

# Reliability Metrics of Explainable CNN based on Wasserstein Distance for Cardiac Evaluation

Yuto Omae (✉ [oomae.yuuto@nihon-u.ac.jp](mailto:oomae.yuuto@nihon-u.ac.jp))

Nihon University

Yohei Kakimoto

Nihon University

Yuki Saito

Nihon University

Daisuke Fukamachi

Nihon University

Koichi Nagashima

Nihon University

Yasuo Okumura

Nihon University

Jun Toyotani

Nihon University

---

## Research Article

### Keywords:

**Posted Date:** August 19th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1965782/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Reliability Metrics of Explainable CNN based on Wasserstein Distance for Cardiac Evaluation

Yuto Omae<sup>1\*</sup>, Yohei Kakimoto<sup>1†</sup>, Yuki Saito<sup>2†</sup>, Daisuke  
Fukamachi<sup>2†</sup>, Koichi Nagashima<sup>2†</sup>, Yasuo Okumura<sup>2†</sup>  
and Jun Toyotani<sup>1†</sup>

<sup>1\*</sup>Department of Industrial Engineering and Management,  
College of Industrial Technology, Nihon University, 1-2-1, Izumi,  
Narashino, Chiba, 275-8575 Japan.

<sup>2</sup>Division of Cardiology, Department of Medicine, Nihon  
University School of Medicine, 30-1, Kami, Ooyaguchi, Itabashi,  
Tokyo, 173-8610 Japan.

\*Corresponding author(s). E-mail(s): [oomae.yuuto@nihon-u.ac.jp](mailto:oomae.yuuto@nihon-u.ac.jp);

<sup>†</sup>These authors contributed equally to this work.

## Abstract

In recent works, convolutional neural networks (CNN) have been used in the non-invasive examination of the cardiac region for estimating pulmonary artery wedge pressure (PAWP) from chest radiographs. Moreover, because CNNs are able to output activated regions, physicians can estimate PAWP along with reasons. However, when new patient radiograph data are fed into the CNN, there is a possibility that activated regions that contain areas other than the cardiac appear. In this case, although we expect a large estimation error, it is not well known. Therefore, we verify this hypothesis by distance theory and statistic approaches. In particular, we build the probability distributions for the cardiac region and the regression activation map (RAM) and measure the similarity between these distributions by Wasserstein distance (WSD). When the CNN estimates PAWP from areas other than the cardiac region, the WSD value is high. Therefore, WSD is a reliability metrics for explainable CNN. We created two groups, normal and anomaly classes, based on WSD values. Chest radiographs which had a high WSD were assigned to the anomaly class, and those with a low WSD were

assigned to the normal class. By comparing the normal and anomaly classes based on the PAWP estimation error, we confirmed that the errors from the anomaly class were higher than those of the normal class. Therefore, physicians need to be aware that there might be large estimation errors when activated regions contain areas other than the cardiac.

## 1 Introduction

Pulmonary artery wedge pressure (PAWP) is an index for representing the cardiac state and is related to heart failure. In general, PAWP is measured by right heart catheterization (RHC) [1]. Because RHC presents various complication risks to patients [2], it is important to develop non-invasive methods for estimating PAWP [1]. PAWP is related to chest radiograph [3]; hence, in general, physicians interpret PAWP by examining X-ray images of patients [4]. However, this method is subjective and depends on the physician's skill level. Therefore, an objective method, convolutional neural network (CNN), has been reported for estimating PAWP from chest radiograph. For example, Hirata et al. [5] developed a classification CNN to detect PAWP over 18 mmHg, and Saito et al. [6] developed a regression CNN to quantitatively estimate PAWP from chest radiographs. Because these CNNs output a class activation map (CAM) [7] or regression activation map (RAM) [8], physicians can detect PAWP and estimate the activated regions, non-invasively.

There are many applications of deep learning within the medical field, such as lung nodule classification [9], skin lesion classification [10], prostate cancer detection [11], glaucoma detection [12], COVID-19 detection [13], and Alzheimer's disease recognition [14], among others. Previous research has also estimated the cardiac state using deep learning, as in the following cases: left cardiac chamber enlargement detection [15], cardiomegaly detection [16], and heart failure detection [17, 18]. Moreover, there are extensive reviews on the techniques and results of inputting chest X-ray images into a CNN [19, 20]. Some of these have the function outputting activated regions as reason of estimation [12–15, 17].

As we can see in these examples, it is important to develop a CNN that can output RAM or CAM. We verified that the RAMs outputted from the CNN developed by Saito et al. [6] contained the cardiac region by estimating PAWP (see Appendix section). However, when inputting chest radiographs taken from new patients into CNN, there is a possibility that activated regions contain not only the cardiac, but other areas. In this case, although we expect large estimation error, there is insufficient analysis regarding this.

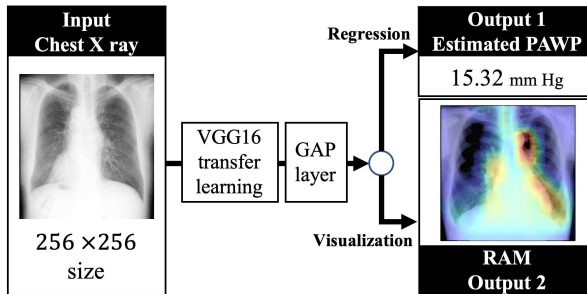
Therefore, we built and verified the following hypothesis: when an activated region contains areas other than the cardiac, the estimation errors are higher than when only the cardiac region is activated. To resolve this, we propose a reliability metrics of an explainable CNN for cardiac evaluation, i.e., an index representing the degree of the activated region containing areas other than the

cardiac. Using this index, we assigned RAMs to the normal or anomaly class. The normal class is the activated region that only contains cardiac, whereas the anomaly class is the activated region that contains other areas as well. We compared the normal and anomaly classes by the estimation error.

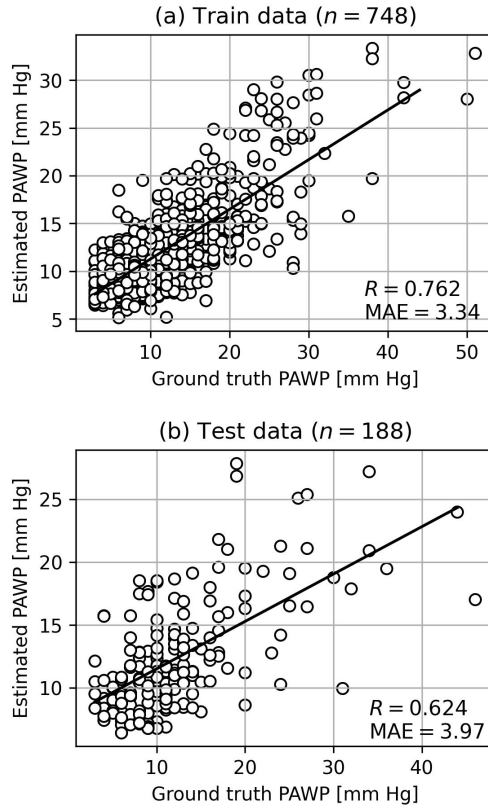
## 2 Developed R-CNN

In this section, we describe the regression CNN (R-CNN) developed by Saito et al. [6], which estimates PAWP from chest radiographs. The model structure is shown in Fig. 1. By inputting a chest radiograph of size  $256 \times 256$ , the R-CNN outputs PAWP, quantitatively. Inputted images are convoluted by VGG16[21]-based convolutional layers, and the convoluted data are flattened by the global average pooling layer (GAP layer) [22]. The model was trained by cross validating 748 samples. The scatter plots representing the reliability of R-CNN are shown in Fig. 2. The ground truth PAWP measured by RHC is on the horizontal axis, and the PAWP estimated by R-CNN is on the vertical axis. For the training samples (748 patients), the correlation coefficient was 0.762 and the mean absolute error (MAE) was 3.34 mmHg. For test samples (188 patients), the correlation coefficient was 0.624 and the MAE was 3.97 mmHg. Because many samples gathered on the diagonal line, we considered developing an R-CNN with a certain reliability. Moreover, the model had the task of outputting RAM. The size of the RAM was  $8 \times 8$ , and was resized to  $256 \times 256$  (input image size). The details of other training conditions are explained in [6].

In this study, an all-analysis procedure, we used this R-CNN and the same dataset from [6]. We confirmed almost all RAMs generated by the R-CNN contained the cardiac region (see Appendix section) using the  $k$ -means clustering approach. However, some of RAMs contained areas other than the cardiac region. We expected such samples would result in large estimation error. To verify the hypothesis, we proposed an index to automatically detect such samples, which is explained in Section 3.



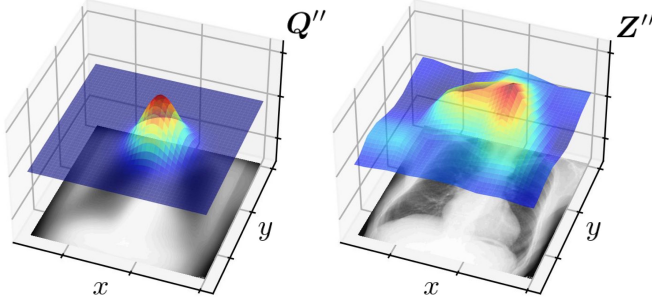
**Fig. 1** R-CNN for estimating PAWP from chest radiographs [6]. The model outputs the activated region of estimation by RAM.



**Fig. 2** Relationship between ground truth PAWP and estimated PAWP. We create them by modifying Fig. 3 in [6].

### 3 Reliability metrics of an explainable CNN for cardiac evaluation

To assign samples to the normal or anomaly class, it is necessary to develop the index for representing the degree where the activated region contains areas other than the cardiac. Therefore, we prepared  $Q''$ , the probability distribution for representing the region of the cardiac, and  $Z''$ , the probability distribution of RAM at estimating PAWP. These distributions are shown in Fig. 3. Subsequently, we calculating the similarity between  $Q''$  and  $Z''$ . When  $Z''$  is similar to  $Q''$ , the activated region for estimating PAWP by R-CNN did not contain areas other than the cardiac. In contrast, when the similarity was low, we determined that the activated region contained not only the cardiac region, but other areas. In this section, we explain the definitions of two distributions ( $Q''$  and  $Z''$ ) and the index of similarity between them. As mentioned in Section 2, the input image size was  $256 \times 256$ , and the RAM size was  $8 \times 8$ .



**Fig. 3** Discrete probability distributions  $Q''$  and  $Z''$ . The distribution  $Q''$  is developed using only the training dataset, and the distribution  $Z''$  is made from R-CNN by inputting a new patient's chest radiograph.

However, to increase abstraction, we represented  $s_i \times s_i$  as the input image size and  $s_o \times s_o$  as the RAM size.

In this study, for simple representation, we used the following abbreviations:

$$\sum_{a=1}^A \sum_{b=1}^B g(a, b) \stackrel{\text{def}}{=} \sum_{a,b}^{A,B} g(a, b). \quad (1)$$

### 3.1 Discrete probability distribution for the cardiac region

Herein, we developed  $Q''$ , the discrete probability distribution for representing the cardiac region. Let a two-dimensional Gaussian distribution on coordinate  $x, y$  be denoted as

$$q(x, y) = \mathcal{N}(x, y; \boldsymbol{\mu}, \boldsymbol{\sigma}), \quad (2)$$

$$\boldsymbol{\mu} = [\mu_x \ \mu_y]^\top, \quad \boldsymbol{\sigma} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix},$$

where  $\boldsymbol{\mu}, \boldsymbol{\sigma}$  represent the mean vector and variance-covariance matrix, respectively; these are parameters of the Gaussian distribution. Next, we denote the matrix  $Q$ , which consists of  $q(x, y)$ , as

$$Q = [q(x, y)] \in \mathbb{R}_{\geq 0}^{s_i \times s_i}, \quad (3)$$

$$\mathbb{R}_{\geq 0} = \{a \in \mathbb{R} \mid a \geq 0\},$$

where the matrix size is the input image, i.e.,  $s_i \times s_i$ .

When adopting the parameters  $\boldsymbol{\mu}$  and  $\boldsymbol{\sigma}$  of the Gaussian distribution, such as when the probabilities of the cardiac region are high, we can regard  $Q$  as a one-channel image as the values of the cardiac region are high. In particular, when the cardiac region on a chest radiograph is stored in a bounding box of  $(x_0, y_0), (x_0, y_1), (x_1, y_0), (x_1, y_1)$ , we adopted  $\boldsymbol{\mu}', \boldsymbol{\sigma}'$  leading to the sum of

probabilities of this region as  $\beta_{\text{th}}$ . In other words, we adopted

$$\begin{aligned} \boldsymbol{\mu}', \boldsymbol{\sigma}' &= \underset{\boldsymbol{\mu}, \boldsymbol{\sigma}}{\operatorname{argmin}} \left( \sum_{y=y_0}^{y_1} \sum_{x=x_0}^{x_1} q(x, y) - \beta_{\text{th}} \right)^2, \\ \text{s.t. } \sigma_{xy} &= 0 \end{aligned} \quad (4)$$

as the parameters of Gaussian distribution for  $\mathbf{Q}$ . Because we defined the rectangle bounding box as the cardiac region, we regarded  $\sigma_{xy}$  as zero.

Next, we compressed the matrix  $\mathbf{Q}$  from  $s_i \times s_i$  to  $s_o \times s_o$  by bi-linear interpolation. Let the matrix calculated by this process be denoted as

$$\mathbf{Q}' = [q'(x, y)] \in \mathbb{R}_{\geq 0}^{s_o \times s_o}. \quad (5)$$

Because of  $\sum_{x,y} q'(x, y) \neq 1$ ,  $\mathbf{Q}'$  does not satisfy the condition of the probability distribution. Therefore, we define

$$\begin{aligned} \mathbf{Q}'' &= [q''(x, y)] \\ &= \frac{\mathbf{Q}'}{\sum_{x,y} q'(x, y)} \in \mathbb{R}_{\geq 0}^{s_o \times s_o}, \text{ i.e., } \sum_{x,y} q''(x, y) = 1. \end{aligned} \quad (6)$$

Although  $\mathbf{Q}''$  is the matrix, because the sum value is 1 and each value adds a real number, we can regard it as the discrete probability distribution.  $\mathbf{Q}''$  has a feature such that the values are high when the coordinates  $x, y$  are near the cardiac region because of the optimization problem defined by Equation (4).

### 3.2 Discrete probability distribution for representing the activated region

Here, we explain the definition of  $\mathbf{Z}''$ , which represents the activated region at estimating PAWP. Let the set of feature maps just before the GAP layer at input of a chest radiograph into the R-CNN be denoted as

$$\begin{aligned} F &= \{\mathbf{f}_k \mid 1 \leq k \leq K, k \in \mathbb{N}\}, \\ \mathbf{f}_k &= [f_k(x, y)] \in \mathbb{R}^{s_o \times s_o}, \end{aligned} \quad (7)$$

where  $K$  is the number of feature maps. We define the value of the coordinates  $x, y$  of the feature map  $k$  as  $f_k(x, y)$ . In this case, the  $K$ -dimensional feature vector by the GAP layer at input of a chest radiograph is

$$\begin{aligned} \mathbf{v} &= [v(k)] \\ &= \left[ \frac{1}{(s_o)^2} \sum_{x,y}^{s_o, s_o} f_k(x, y) \right] \in \mathbb{R}^K. \end{aligned} \quad (8)$$

In other words,  $v(k)$  is an average value of the  $k$ -th feature map  $\mathbf{f}_k$ . Moreover, we define the PAWP estimated by R-CNN as

$$\begin{aligned} y_{\text{est.}} &= \mathbf{w}^\top \mathbf{v} + b, \\ \mathbf{w} &= [w(k)] \in \mathbb{R}^K, \end{aligned} \quad (9)$$

where  $\mathbf{w}$  and  $b$  are parameters obtained by learning the R-CNN. Because the parameter  $w(k)$  is the weight of  $v(k)$ , let the impact of the  $k$ -th feature map  $\mathbf{f}_k$  on estimation be denoted as

$$\mathbf{Z}_k = w(k)[f_k(x, y)] \in \mathbb{R}^{s_o \times s_o}. \quad (10)$$

By adding them all, we obtain RAM, which is defined by

$$\begin{aligned} \mathbf{Z} &= [z(x, y)] \\ &= \sum_k^K \mathbf{Z}_k \in \mathbb{R}^{s_o \times s_o}. \end{aligned} \quad (11)$$

This is the matrix that represents the impact of each coordinate of the feature maps' set  $F$  on the estimated PAWP  $y_{\text{est.}}$ .

However, because of  $\mathbf{Z} \in \mathbb{R}^{s_o \times s_o}$  and  $\sum_{x,y}^{s_o, s_o} z(x, y) \neq 1$ ,  $\mathbf{Z}$  does not satisfy the definition of the discrete probability distribution. Therefore, to satisfy the condition of the discrete probability distribution, we define the standardized matrix as

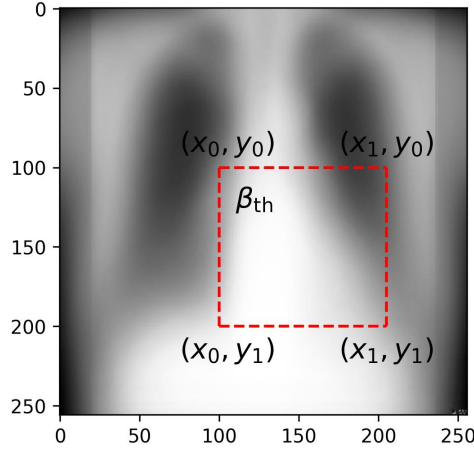
$$\begin{aligned} \mathbf{Z}' &= [z'(x, y)] \\ &= \mathbf{Z} - [z_{\min}] \in \mathbb{R}_{\geq 0}^{s_o \times s_o}, \\ \text{where, } z_{\min} &= \min_{x,y} \{z(x, y) \mid 1 \leq x, y \leq s_o\}. \end{aligned} \quad (12)$$

Moreover, with 1 as the summing value, we define

$$\begin{aligned} \mathbf{Z}'' &= [z''(x, y)] \\ &= \frac{\mathbf{Z}'}{\sum_{x,y}^{s_o, s_o} z'(x, y)} \in \mathbb{R}_{\geq 0}^{s_o \times s_o}, \text{ i.e., } \sum_{x,y}^{s_o, s_o} z''(x, y) = 1. \end{aligned} \quad (13)$$

By these equation transformations,  $\mathbf{Z}''$  is the activated region that estimates PAWP using R-CNN and satisfies the condition of the discrete probability distribution.





**Fig. 4** Average chest radiograph image stacking all training data (748 samples). The red dashed line box is the cardiac region drawn by a cardiovascular physician. The coordinates and  $\beta_{th}$  appear in Equation (4).

### 3.3 Definitions of a reliability metrics and assigning samples to normal and anomaly classes

In this study, we adopted 2-Wasserstein distance (WSD) [23, 24] to represent the similarity metrics between the discrete probability distributions  $\mathbf{Q}''$  and  $\mathbf{Z}''$ . The WSD between  $\mathbf{Z}''$  and  $\mathbf{Q}''$  is defined by

$$D_{ws}(\mathbf{Z}'', \mathbf{Q}'') = (\langle \mathbf{C}, \mathbf{T}^* \rangle)^{1/2}, \quad (14)$$

where  $\langle \mathbf{C}, \mathbf{T}^* \rangle$  indicates the minimum transportation cost from  $\mathbf{Z}''$  to  $\mathbf{Q}''$ , and  $\langle \cdot, \cdot \rangle$  indicates the Frobenius inner product, i.e.,

$$\langle \mathbf{C}, \mathbf{T}^* \rangle = \sum_{i,j}^{s_o^2, s_o^2} c(i, j) t^*(i, j). \quad (15)$$

Moreover, indices  $i$  and  $j$  represent the index of flattened vectors of  $\mathbf{Z}''$  to  $\mathbf{Q}''$  and are defined by

$$\mathbf{Z}_{\text{flat}}'' = [z_{\text{flat}}''(i)] \in \mathbb{R}_{\geq 0}^{s_o^2}, \quad (16)$$

$$\mathbf{Q}_{\text{flat}}'' = [q_{\text{flat}}''(j)] \in \mathbb{R}_{\geq 0}^{s_o^2}. \quad (17)$$

In other words,

$$\begin{aligned} \mathbf{C} &= [c(i, j)] \in \mathbb{R}_{\geq 0}^{s_o^2 \times s_o^2}, \quad c(i, j) = \|\mathbf{m}_i - \mathbf{m}_j\|_2, \\ \mathbf{m}_i, \mathbf{m}_j &\in \{[x \ y]^\top \in \mathbb{N}^2 \mid 1 \leq x, y \leq s_o\}. \end{aligned} \quad (18)$$

The vectors  $\mathbf{m}_i$  and  $\mathbf{m}_j$  represent the coordinates on the distributions  $\mathbf{Z}''$  and  $\mathbf{Q}''$  that correspond to  $z''_{\text{flat}}(i)$  and  $q''_{\text{flat}}(j)$ , respectively. In other words,  $\mathbf{C}$  indicates the matrix representing the Euclid distance from  $\mathbf{m}_i$  on  $\mathbf{Z}''$  to  $\mathbf{m}_j$  on  $\mathbf{Q}''$ . Moreover,  $\mathbf{T}^* = [t^*(i, j)] \in \mathbb{R}_{\geq 0}^{s_o^2 \times s_o^2}$  is the optimal solution of a linear programming problem. This is expressed by

$$\begin{aligned} \mathbf{T}^* &= \underset{\mathbf{T}}{\operatorname{argmin}} \langle \mathbf{C}, \mathbf{T} \rangle, \\ \text{s.t. } \mathbf{T}\mathbf{1} &= \mathbf{Z}''_{\text{flat}}, \quad \mathbf{T}^\top \mathbf{1} = \mathbf{Q}''_{\text{flat}}, \quad \mathbf{T} \geq \mathbf{0}, \end{aligned} \quad (19)$$

where  $\mathbf{1}$  and  $\mathbf{0}$  represent the  $s_o^2$ -dimensional vector of ones and the zero matrix of  $s_o^2 \times s_o^2$  size, respectively.

$\mathbf{Z}''$  represents the activated region at estimating PAWP by R-CNN, and  $\mathbf{Q}''$  is the distribution for the cardiac region. Therefore,  $D_{\text{ws}}(\mathbf{Z}'', \mathbf{Q}'')$  is the reliability metrics of explainable CNN for cardiac evaluation. When  $D_{\text{ws}}(\mathbf{Z}'', \mathbf{Q}'')$  is a small value, we can regard the activated region as containing only the cardiac (normal class). In contrast, when  $D_{\text{ws}}(\mathbf{Z}'', \mathbf{Q}'')$  is a large value, we can regard the activated region as containing not only the cardiac region but other areas as well (anomaly class). In other words, let the class of the  $l$ -th patient's chest radiograph be denoted as

$$c_l = \begin{cases} C_{\text{normal}}, D_{\text{ws}}(\mathbf{Z}'' = \mathbf{Z}''_l, \mathbf{Q}'') \leq \alpha_{\text{th}} \\ C_{\text{anomaly}}, D_{\text{ws}}(\mathbf{Z}'' = \mathbf{Z}''_l, \mathbf{Q}'') > \alpha_{\text{th}} \end{cases}, \quad (20)$$

where  $\mathbf{Z}''_l$  represents the  $l$ -th patient's RAM, and  $\alpha_{\text{th}}$  indicates the threshold.  $C_{\text{normal}}$  represents the normal class, and  $C_{\text{anomaly}}$  represents the anomaly class.

## 4 Experiment

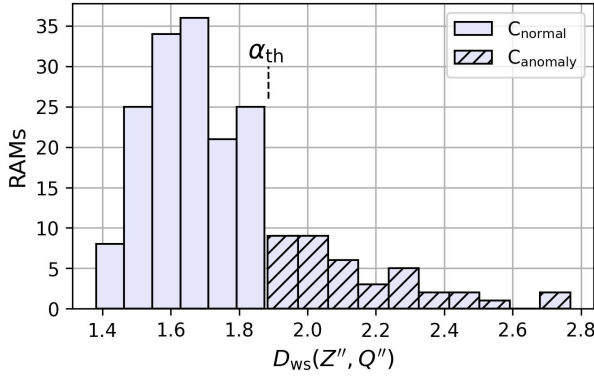
### 4.1 Dataset and ethics approval

We used the same dataset described in Saito et al. [6]. The study was approved by the ethics committee of Nihon University Itabashi Hospital (RK-210112-09) and was performed in accordance with the principles outlined in the Declaration of Helsinki.

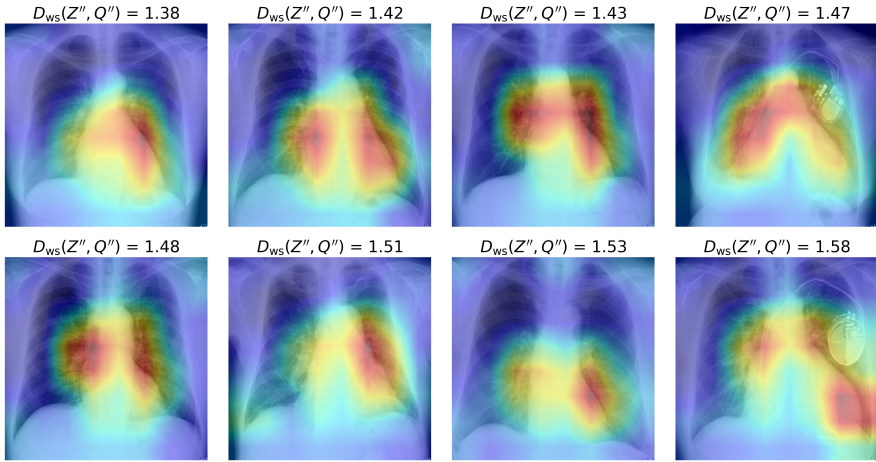
### 4.2 Parameters setting

To calculate  $D_{\text{ws}}(\mathbf{Z}'', \mathbf{Q}'')$ , it is required to decide the parameters  $\beta_{\text{th}}$  and  $\alpha_{\text{th}}$ . Therefore, we created an average chest radiograph by stacking all images of the training data (748 samples). The averaged image is shown in Fig. 4. The bounding box of the red dashed line represents the cardiac region, and it is drawn by a cardiovascular physician. Based on the bounding box, we designed the discrete probability distribution  $\mathbf{Q}''$ .

In particular, we adopted the parameters such that the probability of the cardiac region is 90%; in other words, we solved the optimization problem



**Fig. 5** Histogram  $D_{ws}(\mathbf{Z}'', \mathbf{Q}'')$  of 188 samples in the test dataset. The left side of  $\alpha_{th}$  represents the normal class and the right side of  $\alpha_{th}$  represents the anomaly class.



**Fig. 6** RAMs assigned to the normal class  $C_{normal}$ , i.e.,  $D_{ws}(\mathbf{Z}'', \mathbf{Q}'') \leq \alpha_{th}$ .

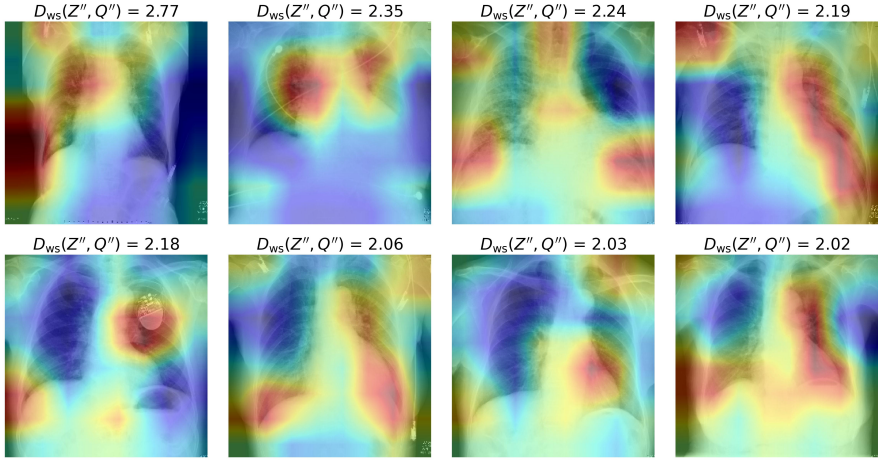
represented in Equation (4). As a result,

$$\beta_{th} = 0.90 \Rightarrow \mu' = [152.5 \ 150.0]^\top, \ \sigma' = \begin{bmatrix} 24.0^2 & 0 \\ 0 & 27.5^2 \end{bmatrix} \quad (21)$$

were adopted.

Next, we decided the parameter  $\alpha_{th}$ , which is the threshold for splitting data into the normal or anomaly class. Therefore,  $D_{ws}(\mathbf{Z}'', \mathbf{Q}'')$  of all the test data (188 samples) were calculated.

The histogram is shown in Fig. 5. We can confirm that the distribution's right side is long. To detect anomaly data, the average value  $E[D_{ws}(\mathbf{Z}'', \mathbf{Q}'')]$



**Fig. 7** RAMs assigned to the anomaly class, i.e.,  $D_{ws}(\mathbf{Z}'', \mathbf{Q}'') > \alpha_{th}$ .

and the variance  $V[D_{ws}(\mathbf{Z}'', \mathbf{Q}'')]$  were calculated. Using them, we adopted

$$\begin{aligned}
 \alpha_{th} &= E[D_{ws}(\mathbf{Z}'', \mathbf{Q}'')] + \frac{1}{2}\sqrt{V[D_{ws}(\mathbf{Z}'', \mathbf{Q}'')]} \\
 &= 1.752 + \frac{0.248}{2} \\
 &= 1.876
 \end{aligned} \tag{22}$$

as the threshold defined in Equation (20). This value is  $\alpha_{th}$  in Fig. 5. Samples of the left side were assigned to the normal class  $C_{normal}$ , whereas samples of the right side were assigned to the anomaly class  $C_{anomaly}$ . The number of samples in the normal class was 149, and the number in the anomaly class was 39 samples.

RAMs assigned to the normal class are shown in Fig. 6. In cases belonging to the normal class, we can check that the activated region that is estimating PAWP by R-CNN contains only the cardiac region (or near there). Next, RAMs assigned to the anomaly class are shown in Fig. 7. In cases belonging to the anomaly class, the activated regions that estimate PAWP contained not only the cardiac region but other areas. Therefore, we consider the proposed index  $D_{ws}(\mathbf{Z}'', \mathbf{Q}'')$  is reliable.

### 4.3 Comparison with estimated PAWP errors of normal and anomaly classes

Here, we verify whether the PAWP estimation errors of the anomaly class are larger than those of the normal class. The average and median of absolute errors of the normal and anomaly classes were calculated, and these are shown in Table 1. We ensured that the absolute error of the anomaly class was larger than that of the normal class.

**Table 1** Absolute error of estimated PAWP of the normal and anomaly classes. Using the test dataset, these errors represent the generalization scores.

	Absolute error [mmHg]		diff.
	Normal class $C_{\text{normal}}$ ( $n = 149$ )	Anomaly class $C_{\text{anomaly}}$ ( $n = 39$ )	
Average	3.60	5.37	+1.77
Median	2.68	5.07	+2.39
Std.	3.85	4.35	-

Next, we described whether the results were statistically significant or not. First, we tested whether the absolute error followed the Gaussian distribution using the Shapiro–Wilk test. Because we examined  $p < .01$ , we carried out the two-sided Mann–Whitney  $U$  test as the non-parametric hypothesis test. Hence, we verified that the difference between the absolute errors of the normal and anomaly classes was statistically significant ( $p < .05$ ). Therefore, we considered the estimation error of the anomaly class to be higher than that of the normal class. In other words, in the case where RAM contains not only the cardiac region, but other areas, we must consider the possibility of a large estimation error.

Next, we carried out a questionnaire survey on 23 cardiology physicians who have medical licenses in Japan. Personal information such as name, age, and gender were not answered as it was an anonymous survey. The question presented was “When using the PAWP estimation system by the CNN, how much estimation error do you allow?”. The answer ranged from 0 to 10 mmHg. Thus, the average allowable error was  $4.04 \pm 1.08$  mmHg. As indicated in Table 1, the average and median absolute errors of the anomaly class were over the allowable error. By contrast, the average and median absolute errors of the normal class were less than the allowable error. This result indicates that when physicians use the R-CNN to detect a patient’s PAWP, it is important to check the RAM and proposed index  $D_{\text{ws}}(\mathbf{Z}'', \mathbf{Q}'')$ .

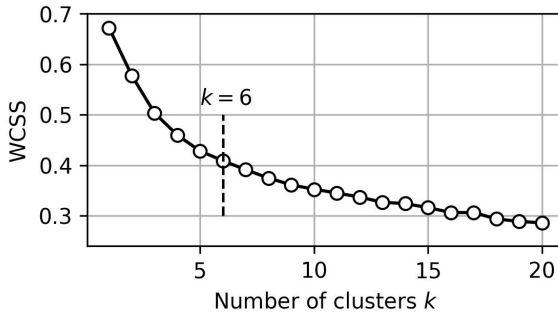
## 5 Conclusions, limitations, and future works

In this study, we proposed an index for representing the degree in which the activated region at estimating PAWP by R-CNN contains not only the cardiac region but other areas. Moreover, we showed a large estimation error when the activated region contained areas other than the cardiac. In conclusion, it is important for physicians to examine RAM.

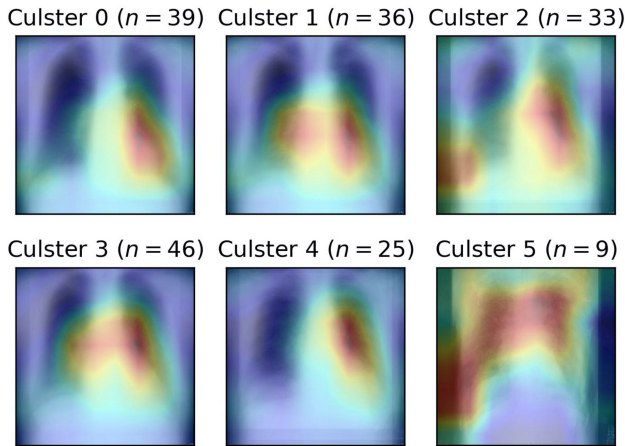
This study had the following limitations, which leads us to future research.

- We used the dataset described in a previous study [6]. The data were collected via a single center study. In applying deep learning to medicine, it is desirable to use dataset collected from multiple centers such as [25–28]. Therefore, as one of our future works, we plan to collect chest radiographs from various medical centers.

- We designed and used  $Q''$  as the probability distribution of the cardiac region indicated in Fig. 4. This is an average cardiac region. However, each patient's cardiac region and the average cardiac region may differ. There exist some studies on the segmentation of the cardiac region [29, 30]. These studies are desirable for using the segmentation model to design the probability  $Q''$ .
- In this study, we indicated the large error, when the activated region contained areas other than the cardiac. Therefore, it is important to develop an R-CNN, which outputs RAMs containing only the cardiac region. As for future works, we plan to develop such an R-CNN by improving the previous model [6].



**Fig. 8** Relationship between the number of clusters  $k$  and WCSS.



**Fig. 9** Centroids of six clusters calculated from the test dataset (188 samples)

## Appendix

### Tendencies and features of RAMs outputted from R-CNN

Herein, we indicate the tendencies and features of RAMs outputted from R-CNN described in [6]. Therefore, we obtained RAMs of the test dataset (188 samples) by inputting chest radiographs into the R-CNN. Because the size of the RAM was  $8 \times 8$ , we transformed it into a 64-dimensional vector. Subsequently, we clustered the RAM using the  $k$ -means method. To decide the number of clusters  $k$ , we carried out the elbow method based on within cluster sum of squares (WCSS) [31]. The relationship between clusters  $k$  and WCSS is shown in Fig. 8. We can see that WCSS rapidly decreases to  $k = 6$ . Therefore, we adopted  $k = 6$  as the number of clusters.

The centroid RAMs of each cluster are shown in Fig. 9. The upper side numbers represent the cluster id and sample size (the total number is 188 samples). We can see that R-CNN estimates PAWP from the cardiac region. However, there are centroids in which the activated regions contain not only

the cardiac but others, as in cluster 2 or 5. These are the features of RAM outputted from R-CNN developed by previous research [6].

## Declarations

### Conflict of interest/Competing interests

The authors declare no conflict of interest.

### Ethics approval

The study was approved by the ethics committee of Nihon University Itabashi Hospital (RK-210112-09) and was performed in accordance with the principles outlined in the Declaration of Helsinki.

### Consent to participate

Not applicable

### Consent for publication

All authors have read and agreed to the published version of the manuscript.

### Availability of data and materials

Data in this research are available upon request to Yuto Omae and Yuki Saito.

### Code availability

Codes in this research are available upon request to Yuto Omae.

### Authors' contributions

Conceptualization, Yuto Omae and Yuki Saito.

Methodology, Yuto Omae, Yohei Kakimoto and Yuki Saito.

Software, Yuto Omae and Yohei Kakimoto.

Validation, Yuto Omae and Yuki Saito.

Discussion in medicine, Yuki Saito, Daisuke Fukamachi, Koichi Nagashima, and Yasuo Okumura.

Discussion in engineering, Yuto Omae, Yohei Kakimoto, and Jun Toyotani.

Writing and original draft preparation, Yuto Omae.

Writing, review and editing, Yuto Omae and Yohei Kakimoto.

Supervision, Yuto Omae and Jun Toyotani.



## References

- [1] Chen, Y., Shlofmitz, E., Khalid, N., Bernardo, N.L., Ben-Dor, I., Weintraub, W.S., Waksman, R.: Right heart catheterization-related complications: a review of the literature and best practices. *Cardiology in Review* **28**(1), 36–41 (2020). <https://doi.org/10.1097/CRD.0000000000000270>
- [2] Hoeper, M.M., Lee, S.H., Voswinckel, R., Palazzini, M., Jais, X., Marinelli, A., Barst, R.J., Ghofrani, H.A., Jing, Z.C., Opitz, C., Seyfarth, H.J., Halank, M., McLaughlin, V., Oudiz, R.J., Ewert, R., Wilkens, H., Kluge, S., Bremer, H.C., Baroke, E., Rubin, L.J.: Complications of right heart catheterization procedures in patients with pulmonary hypertension in experienced centers. *Journal of the American College of Cardiology* **48**(12), 2546–2552 (2006). <https://doi.org/10.1016/J.JACC.2006.07.061>
- [3] Dash, H., Lipton, M.J., Chatterjee, K., Parmley, W.W.: Estimation of pulmonary artery wedge pressure from chest radiograph in patients with chronic congestive cardiomyopathy and ischaemic cardiomyopathy. *Heart* **44**(3), 322–329 (1980). <https://doi.org/10.1136/HRT.44.3.322>
- [4] McDonagh, T.A., Metra, M., Adamo, M., Gardner, R.S., Baumbach, A., Böhm, M.: 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failureDeveloped by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution. *European Heart Journal* **42**(36), 3599–3726 (2021). <https://doi.org/10.1093/EURHEARTJ/EHAB368>
- [5] Hirata, Y., Kusunose, K., Tsuji, T., Fujimori, K., Kotoku, J., Sata, M.: Deep learning for detection of elevated pulmonary artery wedge pressure using standard chest x-ray. *Canadian Journal of Cardiology* **37**(8), 1198–1206 (2021). <https://doi.org/10.1016/J.CJCA.2021.02.007>
- [6] Saito, Y., Omae, Y., Fukamachi, D., Nagashima, K., Mizobuchi, S., Kakimoto, Y., Toyotani, J., Okumura, Y.: Quantitative estimation of pulmonary artery wedge pressure from chest radiographs by a regression convolutional neural network. *Heart and Vessels* (2022). <https://doi.org/10.1007/S00380-022-02043-W/FIGURES/4>
- [7] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, A.T.: Learning deep features for discriminative localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921–2929 (2016)
- [8] Zhiguang Wang, J.Y.: Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. In: *Workshops at the Thirty-second AAAI Conference on Artificial*

Intelligence (2018)

- [9] Nishio, M., Sugiyama, O., Yakami, M., Ueno, S., Kubo, T., Kuroda, T., Togashi, K.: Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. *PLOS ONE* **13**(7), 0200721 (2018). <https://doi.org/10.1371/JOURNAL.PONE.0200721>
- [10] Mahbod, A., Schaefer, G., Wang, C., Dorffner, G., Ecker, R., Ellinger, I.: Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. *Computer Methods and Programs in Biomedicine* **193**, 105475 (2020). <https://doi.org/10.1016/J.CMPB.2020.105475>
- [11] Hao, R., Namdar, K., Liu, L., Haider, M.A., Khalvati, F.: A comprehensive study of data augmentation strategies for prostate cancer detection in diffusion-weighted mri using convolutional neural networks. *Journal of Digital Imaging* 2021 34:4 **34**(4), 862–876 (2021) <https://arxiv.org/abs/2006.01693>. <https://doi.org/10.1007/S10278-021-00478-7>
- [12] Li, L., Xu, M., Liu, H., Li, Y., Wang, X., Jiang, L., Wang, Z., Fan, X., Wang, N.: A large-scale database and a CNN model for attention-based glaucoma detection. *IEEE Transactions on Medical Imaging* **39**(2), 413–424 (2020). <https://doi.org/10.1109/TMI.2019.2927226>
- [13] Kara, M., Öztürk, Z., Akpek, S., Turupcu, A.A., Su, P., Shen, Y.: COVID-19 diagnosis from chest CT scans: a weakly supervised CNN-LSTM approach. *AI* **2**(3), 330–341 (2021). <https://doi.org/10.3390/AI2030020>
- [14] Kavitha, M., Yudistira, N., Kurita, T.: Multi instance learning via deep CNN for multi-class recognition of Alzheimer’s disease. In: 2019 IEEE 11th International Workshop on Computational Intelligence and Applications, pp. 89–94 (2019). <https://doi.org/10.1109/IWCIA47330.2019.8955006>
- [15] Nam, J.G., Kim, J., Noh, K., Choi, H., Kim, D.S., Yoo, S.J., Yang, H.L., Hwang, E.J., Goo, J.M., Park, E.A., Sun, H.Y., Kim, M.S., Park, C.M.: Automatic prediction of left cardiac chamber enlargement from chest radiographs using convolutional neural network. *European Radiology* **31**(11), 8130–8140 (2021). <https://doi.org/10.1007/S00330-021-07963-1/FIGURES/5>
- [16] Que, Q., Tang, Z., Wang, R., Zeng, Z., Wang, J., Chua, M., Gee, T.S., Yang, X., Veeravalli, B.: CardioXNet: automated detection for cardiomegaly based on deep learning. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 612–615 (2018). <https://doi.org/10.1109/EMBC.2018>

8512374

- [17] Matsumoto, T., Kodera, S., Shinohara, H., Ieki, H., Yamaguchi, T., Higashikuni, Y., Kiyosue, A., Ito, K., Ando, J., Takimoto, E., Akazawa, H., Morita, H., Komuro, I.: Diagnosing heart failure from chest x-ray images using deep learning. *International Heart Journal* **61**(4), 781–786 (2020). <https://doi.org/10.1536/IHJ.19-714>
- [18] Yasmin, F., Shah, S.M.I., Naeem, A., Shujauddin, S.M., Jabeen, A., Kazmi, S.: Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future. *Reviews in Cardiovascular Medicine* **22**(4), 1095–1113 (2021). <https://doi.org/10.31083/J.RCM2204121/2153-8174-22-4-1095/FIG1.JPG>
- [19] Feng, Y., Teh, H.S., Cai, Y.: Deep learning for chest radiology: a review. *Current Radiology Reports* **7**(8), 1–9 (2019). <https://doi.org/10.1007/S40134-019-0333-9>
- [20] Anis, S., Lai, K.W., Chuah, J.H., Ali, S.M., Mohafez, H., Hadizadeh, M., Yan, D., Ong, Z.C.: An overview of deep learning approaches in chest radiograph. *IEEE Access* **8**, 182347–182354 (2020). <https://doi.org/10.1109/ACCESS.2020.3028390>
- [21] Theckedath, D., Sedamkar, R.R.: Detecting affect states using vgg16, resnet50 and se-resnet50 networks. *SN Computer Science* **1**(2), 1–7 (2020). <https://doi.org/10.1007/S42979-020-0114-9>
- [22] Al-Sabaawi, A., Ibrahim, H.M., Arkah, Z.M., Al-Amidie, M., Alzubaidi, L.: Amended convolutional neural network with global average pooling for image classification. *International Conference on Intelligent Systems Design and Applications*, 171–180 (2020). [https://doi.org/10.1007/978-3-030-71187-0\\_16](https://doi.org/10.1007/978-3-030-71187-0_16)
- [23] Kolouri, S., Park, S.R., Thorpe, M., Slepcev, D., Rohde, G.K.: Optimal Mass Transport: Signal processing and machine-learning applications. *IEEE Signal Processing Magazine* **34**(4), 43–59 (2017). <https://doi.org/10.1109/MSP.2017.2695801>
- [24] Villani, C.: *Topics in Optimal Transportation* (2003). American Mathematical Soc.
- [25] Betancur, J., Commandeur, F., Motlagh, M., Sharir, T., Einstein, A.J., Bokhari, S., Fish, M.B., Ruddy, T.D., Kaufmann, P., Sinusas, A.J., Miller, E.J., Bateman, T.M., Dorbala, S., Di Carli, M., Germano, G., Otaki, Y., Tamarappoo, B.K., Dey, D., Berman, D.S., Slomka, P.J.: Deep learning for prediction of obstructive disease from fast myocardial perfusion spect: a multicenter study. *JACC: Cardiovascular Imaging* **11**(11), 1654–1663

- (2018). <https://doi.org/10.1016/J.JCMG.2018.01.020>
- [26] Tao, Q., Yan, W., Wang, Y., Paiman, E.H.M., Shamonin, D.P., Garg, P., Plein, S., Huang, L., Xia, L., Sramko, M., Tintera, J., De Roos, A., Lamb, H.J., Van Der Geest, R.J.: Deep learning-based method for fully automatic quantification of left ventricle function from cine MR images: A multivendor, multicenter study. *Radiology* **290**(1), 81–88 (2019). <https://doi.org/10.1148/RADIOL.2018180513/ASSET/IMAGES/LARGE/RADIOL.2018180513.FIG5D.JPEG>
- [27] Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.A., Gaiser, T., Marx, A., Valous, N.A., Ferber, D., Jansen, L., Reyes-Aldasoro, C.C., Zörnig, I., Jäger, D., Brenner, H., Chang-Claude, J., Hoffmeister, M., Halama, N.: Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLOS Medicine* **16**(1), 1002730 (2019). <https://doi.org/10.1371/JOURNAL.PMED.1002730>
- [28] Dong, D., Fang, M.J., Tang, L., Shan, X.H., Gao, J.B., Giganti, F.: Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Annals of Oncology* **31**(7), 912–920 (2020). <https://doi.org/10.1016/J.ANNONC.2020.04.003>
- [29] Novikov, A.A., Lenis, D., Major, D., Hladuvka, J., Wimmer, M., Buhler, K.: Fully convolutional architectures for multiclass segmentation in chest radiographs. *IEEE Transactions on Medical Imaging* **37**(8), 1865–1876 (2018) <https://arxiv.org/abs/1701.08816>. <https://doi.org/10.1109/TMI.2018.2806086>
- [30] Wang, C.: Segmentation of multiple structures in chest radiographs using multi-task fully convolutional networks. *Lecture Notes in Computer Science* **10270 LNCS**, 282–289 (2017). [https://doi.org/10.1007/978-3-319-59129-2\\_24/TABLES/2](https://doi.org/10.1007/978-3-319-59129-2_24/TABLES/2)
- [31] Cui, M.: Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance* **1**(1), 5–8 (2020). <https://doi.org/10.23977/ACCAF.2020.010102>