



# OPEN Extraction and evaluation of features of preterm patent ductus arteriosus in chest X-ray images using deep learning

Phillip Chang<sup>1</sup>, Hyeon Sung Choi<sup>2</sup>, Jimin Lee<sup>1,3,4</sup>✉ & Hyun Ho Kim<sup>5,6</sup>✉

Echocardiography is the gold standard of diagnosis and evaluation of patent ductus arteriosus (PDA), a common condition among preterm infants that can cause hemodynamic abnormalities and increased mortality rates, but this technique requires a skilled specialist and is not always available. Meanwhile, chest X-ray (CXR) imaging is also known to exhibit signs of PDA and is a routine imaging modality in neonatal intensive care units. In this study, we aim to find and objectively define CXR image features that are associated with PDA by training and visually analyzing a deep learning model. We first collected 4617 echocardiograms from neonatal intensive care unit patients and 17,448 CXR images that were taken 4 days before to 3 days after the echocardiograms were obtained. We trained a deep learning model to predict the presence of severe PDA using the CXR images, and then visualized the model using GradCAM++ to identify the regions of the CXR images important for the model's prediction. The visualization results showed that the model focused on the regions around the upper thorax, lower left heart, and lower right lung. Based on these results, we hypothesized and evaluated three radiographic features of PDA: cardiothoracic ratio, upper heart width to maximum heart width ratio, and upper heart width to thorax width ratio. We then trained an XGBoost model to predict the presence of severe PDA using these radiographic features combined with clinical features. The model achieved an AUC of 0.74, with a high specificity of 0.94. Our study suggests that the proposed radiographic features of CXR images can be used as an auxiliary tool to predict the presence of PDA in preterm infants. This can be useful for the early detection of PDA in neonatal intensive care units in cases where echocardiography is not available.

Patent ductus arteriosus (PDA) is a common condition among preterm infants. For example, according to the 2022 Annual Report of the Korean Neonatal Network, 26.3% of low-birth-weight infants (birth weight < 1500 g) exhibited symptomatic PDA. Typically, the ductus arteriosus naturally closes within 72 h after birth in newborns; however, this process can be delayed in preterm infants, leading to hemodynamic abnormalities and increased morbidity and mortality rates among these newborns<sup>1</sup>.

The approach to treating PDA is gradually changing. Surgical treatment was common in the past, but concerns about side effects have led to the recommendation of less invasive treatments. Instead of surgical intervention, conservative fluid therapy, medication, and device closure are now being used to promote natural closure. Throughout this treatment process, clinicians need to frequently evaluate the state of PDA using echocardiography<sup>2</sup>, the gold standard in PDA diagnosis. However, echocardiography is not always available in all neonatal intensive care units (NICUs), and the interpretation of echocardiograms requires a specialist. Therefore, there is a need for an auxiliary diagnostic tool that can be used to evaluate PDA in preterm infants.

The integration of artificial intelligence (AI) technology in NICUs has presented new possibilities. AI research and applications utilizing medical data from critically ill patients are becoming more active in both pediatric

<sup>1</sup>Graduate School of Health Science and Technology, Ulsan National Institute of Science and Technology, 44919 Ulsan, Republic of Korea. <sup>2</sup>Jeonbuk National University School of Medicine, Jeonju 54907, Republic of Korea. <sup>3</sup>Department of Nuclear Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea. <sup>4</sup>Graduate School of Artificial Intelligence, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea. <sup>5</sup>Department of Pediatrics, Jeonbuk National University School of Medicine, Jeonju 54907, Republic of Korea. <sup>6</sup>Research Institute of Clinical Medicine of Jeonbuk National University-Biomedical Research Institute of Jeonbuk National University Hospital, Jeonju 54907, Republic of Korea. ✉email: jimilee@unist.ac.kr; hyunhokim@jbnu.ac.kr

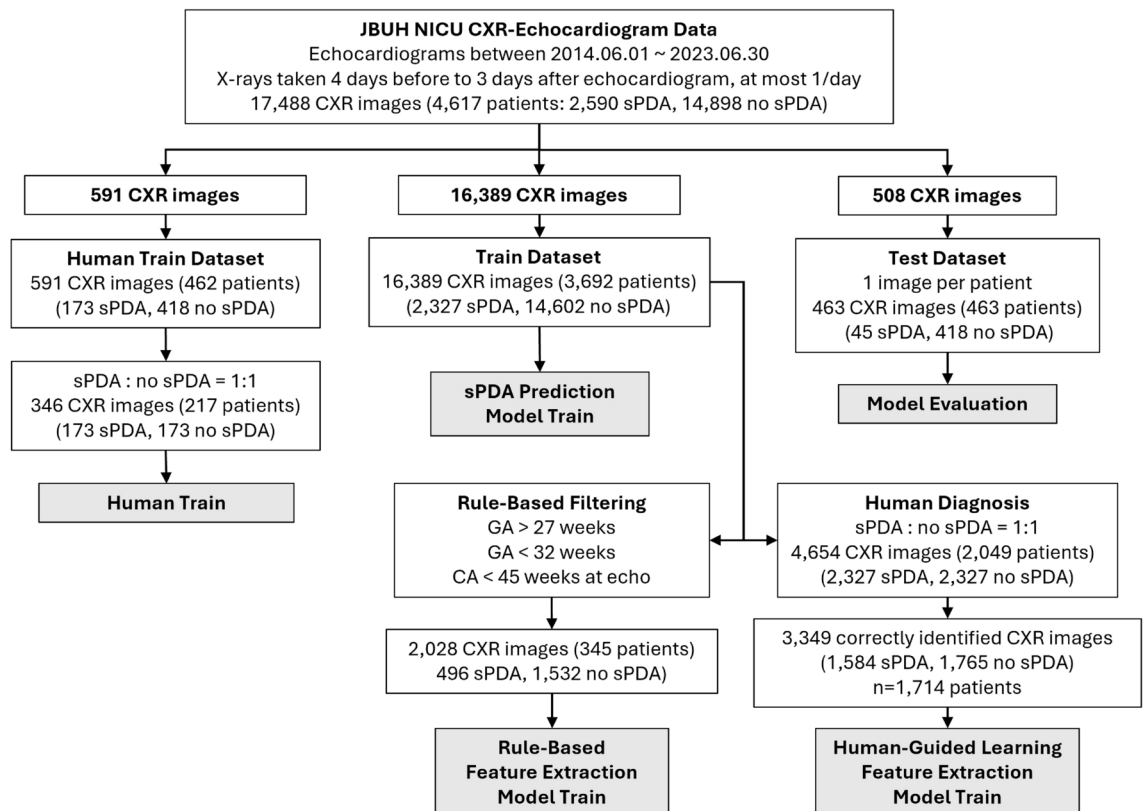
intensive care units<sup>3</sup> and NICUs<sup>4,5</sup>. For example, research has been conducted on models to predict the severity and prognosis of bronchopulmonary dysplasia caused by prolonged mechanical ventilation<sup>6,7</sup>. Also, previous studies have developed predictive models for neonatal respiratory diseases using deep learning algorithms on neonatal chest X-ray (CXR) images<sup>8,9</sup>, which are the primary imaging modality in NICUs due to the convenience of examination and relatively low radiation exposure. In terms of PDA, while qualitative features that can distinguish PDA from CXR images have been suggested<sup>10,11</sup>, at present there are no quantitatively or objectively defined radiographic features of PDA in CXR images of preterm infants.

This study aims to define quantitative features of PDA in CXR images. To do this, we first build a dataset of PDA-labeled CXR images of preterm infants and train a deep learning model, which we then visualize to find the important radiographic features in the CXR images used to detect PDA. Here, we introduce human-guided easy learning, a process that trains the model using manually selected images expected to show PDA-specific features. Then to evaluate the proposed features for diagnostic value, we train a machine learning model that can predict severe PDA in preterm infants using the defined radiographic features combined with clinical data, such as gestational age and birth weight. The results indicate that CXR images have potential as an auxiliary diagnostic tool for symptomatic PDA in preterm infants, which could be particularly valuable in situations where specialists are not available or echocardiography is limited.

## Methods

### Data collection

This study used retrospective data from the NICU of Jeonbuk National University Hospital and was approved by the Institutional Review Board of Jeonbuk National University Hospital (IRB No. 2024-03-027) and conducted following the Declaration of Helsinki. The need for informed consent was waived by the ethics committee/institutional review board of Jeonbuk National University Hospital. The inclusion criteria for the data were as follows: (1) echocardiograms taken from June 1st, 2014, to June 30th, 2023, (2) CXR images taken from 4 days before to 3 days after the echocardiograms, and (3) at most 1 CXR per day. In total the data consisted of 17,486 CXR images from 4617 echocardiogram records (Fig. 1). We did not exclude any cases based on other congenital heart or lung diseases, as this study aimed to identify radiographic features that can be applied in clinical practice regardless of the presence of other diseases.



**Fig. 1.** Flowchart of the data used in this study. The data were collected from the Jeonbuk National University Hospital Neonatal Intensive Care Unit from 2014 to 2023 and labeled by a pediatric cardiologist. The labeling criteria to diagnose severe PDA were as follows: a transductal diameter of 1.5 mm or more, or a left atrial diameter:aortic diameter ratio of 1.4:1 or more. Abbreviations: JBUH, Jeonbuk National University Hospital; NICU, Neonatal intensive care unit; CXR, Chest X-ray; sPDA, severe patent ductus arteriosus.

The definitions of hemodynamically significant PDA may vary among neonatologists, which usually take into account the clinical context<sup>12</sup>. In order to make the labeling more objective and consistent, we defined severe PDA (sPDA) based on the echocardiogram results and used it as the ground truth. The data were labeled by a pediatrician using the following criteria for diagnosing sPDA: a transductal diameter of 1.5 mm or more, or a left atrial diameter:aortic diameter ratio of 1.4:1 or more. Among the 4,617 echocardiogram records, 441 cases (9.6%) were diagnosed with sPDA, and the remaining 4176 cases (90.4%) were labeled as negative. A subset of 463 randomly chosen cases was isolated into a test group to evaluate the two main models used in this work, i.e., a feature extraction model and sPDA prediction model. For an accurate assessment, only one randomly selected image was used per case in the test group.

### Human-guided easy learning

Since the CXR images were collected from an NICU, the images were taken in various conditions, such as with or without electrocardiogram leads. These factors can cause the feature extraction model to learn nonanatomical features that are not medically correlated with sPDA but still affect the model's performance. Therefore, in order to focus on anatomical and medically explainable features, the model should be trained with images that prominently show PDA-related features.

However, selecting such images is contradictory to the purpose of this study, which is to find PDA-related features that have not yet been objectively defined. Instead, we chose an indirect approach to train the feature extraction model. We developed a new training method called 'human-guided easy learning'. A medical professional was first shown 173 CXR images, both positive and negative for sPDA diagnosis, in random order. The ratio of positive to negative images shown to the professional was kept at 1:1 to prevent bias, because humans cannot compensate for dataset imbalances. The professional was trained to diagnose sPDA by being informed of the actual PDA status after each image interpretation. This process was repeated three times. The adequately trained medical professional then diagnosed the presence or absence of sPDA in a total of 4654 images (2327 positive, 2327 negative). Within the images, a total of 3349 correctly diagnosed neonatal CXR images, comprising 1765 true positive and 1584 true negative diagnoses, were used for training the model. These images are considered to be more likely to show prominent PDA-related features, as they were diagnosed correctly by a professional who was trained to focus on PDA-related features and has sufficient medical knowledge to focus on anatomical features.

To evaluate the efficacy of human-guided easy learning, we compared the performance of the feature extraction model trained with human-guided easy learning to that of one trained with a rule-based filtered training dataset. We hypothesized that various gestational ages may show different radiographic characteristics, with the implication that images from a specific gestational age range, especially for very preterm infants who are more likely to manifest sPDA, would be more likely to show prominent PDA-related features. Therefore, the rule-based filtered training dataset was created by selecting cases with gestational ages between 27 and 32 weeks as well as corrected ages of less than 45 weeks at the time of the echocardiogram, for a total of 2028 images (1532 positive, 496 negative).

### Preprocessing

Unlike adult CXR images, neonatal CXR images are not standardized in terms of the area of interest. The images contain not only the chest but also the abdomen, which can introduce unwanted bias or overfitting in the feature extraction model. To address this issue, we developed a chest bounding box detection model to crop the chest area from the original images. We also developed heart and thorax segmentation models to provide the feature extraction model with heart and thorax segmentation masks to help focus on these areas. The chest bounding box detection model and heart and thorax segmentation models were trained with 1,100, 667, and 773 CXR images, respectively, which were randomly selected from the CXR dataset and labeled by a medical professional. The datasets were each divided into 8:1:1 for training, validation, and testing.

To predict the top, bottom, left, and right coordinates of the chest bounding box, we used a modified ConvNeXt<sup>13</sup> model. The model achieved an average intersection over union (IoU) of 0.82 on the test set. For the segmentation tasks, we employed HRNet<sup>14</sup> models. The heart segmentation model achieved an average IoU of 0.88 on the test set, and the thorax segmentation model achieved an average IoU of 0.94 on the test set.

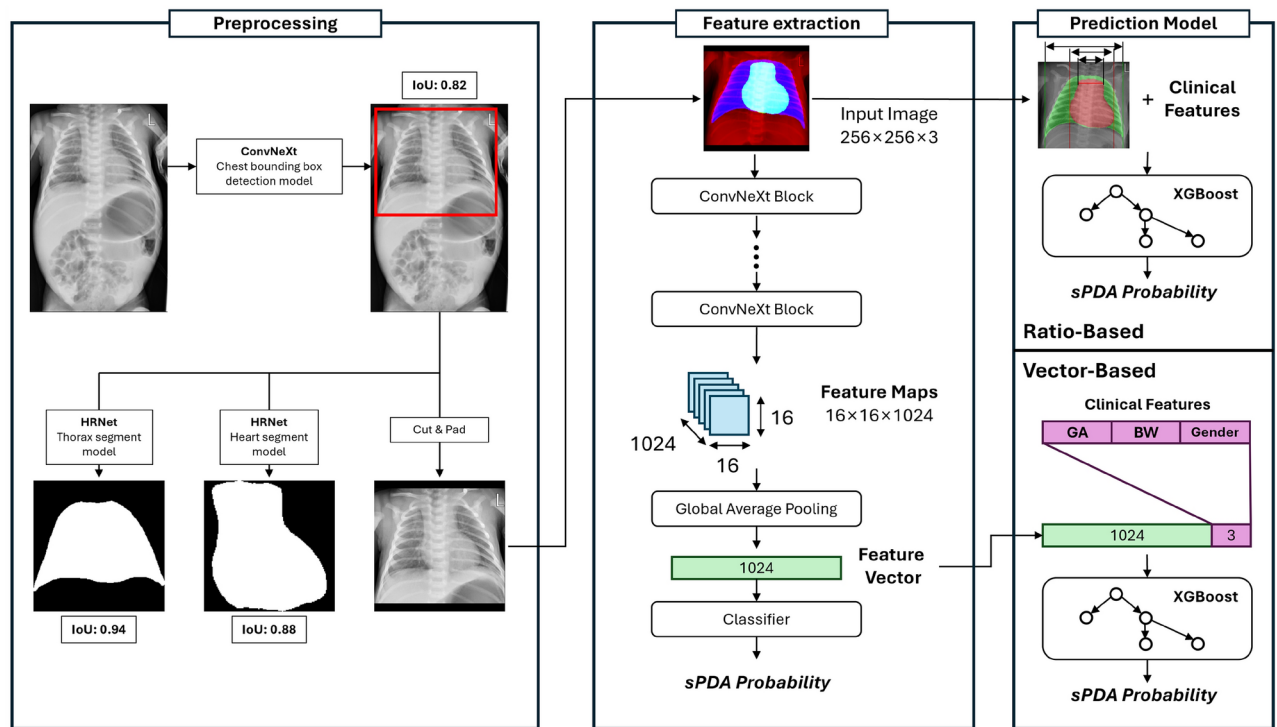
### Feature extraction model training

We also employed the ConvNeXt<sup>13</sup> architecture for the PDA feature extraction model. Unlike the original model that uses RGB channels, we utilized unsegmented images, heart-segmented masks, and thorax-segmented masks as 3-channel inputs. Each image was cropped outside the chest bounding box and resized to 256×256 pixels while maintaining the aspect ratio and padding with zeros where necessary. The model's performance was evaluated using 463 test cases, calculating sensitivity, specificity, and the area under the receiver operating characteristic curve (ROC-AUC).

### Feature evaluation

Following the training of the feature extraction model, we utilized the 1024-dimensional vector preceding the linear classifier layer as the representation of the CXR image features (Fig. 2). These features were subsequently integrated with clinical data, including gestational age, birth weight, and gender, to train the sPDA prediction model. For this task, we employed the XGBoost algorithm<sup>15</sup>, which is a decision tree-based ensemble machine learning algorithm.

In contrast to the feature extraction model training, the entire training dataset was utilized for the sPDA prediction model, as the human-guided learning process was unnecessary. Evaluation of the model was conducted using the isolated test dataset, comprising 463 cases. The model's performance was measured in



**Fig. 2.** Feature extraction and classification pipeline. The global average pooling layer averages the feature maps of the last convolutional layer to generate a feature vector. For example, the 1st element of the feature vector is the average of the 1st feature map.

terms of sensitivity, specificity, and ROC-AUC. A comparison was made between models with and without the incorporation of CXR features to assess the significance and benefits of these features in predicting sPDA.

### Feature analysis

#### Shapley additive explanations (SHAP)

In order to compare the importance of the extracted features and clinical data, we used a method called Shapley Additive Explanations (SHAP)<sup>16,17</sup>. SHAP is a game theory-based interpretation method that explains the output of machine learning models by calculating the contribution of each feature to the prediction. We used absolute SHAP values to quantitatively analyze and find the features crucial for sPDA prediction.

Furthermore, we chose several CXR features that showed high absolute SHAP values and visualized their corresponding channels of the output feature map of the final convolutional layer. With this visualization, we tried to understand what each feature represents.

#### GradCAM++

To develop an integrated view of the extracted features, we employed GradCAM++, which is a visualization technique that highlights the regions of the input image that are important for the model's prediction<sup>18</sup>. Based on the GradCAM++ results and previous feature map visualization, we hypothesized potential radiographic features associated with sPDA. To evaluate the significance of the hypothesized features, we performed a z-test between positive and negative cases across the entire dataset and calculated the corresponding p-values.

Additionally, we trained another XGBoost classifier utilizing the hypothesized features in conjunction with clinical data (the ratio-based model in Fig. 2), comparing its performance to that of a model employing the feature vector by the feature extraction model. Evaluation of this model was also conducted on 463 test cases, with sensitivity, specificity, and ROC-AUC serving as the performance metrics.

## Results

### Model performance

Our PDA feature extraction model achieved a specificity of 0.91, sensitivity of 0.33, and an ROC-AUC of 0.72 on the test set (Table 1). In contrast, when trained with rule-based filtered data, the model achieved a specificity of 0.99, sensitivity of 0.02, and an ROC-AUC of 0.40. The AUC value below 0.5 indicates that the model's performance was worse than random guessing, which suggests that the model trained with control group data failed to learn the features of PDA.

We also investigated the efficacy of providing heart and thorax masks along with raw images. For this we trained the same model with raw images only, heart-segmented images only, and thorax-segmented images only. In all three cases, images were cropped to include the thoracic region. The raw image model and thorax-

	Sensitivity	Specificity	ROC-AUC
CNX (Combined, Human-guided)	0.33	0.91	0.72
CNX (Combined, Rule-based)	0.02	0.99	0.40
CNX (Raw, Human-guided)	0.38	0.77	0.62
CNX (Heart, Human-guided)	0.22	0.90	0.63
CNX (Thorax, Human-guided)	0.49	0.70	0.61
XGB (Vector + Clinical)	0.33	0.95	0.75
XGB (Clinical)	0.29	0.95	0.67
XGB (Vector)	0.33	0.94	0.71
XGB (Ratio + Clinical)	0.42	0.94	0.74
XGB (CTR + Clinical)	0.40	0.93	0.73

**Table 1.** Feature extraction model & sPDA prediction model performance. Sensitivity, specificity, and ROC-AUC of the ConvNeXt-based feature extraction model and the XGBoost-based sPDA prediction model. Combined, Raw, Heart, and Thorax refer to the input images used for training the feature extraction model (see Fig. 2). For the XGB model, Vector refers to the 1024 size feature vectors extracted from the feature extraction model, Clinical refers to the clinical features of gestational age, birth weight, and gender, and Ratio refers to the three ratio-based features that showed statistically significant differences between sPDA-positive and sPDA-negative groups (see Fig. 5). CTR, Cardiothoracic ratio.

segmented image model achieved a sensitivity of 0.38 and 0.49, respectively, which were higher than the default combined model. However, the specificity of the raw image model and thorax-segmented image model was 0.77 and 0.70, respectively, which were significantly lower than the default combined model. In terms of ROC-AUC, all three models performed worse than the default combined model. These experiments show that both segmentation masks and human-guided easy learning data are essential for the feature extraction model to effectively learn the features that differentiate sPDA from non-sPDA cases.

In the case of the sPDA prediction model utilizing the extracted 1024 size feature vector of CXR images from the feature extraction model, it achieved a sensitivity of 0.33, specificity of 0.95, and an ROC-AUC of 0.75 on the test set, which is slightly improved from the feature extraction model. These scores are also better than those from the models trained with only clinical features and only feature vectors, implying that both clinical features and feature vectors contribute to sPDA prediction.

## Feature importance analysis

### SHAP analysis

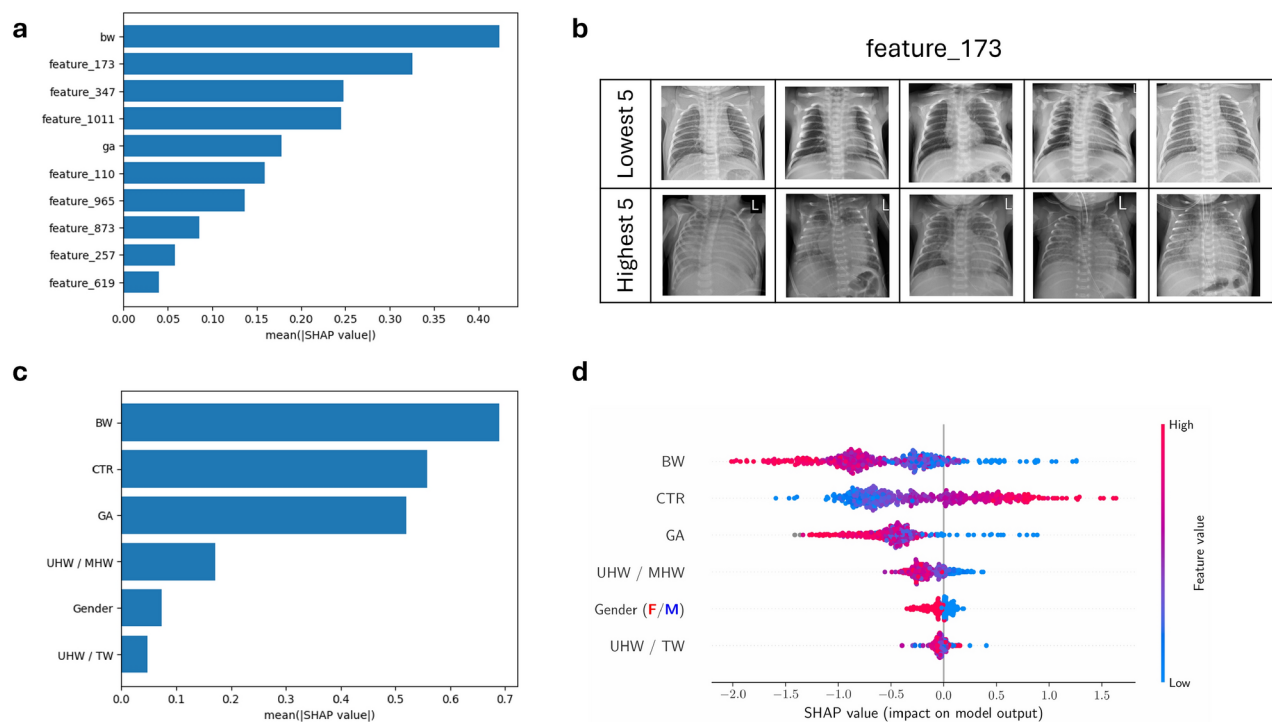
We ranked the importance of the features using the mean absolute SHAP values of the XGBoost model calculated on the test set and compared the top 10 features (Fig. 3a). Birth weight was the most important feature, followed by three extracted features, namely feature\_173, feature\_347, and feature\_1011, followed by gestational age. As an example of an extracted feature with high SHAP values, comparing the images with the highest and lowest feature\_173 values reveals distinct characteristics: images with low feature\_173 values exhibit clear lung fields and normal hearts, while those with high feature\_173 values display more opaque lung fields and larger hearts (Fig. 3b). To understand which regions of the images are related to the features, we retrieved the feature maps of several test images from the last convolutional layer before the global average pooling layer. The channels corresponding to the three features that showed the highest SHAP values are visualized in Fig. 4. The visualization exhibits a contrast between the lung region and the heart region in the feature maps, which implies that the model is focusing on these regions to predict sPDA. However, the visualization is not clear enough to interpret the exact features that the model is focusing on.

### GradCAM++ analysis

We employed GradCAM++ to visualize the regions of the CXR images that the model emphasized when making predictions (Fig. 4). In the GradCAM++ visualizations, the regions around the upper thorax, lower left heart, and lower right lung are prominently highlighted. Notably, in some cases, the highlighted regions extended to include the aortic arch and, in some instances, the left atrium. These findings indicate that the model predominantly concentrated on heart and lung areas during its predictive process.

Based on the highlighted regions identified by GradCAM++ and the feature maps, we hypothesized that the model was focusing on specific radiographic features: the ratio of the upper heart width to the maximum heart width, the cardiothoracic ratio, the ratio of the upper heart width to the thorax width, and the mean normalized brightness of the lung fields. For clarity, we defined the upper heart width as the width of the heart measured at the upper 10% of its total height. The mean normalized brightness of the lung fields was calculated as the average brightness within the thorax area, excluding the heart, with brightness values normalized to a range from 0 to 1 (Fig. 5a).

Among the above features, statistically significant differences were found in three features between sPDA-positive and sPDA-negative groups (Fig. 5b). The mean ratio of the upper heart width to the maximum heart width was 0.45 (95% confidence interval (CI), 0.44–0.47) in sPDA-positive cases, while 0.54 (95% CI, 0.53–0.54) in sPDA-negative cases. The cardiothoracic ratio was 0.57 (95% CI, 0.57–0.58) in sPDA-positive cases, while



**Figure 3.** Results of the analysis of the sPDA prediction model. **a.** Top 10 mean absolute SHAP values of the features, calculated on the test dataset. **b.** Examples of images having the highest and lowest feature\_173 values (i.e., 174<sup>th</sup> element of the feature vector). The images with low feature\_173 values exhibit clear lung fields and normal hearts, while the images with high feature\_173 values exhibit more opaque lung fields and larger hearts. **c.** Mean absolute SHAP values of the ratio-based features on the test dataset. **d.** Beeswarm plot of SHAP values of the features on the test dataset. Negative SHAP values indicate that the feature value impacts the model prediction negatively, while positive SHAP values indicate that the feature value impacts the model prediction positively. The color of the points indicates the feature value. In the case of the gender feature, red indicates female and blue indicates male.

0.55 (95% CI, 0.55–0.55) in sPDA-negative cases. Additionally, the ratio of the upper heart width to the thorax width was 0.27 (95% CI, 0.26–0.28) for sPDA-positive cases, while 0.30 (95% CI, 0.30–0.31) for sPDA-negative cases. The z-test p-values were all less than 0.001. Conversely, the mean normalized brightness of the lung fields showed no significant difference between the two groups, with values of 0.51 (95% CI, 0.50–0.52) for sPDA-positive and 0.51 (95% CI, 0.51–0.52) for sPDA-negative cases.

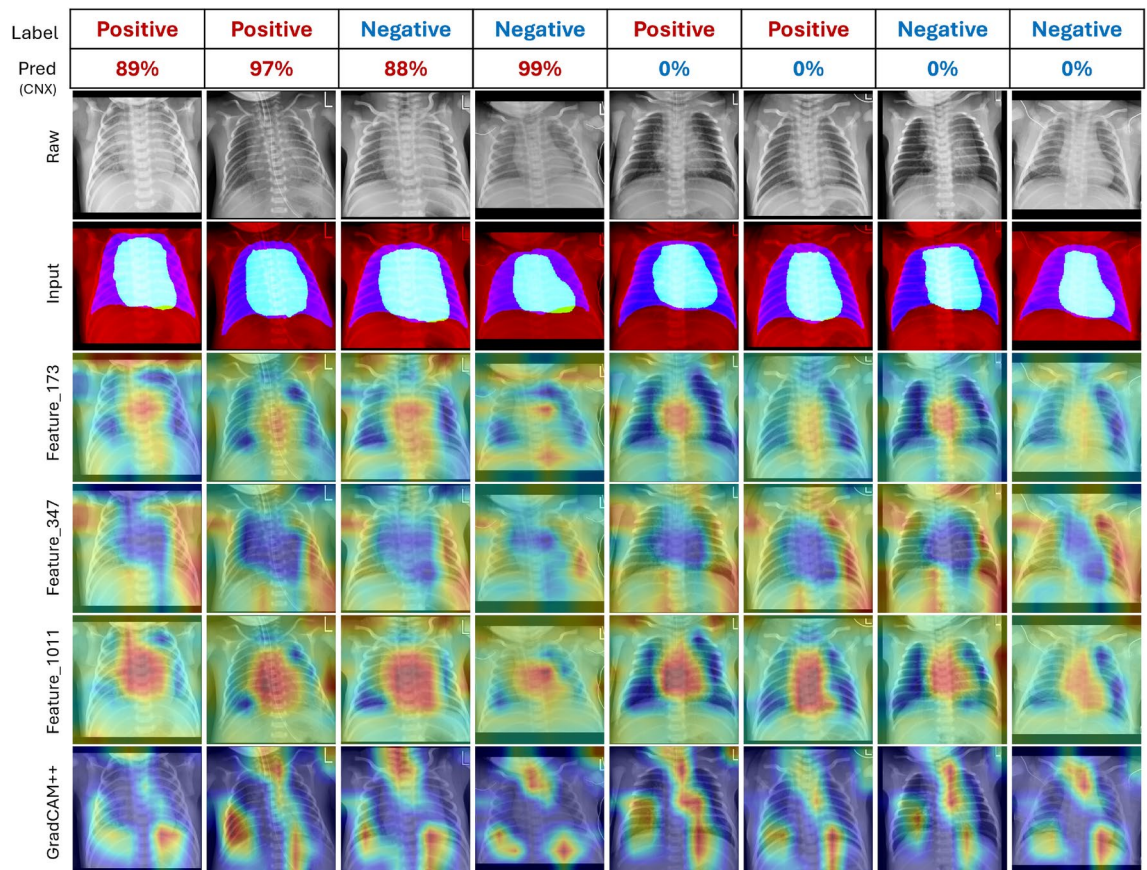
#### Ratio-based sPDA prediction model performance

To evaluate the diagnostic value of the hypothesized features, we trained another sPDA prediction model with the three features that showed statistically significant differences between sPDA-positive and sPDA-negative groups along with the three clinical features. The model achieved a sensitivity of 0.42, specificity of 0.94, and an ROC-AUC of 0.74 on the test set. This result indicates that the hypothesized features are almost as effective as the feature vectors extracted from the feature extraction model in predicting sPDA.

Additional SHAP analysis revealed that the cardiothoracic ratio contributed the most to the model's prediction among the hypothesized features (Fig. 3c, d). This is also consistent with the result of training an sPDA prediction model with only the cardiothoracic ratio and clinical features, which achieved a sensitivity of 0.40, specificity of 0.93, and an ROC-AUC of 0.73 on the test set. A detailed visualization of how each feature contributed to the model's predictions for individual test images is provided in Fig. 6.

## Discussion

The analysis and visualization of the deep learning model for feature extraction that predicts sPDA from CXR images revealed radiographic features relevant to sPDA. The features that were statistically significant between sPDA-positive and -negative groups were the cardiothoracic ratio, upper heart width to thoracic width ratio, and upper heart width to maximum heart width ratio. An XGBoost-based model trained with these three features and clinical data achieved an AUC of 0.74 in predicting sPDA, which is close to the AUC of 0.75 achieved by the model trained with 1024-dimensional features extracted from the deep learning model and clinical data. This suggests that the three hypothesized ratio-based radiographic features play a crucial role in predicting sPDA. Furthermore, the histograms of the three features showed a significant overlap between the sPDA-positive and -negative groups, indicating that complex relationships among these features are more important than each feature alone for predicting sPDA. Therefore, instead of setting diagnostic thresholds for each feature, we suggest



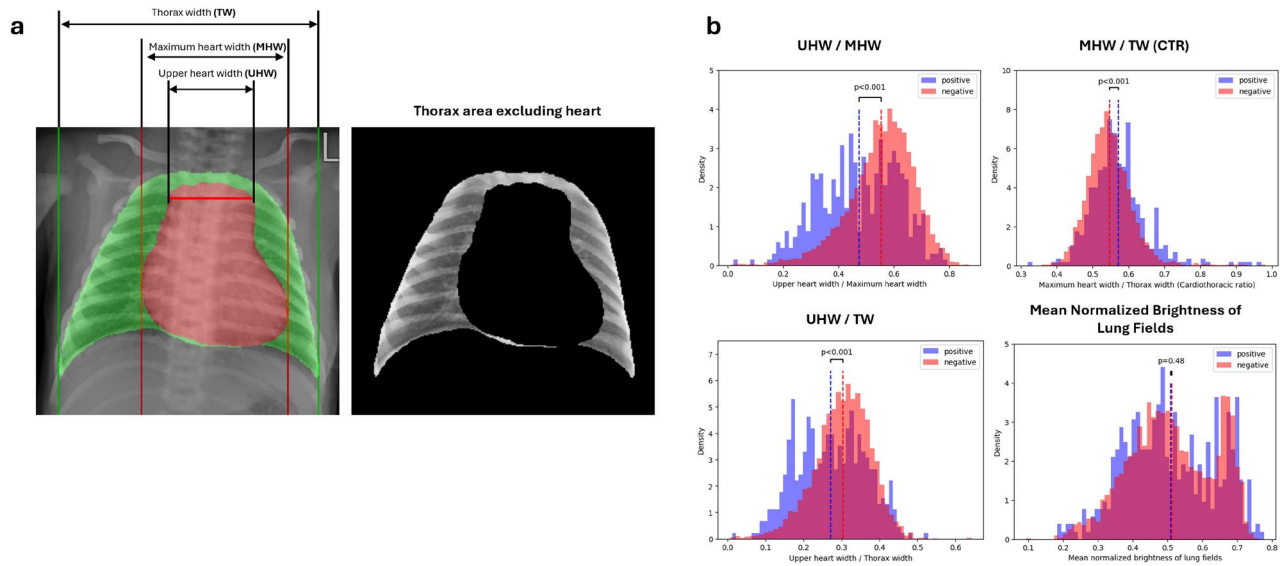
**Fig. 4.** Examples of feature maps of the last convolutional layer before the global average pooling layer and GradCAM++ visualizations of the test images. The visualized feature maps correspond to the three feature values that showed the highest SHAP values (feature\_173, feature\_347, and feature\_1011). Label refers to the sPDA label of the test image, and Pred refers to the predicted probability of sPDA, calculated by the CNX-based feature extraction model.

that clinicians measure the features from CXR images and input them along with clinical data into the prediction model to calculate the probability of sPDA. As shown in Fig. 6, our prediction model can provide SHAP values to explain the prediction results, which can help clinicians understand the decision-making processes of the model and aid in their clinical judgments.

Among the three radiographic features, the cardiothoracic ratio was the most important in predicting sPDA. More specifically, the cardiothoracic ratio was positively correlated with sPDA, which is consistent with clinical knowledge that PDA patients may demonstrate cardiomegaly due to left atrial and left ventricular enlargement<sup>19,20</sup>. Meanwhile, unlike the cardiothoracic ratio, both the upper heart width to thoracic width ratio and the upper heart width to maximum heart width ratio were negatively correlated with sPDA. Enlargement of the left atrium and left ventricle in PDA patients may cause the lower portion of the heart to widen, leading to a decrease in the upper heart width to maximum heart width ratio. However, since the upper heart width to thoracic width ratio was also negatively correlated with sPDA, the upper portion of the heart may be relatively narrower in CXR images of sPDA-positive compared to sPDA-negative patients. This may be due to organs in the mediastinum area that are overlapped with the upper portion of the heart, especially the thymus that is commonly found in CXR images of infants<sup>21</sup>. Since our heart segmentation model cannot differentiate between the heart and other organs in the mediastinum area, other organs may have been included in the upper heart width measurement, leading to an increase in the upper heart width to thoracic width ratio in the CXR images of sPDA-negative patients.

It is also worth noting that normalized lung brightness was not significantly different between sPDA-positive and -negative groups. This is contrary to the expectation that PDA patients may have increased lung brightness due to pulmonary edema caused by severe volume overload<sup>20</sup>. The lack of significance found here may be due to the presence of other lung-related diseases that affect lung brightness, such as respiratory distress syndrome or bronchopulmonary dysplasia, which are relatively common in NICU preterm infants<sup>21</sup>.

Although the features extracted from the deep learning model achieved a slightly higher AUC than the three ratio-based radiographic features, the latter can be measured by radiologists without using deep learning models. This is advantageous because human-measurable features are immune to cross-domain generalization issues that may arise from using deep learning models. For example, our feature extraction model was trained



**Fig. 5.** Definition and analysis of the hypothesized radiographic features. **(a)** CXR images highlighting the proposed feature definitions. The upper heart width is defined as the width of the heart measured at the upper 10% of its total height. The widths are measured based on the heart and thorax masks generated by the segmentation models. **(b)** Histograms of the features in sPDA-positive and sPDA-negative groups. For accurate comparison, only one image was randomly selected from each patient. The p-values were calculated from z-tests.

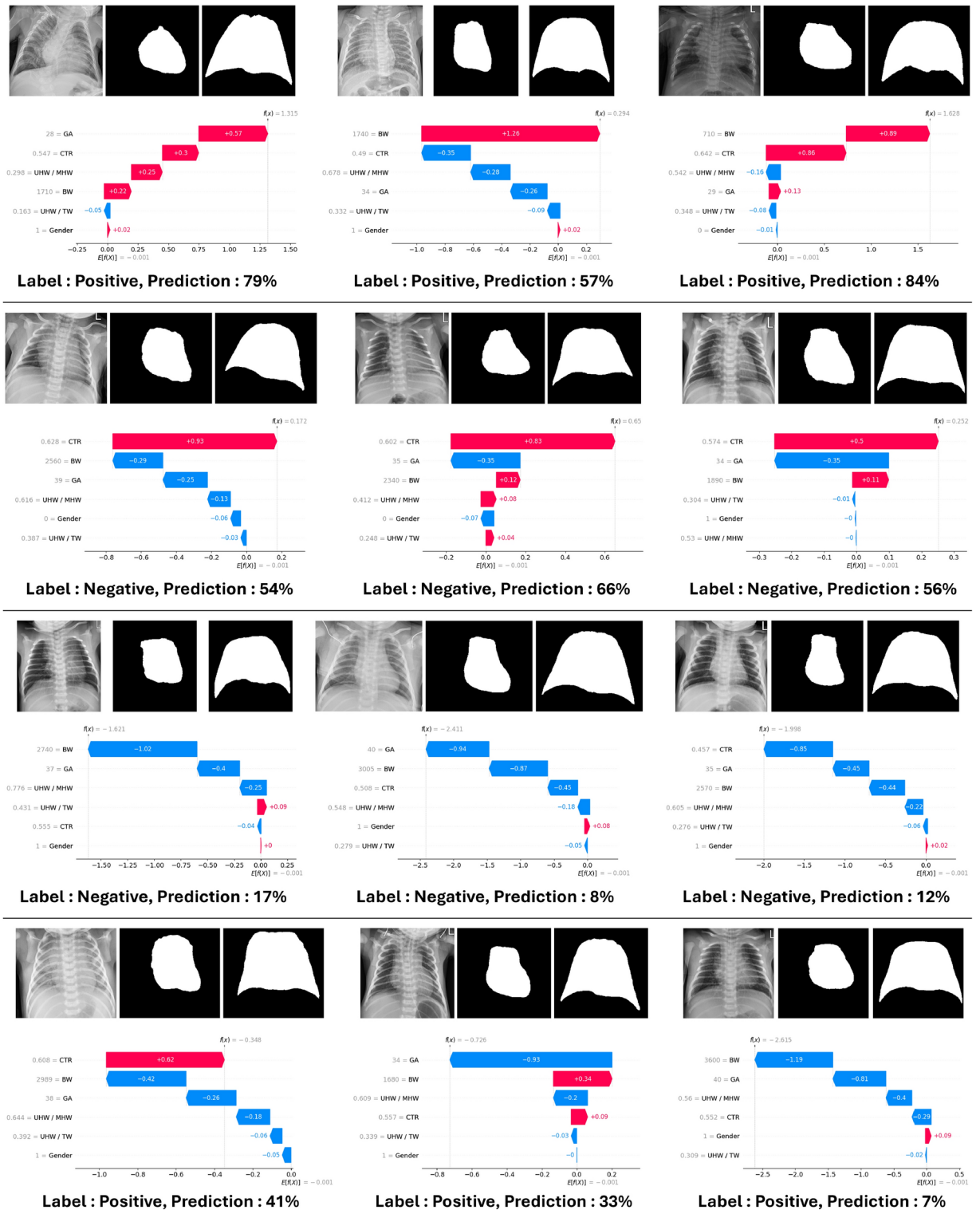
with CXR images from a single institution, which may not guarantee its feature extraction performance on images from other institutions. In contrast, the ratio-based features can be measured consistently across different institutions, making them more generalizable. Furthermore, the three radiographic features are interpretable and can provide insights into the pathophysiology of PDA.

One of the limitations of our study is that every model that predicted sPDA showed relatively low sensitivity and high specificity, which is not ideal for screening purposes. This result is in line with known challenges in diagnosing PDA using CXR images, which may show nonspecific findings depending on the severity of the disease<sup>19,20</sup>. Furthermore, while the prediction model can rule in sPDA with high specificity, positive prediction does not rule out the possibility of other congenital heart diseases that may occur with PDA, such as tetralogy of Fallot or ventricular septal defect. Therefore, positive predictions should be followed by echocardiography for accurate diagnosis.

Due to the limitations of our data, we were not able to utilize various other clinical data that may be relevant to sPDA, such as vital signs, prenatal history, or laboratory results. A recent study showed that employing various clinical factors can help diagnose sPDA with relatively high sensitivity<sup>22</sup>. Therefore, future studies should consider incorporating various clinical data to improve the sensitivity of the prediction model. We also suggest that future studies focus on finding other rule-in factors of sPDA instead of searching for a single rule-out factor, as the latter may be difficult to find since a large portion of the sPDA-positive group may not show any symptoms or signs of sPDA. We expect that with enough rule-in factors, the sensitivity of the prediction model can be improved to a level that it can be used in practice for screening purposes.

Despite the above limitations, based on the results showing that the radiographic features can improve diagnostic performance compared to using clinical data alone, our prediction model can be used to identify specific sPDA patients who were not detected by clinical data. We believe that the prediction model can be useful in rule-in scenarios, such as when prioritizing patients when echocardiogram resources are limited. The ratio-based radiographic features require only basic radiographic knowledge and can be easily measured by doctors, making the prediction model more accessible in such scenarios. Moreover, this study demonstrated that specific radiographic features for sPDA exist in CXR images, which shows the potential of using CXR images not only for sPDA diagnosis but also for postoperative follow-up. Considering that postoperative follow-up for sPDA patients requires frequent echocardiograms, substituting some of them with CXR-based follow-up can help reduce patient costs.

Finally, our study demonstrated a method to analyze and visualize deep learning models to understand the radiographic features relevant to a specific disease. This approach can be applied to other diseases lacking quantitatively defined radiographic features but known to exhibit characteristic findings. By training a deep learning model to predict the disease and then analyzing the model, the radiographic features relevant to the disease can be objectively identified, evaluated, and further investigated for clinical significance.



**Fig. 6.** Waterfall plots of the ratio-based sPDA prediction model. The plots visualize the impact of each feature on the model's prediction for each test image. The bottom of the plot represents the expected value of the model prediction, and each row represents how the features contribute to the prediction for each test image. Red indicates positive and blue indicates negative contribution to the output of the model. The x-axis represents the log-odds, so negative values indicate a probability of sPDA less than 0.5, and positive values indicate a probability of sPDA greater than 0.5. Abbreviations: CTR, cardiothoracic ratio; Gender: 0 is female, 1 is male; BW, Birth weight (g); GA, Gestational age (weeks); UHW, Upper heart width; MHW, Maximum heart width; TW, Thorax width.

## Data availability

Additional anonymized data are available from the corresponding authors upon reasonable request.

Received: 3 September 2024; Accepted: 8 November 2024

Published online: 26 November 2024

## References

- Soliman, R. M., Mostafa, F. A., Abdelmassih, A., Sultan, E. & Mosallam, D. Patent ductus arteriosus in preterm infants; experience of a tertiary referral neonatal intensive care unit: prevalence, complications, and management. *Egypt. Pediatr. Assoc. Gazette* **68**, 1–9. <https://doi.org/10.1186/s43054-020-00046-8> (2020).
- Tavera, M. C. et al. Role of echocardiography on the evaluation of patent ductus arteriosus in newborns. *J. Maternal-Fetal Neonatal Med.* **22**, 10–13. <https://doi.org/10.1080/14767050903198181> (2009).
- Adegboro, C. O., Choudhury, A., Asan, O. & Kelly, M. M. Artificial intelligence to improve health outcomes in the nicu and picu: a systematic review. *Hosp. Pediatr.* **12**, 93–110. <https://doi.org/10.1542/hpeds.2021-006094> (2022).
- Beam, K., Sharma, P., Levy, P. & Beam, A. L. Artificial intelligence in the neonatal intensive care unit: the time is now. *J. Perinatol.* **44**, 131–135. <https://doi.org/10.1038/s41372-023-01719-z> (2024).
- Chioma, R. et al. Applications of artificial intelligence in neonatology. *Appl. Sci.* **13**, 3211. <https://doi.org/10.3390/app13053211> (2023).
- Xing, W. et al. Early severity prediction of bpd for premature infants from chest x-ray images using deep learning: A study at the 28th day of oxygen inhalation. *Comput. Methods Programs Biomed.* **221**, 106869. <https://doi.org/10.1016/j.cmpb.2022.106869> (2022).
- Patel, M., Sandhu, J. & Chou, F.-S. Developing a machine learning-based tool to extend the usability of the nichd bpd outcome estimator to the asian population. *Plos one* **17**, e0272709. <https://doi.org/10.1371/journal.pone.0272709> (2022).
- Yildirim, A. E. & Canayaz, M. A novel deep learning-based approach for prediction of neonatal respiratory disorders from chest x-ray images. *Biocybern. Biomed. Eng.* **43**, 635–655. <https://doi.org/10.1016/j.bbe.2023.08.004> (2023).
- Maeda, R. et al. Predicting the severity of neonatal chronic lung disease from chest x-ray images using deep learning. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* 1543–1547 (IEEE, 2022). <https://doi.org/10.1109/SMC53654.2022.9945486>.
- Chakkarapani, A. A. & Gupta, S. Preterm patent ductus arteriosus: what the research tells us. *Paediatr. Child Health* **29**, 257–262. <https://doi.org/10.1016/j.paed.2019.03.002> (2019).
- Weerakkody, Y. Patent ductus arteriosus | Radiology Reference Article | Radiopaedia.org. <https://radiopaedia.org/articles/patent-ductus-arteriosus>. <https://doi.org/10.5334/rid-7355> (2022).
- Dice, J. E. & Bhatia, J. Patent ductus arteriosus: an overview. *J. Pediatr. Pharmacol. Therapeut. JPPT* **12**, 138–146. <https://doi.org/10.5863/1551-6776-12.3.138> (2007).
- Liu, Z. et al. A ConvNet for the 2020s. <https://doi.org/10.48550/arXiv.2201.03545> (2022).
- Sun, K. et al. High-Resolution Representations for Labeling Pixels and Regions. [arXiv:1904.04514](https://arxiv.org/abs/1904.04514) [cs] (2019).
- Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (2016). <https://doi.org/10.1145/2939672.2939785>.
- Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, vol. 30 (Curran Associates, Inc., 2017).
- Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67. <https://doi.org/10.1038/s42256-019-0138-9> (2020).
- Chattopadhyay, A., Sarkar, A., Howlader, P. & Balasubramanian, V. N. Grad-CAM++: improved visual explanations for deep convolutional networks. <https://doi.org/10.1109/WACV.2018.00097> (2018).
- Schneider, D. J. & Moore, J. W. Patent ductus arteriosus. *Circulation* **114**, 1873–1882. <https://doi.org/10.1161/CIRCULATIONAH.105.592063> (2006).
- Webb, W. R. & Higgins, C. B. *Thoracic Imaging: Pulmonary and Cardiovascular Radiology* (Lippincott Williams & Wilkins, 2011).
- Lee, E. Y. *Pediatric Thoracic Imaging* (Wolters Kluwer Health, 2018).
- Na, J. Y. et al. Artificial intelligence model comparison for risk factor analysis of patent ductus arteriosus in nationwide very low birth weight infants cohort. *Sci. Rep.* **11**, 22353. <https://doi.org/10.1038/s41598-021-01640-5> (2021).

## Acknowledgements

This paper was supported in part by the Fund of Biomedical Research Institute, Jeonbuk National University Hospital, in part by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2023-00236157), in part by an Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. RS-2020-II201336, Artificial Intelligence Graduate School Program (UNIST)), and in part by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. RS-2024-00344958).

## Author contributions

Conceptualization, Phillip Chang, Jimin Lee and Hyun Ho Kim; Data curation, Hyun Ho Kim; Funding acquisition, Jimin Lee and Hyun Ho Kim; Investigation, Phillip Chang and Hyeon Sung Choi; Methodology, Phillip Chang and Jimin Lee; Project administration, Jimin Lee and Hyun Ho Kim; Software, Phillip Chang; Supervision, Jimin Lee and Hyun Ho Kim; Validation, Phillip Chang; Visualization, Phillip Chang and Hyeon Sung Choi; Writing-original draft, Phillip Chang, Hyeon Sung Choi, Jimin Lee and Hyun Ho Kim; Writing-review & editing, Phillip Chang, Jimin Lee and Hyun Ho Kim.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to J.L. or H.H.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024