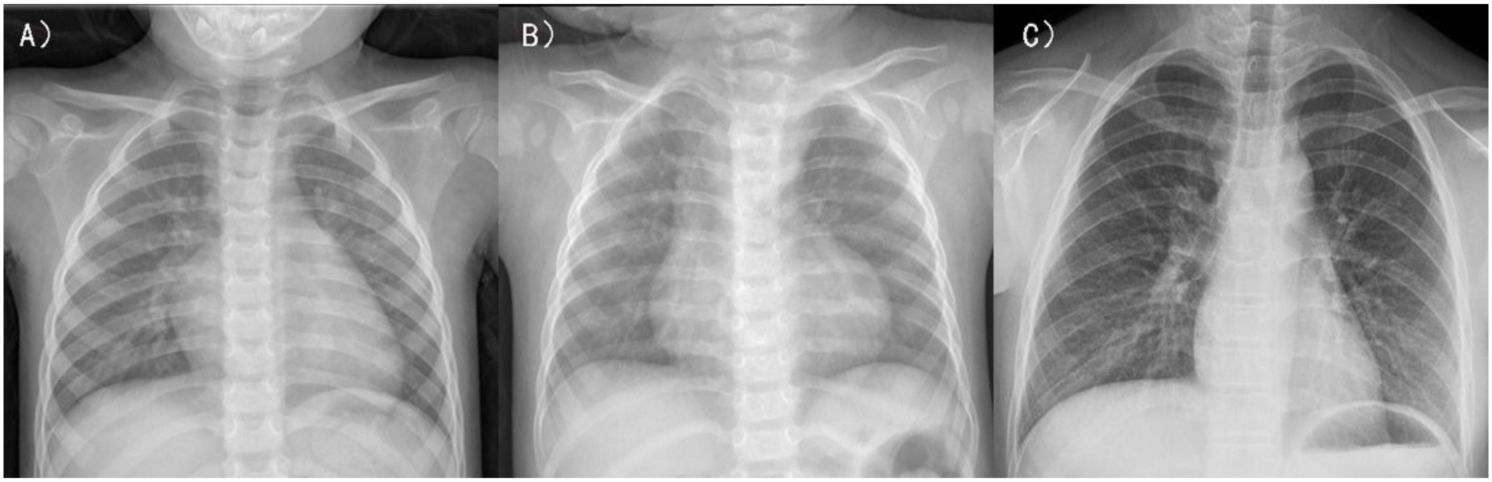

Figure 5

Predicting cases of CNN. (A) the chest radiograph of a CHD patient with increased PBF, the probability of CHD was 98.9%, the probability of PBF (normal, 2.7%; increased, 96.3%; decreased, 1.0%); (B) the chest radiograph of a CHD patient with decreased PBF, the probability of CHD was 99.5%, the probability of PBF (normal, 6.2%; increased, 23.1%; decreased, 70.7%); (C) the chest radiograph of a patient without CHD, the probability of CHD was 7.6%, the probability of PBF (normal, 81.3%; increased, 18.7%; decreased, <0.01%); CNN, convolution neural network; PBF, pulmonary blood flow;