ICIC Express Letters
Part B: Applications
Volume **14**, Number **3**, March **2023**

ICIC International ©2023 ISSN 2185-2766
pp. **271–277**

THREE-STATE CLASSIFICATION OF PULMONARY ARTERY WEDGE PRESSURE FROM CHEST X-RAY IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

Tomoki Miura[1], Yuto Omae[1], Yuki Saito[2], Daisuke Fukamachi[2]
Koichi Nagashima[2], Yasuo Okumura[2], Yohei Kakimoto[1]
and Jun Toyotani[1]

[1]College of Industrial Technology
Nihon University
1-2-1, Izumi, Narashino, Chiba 275-8575, Japan
oomae.yuuto@nihon-u.ac.jp

[2]Division of Cardiology
Department of Medicine
School of Medicine
Nihon University
30-1, Ohyaguchi-kamicho, Itabashi-ku, Tokyo 173-8610, Japan
saito.yuki@nihon-u.ac.jp

Received September 2022; accepted December 2022

Abstract. *The pulmonary artery wedge pressure (PAWP) is an index used to evaluate pulmonary congestion caused by heart failure. In a previous study, a convolutional neural network (CNN) was used to estimate PAWP from chest X-ray images in binary states. This study is beneficial for medicine; however, there is a need to estimate PAWP in more detail. Therefore, we developed a CNN that outputs three classes depending on the PAWP (normal class: less than 12 mmHg; anomaly1 class: between 12 and 18 mmHg; anomaly2 class: 18 mmHg or more). The experiment used data of 936 patients, which were divided into training (80%) and test data (20%). Moreover, a validation dataset (20%) was extracted from the training dataset to tune the hyperparameters (learning rate and number of epochs). As a result of learning the CNNs, the optimal learning rate and epochs were $10^{-5.5}$ and 96, respectively. The accuracy of the test data was approximately 63%. The accuracy of the normal class was sufficient; however, those of anomaly1 and anomaly2 classes were insufficient. Therefore, the estimation accuracy must be improved in future work.*
Keywords: Deep learning, Convolutional neural network, Pulmonary artery wedge pressure

1. **Introduction.** Heart failure is associated with a high risk of death; therefore, its early detection is important. Pulmonary artery wedge pressure (PAWP) data obtained by right heart catheterization (RHC) is essential to evaluate the severity of heart failure. However, the procedure of obtaining RHC data is invasive and several complications have been reported [1]. As a safety method, experienced physicians can roughly diagnose PAWP states from the chest X-ray images of patients. Although this method [2] is commonly employed, it is subjective and not objective. We presumed that deep learning can be employed for estimating PAWP objectively.

Recently, there have been many applications of deep learning in the medical field owing to the growing image recognition capability of artificial intelligence.

DOI: 10.24507/icicelb.14.03.271

Jain et al. [3] detected COVID-19 from chest X-ray images of COVID-19 infected patients and healthy patients using deep learning. To do this, they performed data augmentation and developed CNNs. They compared the accuracies of Inception V3, Xception, and ResNeXt models, and discovered that the Xception model showed the best accuracy for detecting COVID-19 in infected patients.

Moreover, deep learning models have been used for detecting heart failure [4], hypertrophic hearts [5], and cardiac chamber enlargement [6] from chest X-ray images. Novikov et al. [7] implemented multi-class segmentation using a CNN from chest X-ray images. Betancur et al. [8] developed a method for predicting obstructive disease from myocardial perfusion imaging using deep learning. Tao et al. [9] developed a deep learning-based method for fully automatic quantification of the left ventricle function from short-axis cine MR images. Zhao et al. [10] proposed a new finger vein recognition algorithm based on an improved CNN and curvature gray feature decomposition. Experimental results indicated that their method is more effective and better than traditional schemes and the previous method.

As mentioned above, there have been many cases of deep learning applications in medicine. Furthermore, there is a review paper on deep learning in medicine (e.g., [11]). Because machine learning and medicine have a high affinity, we considered applying deep learning to estimating PAWP; some studies have already employed this principle. For example, Hirata et al. [12] constructed a CNN to estimate two classes (normal: 18 mmHg or less; anomaly: more than 18 mmHg) of PAWP. This study is valuable; however, we consider that PAWP states must be estimated in more detail. Saito et al. [13] developed a regression CNN to estimate PAWP in the form of a real number. A statistically significant correlation coefficient between the ground truth and estimated PAWP was reported; however, there was a certain level of error. To determine heart states, we consider not only methods for estimating PAWP by regression but also by multiclass recognition. Although the previous study [12] is important and beneficial, wherein more than 18 mmHg of PAWP indicated an abnormal condition, we considered that detecting a slightly abnormal condition (between 12 and 18 mmHg) can be crucial to provide early treatment. Therefore, we added a new anomaly class from 12 to 18 mmHg as a detection target and aimed to construct a CNN that estimates and classifies PAWP states in three classes.

This paper is organized as follows. The proposed method is detailed in Section 2. Particularly, the dataset used in the experiment is reviewed in Section 2.1 and the structure of the developed CNN is explained in Section 2.2. The experiment is described in Section 3. Specifically, the experimental objective and outline are presented in Section 3.1, and the results and discussions are provided in Section 3.2. Section 4 concludes the paper and offers directions for future research.

2. **Method.**

2.1. **Dataset.**

In this study, we used a dataset from a previous study [12]. The chest X-ray image size was 256×256. The structure of the split data is shown in Figure 1. All the data were split into training (80%) and test data (20%). Moreover, for tuning the hyperparameters, 20% of the training data were used as the validation data.

In this study, we denote the three classes as

$$C = \{\mathrm{C}_{\mathrm{norm}}, \mathrm{C}_{\mathrm{ano1}}, \mathrm{C}_{\mathrm{ano2}}\}. \tag{1}$$

Let us denote the class the c_k of the kth patient's chest radiograph as

$$c_k = \begin{cases} \mathrm{C}_{\mathrm{norm}} & \left(y^{\mathrm{rhc}} < 12\right) \\ \mathrm{C}_{\mathrm{ano1}} & \left(12 \leq y^{\mathrm{rhc}} < 18\right), \\ \mathrm{C}_{\mathrm{ano2}} & \left(18 \leq y^{\mathrm{rhc}}\right) \end{cases} \tag{2}$$

where y^{rhc} represents the ground truth PAWP measured by RHC in [mmHg]. $\mathrm{C}_{\mathrm{norm}}$ represents the normal class and $\mathrm{C}_{\mathrm{ano1}}$ and $\mathrm{C}_{\mathrm{ano2}}$ indicate high PAWP, that is, a high risk of the heart failure. In the main training data, the numbers of $\mathrm{C}_{\mathrm{norm}}$, $\mathrm{C}_{\mathrm{ano1}}$, and $\mathrm{C}_{\mathrm{ano2}}$ were 331, 161, and 106 patients, respectively. In the validation data, the numbers of $\mathrm{C}_{\mathrm{norm}}$, $\mathrm{C}_{\mathrm{ano1}}$, and $\mathrm{C}_{\mathrm{ano2}}$ were 90, 40, and 20 patients, respectively. In the test data, the numbers of $\mathrm{C}_{\mathrm{norm}}$, $\mathrm{C}_{\mathrm{ano1}}$, and $\mathrm{C}_{\mathrm{ano2}}$ were 107, 52, and 29 patients, respectively. There are features in which the number of normal classes is large; in contrast, the number of anomaly classes is small.

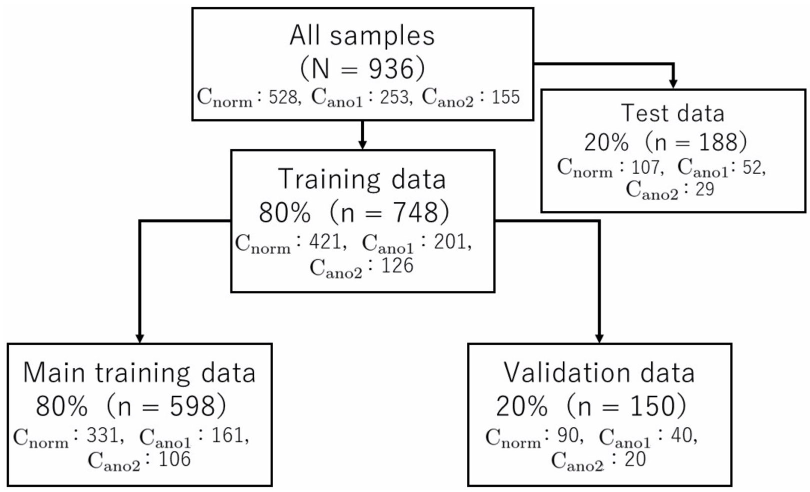

FIGURE 1. Tree of splitting data

2.2. **Structure of CNN.** The structure of the CNN is illustrated in Figure 2. The CNN consists of three convolutional layers, two pooling layers, and a fully connected layer. The size of the input image for the input layer was 256×256. Using the first convolutional layer, the input image is convolved by 3×3 kernel filters. The number of filters used for convolution was 32 in the second layer and 64 in the fourth and sixth layers. The pooling layers extract the maximum value via a 2×2 kernel filter. The rectified linear unit (ReLU) was adopted as the activation function for all layers, except for the output layer.

The ReLU function outputs zero when the input value is zero or less. In contrast, when the input value is greater than zero, the ReLU is the identity map. Therefore, the ReLU

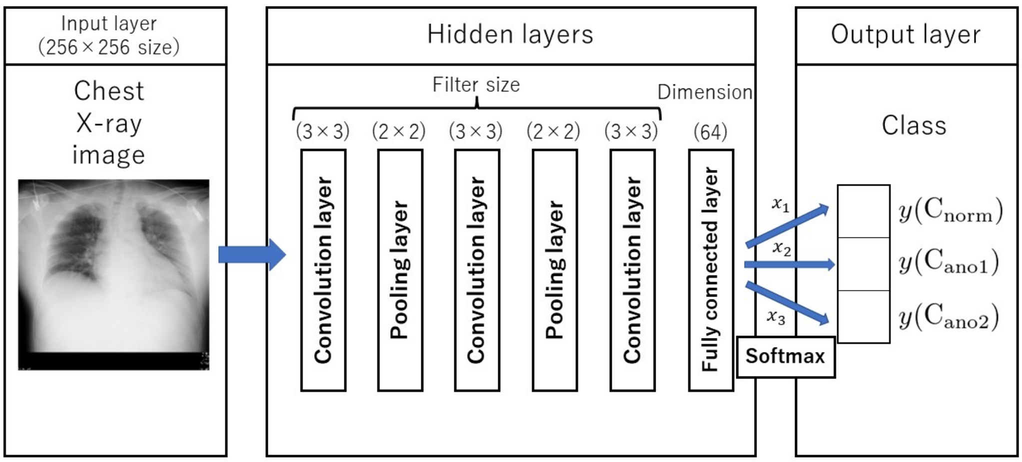

FIGURE 2. Structure of CNN

function is defined as follows:

$$f^{\mathrm{ReLU}}(z) = \begin{cases} 0 \ (z < 0) \\ z \ (z \geq 0) \end{cases}, \text{ i.e., } f^{\mathrm{ReLU}} : (-\infty, \infty) \to [0, \infty). \tag{3}$$

Compared with other activation functions, such as the sigmoid function, this function has the effect of reducing the vanishing gradient problem during backpropagation. Backpropagation is a method for updating the weights and bias parameters from the output layer to the input layer.

Next, we explain the transformation of the feature vector during classification. When inputting chest X-ray images into the CNN,

$$\boldsymbol{x}^{\mathrm{fcl}} = \left[x_1^{\mathrm{fcl}} \ x_2^{\mathrm{fcl}} \ \cdots \ x_{64}^{\mathrm{fcl}}\right]^\top \tag{4}$$

is generated from the fully connected layer shown in Figure 2, which is a 64-dimensional feature vector. This is compressed as a three-dimensional vector by

$$\boldsymbol{x} = [x_1 \ x_2 \ x_3]^\top = f^{\mathrm{ReLU}}\left(\boldsymbol{w}\boldsymbol{x}^{\mathrm{fcl}} + \boldsymbol{b}\right), \tag{5}$$

where

$$\boldsymbol{w} = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} \\ \vdots & \vdots & \vdots \\ w_{64,1} & w_{64,2} & w_{64,3} \end{bmatrix}, \ \boldsymbol{b} = [b_1 \ b_2 \ b_3]^\top \tag{6}$$

are the parameters learned by backpropagation. By inputting the vector $\boldsymbol{x}$ into the Softmax function, we obtain the discrete probability distribution of the output classes. In other words,

$$\begin{aligned} &\boldsymbol{y} = [y(\mathrm{C_{norm}}) \ y(\mathrm{C_{ano1}}) \ y(\mathrm{C_{ano2}})]^\top = \frac{1}{\sum_{i \in C} \exp(x_i)} [\exp(x_1) \ \ \exp(x_2) \ \ \exp(x_3)]^\top, \\ &\text{i.e., } y(\mathrm{C_{norm}}) + y(\mathrm{C_{ano1}}) + y(\mathrm{C_{ano2}}) = 1, \end{aligned} \tag{7}$$

where $y(i \in C)$ is the probability of class $i \in C$. The classification results are defined as follows:

$$\begin{aligned} \max(\boldsymbol{y}) = y(\mathrm{C_{norm}}) &\Rightarrow \text{Estimation class: } \mathrm{C_{norm}}, \\ \max(\boldsymbol{y}) = y(\mathrm{C_{ano1}}) &\Rightarrow \text{Estimation class: } \mathrm{C_{ano1}}, \\ \max(\boldsymbol{y}) = y(\mathrm{C_{ano2}}) &\Rightarrow \text{Estimation class: } \mathrm{C_{ano2}}. \end{aligned} \tag{8}$$

Next, we explain the cost function and the method of adding labels. In this study, we adopted cross-entropy as the loss function [14], which is defined as

$$H = -\sum_{i \in C} p(i) \log y(i), \tag{9}$$

where $y(i)$ is the estimated probability of class $i \in C$ calculated using Equation (7). $p(i)$ is the ground-truth one-hot vector of the actual class $i \in C$, which is defined as

$$\begin{aligned} \text{True class: } \mathrm{C_{norm}} &\Rightarrow [p(\mathrm{C_{norm}}) \ p(\mathrm{C_{ano1}}) \ p(\mathrm{C_{ano2}})]^\top = [1 \ 0 \ 0]^\top, \\ \text{True class: } \mathrm{C_{ano1}} &\Rightarrow [p(\mathrm{C_{norm}}) \ p(\mathrm{C_{ano1}}) \ p(\mathrm{C_{ano2}})]^\top = [0 \ 1 \ 0]^\top, \\ \text{True class: } \mathrm{C_{ano2}} &\Rightarrow [p(\mathrm{C_{norm}}) \ p(\mathrm{C_{ano1}}) \ p(\mathrm{C_{ano2}})]^\top = [0 \ 0 \ 1]^\top. \end{aligned} \tag{10}$$

A large cross-entropy H indicates large errors. Therefore, the machine learning method employed in this study aims to obtain parameters that result in a low H. We adopted Adam as the optimization algorithm [15].

To search for a desirable model, multiple values of learning rates and epochs were verified. In particular, $10^{-5.5}$, $10^{-6.0}$ and $10^{-6.5}$ were adopted as the learning rates, and the number of epochs was ranged from 1 to 200. We used the method of choosing the learning rate and epochs, which resulted in the highest validation accuracy of the developed models.

3. Experiment.

3.1. Objective and outline. We constructed a CNN to classify PAWP from chest X-ray images into one of three classes, and conducted an experiment to evaluate the generalization of the model. We used the dataset presented in Section 2.1, and structure of CNN and learning condition presented in Section 2.2. The following software environment was used for machine learning: Ubuntu 20.04.3 LTS, Python 3.9.4, and Keras 2.7.0. The hardware environment was as follows: 12 GB RAM, Intel Core i7-9750H CPU (base frequency: 2.6 GHz, max turbo frequency: 4.50 GHz), and NVIDIA GeForce GTX 1660 Ti GPU (clock speed: 1770 MHz, Memory: 6 GB). To avoid overflow, we conducted minibatch learning using a batch size of 32.

3.2. Results and discussions. Among all CNNs, we adopted the one which had a learning rate of $10^{-5.5}$ and 96 epochs that obtained the highest accuracy using the validation dataset. Estimation and reliability scores were calculated using the main training, validation, and test data. The values are listed in Table 1. When comparing the reliability scores among the three datasets, the differences were small. Therefore, we concluded that the adopted CNN did not overfit.

TABLE 1. Reliability of each dataset (learning rate: $10^{-5.5}$, epochs: 96)

	Main training data	Validation data	Test data
Accuracy	0.699	0.687	0.633
Precision	0.694	0.666	0.544
Recall	0.598	0.544	0.536

The confusion matrices for each dataset are shown in Figure 3. First, we describe the confusion matrix of the main training data, which is shown in Figure 3(a). The number of correct answers for C_{norm} was 312 out of 331. The recall of C_{norm} was $312/(312+7+12) \simeq 0.943$, which is high. In the case of C_{ano1}, recall was $46/(99+46+16) \simeq 0.286$. Although the true class is C_{ano1}, the cases of estimating it as C_{norm} include many misclassifications. In the case of C_{ano2}, recall was $60/(33+13+60) \simeq 0.566$. We consider this to be a high value compared with C_{ano1}.

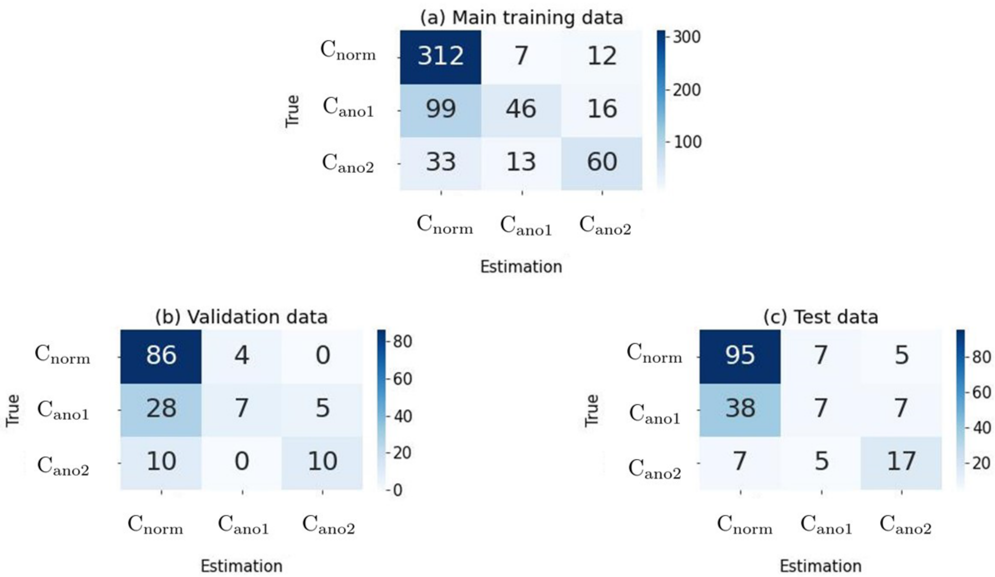

FIGURE 3. Confusion matrices

The estimation tendencies of the validation and test datasets (Figures 3(b) and 3(c)) are similar to those of the main training dataset. It is characteristic that the CNN estimates C_{ano1} as C_{norm} for all datasets. Based on the results obtained in this study, we consider that the detection of the slightly cardiac anomaly state should be improved. We consider that data augmentation can improve performance. Data augmentation involves adding new processed data to enable effective machine learning using few data [16]. A survey of this paper indicated that accuracy could be improved via data augmentation. Therefore, we will employ this method in a future work.

4. **Conclusion.** In this study, we constructed a CNN that classifies PAWP states in three classes (C_{norm}, C_{ano1}, and C_{ano2}). Although PAWP is measured using RHC, this method has some risk of complications [1]. Hence, objective and noninvasive methods for measuring PAWP are required. Therefore, we developed a CNN for the three-state classification of PAWP from chest X-ray images. Consequently, the accuracy of the test data that was not used for parameter learning was 63.3%. However, the detection of the slightly anomalous state C_{ano1} must be improved. We consider that the low accuracy of C_{ano1} is caused by the small sample size. Therefore, we aim to improve the accuracy of C_{ano1} using data augmentation in the future.

Acknowledgement. This 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.

REFERENCES

[1] M. M. Hoeper, S. H. Lee, R. Voswinckel, M. Palazzini et al., Complications of right heart catheterization procedures in patients with pulmonary hypertension in experienced centers, *Journal of the American College of Cardiology*, vol.48, no.12, pp.2546-2552, 2006.

[2] A. T. McDonagh, A. Theresa, M. Metra, M. Adamo, S. R. Gardner et al., 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed 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 of the heart failure association (HFA) of the ESC, *European Heart Journal*, vol.42, no.36, pp.3599-3726, 2021.

[3] R. Jain, M. Gupta, S. Taneja and D. J. Hemanth, Deep learning based detection and analysis of COVID-19 on chest X-ray images, *Applied Intelligence*, vol.51, pp.1690-1700, DOI: 10.1007/s10489-020-01902-1, 2021.

[4] T. Matsumoto, S. Kodera, H. Shinohara, H. Ieki et al., Diagnosing heart failure from chest X-ray images using deep learning, *International Heart Journal*, vol.61, no.4, pp.781-786, DOI: 10.1536/ihj.19-714, 2020.

[5] Q. Que, Z. Tang, R. Wang, Z. Zeng et al., CardioXNet: Automated detection for cardiomegaly based on deep learning, *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, https://ieeexplore.ieee.org/document/8512374, 2022.

[6] J. G. Nam, J. Kim, K. Noh, H. Choi et al., Automatic prediction of left cardiac chamber enlargement from chest radiographs using convolutional neural network, *European Radiology*, vol.31, pp.8130-8140, DOI: 10.1007/s00330-021-07963-1, 2021.

[7] A. A. Novikov, D. Lenis, D. Major, J. Hladůvka et al., Fully convolutional architectures for multi-class segmentation in chest radiographs, *IEEE Trans. Medical Imaging*, vol.37, no.8, pp.1865-1876, DOI: 10.1109/TMI.2018.2806086, 2018.

[8] J. Betancur, F. Commandeur, M. Motlagh, T. Sharir et al., Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT: A multicenter study, *JACC: Cardiovascular Imaging*, vol.11, no.11, pp.1654-1663, DOI: 10.1016/j.jcmg.2018.01.020, 2018.

[9] Q. Tao, W. Yan, Y. Wang, E. H. M. Paiman et al., *Deep Learning-Based Method for Fully Automatic Quantification of Left Ventricle Function from Cine MR Images: A Multivendor, Multicenter Study*, The Radiological Society of North America, https://pubs.rsna.org/doi/10.1148/radiol.2018180513, Accessed on June 13, 2022.

[10] J.-Y. Zhao, J. Gong, S.-T. Ma and Z.-M. Lu, Curvature gray feature decomposition based finger vein recognition with an improved convolutional neural network, *International Journal of Innovative Computing, Information and Control*, vol.16, no.1, pp.77-90, DOI: 10.24507/ijicic.16.01.77, 2020.

[11] E. Çallı, E. Sogancioglu, B. V. Ginneken, K. G. Leeuwen et al., Deep learning for chest X-ray analysis: A survey, *Medical Image Analysis*, vol.72, DOI: 10.1016/j.media.2021.102125, 2021.
[12] Y. Hirata, K. Kusunose, T. Tsuji, K. Fujimori et al., Deep learning for detection of elevated pulmonary artery wedge pressure using standard chest X-ray, *Canadian Journal of Cardiology*, vol.37, no.8, pp.1198-1206, 2021.
[13] Y. Saito, Y. Omae, D. Fukamachi, K. Nagashima et al., Quantitative estimation of pulmonary artery wedge pressure from chest radiographs by a regression convolutional neural network, *Heart and Vessels*, vol.37, pp.1387-1394, DOI: 10.1007/s00380-022-02043-w, 2022.
[14] P.-T. de Boer, D. P. Kroese, S. Mannor and R. Y. Rubinstein, A tutorial on the cross-entropy method, *Annals of Operations Research*, vol.134, pp.19-67, DOI: 10.1007/s10479-005-5724-z, 2005.
[15] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv.org*, arXiv: 1412.6980, 2015.
[16] C. Shorten and T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, *Journal of Big Data*, vol.6, no.60, DOI: 10.1186/s40537-019-0197-0, 2019.