

Developing a Novel Deep Learning Model for Pediatric Chest Radiograph Analysis

Heeyoon Choi¹ and Gregory Rose[#]

¹Shanghai American School Puxi, China

[#]Advisor

ABSTRACT

Pediatric thoracic diseases pose significant challenges to healthcare systems worldwide, particularly in regions with limited access to specialized medical resources. Lower respiratory infections, in particular, contribute significantly to childhood morbidity and mortality. This research aims to address the diagnostic challenges associated with pediatric thoracic diseases by harnessing the power of artificial intelligence. The study utilizes the PediCXR dataset, composed of pediatric chest radiographs, and employs a Faster R-CNN model pretrained on ImageNet data for lesion detection and classification. The model is trained to identify abnormal lesions indicative of various thoracic diseases, with a focus on achieving high accuracy and minimizing loss. Experimental results demonstrate the model's capability to achieve an accuracy exceeding 90% and a loss under 0.1, meeting predefined objectives.

Introduction

Pediatric thoracic diseases remain a large concern worldwide, with lower respiratory infections standing as the fifth leading cause of death for children under five (GBD 2015 LRI Collaborators 2023). The impact of these diseases is felt strongly in developing countries with poor health infrastructure. In these countries, lack of information access for healthcare workers often leads to the use of unreliable sources, such as personal observation and colleague advice, when handling diagnoses or treatments (Pakenham-Walsh and Bukachi 2024). Moreover, the gross lack of healthcare workers impedes delivering quality care for each patient. The World Health Organization (WHO) projects that by 2030, there will be a worldwide shortage of 10 million health workers; this shortage will have the greatest impact on low- and lower-middle income countries (Health Workforce 2024). As developed countries face personnel shortages, they turn to recruiting workers from lower income countries, leaving developing nations with chronic understaffing (United Nations 2024). These issues can result in decreased quality of diagnoses: according to the WHO, healthcare workers in seven low- and middle-income African countries only made accurate diagnoses one-third to three-quarters of the time. Particularly as acute pediatric respiratory infections increase with natural disasters, the efficiency and accuracy of diagnoses by healthcare workers can be even more negatively affected (Leggiadro 2024). Under such circumstances, diseases that could have been easily treated in developed regions may be unrecognized, misdiagnosed, or diagnosed only after a long delay. Considering these problems, this study aims to harness the use of artificial intelligence, specifically deep learning detection and classification, to create a deep learning model that can improve the efficiency and accuracy of pediatric thoracic disease diagnosis in areas with underdeveloped health infrastructure through chest radiograph analysis.

Related Work

Previous studies on the applications of artificial intelligence on pediatric chest radiographs heavily focus on pneumonia. Over 65% of all studies published on the use of AI on pediatric chest radiographs between 1990 and 2021 concentrate on pneumonia detection and classification (Padash et al 2024). Barakat et al. focused on increasing the accuracy of pediatric pneumonia detection through machine learning. Ravi et al. developed a more cost-effective model for detection of pneumonia, additionally improving generalizability and decreasing sensitivity to data imbalances. Tang et al. used deep convolutional neural networks (CNN) to detect pneumonia from pediatric chest radiographs. Other studies such as E et al. have focused on differentiating the etiology of pneumonia, either bacterial or viral. Most studies have also focused on the development of binary classification models; however, multiple classification regarding pediatric pneumonia has also been demonstrated. Rahman et al. used CNN models to classify pneumonia etiology, using a total of three classes of normal, bacterial, and viral. Chen et al. classified pneumonia radiographs into three classes of different pneumonia end points. Multiple classification of pediatric thoracic diseases other than pneumonia have also been conducted, but have not yet achieved high performance. Chen et al. used a five-fold image classification based on CNN to classify pediatric radiographs into five categories; however, the highest accuracy among the five classes is 85.71%, significantly lower than the 92.47% accuracy of the model when detecting abnormal radiographs using a binary algorithm.

As seen with Rahman et al., Chen et al., and Chen et al., CNN models are often utilized for image classification. CNN models have successfully been employed for multiple classification of adult radiographs. Türk and Kökver used a three-channel fusion CNN model to classify adult chest radiographs into five classes, of which they obtained an accuracy value of 91.71%.

Overall, the AI diagnosis of pediatric thoracic diseases excluding pneumonia has largely remained unexplored. While pneumonia is common, the improvement of diagnosis of other diseases is also needed. In situations lacking healthcare personnel, the accuracy and efficiency of diagnoses could be profoundly improved by a deep learning model with wider coverage than just pneumonia. Healthcare workers could also benefit from the model identifying the specific location of lesions. This study aims to develop an accurate and reliable multiple classification and object detection model for a wider range of pediatric thoracic lesions, not constricted to pneumonia. Additionally, this study aims to classify such lesions with an accuracy greater than 90% and total loss under 0.1.

Methodology

Dataset

This study used the dataset PediCXR (Pham et al. 2022). This dataset is an open dataset of pediatric chest radiographs which contains a training set of 7728 images and a test set of 1397 images. The dataset provided annotations for local labels, abnormal findings, and global labels, diagnoses. The PediCXR dataset has a total of 36 local labels; however, only five classes had more than 100 images, and one class had over 1000 images, indicating significant class imbalances. The dataset has 15 global labels, of which the normal category included over 60% of images, again showing class imbalances. This study will focus on using the local labels to locate abnormal lesions on the radiographs. The annotations for the local labels provided bounding box coordinates, which will be used in this study.

Procedure

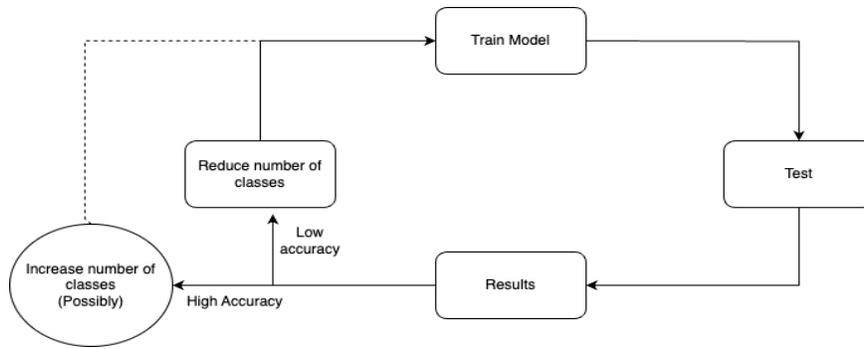


Figure 1. An illustration to show the overall workflow of the study.

The purpose of the study is to develop a model that can locate and detect a wide range of lesions; hence, the study is set up to conduct experimentation to determine the greatest number of classes that could possibly be used without compromising a 90% accuracy. The number of classes will be continued to increase until the accuracy is compromised.

Based on the classes that were used, the PediCXR dataset was split to only include data regarding the specific classes. This dataset then underwent data augmentation to offset the effects of data imbalance and scarcity. This new dataset was further split into a train dataset, a validation dataset, and a test dataset. The test dataset was the test dataset originally provided by PediCXR. The originally provided train dataset was split into a train dataset and a validation dataset, with a ratio of 8:2. For one epoch, the model would first train itself on the train dataset, and then test itself on the validation dataset. If the total loss decreased, then the model would update itself to the newly trained version and then start the next epoch. If total loss increased, then the newly trained version would be abandoned, and the same model would start the next epoch. The model would finally be tested on the test dataset. The individual images in the dataset first were windowed, then some of these images were augmented with a random probability and added to the dataset.

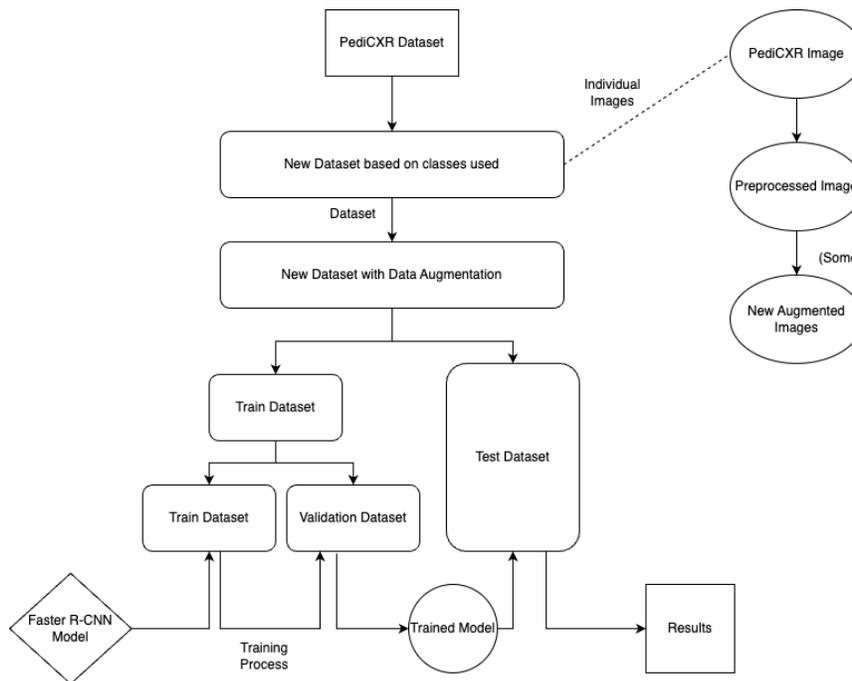


Figure 2. An illustration to show the training process.

Model

This study uses a Faster R-CNN model imported from detectron2. This model was pretrained on ImageNet data. The first major component of the Faster R-CNN is the Region Proposal Network (RPN). This network takes an image input and determines areas of the image to be “proposals.” The RPN uses anchor boxes of varying sizes and ratios to locate the proposal regions. These boxes are distributed equally across the image. The proposed regions go through two processes, classification, and regression. Classification involves determining if the region contains the proposed object. Regression refines the size and location of the boxes to better locate the proposed object. The region proposed by the RPN then undergo Region of Interest (RoI) Pooling. In this stage, each proposed region is fixed to a controlled size, preparing for subsequent further classification and regression. Each region proposed by RPN is a different size, but RoI Pooling standardizes the sizes to the standard size of the network. After RoI Pooling, each proposed region is sent to a network for two processes, classification, and regression, similar to that during the stage of the RPN network. Classification identifies the object of interest in the proposed region. Regression refines the size and location of the boxes, and this is an important step determining the accuracy of bounding boxes. The final output of the model is a combination of the outputs of the classification and regression networks. The output contains both the object’s category and location, along with the confidence of the model.

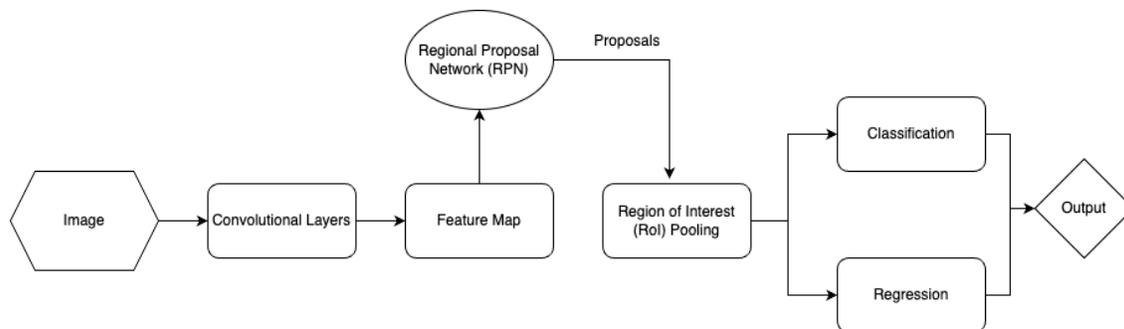


Figure 3. An illustration to show the mechanism of the Faster R-CNN model.

Windowing

Images in the dataset were preprocessed through windowing. The purpose of windowing is to increase contrast and highlight the manifestations of lesions. Images were first converted to a numpy array. These values were then rescaled into Hounsfield units through the pydicom tags rescale intercept and rescale slope. The equation $U=m*SV+b$ was used, where U represents the output, SV represents the stored value, m represents the rescale slope, and b represents the rescale intercept. Then, a window width and window level were set. The window width is the range of brightness displayed in the image, and it is inversely correlated with contrast. Window level is the midpoint of the range of brightness. Such measures were taken to standardize the images in the dataset and also to highlight the manifestations of lesions. Especially with medical imaging such as chest radiographs, subtle differences could play a large role in diagnosis, so appropriate windowing measures can enhance the performance of the model.

Data Augmentation

Data scarcity is a major challenge for deep learning applications using pediatric data. Data augmentation is a method of overcoming data shortages by modifying existing images and adding them to the dataset as new images. Augmentation modifications such as geometric transformations have been found to increase detection accuracy of chest radiographs (Elgendi et al. 2021). The majority of classes in the PediCXR do not reach 100 images. The class with the most images is peribronchovascular interstitial opacity (PIO), but there still is only 1358 train images. Considering such data shortage, data augmentation was adopted. Images were modified in various ways, including random contrast, random brightness, random gamma value, horizontal flip, random rotation, random gaussian noise, and random gaussian blur. Newly created images were added back to the dataset.

Calculation of Loss

The loss function of Faster R-CNN indicates the performance of the model. In the training process, the model automatically adjusts its parameters to reduce the loss. Faster R-CNN generates loss during two stages, the RPN stage and the classification and regression stage. RPN loss consists of anchor box classification loss and anchor box regression loss. Anchor box classification loss evaluates whether the anchor boxes accurately classify the object (foreground) and the background. For this purpose, the following binary cross-entropy loss equation is used.

Equation 1: binary cross-entropy loss

$$L_{cls} = -\frac{1}{N_{cls}} \sum_i [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

y_i has the value of 1 if the anchor box i includes the object and has the value of 0 if it does not include the object. p_i is the model's prediction of the probability of the anchor box containing the object. N_{cls} is the number of anchor boxes used in classification.

Anchor box regression loss evaluates how accurately the RPN locates the object using the anchor box. The Smooth L1 loss equation, which is based on the distance from the actual object location and the predicted object location by the RPN, is used to calculate regression loss. The following equation is used for anchor box regression loss.

Equation 2: anchor box regression loss

$$L_{reg} = \frac{1}{N_{reg}} \sum_i \text{smooth}_{L1}(t_i - \hat{t}_i)$$

smooth_{L1} represents the Smooth L1 loss function imported from pytorch v7. t_1 represents the actual location of the bounding box, while \hat{t}_1 represents the model's predicted location of the bounding box. N_{reg} is the number of anchor boxes used in regression. During the classification and regression stage, the network adjusts bounding boxes to the according proposed regions. Loss during this stage is comprised of classification loss and bounding box regression loss. Classification loss evaluates how accurately the network classifies each object. As this study aims to conduct multiple classification, the following cross-entropy loss equation will be used in this stage.

Equation 3: anchor box regression loss

$$L_{cls}^{roi} = -\log(p_{class})$$

p_{class} represents the probability that the model predicts the object to be of the according class. Bounding box regression loss evaluates the bounding box regression process of refining the bounding box to match the final object's size and location. Similar to the anchor box regression loss, bounding box regression loss also uses the Smooth L1 loss equation. The model automatically aims to reduce this loss by minimizing the distance between the actual object and the bounding box. The following equation is used in the calculation of bounding box regression loss.

Equation 4: bounding box regression loss

$$L_{reg}^{roi} = smooth_{L1}(t_i - \hat{t}_i)$$

$smooth_{L1}$ represents the Smooth L1 loss function. t_i represents the actual location of the bounding box, while \hat{t}_i represents the model's predicted location of the bounding box.

The final loss is calculated by combining the loss values of the four components above.

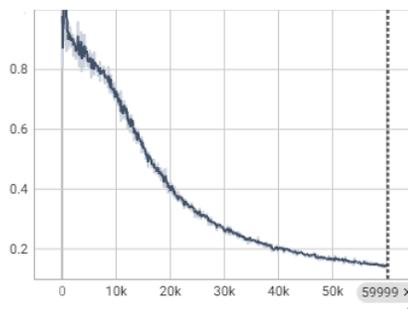
Equation 5: final loss

$$L = \lambda_1 L_{cls} + \lambda_2 L_{reg} + \lambda_3 L_{cls}^{roi} + \lambda_4 L_{reg}^{roi}$$

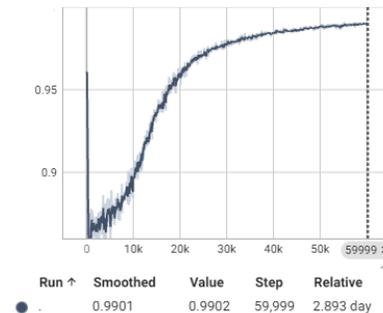
This loss function indicates the proficiency of the model, and each of $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are parameters that control the weight of the different loss values. This study used the default values of the parameters in pytorch. The aim of the model is to reduce the total loss. Through this process, the model automatically improves its ability to accurately classify and locate the object of interest.

Results and Discussion

The maximum number of classes that could be classified without compromising the 90% accuracy was three classes. These classes were PIO, bronchial thickening, and reticulonodular opacity. These three classes were the classes with the greatest amount of data. The model was trained for 100,000 epochs. As seen in Graph 1, the total loss approaches 0.1. Graph 2 also shows the total accuracy approaching 99.01%.



(a)



(b)

Figure 4. Learning curve. (a): A graph showing the total loss of the model and (b): A graph showing the accuracy of the model.

Sample model predictions were taken, showing the answer image on the right and the model prediction on the left. The model was able to accurately identify the location and type of the lesion for all these images, and it generally had a confidence of greater than 75%.

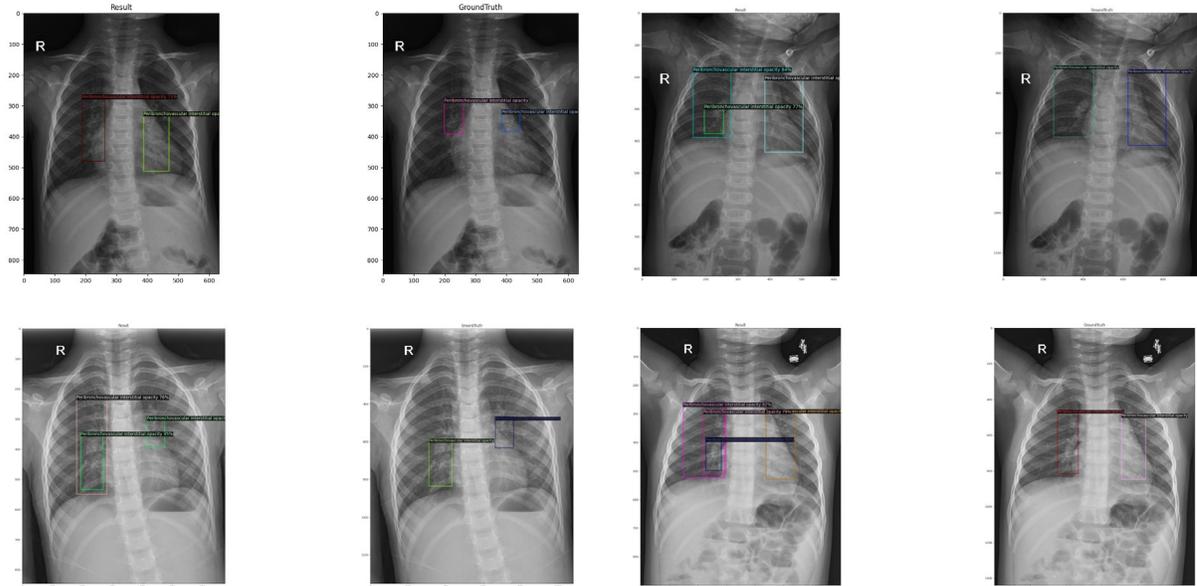


Figure 5. A series of images showing samples of model predictions.

The model met the objectives of 0.1 loss and greater than 90% accuracy. The model's high performance indicates its viability to be utilized in the field. The image samples further corroborate that the model has potential to aid healthcare workers in the diagnosis of pediatric thoracic diseases through identifying abnormal lesions. For example, lung opacity may indicate fluid in air spaces, thickening of air space walls, inflammation, and more (Türk and Kökver). In underdeveloped regions with highly strained medical facilities, this model could be of aid by reducing burden and saving time of healthcare workers while not greatly increasing the cost of healthcare or demanding advanced medical supplies.

However, the range of lesions this model can detect is not very large, contrary to the original objectives. Data shortage remains a large challenge for studies on pediatric data. In the PediCXR dataset, the classes that were used in the creation of this model (PIO, bronchial thickening, reticulonodular opacity) all had over 500 images, and PIO had over 1500 images total. The class with the fourth-greatest amount of data was consolidation, but this class only had a little over 200 images. Most of the other classes in the dataset have less than 100 images. Hence, the decreased accuracy of the model with increased classes may not be because multiple classification is challenging, but because there is insufficient data for the other classes. When compared to adult data, open access pediatric data is much more challenging to find. With sufficient data, multiple classification with high accuracy and low loss is possible. Studies on adult chest radiographs have been able to pool data from various datasets to develop a CNN model for five classes and achieve high accuracy (Malik et al. 2024). More attention will be needed towards the creation of open access pediatric databases to further the development of pediatric-related AI models.

Further research will be focused on overcoming this challenge of data scarcity. As demonstrated with the Malik et al. study, pooling data from multiple sources can create a larger dataset with less class imbalance. Pooling data from various pediatric chest radiograph datasets will be attempted. Pediatric data could also be

extracted from large datasets with data from multiple age groups. After constructing the larger dataset, the same experimental process of training the model with an increasing number of classes will be attempted. Afterwards, external validation with a different dataset will be attempted. Another future objective is to create an accessible version of the model, such as a computer application.

Conclusion

In conclusion, this research has demonstrated the feasibility and potential of leveraging artificial intelligence to aid in the diagnosis of pediatric thoracic diseases through chest radiograph analysis. The developed model, based on a Faster R-CNN architecture trained on the PediCXR dataset, has achieved remarkable performance, meeting the predefined objectives of a loss under 0.1 and an accuracy exceeding 90%. While the model's high performance underscores its utility in identifying abnormal lesions on pediatric chest radiographs, it is acknowledged that the range of detectable lesions remains somewhat limited. This limitation is attributed to the challenge of data scarcity within pediatric datasets, as evidenced by the class imbalance observed in the PediCXR dataset. Future research I will focus on overcoming data scarcity by pooling data from multiple sources and potentially extracting pediatric data from larger datasets containing multiple age groups. By constructing a more extensive and diverse dataset, it is anticipated that the model's capabilities can be enhanced, enabling it to detect a broader spectrum of lesions with higher accuracy.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

- Barakat, Natali, et al. "A Machine Learning Approach on Chest X-rays for Pediatric Pneumonia Detection." *Digital Health*, vol. 9, 4 June 2023, p. 20552076231180008. PubMed Central, <https://doi.org/10.1177/20552076231180008>. Accessed 7 Nov. 2023.
- Chen, Kai-Chi, et al. "Diagnosis of Common Pulmonary Diseases in Children by X-ray Images and Deep Learning." *Scientific Reports*, vol. 10, 2020, <https://doi.org/10.1038/s41598-020-73831-5>. Accessed 7 Nov. 2023.
- Chen, Yiyun, et al. "Deep Learning for Classification of Pediatric Chest Radiographs by WHO's Standardized Methodology." *PLOS-ONE*, 21 June 2021, <https://doi.org/10.1371/journal.pone.0253239>. Accessed 7 Nov. 2023.
- E, Longjiang, et al. "Image-based Deep Learning in Diagnosing the Etiology of Pneumonia on Pediatric X-rays." *Pediatric Pulmonology*, vol. 56, no. 5, 19 Mar. 2021, pp. 1036-44. Wiley Online Library, <https://doi.org/10.1002/ppul.25229>. Accessed 7 Nov. 2023. Abstract.
- Elgendi, Mohamed, et al. "The Effectiveness of Image Augmentation in Deep Learning Networks for Detecting COVID-19: A Geometric Transformation Perspective." *Frontiers in Medicine*, vol. 8, 1 Mar. 2021, <https://doi.org/10.3389/fmed.2021.629134>. Accessed 2 Jan. 2024.
- "55 Countries Face a Health Worker Crunch Linked to COVID-19: WHO." United Nations, 14 Mar. 2023, news.un.org/en/story/2023/03/1134562. Accessed 1 Jan. 2024.
- GBD 2015 LRI Collaborators. "Estimates of the Global, Regional, and National Morbidity, Mortality, and Aetiologies of Lower Respiratory Tract Infections in 195 Countries: A Systematic Analysis for the Global Burden of Disease Study 2015." *The Lancet Infectious Diseases*, vol. 17, no. 11, Nov. 2017, pp. 1133-61, [https://doi.org/10.1016/S1473-3099\(17\)30396-1](https://doi.org/10.1016/S1473-3099(17)30396-1). Accessed 7 Nov. 2023.

Goldberger, A., et al. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* [Online]. 101 (23), pp. e215–e220." (2000).

"Health Workforce." World Health Organization, www.who.int/health-topics/health-workforce#tab=tab_1. Accessed 1 Jan. 2024.

Leggiadro, Robert J. "Epidemics after Natural Disasters." *The Pediatric Infectious Disease Journal*, vol. 26, no. 6, June 2007, p. 552, <https://doi.org/10.1097/INF.0b013e318054e34c>. Accessed 1 Jan. 2024.

"Low Quality Healthcare Is Increasing the Burden of Illness and Health Costs Globally." World Health Organization, 5 July 2018, www.who.int/news/item/05-07-2018-low-quality-healthcare-is-increasing-the-burden-of-illness-and-health-costs-globally. Accessed 1 Jan. 2024.

Malik, Hassaan, et al. "CDC_Net: Multi-classification Convolutional Neural Network Model for Detection of COVID-19, Pneumothorax, Pneumonia, Lung Cancer, and Tuberculosis Using Chest X-rays." *Multimedia Tools and Applications*, vol. 82, 20 Sept. 2022, pp. 13855-80, <https://doi.org/10.1007/s11042-022-13843-7>. Accessed 3 Jan. 2024.

Nguyen, Ha Quy, et al. "VinDr-CXR: An open dataset of chest X-rays with radiologist annotations" (version 1.0.0). *PhysioNet* (2021), <https://doi.org/10.13026/3akn-b287>.

Nguyen, Ha Q., et al. "VinDr-CXR: An Open Dataset of Chest X-rays with Radiologist's Annotations." *Scientific Data*, vol. 9, 2022, <https://doi.org/10.1038/s41597-022-01498-w>. Accessed 7 Nov. 2023.

Padash, Sirwa, et al. "Pediatric Chest Radiograph Interpretation: How Far Has Artificial Intelligence Come? A Systematic Literature Review." *Pediatric Radiology*, vol. 52, 23 Apr. 2022, pp. 1568-80, <https://doi.org/10.1007/s00247-022-05368-w>. Accessed 2 Jan. 2024.

Pakenham-Walsh, Neil, and Frederick Bukachi. "Information Needs of Health Care Workers in Developing Countries: A Literature Review with a Focus on Africa." *Human Resources for Health*, vol. 7, no. 30, 8 Apr. 2009. PubMed Central, <https://doi.org/10.1186/1478-4491-7-30>. Accessed 1 Jan. 2024.

Pham, Hieu Huy, et al. "VinDr-PCXR: An open, large-scale pediatric chest X-ray dataset for interpretation of common thoracic diseases" (version 1.0.0). *PhysioNet* (2022), <https://doi.org/10.13026/k8qc-na36>.

Pham, Hieu H., et al. "PediCXR: An Open, Large-scale Chest Radiograph Dataset for Interpretation of Common Thoracic Diseases in Children." *Scientific Data*, vol. 10, 2023, <https://doi.org/10.1038/s41597-023-02102-5>. Accessed 7 Nov. 2023.

Rahman, Tawsifur, et al. "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray." *Applied Sciences*, vol. 10, no. 9, 6 May 2020, p. 3233, <https://doi.org/10.3390/app10093233>. Accessed 1 Jan. 2024.

Ravi, Vinayakumar, et al. "A Cost-sensitive Deep Learning-based Meta-classifier for Pediatric Pneumonia Classification Using Chest X-rays." *Expert Systems*, vol. 39, no. 7, 23 Mar. 2022. Wiley Online Library, <https://doi.org/10.1111/exsy.12966>. Accessed 7 Nov. 2023. Abstract.

Tang, Yu-Xing, et al. "Automated Abnormality Classification of Chest Radiographs Using Deep Convolutional Neural Networks." *npj Digital Medicine*, vol. 3, 14 May 2020, <https://doi.org/10.1038/s41746-020-0273-z>. Accessed 1 Jan. 2024.

Türk, Fuat, and Yunus Kökver. "Detection of Lung Opacity and Treatment Planning with Three-Channel Fusion CNN Model." *Arab Journal for Science and Engineering*, 14 Apr. 2023, pp. 1-13. PubMed Central, <https://doi.org/10.1007/s13369-023-07843-4>. Accessed 13 Nov. 2023.