REVIEW

Pediatric chest radiograph interpretation: how far has artificial intelligence come? A systematic literature review

Sirwa Padash[1,2] · Mohammad Reza Mohebbian[3] · Scott J. Adams[1] · Robert D. E. Henderson[1] · Paul Babyn[1]

Abstract

Most artificial intelligence (AI) studies have focused primarily on adult imaging, with less attention to the unique aspects of pediatric imaging. The objectives of this study were to (1) identify all publicly available pediatric datasets and determine their potential utility and limitations for pediatric AI studies and (2) systematically review the literature to assess the current state of AI in pediatric chest radiograph interpretation. We searched PubMed, Web of Science and Embase to retrieve all studies from 1990 to 2021 that assessed AI for pediatric chest radiograph interpretation and abstracted the datasets used to train and test AI algorithms, approaches and performance metrics. Of 29 publicly available chest radiograph datasets, 2 datasets included solely pediatric chest radiographs, and 7 datasets included pediatric and adult patients. We identified 55 articles that implemented an AI model to interpret pediatric chest radiographs or pediatric and adult chest radiographs. Classification of chest radiographs as pneumonia was the most common application of AI, evaluated in 65% of the studies. Although many studies report high diagnostic accuracy, most algorithms were not validated on external datasets. Most AI studies for pediatric chest radiograph interpretation have focused on a limited number of diseases, and progress is hindered by a lack of large-scale pediatric chest radiograph datasets.

Keywords Artificial intelligence · Chest · Pediatric · Deep learning · Pneumonia · Radiography

Introduction

Chest radiography is one of the most frequently used imaging modalities in children. It is often used to assess disorders of the pediatric lung parenchyma, airways, heart and mediastinum. Chest radiographs play a significant role in screening and diagnosing disease across all pediatric care settings, including the emergency department and inpatient settings such as neonatal and pediatric intensive care units. Chest radiographs also play a critical role in assessing support devices, including chest tube and catheter positions, and are vital in diagnosing a number of life-threatening conditions and their complications, such as pneumonia, pneumothorax and diaphragmatic hernia [1].

It is well known that the interpretation of pediatric chest radiographs differs from that in adult chest radiographs. There is a varying radiographic appearance of the normal growing child, ranging from tiny premature neonates to adolescents. Pediatric radiography often utilizes differing techniques of acquisition, and frequently, differing pathologies are encountered that are not commonly seen in adults. For example, the developmental appearance of the normal thymus in young infants can mimic a parenchymal lung infection or mediastinal mass to the unwary. Additionally, children often present with differing appearances of common pathologies encountered throughout life, such as the "round" pneumonia seen in younger children. Radiographic findings in children also have broader differential diagnoses, such as the inclusion of congenital lung disorders and other malformations [2]. These differences in radiographic interpretation pose additional challenges for any artificial intelligence (AI) models developed for chest radiography.

✉ Sirwa Padash
Sirwa.Padash@gmail.com

1 Department of Medical Imaging,
University of Saskatchewan, 103 Hospital Drive,
Saskatoon, Saskatchewan S7N 0W8, Canada

2 Department of Radiology, Mayo Clinic, Rochester, MN, USA

3 Department of Electrical and Computer Engineering,
University of Saskatchewan, Saskatoon,
Saskatchewan, Canada

Background

The application of AI and its subset of machine learning (ML), especially deep learning (DL), is of growing clinical, research and commercial interest within medical imaging. Initial applications of AI have spanned all imaging modalities and now include the detection of stroke and acute intracranial hemorrhage with CT, the assessment of breast lesions with mammography and US, the detection of fractures with wrist radiographs and the assessment of a number of lung and cardiac abnormalities with chest radiographs [3]. There are many promising benefits of incorporating AI into the interpretation of chest radiographs, such as the potential for improved detection and diagnosis, quantification, triage and streamlined workflow [4, 5]. Assessment of chest imaging represents one of the most rapidly growing research areas, with determination of the presence of lung nodules, tuberculosis, pneumonia, pneumothorax and cardiomegaly being among the most common applications [4, 6]. Several excellent review articles have detailed AI and DL techniques and their applications for chest imaging; however, these reviews have primarily focused on adult chest radiograph applications [4, 7] without a dedicated summary of related benefits and limitations for pediatric patients.

Machine learning models or algorithms are trained with datasets. The designed model uses the training dataset to learn the data characteristics and then predicts the target task on new input data [8]. Dataset characteristics including size and domain, availability of labels and annotated data, class type, data balance (frequency of each class in the dataset) and the quantity and quality of features have a critical impact on the model's performance and can be a source of bias [9]. AI training is typically done with supervision provided by labeled images. Categorical labels might be clinically based, such as the presence or absence of pneumonia, or related to radiographic findings such as consolidation or pneumothorax. Categorical labels can be derived either by separate review from radiologists or other experts or derived by natural language processing (NLP) of radiologist reports. Annotated images where individual lesions are marked with boundary boxes placed on the image provide better information for object localization and class recognition, allowing the location and size of any lesions, such as lung consolidation, to be captured. Annotated images provide a better opportunity to develop explainable AI to identify where an algorithm localizes pathology such as pneumonia using methods such as Gradient-weighted Class Activation Mapping (Grad-CAM) [10] and SHapley Additive exPlanations (SHAP) [11]. However, annotation generally requires significant expert input and time.

This review article summarizes the available literature on the application of AI to pediatric chest radiograph interpretation. Given that datasets are critical to algorithm training, testing and validation, this paper also compiles data on all publicly available chest radiograph datasets to determine pediatric image availability and potential utility and limitations for pediatric AI-related studies. We hope this paper helps researchers understand the state of AI for pediatric chest radiograph interpretation and catalyzes fruitful progress in this area.
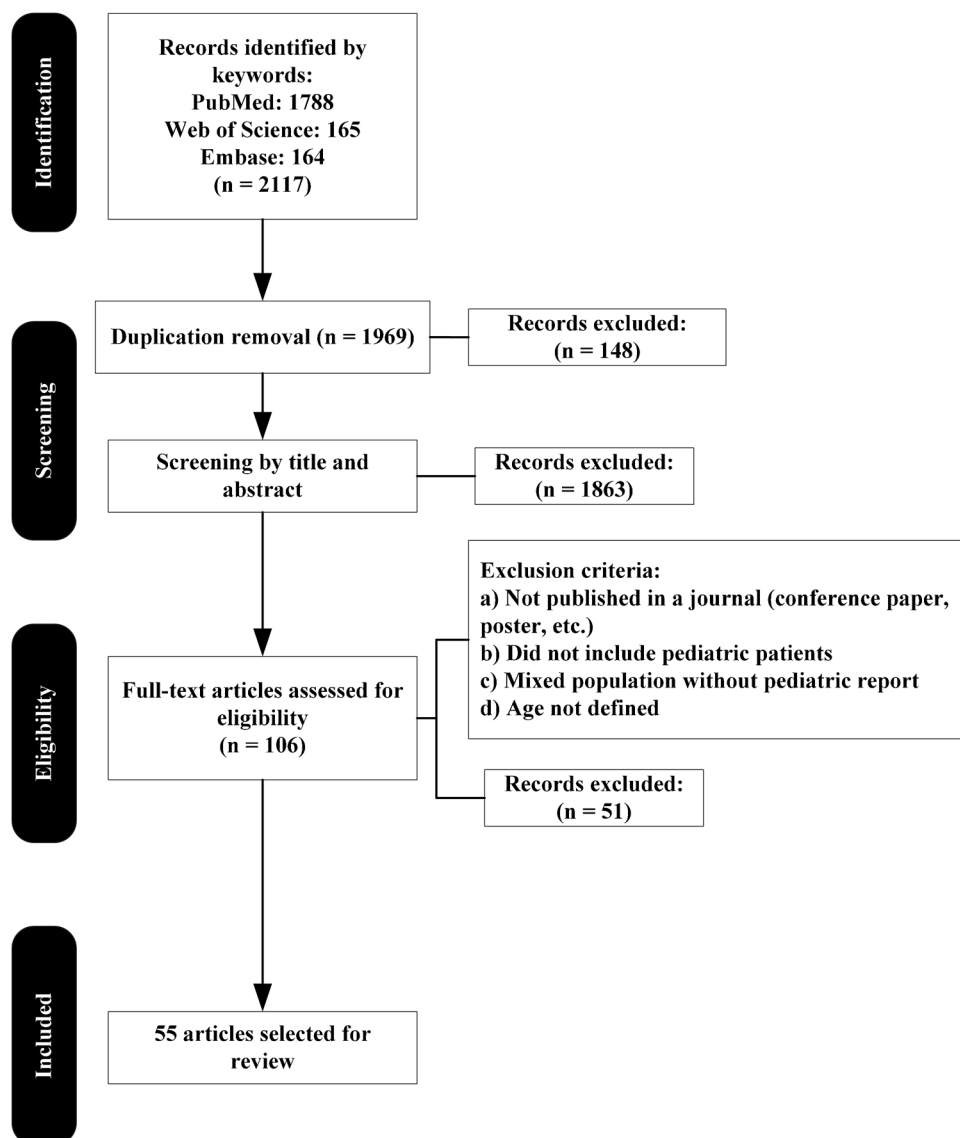
Methods

Search strategy

We performed a systematic literature review to capture relevant studies that either designed or evaluated AI for chest radiograph interpretation in the pediatric population. We utilized the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy (PRISMA-DTA) checklist [12]. Figure 1 shows the study selection diagram. We conducted the literature search in the PubMed, Embase and Web of Science databases from Jan. 1, 1990, to July 5, 2021. Each database was queried with the following keywords: ("artificial intelligence" OR "machine learning" OR "neural networks" OR "deep learning") AND ("pediatric" OR "child" OR "toddler" OR "infant" OR "neonate") AND ("radiograph" OR "chest film" OR "X-rays"). Only articles written in English were considered. Two reviewers independently searched the databases to capture eligible studies (S.P. with 2 years of research fellowship experience and M.R.M., AI-electrical engineer with 6 years of research experience). They also reviewed references from selected studies to identify further relevant studies. In addition, we identified all publicly available chest radiograph datasets by analyzing the publications found in the literature search, as well as a narrative review of the databases. We also reviewed each dataset and contacted the dataset's provider to clarify some of the dataset's properties and availability.

Study selection

Inclusion criteria were as follows: (a) studies related to developing or evaluating AI for chest radiograph interpretation and (b) articles that included a pediatric population either alone or in combination with adult patients. We screened the retrieved articles based on the title and the abstract and reviewed full texts of selected articles. Pediatric patients were defined as patients younger than 18 years based on the World Health Organization (WHO) definition of children [13]. Studies differed to some extent in their definition or inclusion of pediatric age ranges; therefore, for clarity, the tables of studies included in this review include descriptions of the age ranges. We excluded articles if they

Fig. 1 Study selection diagram

```
Identification    Records identified by
                  keywords:
                  PubMed: 1788
                  Web of Science: 165
                  Embase: 164
                  (n = 2117)
                        │
                        ▼
Screening         Duplication removal (n = 1969) ──── Records excluded:
                                                      (n = 148)
                        │
                        ▼
                  Screening by title and ──── Records excluded:
                  abstract                    (n = 1863)
                        │
                        ▼
Eligibility       Full-text articles assessed for ──── Exclusion criteria:
                  eligibility                          a) Not published in a journal (conference paper,
                  (n = 106)                            poster, etc.)
                                                       b) Did not include pediatric patients
                                                       c) Mixed population without pediatric report
                                                       d) Age not defined

                                                  ──── Records excluded:
                                                       (n = 51)
                        │
                        ▼
Included          55 articles selected for
                  review
```

did not include pediatric patients, contained both adult and pediatric populations without separately reporting evaluation metrics for the pediatric group, did not define the age of the target population, or were not published in peer-reviewed journals (i.e. conference papers, posters, etc.).

Data abstraction

The data collected from each article included year of publication, the country in which the study originated, study design, reference standard, sample size and age distribution. In cases for which this information was not publicly available, the dataset's provider was contacted to determine dataset characteristics and availability.

The names of all chest radiograph datasets utilized in the selected articles were also collated. Given that the datasets usually have an associated article to introduce the dataset,

we reviewed the articles, access sources and metadata files of each dataset to understand the quantity and quality of the dataset, including patient inclusion criteria, available annotations and labeling, chest radiograph acquisition technique (anteroposterior or lateral views) and age range of the included patients. We collected labels used for the dataset, along with the definition of the ground truth and whether abnormal regions on each image were annotated.

Results

The initial search retrieved 2,117 articles from the databases. Because of duplication, we removed 148 articles. After assessing the title and abstract, we chose 106 papers for full-text review. We subsequently excluded 51 papers

because they did not meet inclusion criteria, resulting in 55 papers being included in the study (Fig. 1).

Review of publicly available chest radiograph datasets

The 55 articles that were eligible for inclusion in the study used a total of 20 proprietary and 9 publicly available chest radiograph datasets. We found an additional seven publicly available chest radiograph databases and multiple coronavirus disease 2019 (COVID-19) datasets from other sources. Of the 16 publicly available datasets, 2 were solely pediatric [14–16] and 7 had mixed pediatric and adult populations; the others included only adult populations [17–24]. We identified a total of 20,337 unique pediatric images (5,948 normal and 14,389 abnormal) from 7 datasets, including Guangzhou Women and Children's Medical Center (GWCMC) [14], Pneumonia Etiology Research for Child Health (PERCH), PadChest, the National Institutes of Health (NIH) NIH-14, the National Institute of Allergy and Infectious Diseases (NIAID) TB dataset, the Shenzhen database and the Montgomery County datasets. In addition, the Radiological Society of North America (RSNA) and Society for Imaging Informatics in Medicine (SIIM)–American College of Radiology (ACR) datasets include 1,276 and 617 pediatric images, respectively, which are reannotated images from the NIH-14 dataset. Figure 2 and Online Supplementary Material 1 show the age distribution in the available pediatric chest radiographs. The most frequent labels were pneumonia, infiltration, consolidation and cardiomegaly, representing 58% of the abnormal labels. Further details of the publicly available chest radiograph datasets are presented in Online Supplementary Material 2 [14–23].
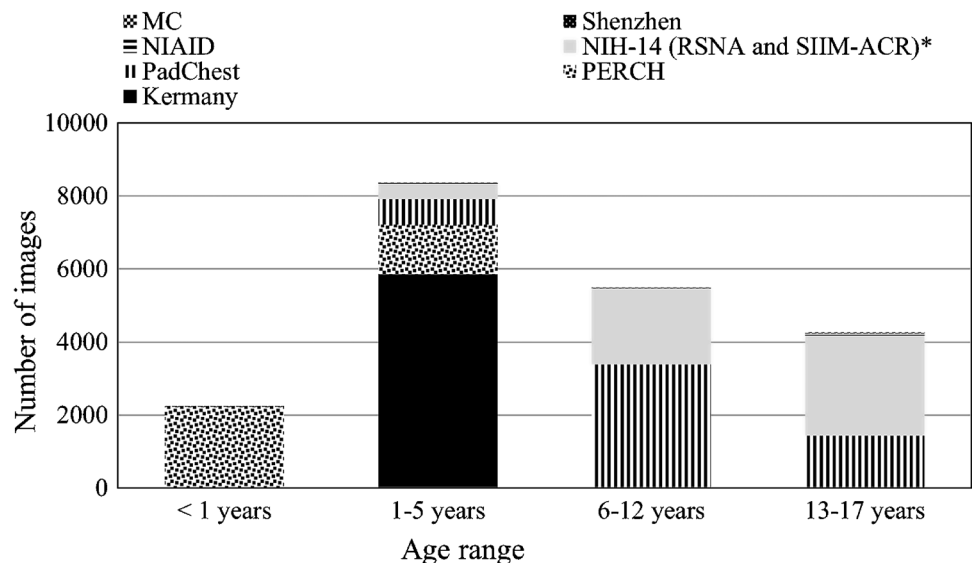
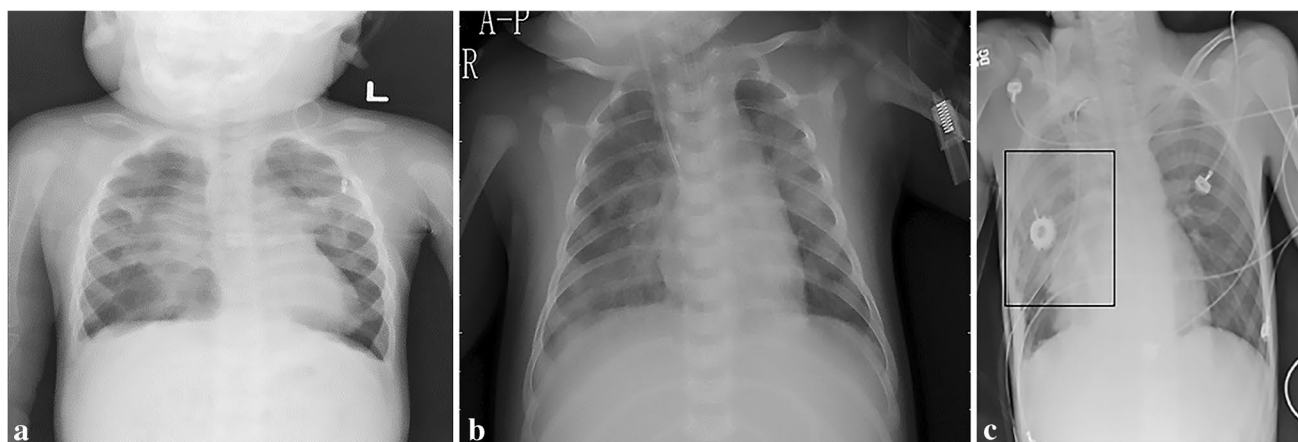## Guangzhou Women and Children's Medical Center (GWCMC) dataset

This dataset, also known as the Kermany dataset, is from Guangzhou Women and Children's Medical Center in Guangzhou, Guangdong, China [14]. This dataset consists of 5,856 anteroposterior (AP) chest radiographs from children ages 1–5 years. The dataset includes three labels: normal, bacterial pneumonia or viral pneumonia, including 5,232 and 624 training and test samples, respectively. Two physicians labeled all images, with a third physician verifying all test dataset labels. It is not clear what, if any, additional clinical criteria were used to determine these labels (Fig. 3).

Pneumonia Etiology Research for Child Health (PERCH) dataset

The PERCH project attempted to identify the etiology of severe childhood pneumonia requiring hospital admission. Children who were hospitalized with pneumonia were included from nine sites in seven countries (Bangladesh, The Gambia, Kenya, Mali, South Africa, Thailand and Zambia) between Aug. 15, 2011, and Jan. 30, 2014 [15, 16, 25]. The data are held by Johns Hopkins University and are available for research purposes under a data-sharing agreement [26]. The PERCH chest radiograph dataset consists of 3,587 images that were randomly interpreted and labeled by two members of a reading panel. This panel included 14 radiologists and pediatricians from 7 countries trained with WHO methodology. In cases of disagreement, a second group of arbitrators consisting of two experienced radiologists who were unaware of the initial assessment made the final consensus by discussion. All image readers were blinded to the clinical and demographic information associated with each

Fig. 2 The age distributions of the pediatric population for the seven pediatric and mixed adult datasets containing unique pediatric images. *Note that the Radiological Society of North America (RSNA) and Society for Imaging Informatics in Medicine (SIIM)–American College of Radiology (ACR) datasets are subsets of the National Institutes of Health NIH-14 dataset

Fig. 3 Example images from two datasets containing pediatric patients illustrate the variation in labeling used. **a** Anteroposterior chest radiograph of a 1-year-old boy from the National Institutes of Health (NIH) NIH-14 dataset demonstrates extensive perihilar infiltrates without effusion. This radiograph has two labels in the NIH-14 dataset: infiltration and effusion (identification number 6649). **b** Anteroposterior chest radiograph of a child age 1–5 years from the Guangzhou Women and Children's Medical Center (GWCMC) dataset. This radiograph is labeled as bacterial pneumonia in the dataset. Note the presence of support tubes and lines that are not labeled (identification number person1946_bacteria_4875-Kermany). **c** Anteroposterior chest radiographs of a boy age 13 years from the Radiological Society of North America (RSNA) dataset. This radiograph is labeled as lung opacity and the bounding box localizes the pathology (identification number: 089a996e-425c-4,311-b473-6948b3eb1060)

chest radiograph, and readers did not interpret chest radiographs from their site. The chest radiographs were assigned with five labels, including (1) consolidation only (alveolar consolidation, including pleural effusion if present), (2) other infiltrate only, (3) both consolidation and other infiltrate, (4) normal (no consolidation or infiltrate) and (5) uninterpretable for consolidation and/or other infiltrate [15, 16].

PadChest dataset

This dataset contains 160,861 chest radiographs of 67,000 patients from the Hospital Universitario de San Juan, Alicante, Spain. Of the 160,861 chest radiographs, 5,533 are of patients younger than 18 years [17]. Cases are categorized into a hierarchical structure and annotated with 174 different radiographic findings, 19 differential diagnoses and 104 anatomical locations. NLP was used to label 73% of the cases while trained physicians manually labeled 27% from radiology reports. This dataset contains the largest number of manual labels among the large public datasets. PadChest provides some metadata, including patients' year of birth, gender, image projection, positioning, type and manufacturer of the equipment used to acquire the chest radiograph, and annotation method (manually or by NLP).

National Institutes of Health Clinical Center (NIH)-14 dataset

The National Institutes of Health (NIH) Clinical Center released a chest radiograph dataset with 112,120

frontal-view chest radiograph images of 32,717 patients from the NIH Clinical Center [18]. There are 5,242 pediatric images in this collection, spanning ages 1–17 years (Fig. 3). Fourteen keywords of common radiologic findings were searched within the picture archiving and communication system (PACS) to extract all related radiologic reports and their associated images. NLP was employed to label the images based on the text of the radiology report. The dataset has 993 images annotated with bounding boxes to show 1,600 regions of interest for 8 labels.

Radiological Society of North America (RSNA) pneumonia dataset

The RSNA's AI challenge in 2018 focused on pneumonia detection on chest radiographs. To facilitate this challenge, 30,227 chest radiographs from the NIH-14 dataset were relabeled by 18 radiologist volunteers from the Society of Thoracic Radiology (STR) and the RSNA. The test set was labeled based on the majority assessments of three radiologists (two from STR and one from RSNA). The pediatric subset of the RSNA dataset includes normal cases ($n=462$), cases with a lung opacity ($n=1,074$) and cases with no lung opacity but another abnormality ($n=429$), for a total of 1,965 pediatric images. Cases with lung opacities were manually labeled with bounding boxes to highlight regions of opacity [19, 27] (Fig. 3). This dataset has an age range of 1–17 years. Training and test sets are provided by the challenge holder.

Society for Imaging Informatics in Medicine (SIIM)–American College of Radiology (ACR) pneumothorax segmentation dataset

The SIIM and the ACR created a dataset for a pneumothorax detection challenge. The dataset is a subset of the NIH-14 dataset and consists of a 12,047-image training set. Because the dataset has been published for competition purposes, the test set is not publicly available. Six radiologists from SIIM and 13 radiologists from ACR were involved in labeling each image manually [21, 28]. The dataset contains 617 pediatric cases ranging 1–17 years old. This dataset also provides training and test sets.

National Institute of Allergy and Infectious Diseases (NIAID) tuberculosis dataset

The United States National Institute of Allergy and Infectious Diseases (NIAID) led the development of the Tuberculosis Portals Program, an international collaboration for tuberculosis data-sharing and analysis to promote tuberculosis research. It has a database containing clinical, imaging and bacterial genomic information from drug-sensitive and resistant tuberculosis cases sourced from 40 sites across different countries. The dataset contains 6,251 chest radiographs and CT images from members of the collaboration. Among these, 4,656 images are unique images from 3,149 unique patients and have been labeled manually or by a convolutional neural network (CNN) classifier. There are 71 chest radiographs from 45 pediatric patients in total [23].

Shenzhen dataset

This dataset was obtained from the outpatient clinics of Shenzhen No. 3 People's Hospital, Guangdong Medical College, Shenzhen, China, over 1 month in 2012. Associated available metadata include gender, age, radiologic diagnosis and, in some cases, the location of the abnormality.

However, the annotation method and ground truth of labeling were not reported [24]. The dataset includes 662 chest radiographs, including 31 pediatric chest radiographs (from children ages 2 months to 17 years).

Montgomery County dataset

This dataset was collected from the tuberculosis screening program in Montgomery County, MD, United States. It comprises 58 chest radiographs of tuberculosis and 80 normal chest radiographs, with only 17 images of pediatric patients (ranging 4–17 years old). Images in the dataset are associated with a segmentation mask to illustrate the involved lung region as well as metadata including gender, age, abnormality seen in the lung, diagnosis and location of abnormalities. Some cases also contain labels indicating chronic or acute tuberculosis patterns and the treatment received and duration [24].

Review of artificial intelligence studies for pediatric chest radiographs

The comprehensive literature review showed that 36 (65%) of the 55 eligible studies published between 1990 and 2021 focused predominantly on pneumonia. Of the studies focusing on pneumonia, 32 used publicly available datasets and 4 used private datasets. Beyond pneumonia, other studies have focused on detecting and classifying thoracic diseases such as hyaline membrane disease and meconium aspiration syndrome, quantitative and scoring prediction, catheter and tube detection and classification, and lung segmentation (Fig. 4).

Studies to detect pneumonia on pediatric chest radiographs

Online Supplementary Materials 3 and 4 summarize reviewed papers pertaining to detecting pneumonia on pediatric chest radiographs [14, 26, 29–62]. Most early studies

Fig. 4 Chart shows a summary of conditions studied in the 55 selected pediatric artificial intelligence and chest radiograph articles, listing the total numbers and percentages

- Pneumonia detection and classification: 36
- Lung region detection: 7
- Multiple pulmonary disease detection: 5
- Quantitative and scoring prediction: 3
- Catheter and tube detection: 2
- Consolidation: 2

Total number: 55
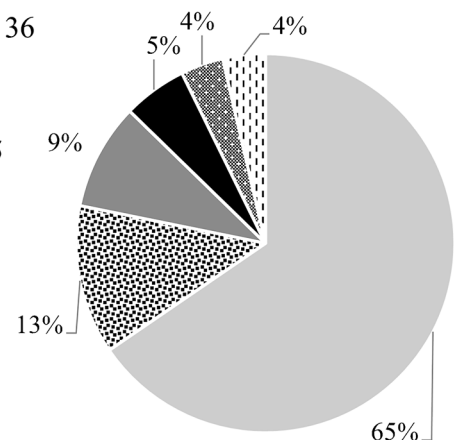
on the classification of pediatric chest radiographs as pneumonia being present vs. absent used conventional texture analysis and conventional machine learning classifiers such as k-nearest neighbors [29] and support vector machines [30].

More recently, deep learning has become the dominant ML approach for chest radiograph classification for pneumonia, with a variety of CNNs used over the last few years. While larger datasets have generally been required for training deep learning algorithms compared to conventional texture analysis, many deep learning algorithms have achieved higher performance. Kermany et al. [14] employed an Inception v3 architecture to classify optical coherence tomography (OCT) images of the retina and subsequently demonstrated the generalizability of this technique for the classification of pediatric chest radiographs. They achieved area under the receiver operating characteristic curve (AUC) of 0.968 for classifying radiographs as pneumonia present vs. absent. The GWCMC dataset was first released in this project.

Since the GWCMC dataset was released, multiple additional studies have used this dataset for classifying pediatric chest radiographs as pneumonia present vs. absent, or bacterial pneumonia vs. viral pneumonia vs. normal, using different architectures such as AlexNet, ResNet, DenseNet, SqueezeNet, VGG16, VGG19, ResNet-18, ResNet-50 and Inception v3 (Online Supplementary Material 4) [14, 26, 33–62].

In addition to pre-training models with general image datasets such as ImageNet (which is commonly employed in many algorithms in medical imaging [63]), pre-training models with large datasets of adult chest radiographs provide a promising approach to reducing the number of pediatric chest radiographs required for training [37, 54]. Tang et al. [54] used GoogleNet (Inception v3) to classify pediatric chest radiographs from the GWCMC dataset as pneumonia present vs. absent. When trained on an adult dataset (NIH ChestX-ray 14), the model achieved an AUC of 0.916 in the test subset of the GWCMC dataset. When the same model was trained exclusively on a subset of the GWCMC dataset, the model achieved an AUC of 0.975, and when the same architecture was pre-trained on NIH ChestX-ray 14 and then fine-tuned on a training subset of the GWCMC dataset using a transfer learning approach, AUC increased to 0.985.

Ensemble approaches have also shown excellent results. Chouhan et al. [45] individually trained AlexNet, DenseNet121, Inception v3, GoogleNet 50 and ResNet-18 on a training subset from the GWCMC dataset and subsequently developed an ensemble model by majority voting; accuracy of 96.4% was achieved, higher than each of the individual models. A similar approach was utilized by Hashmi et al. [46], who achieved an accuracy of 98.4% and an AUC of 0.999. Other groups have similarly combined

well-known CNNs and conventional machine learning classifiers, as well as combined handcrafted features from texture analysis and features extracted from CNNs, with promising results [60, 61].

Qu et al. [55] highlighted the effect of class imbalance on pneumonia present vs. absent classification performance. They utilized ResNet-18 and showed that a close-to-balanced training set (with relatively equal numbers of images with pneumonia present or absent) resulted in the best performance. Furthermore, oversampling and under-sampling methods were successful in mitigating class imbalance.

To capture multi-scale image features while minimizing computational costs, Hu et al. [38] developed a multi-kernel depth-wise convolution (MD-Conv). In addition to assessing the network using NIH-14, they assessed the network for classifying pediatric chest radiographs using the GWCMC dataset and achieved an AUC of 0.98, slightly higher than the original paper of Kermany et al. [14], which achieved an AUC of 0.96.

Studies using artificial intelligence to detect other pathologies

Few AI studies have considered the broad range of pathologies other than pneumonia that are common on pediatric chest radiographs (Online Supplementary Materials 5 and 6) [64–82]. A pioneering study by Gross et al. [64] first implemented a neural network to assist in neonatal chest radiograph interpretation. Zaglam et al. [65] and Chen et al. [66] used machine learning models to classify chest radiographs of children for respiratory distress syndrome, bronchiolitis/bronchitis, bronchopneumonia/interstitial pneumonitis, lobar pneumonia or pneumothorax.

Artificial intelligence has also been used to automate the quantitative scoring of pediatric chest radiographs. Toba et al. [71] achieved an accuracy of 86% and AUC 0.88 in predicting the pulmonary-to-systemic blood flow ratio based on chest radiographs. These performance metrics were equal to or higher than three pediatric cardiologists' predictions. Predicting Brasfield chest radiographic score in children with cystic fibrosis is another AI use case that has been shown to achieve similar performance to radiologist scores [72]. A combination of an artificial neural network and texture analysis was performed by Clark et al. [76] to classify pediatric chest radiographs of children with cystic fibrosis. A strength of the study is its comparison to CT images as a gold standard, but a substantial limitation — both for training and testing the algorithm — is the small dataset.

Artificial intelligence for catheter and tube assessment — particularly for neonatal radiographs — is another important yet underdeveloped area for AI in pediatric imaging. Various techniques, including image segmentation using deep learning [73] and image classification based on a CNN [80],

are used to detect and classify tubes and catheters. AI has also been used extensively for lung segmentation, both as a pre-processing step to increase the accuracy of other classification models focused on lung pathology, as well as a method of estimating lung capacity. Mansoor et al. [74] developed a deep-learning-based method of lung segmentation generalizable to both adult and pediatric chest radiographs for the purpose of identifying lung fields. Their algorithm achieved Dice similarity coefficients, which measure the overlap between the ground truth and model segmented regions, ranging from 0.969 to 0.975. Other pediatric chest radiograph datasets, including the GWCMC dataset and a pediatric chest radiograph dataset from the University of Michigan, have also been used in training and testing algorithms for lung segmentation [66, 83]. It has been shown that applying age-specific models (using age groupings of 0–23 months, 2–10 years and 11–18 years), instead of adopting a model for all ages, improves algorithm performance [84].

Beyond image interpretation, efforts have focused on applications of AI that might be useful in PACS. Hržić et al. [77] used a transfer-learning approach based on ResNet, VGG16, and AlexNet to correct the orientation and alignment of radiographic images, achieving 99.3% accuracy. Kim et al. [69] performed classification of anteroposterior and posteroanterior chest radiographs by training a CNN. The CNN that was trained on both adult and pediatric chest radiographs achieved AUCs ranging from 0.985 to 0.999 when tested on three pediatric datasets — two of which were external. The CNN trained using only pediatric chest radiographs achieved an AUC of only 0.997, possibly because of the substantially smaller training dataset on which the CNN was trained.

Discussion

Chest radiography is one of the most frequently performed imaging tests, being easy to perform, low in cost and readily available even in low-resource settings. It provides important information regarding pediatric thoracic pathology; however, chest radiographs can be challenging to interpret, with significant variability and discrepancy in findings often noted. Machine learning, especially deep learning, is a rapidly developing field with great promise for medical imaging, offering the potential for enhanced accuracy, faster diagnoses and standardized interpretations. Specific AI applications for chest radiography have rapidly proliferated within the last 5 years, with several commercial products for abnormality detection, including detection of lung nodules and cancer, pneumonia and pneumothorax, among others now available

[3]. Despite this activity, there has been limited application of machine learning to pediatric chest radiography.

The availability of appropriate pediatric image datasets is critical for training, testing and validating AI algorithms in pediatric radiology. Indeed, the performance of any model is dependent on the dataset used for training. Our systematic review found a paucity of gold standard open-source datasets for pediatric ML training and testing. We found a limited number of pediatric-specific datasets with an overall quite limited number and age-limited pediatric images available. The development and public release of the GWCMC dataset [14] can be considered a turning point that has brought AI researchers' attention to the pediatric population. The GWCMC dataset has resulted in the generation of multiple studies predominantly exploring the diagnosis of pediatric pneumonia using deep learning. However, several studies had been conducted on pediatrics before the release of this dataset; most of these studies utilized traditional machine learning techniques and generally small size and proprietary datasets. In the GWCMC dataset, the ground truth was established using only radiologic features, without using clinical and paraclinical information. On the other hand, the PERCH chest radiograph dataset has strength in this area and includes infants, with well-defined patient inclusion criteria, multi-institutional data sources and associated clinical and paraclinical information [15, 16]. It is hoped that making this dataset available for AI research and development will mark another milestone in pediatric imaging AI development. There is a need for additional open-source pediatric-specific datasets that include a broader range of children across different age ranges and across additional pediatric conditions and diseases.

Important considerations for datasets include whether they were retrospectively or prospectively collected, the reference standards used, and how labels and annotations were obtained. For example, labels in the NIH training set were obtained by natural language processing, which is considered to represent weak or noisy supervision in the training process. Labeling images based on NLP from radiology reports is intrinsically associated with uncertainty [85] because radiologists might mention a disease only as a possibility, not as a definitive diagnosis (Fig. 3). The assembled datasets sometimes have poorly defined criteria for the presence or absence of the disease and might not reflect the uncertainty of radiologists' diagnoses for some cases. It can be costly and time-consuming to create large-scale, richly labeled datasets because this generally requires extensive effort by experts. Appropriate datasets are not yet available for many diseases, especially for the pediatric population. Only a small pool of pediatric imagers is available, making obtaining these datasets a multi-institutional challenge. The use of transfer learning might assist in addressing the small

size of available pediatric datasets and help reduce computation time for training models.

Our literature review on AI for pediatric chest radiograph interpretation found pneumonia detection to be by far the most common task for which AI algorithms have been developed. This is a particularly important use case in pediatric radiology, given that pneumonia is one of the leading causes of morbidity and mortality in children, accounting for an estimated 900,000 deaths worldwide annually [86, 87]. The diagnosis of pneumonia is essentially clinical, and radiographic findings such as consolidation, infiltration or pleural effusion only support the diagnosis [88]. However, many images of the large-scale public chest radiograph datasets have been labeled with just "pneumonia" [17, 18, 85]. When labeling images with clinical diagnoses rather than radiographic observations, it is important to ensure that both radiographic observations and clinical information are incorporated when images are labeled, or that the images are obtained from a population of children with a high pre-test probability of having that diagnosis. In addition, many algorithms in our review were trained only by the image features present without incorporating any clinical information as inputs. The drawback of this approach is that the AI algorithm might provide a clinical diagnosis that does not fit with the child's clinical presentation as an output, even if it is consistent with the imaging findings. Therefore, we recommend that researchers consider the distinction between *radiographic observations* and *clinical diagnoses* when labeling images in future chest radiograph dataset construction.

The challenges of developing pediatric AI models are not just limited to the lack of appropriate datasets. Recognition of what is considered normal is defined differently in the various datasets identified in this review. The chest wall and lungs, ossification centers and cardiothymic width grow and change during normal development. This makes some tasks such as lung segmentation more difficult for AI models, and separate models for specific age groups might be required. Children with medical devices, including a variety of chest tubes and catheters and pacemakers, are often included within the group called normal. In a scenario that a chest radiograph demonstrates catheters, tubes or other medical devices, the model might easily find the device as an abnormality, leading to a high false-positive rate. Care must be taken to mitigate the bias that this can introduce into the results.

Another challenge is the need for the integration of clinical data. For example, neonatal pneumonia and meconium aspiration syndrome can have similar radiographic observations, and only knowledge of a specific patient's clinical data makes differentiation possible [89]. An additional concern is a fact that specific pediatric age groups have differing disease predispositions. For example, transient tachypnea of the newborn is diagnosed only in neonates [90]. Studies have demonstrated that AI models might show false positives in cases with insufficient inflation, inappropriate positioning, non-standard exposure, clothing, or external or implanted medical devices [5, 91]. These challenges can be particularly problematic in pediatric populations because providing optimum conditions for taking high-quality images in pediatrics can be difficult [72]. Different radiologists have varying levels of sensitivity and specificity, which they alter to fit the clinical situation. In the diagnosis of lung nodules in children with a known malignancy, for example, high sensitivity is essential, but sensitivity can be reduced to boost specificity in the identification of incidental pulmonary nodules in asymptomatic individuals. In children with established immunodeficiency, high sensitivity is essential, while a greater specificity is preferred in the case of community-acquired respiratory infections and fevers of unknown origin. Similarly, different operating points could be defined for AI algorithms based on clinical information, though this work has not been undertaken.

Deep learning models in radiology frequently suffer from generalization issues because of large source and target domain divergence and sometimes need fine-tuning for target domains, including patient age, as in pediatrics. For example, Tang et al. [54] showed that there is a significant domain shift between adult and pediatric datasets, as demonstrated by substantially higher AUC values for a model trained on a pediatric dataset compared to the same CNN architecture trained on an adult dataset when both models were tested on a pediatric dataset. Transfer learning can learn the common characteristics of both domains, leading to a better initialization of the model parameters and faster training.

Some reporting guidelines have recently proposed specific AI extensions to develop and report AI-related studies, including the Standards for Reporting of Diagnostic Accuracy Studies (STARD) [92], Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) [93] and Checklist for Artificial Intelligence in Medical Imaging (CLAIM) [94]. These guidelines provide enhanced transparency and reproducibility and help readers assess the generalizability of results. Many studies included in this systematic review lack at least some of the reporting items recommended in these guidelines. In this review we found a wide variety of machine learning architectures used with significant study heterogeneity in methodologies, sizes of datasets for training and testing, validation techniques, reference standards, terminology and reporting, which is not surprising given the time period covered by this review and the recent technological progress. Many studies reported high diagnostic accuracy; however, there was often poor study design, which can lead to bias and overestimation of the accuracy of these algorithms. Some of these

study limitations are noted in the tables provided in Online Supplementary Material. For example, many of the studies summarized here did not report their performance indices' average and standard deviation. Furthermore, according to the prevalence of abnormalities, some indices like AUC, F1-score and Matthews correlation coefficient (MCC) [95] should be reported in addition to accuracy but are often not reported. Clinical trials involving AI and DL interventions with rigorous standardized reporting guidelines are expected to be valuable in determining the future clinical implications of algorithms in real life [54]. According to the STARD checklist, confidence intervals give information about a range within which the real value is likely to be found, as well as the direction and strength of the observed effect. This allows for conclusions to be formed regarding the study's statistical validity and clinical usefulness. In addition, full access to the dataset and code is unavailable for the vast majority of AI studies, which limits the reproducibility of AI research that has been published [95]. These limitations in pediatric AI research make progress in this domain more challenging.

The performance of DL models as currently reported in the literature might not reflect the actual performance of the models in clinical practice. For example, despite recommendations that an external test dataset be used, almost all of the studies identified in this review used the same dataset for training, testing and validation. Algorithm performance might be significantly different when used in clinical settings with different patient populations, imaging equipment and acquisition parameters. Further, the performance of most algorithms included in our review was not compared with the performance of radiologists.

In radiology, AI products can be used for triage or alongside a radiologist as a second reader [96]. While commercial software is increasingly becoming available for adult imaging, our review highlights that progress for pediatric chest radiographs interpretation lags. Many commercially available United States Food and Drug Administration (FDA) approved/cleared tools are not at this time applicable to pediatric populations [97, 98]. In addition to common use cases such as pneumonia detection, there are a number of distinct use cases for AI development in pediatric imaging, such as neonatal catheter and tube assessment and cystic fibrosis scoring. How best to integrate the output of AI into clinical practice and workflow remain unanswered questions.

There are many opportunities to develop AI for pediatric imaging. Use cases range from correctly orienting images in PACS, flagging abnormal images to prioritize them for interpretation by a radiologist (particularly for life-threatening conditions like pneumothorax), automatic labeling of images to support the creation of annotated datasets for development of other AI algorithms, automatically detecting radiographic features ("lung sequencing") and localizing

potential abnormalities in the various lobes, performing automatic scoring of diffuse lung abnormalities (such as for cystic fibrosis) and providing preliminary interpretations to referring clinicians in resource-limited settings where many radiographs go unreported.

This systematic review has several limitations. First, our literature search strategy did not encompass conference papers or articles that were only included in open-access repositories such as arXiv, resulting in some pediatric AI studies likely being missed. Second, while we performed a systematic review, we did not perform a meta-analysis as part of this study. This is a result of the rapid evolution of AI approaches over the last decade, the fact that most studies did not assess AI algorithms on external datasets, and the substantial heterogeneity of the studies included (including heterogeneity in terms of patient population, classification approaches used and classification tasks). Third, while we considered the strengths and limitations of each study included in the systematic review, we did not use a validated checklist to assess study quality. As AI in pediatric chest radiographs interpretation continues to advance and a growing body of clinical evidence is established, formally assessing study quality will be increasingly important as AI is considered for clinical use in pediatric radiology.

Conclusion

In this article, we presented a systematic review of the application of AI to pediatric chest radiograph interpretation. Although ML in pediatric radiology holds great promise, it clearly remains in its infancy and has a long way to go before adding value to the pediatric radiology workflow. Designing appropriate AI models requires significant work with reliable features derived from diagnostic images. Radiologists with knowledge of AI bring ideas to solve problems and act as a link between radiology and biomedical engineering. For AI to progress in pediatric radiology, they suggest improving the awareness of radiologists about AI science, providing better publicly available image datasets, and creating developmental research to design AI models and validate AI's utility and reliability in the clinical workflow.

Declarations

Conflicts of interest None

References

1. Arthur R (2001) The neonatal chest X-ray. Paediatr Respir Rev 2:311–323
2. Arthur R (2003) Interpretation of the paediatric chest X-ray. Curr Paediatr 13:438–447
3. Benjamens S, Dhunnoo P, Meskó B (2020) The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 3:118
4. Çallı E, Sogancioglu E, van Ginneken B et al (2021) Deep learning for chest X-ray analysis: a survey. Med Image Anal 72:102125
5. Kao EF, Liu GC, Lee LY et al (2015) Computer-aided detection system for chest radiography: reducing report turnaround times of examinations with abnormalities. Acta Radiol 56:696–701
6. Rezazade Mehrizi MH, van Ooijen P, Homan M (2020) Applications of artificial intelligence (AI) in diagnostic radiology: a technography study. Eur Radiol 31:1805–1811
7. Qin C, Yao D, Shi Y, Song Z (2018) Computer-aided detection in chest radiography based on artificial intelligence: a survey. Biomed Eng Online 17:1–23
8. Moore MM, Slonimsky E, Long AD et al (2019) Machine learning concepts, concerns and opportunities for a pediatric radiologist. Pediatr Radiol 49:509–516
9. Willemink MJ, Koszek WA, Hardell C et al (2020) Preparing medical imaging data for machine learning. Radiology 295:4–15
10. Selvaraju RR, Cogswell M, Das A et al (2020) Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis 128:336–359
11. Lundberg S, Lee S-I (2017) A unified approach to interpreting model predictions. arXiv:1705.07874
12. McInnes MDF, Moher D, Thombs BD et al (2018) Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA 319:388–396
13. World Health Organization (n.d.) Adolescence: a period needing special attention. Recognizing adolescence. WHO website. https://apps.who.int/adolescent/second-decade/section2/page1/recognizing-adolescence.html. Accessed 25 Jan 2022
14. Kermany DS, Goldbaum M, Cai W et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172:1122–1131.e9
15. Pneumonia Etiology Research for Child Health (PERCH) Study Group (2019) Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. Lancet 394:755–779
16. Fancourt N, Knoll MD, Baggett HC et al (2017) Chest radiograph findings in childhood pneumonia cases from the multisite PERCH study. Clin Infect Dis 64:S262–S270
17. Bustos A, Pertusa A, Salinas JM, de la Iglesia-Vayá M (2020) PadChest: a large chest X-ray image dataset with multi-label annotated reports. Med Image Anal 66:101797
18. Wang X, Peng Y, Lu L et al (2017) ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Piscataway
19. Radiological Society of North America (2018) RSNA pneumonia detection challenge (2018). https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/RSNA-Pneumonia-Detection-Challenge-2018. Accessed 26 Feb 2022
20. Pan I, Cadrin-Chênevert A, Cheng PM (2019) Tackling the Radiological Society of North America pneumonia detection challenge. AJR Am J Roentgenol 213:568–574
21. Society for Imaging Informatics in Medicine (n.d.) SIIM–ACR pneumothorax segmentation. Kaggle website. https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation/overview/description. Accessed 3 Jun 2021
22. Society for Imaging Informatics in Medicine (n.d.) The pneumothorax challenge. SIIM website. https://siim.org/page/pneumothorax_challenge. Accessed 21 Jun 2021
23. National Institute of Allergy and Infectious Diseases (2022) Tuberculosis (TB) data science for public health impact. National Institutes of Health website. https://tbportals.niaid.nih.gov/. Accessed 7 Jun 2021
24. Jaeger S, Candemir S, Antani S et al (2014) Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med Surg 4:475–477
25. Fancourt N, Knoll MD, Barger-Kamate B et al (2017) Standardized interpretation of chest radiographs in cases of pediatric pneumonia from the PERCH study. Clin Infect Dis 64:S253–S261
26. Chen Y, Roberts CS, Ou W et al (2021) Deep learning for classification of pediatric chest radiographs by WHO's standardized methodology. PLoS One 16:e0253239
27. Shih G, Wu CC, Halabi SS et al (2019) Augmenting the National Institutes of Health chest radiograph dataset with expert annotations of possible pneumonia. Radiol Artif Intell 1:e180041
28. Tolkachev A, Sirazitdinov I, Kholiavchenko M et al (2021) Deep learning for diagnosis and segmentation of pneumothorax: the results on the Kaggle competition and validation against radiologists. IEEE J Biomed Health Inform 25:1660–1672
29. Oliveira LLG, E Silva SA, Ribeiro LHV et al (2008) Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. Int J Med Inform 77:555–564
30. Mahomed N, van Ginneken B, Philipsen RHHM et al (2020) Computer-aided diagnosis for World Health Organization–defined chest radiograph primary-endpoint pneumonia in children. Pediatr Radiol 50:482–491
31. Sousa RT, Marques O, Soares FAAMN et al (2013) Comparative performance analysis of machine learning classifiers in detection of childhood pneumonia using chest radiographs. Procedia Comput Sci 18:2579–2582
32. Longjiang E, Zhao B, Liu H et al (2020) Image-based deep learning in diagnosing the etiology of pneumonia on pediatric chest X-rays. Pediatr Pulmonol 56:1036–1044
33. Rajaraman S, Candemir S, Kim I et al (2018) Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. Appl Sci 8:1715
34. Rahman T, Chowdhury MEH, Khandakar A et al (2020) Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. Appl Sci 10:3233
35. Ferreira JR, Armando Cardona Cardenas D, Moreno RA et al (2020) Multi-view ensemble convolutional neural network to improve classification of pneumonia in low contrast chest X-ray images. Annu Int Conf IEEE Eng Med Biol Soc 2020:1238–1241
36. Costa NJ, Sousa JVM, Santos DB et al (2020) Classification of X-ray images for detection of childhood pneumonia using pre-trained neural networks. Brazilian J Appl Comput 12:132–141
37. Nguyen HT, Tran TB, Luong HH et al (2020) Viral and bacterial pneumonia diagnosis via deep learning techniques and model explainability. Int J Adv Comput Sci Appl 11:667–675
38. Hu M, Lin H, Fan Z et al (2020) Learning to recognize chest-Xray images faster and more efficiently based on multi-kernel depthwise convolution. IEEE Access 8:37265–37274
39. Masud M, Bairagi AK, Nahid A et al (2021) A pneumonia diagnosis scheme based on hybrid features extracted from chest radiographs using an ensemble learning algorithm. J Healthc Eng 2021:8862089

40. Alqudah AM, Qazan S, Masad IS (2021) Artificial intelligence framework for efficient detection and classification of pneumonia using chest radiography images. J Med Biol Eng 41:599–609

41. Stephen O, Sain M, Maduh UJ, Jeong D-U (2019) An efficient deep learning approach to pneumonia classification in healthcare. J Healthc Eng 2019:4180949

42. Chakraborty P, Tharini C (2020) Pneumonia and eye disease detection using convolutional neural networks. Eng Technol Appl Sci Res 10:5769–5774

43. Jain R, Nagrath P, Kataria G et al (2020) Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning. Measurement 165:108046

44. Hidayatullah RC (2020) Convolutional neural network architecture and data augmentation for pneumonia classification from chest X-rays images. Int J Innov Sci Res Technol 5:158–164

45. Chouhan V, Singh SK, Khamparia A et al (2020) A novel transfer learning based approach for pneumonia detection in chest X-ray images. Appl Sci 10

46. Hashmi MF, Katiyar S, Keskar AG et al (2020) Efficient pneumonia detection in chest Xray images using deep transfer learning. Diagnostics 10:1–23

47. Habib N, Hasan MM, Reza MM, Rahman MM (2020) Ensemble of CheXNet and VGG-19 feature extractor with random forest classifier for pediatric pneumonia detection. SN Comput Sci 1:1–9

48. Mittal A, Kumar D, Mittal M et al (2020) Detecting pneumonia using convolutions and dynamic capsule routing for chest X-ray images. Sensors 20:1–30

49. Luján-García JE, Yáñez-Márquez C, Villuendas-Rey Y, Camacho-Nieto O (2020) A transfer learning method for pneumonia classification and visualization. Appl Sci 10

50. Yue Z, Ma L, Zhang R (2020) Comparison and validation of deep learning models for the diagnosis of pneumonia. Comput Intell Neurosci 2020:8876798

51. Shah U, Abd-Alrazeq A, Alam T et al (2020) An efficient method to predict pneumonia from chest X-rays using deep learning approach. Stud Health Technol Inform 272:457–460

52. Al Nahid A, Sikder N, Bairagi AK et al (2020) A novel method to identify pneumonia through analyzing chest radiographs employing a multichannel convolutional neural network. Sensors 20:1–18

53. Liang G, Zheng L (2020) A transfer learning method with deep residual network for pediatric pneumonia diagnosis. Comput Methods Prog Biomed 187:104964

54. Tang Y-X, Tang Y-B, Peng Y et al (2020) Automated abnormality classification of chest radiographs using deep convolutional neural networks. NPJ Digit Med 3:70

55. Qu W, Balki I, Mendez M et al (2020) Assessing and mitigating the effects of class imbalance in machine learning with application to X-ray imaging. Int J Comput Assist Radiol Surg 15:2041–2048

56. Akgundogdu A (2020) Detection of pneumonia in chest X-ray images by using 2D discrete wavelet feature extraction with random forest. Int J Imaging Syst Technol 31:82–93

57. Dey N, Zhang Y, Rajinikanth V et al (2021) Customized VGG19 architecture for pneumonia detection in chest X-rays. Pattern Recogn Lett 143:67–74

58. das Chagas, JVS, de A Rodrigues D, Ivo RF et al (2021) A new approach for the detection of pneumonia in children using CXR images based on an real-time IoT system. J Real Time Image Process 18:1099–1114

59. Gm H, Gourisaria MK, Rautaray SS, Pandey M (2021) Pneumonia detection using CNN through chest X-ray. J Eng Sci Technol 16:861–876

60. Yu X, Wang S, Zhang Y (2021) CGNet: a graph-knowledge embedded convolutional neural network for detection of pneumonia. Inf Process Manag 58:102411

61. Eid MM, Elawady YH (2021) Efficient pneumonia detection for chest radiography using ResNet-based SVM. Eur J Electr Eng Comput Sci 5:1–8

62. Salehi M, Mohammadi R, Ghaffari H et al (2021) Automated detection of pneumonia cases using deep transfer learning with paediatric chest X-ray images. Br J Radiol 94:20201263

63. Morid MA, Borjali A, Del Fiol G (2021) A scoping review of transfer learning research on medical image analysis using ImageNet. Comput Biol Med 128:104115

64. Gross GW, Boone JM, Greco-Hunt V, Greenberg B (1990) Neural networks in radiologic diagnosis. II. Interpretation of neonatal chest radiographs. Investig Radiol 25:1017–1023

65. Zaglam N, Jouvet P, Flechelles O et al (2014) Computer-aided diagnosis system for the acute respiratory distress syndrome from chest radiographs. Comput Biol Med 52:41–48

66. Chen KC, Yu HR, Chen WS et al (2020) Diagnosis of common pulmonary diseases in children by X-ray images and deep learning. Sci Rep 10:1–9

67. Radwan AA, El-Bakry HM, El Hadad HM (2011) A new expert system for pediatric respiratory diseases by using neural networks. Int J Comput Sci Inform Secur Sci Inf Secur 9:11–17

68. Candemir S, Antani S, Jaeger S et al (2015) Lung boundary detection in pediatric chest X-rays. Proc SPIE 9418

69. Kim TK, Yi PH, Wei J et al (2019) Deep learning method for automated classification of anteroposterior and posteroanterior chest radiographs. J Digit Imaging 32:925–930

70. Longjiang E, Zhao B, Guo Y et al (2019) Using deep-learning techniques for pulmonary-thoracic segmentations and improvement of pneumonia diagnosis in pediatric chest radiographs. Pediatr Pulmonol 54:1617–1626

71. Toba S, Mitani Y, Yodoya N et al (2020) Prediction of pulmonary to systemic flow ratio in patients with congenital heart disease using deep learning-based analysis of chest radiographs. JAMA Cardiol 5:449–457

72. Zucker EJ, Barnes ZA, Lungren MP et al (2020) Deep learning to automate Brasfield chest radiographic scoring for cystic fibrosis. J Cyst Fibros 19:131–138

73. Yi X, Adams S, Babyn P, Elnajmi A (2020) Automatic catheter and tube detection in pediatric X-ray images using a scale-recurrent network and synthetic data. J Digit Imaging 33:181–190

74. Mansoor A, Cerrolaza JJ, Perez G et al (2020) A generic approach to lung field segmentation from chest radiographs using deep space and shape learning. IEEE Trans Biomed Eng 67:1206–1220

75. Reamaroon N, Sjoding MW, Derksen H et al (2020) Robust segmentation of lung in chest X-ray: applications in analysis of acute respiratory distress syndrome. BMC Med Imaging 20:1–13

76. Clark AR, Her EJ, Metcalfe R, Bymes CA (2021) Could automated analysis of chest X-rays detect early bronchiectasis in children? Eur J Pediatr 180:3171–3179

77. Hržić F, Tschauner S, Sorantin E, Štajduhar I (2021) XAOM: a method for automatic alignment and orientation of radiographs for computer-aided medical diagnosis. Comput Biol Med 132:104300

78. Liz H, Sánchez-Montañés M, Tagarro A et al (2021) Ensembles of convolutional neural network models for pediatric pneumonia diagnosis. Futur Gener Comput Syst 122:220–233

79. Henderson RDE, Yi X, Adams SJ, Babyn P (2021) Automatic classification of multiple catheters in neonatal radiographs with deep learning. J Digit Imaging 34:888–897

80. Behzadi-Khormouji H, Rostami H, Salehi S et al (2020) Deep learning, reusable and problem-based architectures for detection of consolidation on chest X-ray images. Comput Methods Prog Biomed 185:105162

81. Xi Y, Zhong L, Xie W et al (2021) View identification assisted fully convolutional network for lung field segmentation of frontal and lateral chest radiographs. IEEE Access 9:59835–59847

82. Yi PH, Wei J, Kim TK et al (2021) Radiology "forensics": determination of age and sex from chest radiographs using deep learning. Emerg Radiol 28:949–954

83. Yi X, Adams SJ, Henderson RDE, Babyn P (2020) Computer-aided assessment of catheters and tubes on radiographs: how good is artificial intelligence for assessment? Radiol Artif Intell 2:e190082

84. Sagi O, Rokach L (2018) Ensemble learning: a survey. Wiley Interdiscip Rev Data Min Knowl Discov 8:1–18

85. Irvin J, Rajpurkar P, Ko M et al (n.d.) CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. Stanford ML Group. https://stanfordmlgroup.github.io/competitions/chexpert/. Accessed 24 Sep 2020

86. Black RE, Cousens S, Johnson HL et al (2010) Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet 375:1969–1987

87. Fischer Walker CL, Rudan I, Liu L et al (2013) Global burden of childhood pneumonia and diarrhoea. Lancet 381:1405–1416

88. World Health Organization (2001) Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children/World Health Organization pneumonia vaccine trial investigators' group. World Health Organization, Geneva

89. Jana M, Bhalla AS, Gupta AK (2016) Approach to pediatric chest radiograph. Indian J Pediatr 83:533–542

90. Gallacher DJ, Hart K, Kotecha S (2016) Common respiratory conditions of the newborn 12:30–42

91. Singh R, Kalra MK, Nitiwarangkul C et al (2018) Deep learning in chest radiography: detection of findings and presence of change. PLoS One 13:1–12

92. Cohen JF, Korevaar DA, Altman DG et al (2016) STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 6:1–17

93. Collins GS, Reitsma JB, Altman DG, Moons KGM (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). The TRIPOD statement. Circulation 131:211–219

94. Mongan J, Moy L, Kahn CE (2020) Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2:e200029

95. Boughorbel S, Jarray F, El-Anbari M (2017) Optimal classifier for imbalanced data using Matthews correlation coefficient metric. PLoS One 12:1–17

96. Tang A, Tam R, Cadrin-Chênevert A et al (2018) Canadian Association of Radiologists white paper on artificial intelligence in radiology. Can Assoc Radiol J 69:120–135

97. Adams SJ, Henderson RDE, Yi X, Babyn P (2021) Artificial intelligence solutions for analysis of X-ray images. Can Assoc Radiol J 72:60–72

98. van Leeuwen KG, Schalekamp S, Rutten MJCM et al (2021) Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol 31:3797–3804