2022 IEEE International Conference on
Systems, Man, and Cybernetics (SMC)

Predicting the severity of Neonatal Chronic Lung Disease from chest X-ray images using deep learning

line 1: 1st Ryunosuke Maeda
line 2: *University of Hyogo*
line 3: *Graduate School of Data Science*
line 4: Hyogo-ken, Japan
line 5: mae.ryu.study@gmail.com

line 1: 2nd Daisuke Fujita
line 2: *University of Hyogo*
line 3: *Graduate School of Data Science*
line 4: Hyogo-ken, Japan
line 5: dai_fujita@eng.u-hyogo.ac.jp

line 1: 3rd Kosuke Tanaka
line 2: *Saitama Medical University*
line 3: *Saitama Medical Center Pediatrics*
line 4: Saitama-ken, Japan
line 5: kotanaka0403@gmail.com

line 1: 4th Jyunichi Ozawa
line 2: *Saitama Medical University*
line 3: *Saitama Medical Center Pediatrics*
line 4: Saitama-ken, Japan
line 5: ozajun321@gmail.com

line 1: 5th Mitsuhiro Haga
line 2: *Saitama Medical University*
line 3: *Saitama Medical Center Pediatrics*
line 4: Saitama-ken, Japan
line 5: haga_m@saitama-med.ac.jp

line 1: 6th Haoyuki Miyahara
line 2: *Saitama Medical University*
line 3: *Saitama Medical Center Pediatrics*
line 4: Saitama-ken, Japan
line 5: ma03084ma@yahoo.co.jp

line 1: 7th Fumihiko Nanba
line 2: *Saitama Medical University*
line 3: *Saitama Medical Center Pediatrics*
line 4: Saitama-ken, Japan
line 5: nambaf@saitama-med.ac.jp

line 1: 8th Syoji Kobashi
line 2: *University of Hyogo*
line 3: *Graduate School of Data Science*
line 4: Hyogo-ken, Japan
line 5: kobashi@ieee.org

2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC) | 978-1-6654-5258-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SMC53654.2022.9945486

Abstract—**Chronic lung disease (CLD) is the most common and serious lung disease in premature infants. In this study, we predict the severity (mild or severe) of neonatal chest X-ray images using a convolutional neural network (CNN) to enable early intervention to provide personalized treatment and improve prognosis. Thirty subjects were tested in a leave-one-out cross validation experiment using 30 chest X-ray images of 11 patients with mild disease and 19 patients with severe disease at 7 days of age. To improve the prediction accuracy, we proposed to limit the input image of the CNN to the lung field region and to use a pre-training model for transfer learning. Four different experiments were conducted, comparing the results with different input images (whole image or lung field region) and with and without transfer learning. The results showed that the best accuracy was obtained when the entire image was used as input and no transfer learning was performed, with an Accuracy of 0.667.**

Keywords—deep learning, medical imaging, chronic lung disease, yolo, convolutional neural networks, transfer learning

I. Introduction

Neonatal chronic lung disease (CLD) is the most common and serious lung disease in premature infants, diagnosed in neonates requiring respiratory support at 28 days of age [1], and the incidence of CLD is as high as 61.2% in infants weighing less than 1000 g at birth [2]. Although the survival rate of very preterm infants with a gestational age of less than 28 weeks has improved with the development of neonatal care, the incidence of CLD due to prematurity is on the rise. In addition, the detailed pathogenesis of the disease has not been elucidated, and there is no effective treatment [3]. Therefore, CLD requires intensive care, including long-term respiratory management, and as many as 150-200 cases of prolonged hospitalization for more than 6 months after birth have been identified annually [4]. Therefore, prediction of CLD severity is expected to enable early intervention such as preventive treatment suited to individual patients, shortening treatment and improving prognosis.

Several studies have been conducted to predict the severity of CLD using premature infant patient information, and many factors have been identified as CLD risk factors, including low birth weight, low gestational age, male, patent ductus arteriosus (PDA), sepsis, and mechanical ventilation [5][6][7][8]. Studies have included postnatal age in these risk factors to predict CLD severity prior to CLD diagnosis [9], but chest radiographs have not been used to predict CLD severity.

Currently, the majority of preterm infants born at less than 37 weeks' gestation who are admitted to the neonatal intensive care unit (NICU) are born without adequate lung function, and chest radiographs are frequently taken to confirm lung function status [10]. In recent years, several studies have been conducted on image classification using convolutional neural networks (CNN), which can classify images with a high accuracy of 97.7%, suggesting the possibility of identifying pathogens based on image features[11].

This study aims to improve the accuracy of predicting the severity of CLD by analyzing chest X-ray images taken in NICUs. Proposal 1 automatically extracts lung field regions. In the proposed method 2, a pre-training model for pneumonia prediction is created using open COVID-19 data. The pre-training model is re-learned using neonatal data to obtain a prediction model for CLD severity. The prediction accuracy is then compared by changing the image range (entire image or lung field region) and whether or not the model is relearned.

II. Materials and Methods

A. Dataset

This study will focus on chest X-ray images of 30 patients taken at the NICU of Saitama Medical School General Medical Center. Each image includes the severity of the CLD and the date and time the image was taken. There were a total of 30 cases, 11 with mild disease and 19 with severe disease.

CLD severity is defined as follows [12] :

mild : Requires oxygen supplementation for at least 28 days after birth and at 36 weeks of age after last menstrual period or at discharge

moderate : Requires oxygen supplementation for at least 28 days after birth and treatment with less than 30% oxygen at 36 weeks of age after last menstrual period

severe : Requires oxygen supplementation for at least 28 days after birth and treatment with oxygen or positive pressure of at least 30% at 36 weeks of age after last menstrual period

B. CNN

Convolutional Neural Network (CNN) are a means of automatically extracting discriminative features from images, and are now rapidly gaining popularity as a method for analyzing medical images. CNN contain many layers that transform their input with convolution filters of a small extent[13].

C. Transfer Learning

Transfer learning is a method of transferring knowledge extracted from one domain (source domain) to another domain (target domain). One of the most common transfer learning approaches used in the analysis of chest X-ray images is the use of prior learning [14]. Pre-training on a large dataset and using the learned weights as initial values for subsequent tasks to fine-tune the network [15]. In transfer learning, only the final layer of the network is relearned; prior weights are not adjusted. In this approach, useful low-level features are learned from data in the source domain, allowing neural networks to be trained for new tasks using relatively small data sets.

D. YOLO

You Only Look Once(YOLO) series is one of the object recognition models with a real-time object detection algorithm that generates features from input images and feeds them to the prediction system, which draws a box around the object and predicts its class. YOLOv5, developed based on YOLOv3 [16], is currently the latest version. An overview of the YOLOv5 network [17] is shown in Figure 1. As shown in Figure 1, it consists of three major networks, The backbone is a convolutional neural network that aggregates and forms image features, the neck combines those image features, and the head uses the features for bounding box generation and class prediction.

YOLO uses a confidence score as an evaluation index. This confidence score reflects how confident the model is that the box contains the object and how accurate the predicted box is [18]. The confidence level is defined by the following equation.

$$Confidence = P_r(Object) \cdot IOU_{pred}^{truth} \quad (1)$$

$P_r(Object)$ is the probability that the object exists, and IOU_{pred}^{truth}(intersection over union; IOU) is a measure of how much the predicted and correct regions overlap, which is the common part of the two regions divided by the union set. The object probability for each class, $P_r(Class_i|Object)$ is predicted and multiplied by the confidence level in equation (1) to indicate the degree to which objects are present in the box and the predicted probability of the class of objects there, which is the confidence score. The confidence score is defined by the following equation.

$$P_r(Class_i|Object) \cdot P_r(Object) \cdot IOU_{pred}^{truth} = P_r(Class_i) \cdot IOU_{pred}^{truth} \quad (2)$$

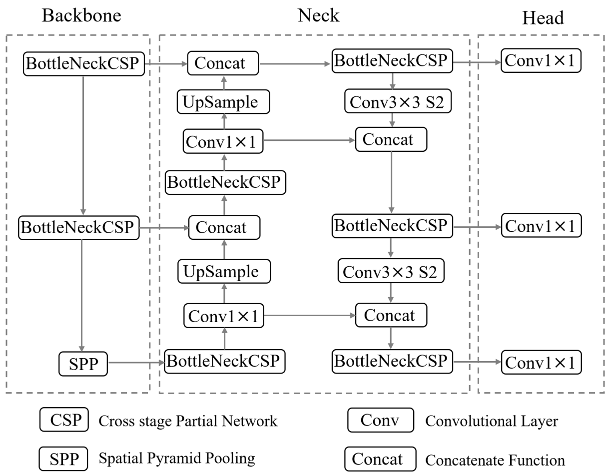


Figure 1 YOLOv5 network overview

III. PROPOSED METHOD

In Chapter Ⅲ, we present the proposed two-step method to improve the accuracy of predicting CLD severity, and the procedure for comparing the prediction accuracy using these methods in practice. The first proposed method is the automatic extraction of lung field regions, and the second method is the creation of a model using chest X-ray images for pre-training to predict pneumonia. Comparison is made in predicting the severity of CLD using lung field images and the pre-training model.

A. Proposed method 1: Automatic lung field area extraction method

Since CLD is a lung disease, it is assumed that images outside of the lung field region are unnecessary. Therefore, we propose to improve the accuracy of CLD severity prediction by detecting lung field regions from chest X-ray images by object detection, extracting lung field regions using their coordinates, and using the extracted images as input images for CNN. The lung field region is automatically extracted as follows.

[Step 1] Annotation of lung field area

[Step 2] Creation of a learning model for object detection in the lung field region from YOLOv5

[Step 3] Input chest X-ray images to the created training model and generate bounding boxes

[Step 4] Crop the lung field area from the generated bounding box coordinates

B. Proposed Method 2: Prior Learning Modeling by Predicting Pneumonia (chest X-ray images of adults)

Since there is insufficient neonatal data available for this study, we created a pre-training model using chest X-ray images of adults and transferred the model to neonatal data to improve the accuracy of predicting CLD severity. Therefore, we create a model for 2-class classification (pneumonia or normal) using chest X-ray images[19] taken during the

COVID-19 examination published on Kaggle by using CNN. The network uses AlexNet [20] with a total of eight layers, consisting of five convolutional layers and three all-coupled layers.

C. *Prediction of CLD severity*

The chest X-ray image of a neonate is input into the CNN to classify CLD severity as mild or severe. In this study, we use four different methods to predict CLD severity and compare the accuracy of each method.

1. learning CNN by whole chest X-ray images
2. Learning CNN by lung field region images
3. Transfer learning of a pre-training model using the entire chest X-ray images
4. Transfer learning of a pre-training model using lung field images

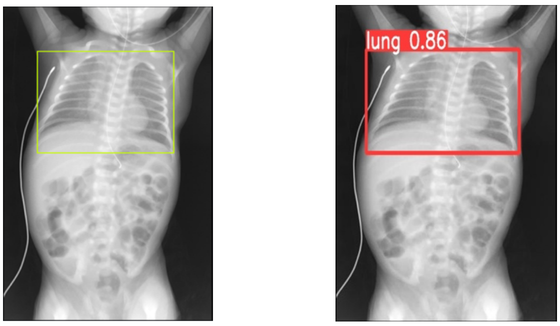
It monitors the loss value of the validation data and halves the learning rate if it is not updated for 10 epochs, and terminates learning if it is not updated for 30 epochs.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. *Automatic lung field extraction results*

A learning model was created from YOLOv5 by randomly classifying all 97 chest X-ray images from 30 subjects (68 training data, 19 validation data, and 10 evaluation data) to generate a bounding box for the lung field region. An example of annotated image is shown in Figure 2(a), and an example of bounding box generation is shown in Figure 2(b). Figure 3 shows the result of extracting lung field region images.

The results of automatic detection of lung field regions using YOLOv5 showed that no bounding boxes were generated that deviated significantly from the lung field regions, and all evaluation data showed a reliability score of 0.82 or higher. Based on these results, we consider that the automatic lung field region detection model created in this study is sufficiently accurate. However, there are annotation concerns apart from the accuracy of the model. As a preprocessing step for the training images, the authors, who are not physicians, manually annotated the lung field regions on the chest X-ray images. Therefore, it may differ from the actual lung field area. Therefore, in order to achieve more accurate automatic extraction of the lung field area, we believe it is necessary to ask the physician to make a judgment on the annotated image and the image after automatic lung field area detection.



(a)annotation (b)bounding box

Figure 2 Example of automatic lung field area detection

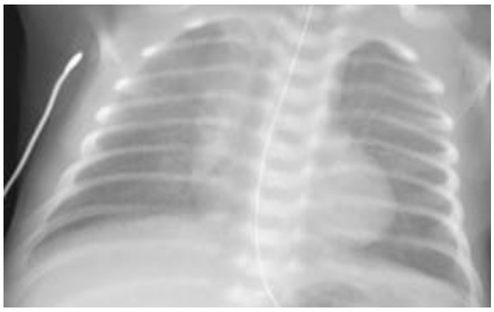


Figure 3 Example of lung field area image

B. *Prior Learning Model Accuracy by Predicting Pneumonia (chest X-ray images of adults)*

Prediction of pneumonia was performed using 5910 chest X-ray images of adults (4334 pneumonia cases and 1576 normal cases) with a ratio of training data, validation data, and evaluation data of 7:2:1. The learning curve for loss during learning is shown in Figure 4(a), and the learning curve for accuracy is shown in Figure 4(b). Here, the number of epochs is divided in half for visual clarity. Table 1 shows the confusion matrix of the results predicted using the evaluation data. In addition, *Accuracy*, *Precision*, *Recall*, and *F1_score* are defined as evaluation indices as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

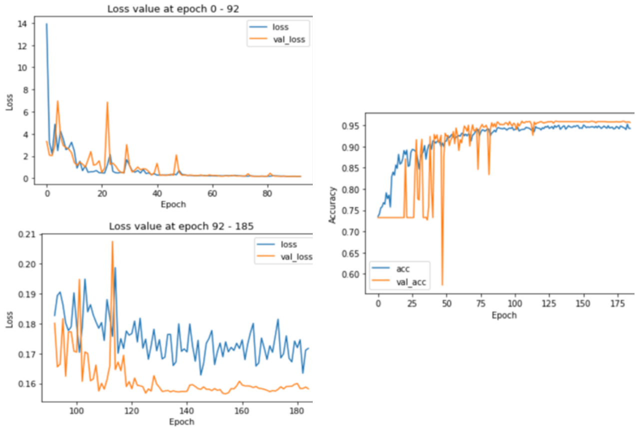
$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1_{score} = \frac{2Precision \cdot Recall}{Precision + Recall} \tag{6}$$

TP indicates the number of cases in which the pneumonia image was correctly predicted as pneumonia, *TN* indicates the number of cases in which the normal image was correctly predicted as normal, *FP* indicates the number of cases in which the pneumonia image was incorrectly predicted as normal, and *FN* indicates the number of cases in which the normal image was incorrectly predicted as pneumonia. Table 2 shows the confusion matrix in Table 1 and the evaluation indices obtained from equations (3), (4), (5), and (6). The AUC was 0.93.

A model was created by predicting pneumonia using chest X-ray images of adults for preliminary learning. Transfer training of this model on neonatal data showed no improvement in accuracy. A pre-training model using chest X-ray images was created for the target region of this study. To see if there is a problem in this respect, it is necessary to use a pre-training model on a large ImageNet, which is not a chest X-ray, and compare it with the results of this experiment. In transfer learning, only the final layer is re-trained, and the weights of the other trained networks are fixed. To create a CLD prediction model suitable for neonatal data, we consider the application of fine tuning to relearn the weights of the entire model, including the feature extraction part.



(a) loss (b) accuracy

Figure 4 Learning curve

Table 1 Pneumonia Prediction Results

		Predicted value	
		pneumonia	normal
true value	pneumonia	373	17
	normal	58	176

Table 2 Pneumonia Prediction Results (Assessment Indicators)

Accuracy	0.880
Precision	0.865
Recall	0.956
F1_score	0.908

C. *CLD severity prediction results*

Thirty chest X-ray images taken on the seventh day after birth from 30 subjects were used as training data (24 images), validation data (5 images), and evaluation data (1 image), and a total of 30 times of training and evaluation were used to predict CLD severity using a method called Leave One Out Cross Validation.

Experiments were conducted with two classes of CNN outputs, mild or severe, and the confusion matrices resulting from the prediction of the evaluation data are shown in Tables 3, 4, 5, and 6, respectively. The evaluation indices obtained from Tables 3, 4, 5, and 6 and Equations (3), (4), (5), and (6) are shown in Table 7. The average number of epochs required during the study is shown in Table 8.

Comparing the results of the four different experiments, no significant differences were found in the prediction of CLD severity by changing the range of images. Although the average number of epochs during learning was significantly reduced by transfer learning, the accuracy was not improved. From the confusion matrices presented in Tables 5, 6, 7, and 8, it is often the case that the predicted results are severe. This is due to the large number of severe cases in the data used in this study. Therefore, we consider that the biased prediction results can be improved by changing the loss function to take unbalanced data into account, by adding more light cases and reducing heavy cases, and by achieving a state of equilibrium for each number of data. In this study, a bounding box was generated and the lung field area was automatically extracted by clipping the bounding box. However, although this method increased the proportion of images in which the lung fields occupied a larger percentage of the total images in which images other than the lung fields were visible, we believe that automatic extraction limited to the lung fields is necessary to make the prediction of CLD severity more precise.

Table 3 CLD severity prediction results (whole image)

		Predicted value	
		severe	mild
true valu	severe	15	4
	mild	6	5

Table 4 Results of CLD severity prediction (lung field area)

		Predicted value	
		severe	mild
true valu	severe	18	1
	mild	11	0

Table 5 CLD severity prediction results (whole image, transfer study)

		Predicted value	
		severe	mild
true valu	severe	14	5
	mild	5	6

Table 6 Results of CLD severity prediction (lung field area, metastasis study)

		Predicted value	
		severe	mild
true valu	severe	12	7
	mild	8	3

Table 7 CLD severity prediction results (evaluation index)

	Whole image	lung field area	Whole Image Transfer Learning	Lung Field Area Transfer Learning
Accuracy	0.667	0.600	0.633	0.500
Precision	0.714	0.621	0.737	0.600
Recall	0.789	0.947	0.737	0.632
F1_score	0.750	0.750	0.737	0.616

Table 8 Comparison of average number of epochs

	Whole image	lung field area	Whole Image Transfer Learning	Lung Field Area Transfer Learning
Average number of epochs	151	157	86	90

V. Conclusion

CLD is the most common and serious lung lesion in premature infants and tends to cause developmental problems in adulthood, not only in cardiopulmonary function but also in the brain, vision, and hearing. It has been suggested that early intervention and treatment of CLD may lead to a decrease in the number of such developmentally disabled children. In this paper, we propose two methods to improve the accuracy of CLD severity prediction in order to enable early intervention for individual patients and to improve prognosis. The proposed method 1 automatically extracts the lung field region to limit the range of images to be input to the CNN, and it showed a confidence level of 0.82 or higher for all images in the evaluation data. In the proposed method 2, a model was created by predicting pneumonia using chest X-ray images for pre-training.

Experiments were conducted on 30 cases of neonatal chest X-ray images using lung field region images and pre-training models generated by the proposed methods 1 and 2. The prediction accuracy of the method without the pre-training model using the entire chest X-ray image as input to the CNN was 0.677, which is better than the other three methods. This result is not sufficient compared to the prediction of CLD severity using only patient information in a previous study [9].

Future work includes creating a learning model that takes into account unbalanced data, modifying parameters, and improving prediction accuracy by combining the features obtained from neonatal chest X-rays in this study with individual patient information such as gender and gestational age.

Acknowledgment

We thank Associate Professor Fumihiko Namba, the principal investigator of "Saitama Medical University General Medical Center," and other researchers for providing the chest X-ray images used in this study and for sharing their medical knowledge on chronic lung diseases in neonates. I would also like to thank Professor Syoji Kobashi, Associate Professor Manabu Nii, Assistant Professor Daisuke Fujita, and other members of the Data Science Research Group for their guidance throughout the project.

References

[1] J. D. Horbar et al. "Trends in mortality and morbidity for very low birth weight infants, 1991–1999," *Pediatrics*, vol. 110, pp.143-151, 2002.

[2] Japan Ministry of Health, Labour and Welfare, "Results of national survey," pp. 112, 2010.

[3] S. Ohnishi. "Manseihaishikkan・Noukyoketushikkannitaisuru Stem cell therapy,"*Journal of Japan society of Perinatal and Neonatal Medicine*, Vol. 56, No. 4, pp. 590-594, 2021(in Japanese).

[4] Y. Ogawa "Manseihaishikkannokanritoyobounikansurukenkyu," *heiseininendokouseisyousinsinsyougaikenkyuhankenkyuhoukokusyo*, pp. 43-50, 1991 (in Japanese).

[5] M. A. Rojas et al. "Changing trends in the epidemiology and pathogenesis of neonatal chronic lung disease," *J Pediatr,* vol. 126, no. 4, pp.605-610, 1995.

[6] D. D. Marshall et al. "Risk factors for chronic lung disease in the surfactant era: a North Carolina population-based study of very low birth weight infants. North Carolina Neonatologists Association," *Pediatrics*, vol. 104, no. 6, pp. 1345-1350, 1999.

[7] W. Oh et al "Association between fluid intake and weight loss during the first ten days of life and weight loss during the first ten days of life and risk of bronchopulmonary dysplasia in extremely low birth weight infants," *J. Pediatr*, vol. 147, no. 6, pp. 786-790, 2005.

[8] N. Ambalavanan et al. "Predictors of death or bronchopulmonary dysplasia in preterm infants with respiratory failure," *J. Perinatol*, vol. 28, pp. 420-426, 2008.

[9] M. M. Laughon et al. "Prediction of bronchopulmonary dysplasia by postnatal age in extremely premature infants," *American journal of respiratory and critical care medicine*, vol. 183, no. 12, 2011.

[10] S. Tani et al. "Portable radiography in NICU : Factors on density variation," *Japanese Journal of Radiological Technology*, Vol 49, No.9, pp. 1663, 1993.

[11] T. Dyer et al. "Diagnosis of normal chest radiographs using an autonomous deep learning algorithm," *Clinical radiology*, vol. 76, no. 6, pp. 9-15, 2021.

[12] A. Richard et al. "Validation of the National Institutes of Health Consensus Definition of Bronchopulmonary Dysplasia," *Pediatrics*, vol. 116, no. 6, pp. 1353-1360, 2005.

[13] L. Geert et al., "*A survey on deep learning in medical image analysis*", Medical Image Analysis, vol. 42, pp.60-88, 2017.

[14] E. Sogancioglu et al. "Deep Learning for Chest X-ray Analysis: A Survey," *Medical Image Analysis*, vol. 72, 2021.

[15] J. Yosinski et al. "How transferable are features in deep neural networks?," *Advances in Neural Information Processing Systems*, vol. 27, pp. 3320-3328, 2014.

[16] J. Redmon et al. "YOLOv3: An Incremental Improvement," C*omputer Vision and Pattern Recognition*, 2018.

[17] X. Renjie et al. "A Forest Fire Detection System Based on Ensemble Learning," *Forests,* vol. 12, no. 2, 2021.

[18] J. Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.

[19] kaggle.com,"Corona Hack – Chest X-Ray - Dataset," https://www.kaggle.com/praveengovi/coronahack-chest-xraydataset

[20] A. Krizhevsky et al., "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 2012.