This is the published version of the paper.

This version of the publication may differ from the final published version.

Efficacy of Artificial Intelligence in the Categorisation of Paediatric Pneumonia on Chest Radiographs: A Systematic Review

Erica Louise Field [1], Winnie Tam [2,*], Niamh Moore [1] and Mark McEntee [1]

1   Discipline of Medical Imaging and Radiation Therapy, University College Cork, College Road, T12 K8AF Cork, Ireland
2   Department of Midwifery and Radiography, University of London, Northampton Square, London EC1V 0HB, UK
*   Correspondence: winnie.tam@city.ac.uk

Abstract: This study aimed to systematically review the literature to synthesise and summarise the evidence surrounding the efficacy of artificial intelligence (AI) in classifying paediatric pneumonia on chest radiographs (CXRs). Following the initial search of studies that matched the pre-set criteria, their data were extracted using a data extraction tool, and the included studies were assessed via critical appraisal tools and risk of bias. Results were accumulated, and outcome measures analysed included sensitivity, specificity, accuracy, and area under the curve (AUC). Five studies met the inclusion criteria. The highest sensitivity was by an ensemble AI algorithm (96.3%). DenseNet201 obtained the highest level of specificity and accuracy (94%, 95%). The most outstanding AUC value was achieved by the VGG16 algorithm (96.2%). Some of the AI models achieved close to 100% diagnostic accuracy. To assess the efficacy of AI in a clinical setting, these AI models should be compared to that of radiologists. The included and evaluated AI algorithms showed promising results. These algorithms can potentially ease and speed up diagnosis once the studies are replicated and their performances are assessed in clinical settings, potentially saving millions of lives.

Keywords: artificial intelligence (AI); deep learning (DL); paediatric pneumonia; chest radiograph; computer-aided detection (CAD)

1. Introduction

Pneumonia is one of the leading causes of global mortality and morbidity [1] and the leading cause of death among children under five years of age [2]. The chest X-ray (CXR) is the primary diagnostic tool in both the detection and diagnosis of paediatric pneumonia [1] due to the unspecific and subjective signs and symptoms of the infection [3], while sputum cultures are often extremely difficult to ascertain [4]. A lack of expert radiologists, particularly in resource-constrained countries, where paediatric pneumonia is endemic with shockingly high mortality rates [5], might have significantly contributed to the high mortality rates.

Pneumonia, as a whole, often manifests on a chest radiograph as areas of increased opacity [6]. Bacterial and viral pneumonia, the two most common aetiologies, have different appearances on CXRs and have different treatment regimes [7]. Bacterial pneumonia often manifests as a lobar and focal consolidation, whereas viral pneumonia can present as an interstitial pattern. Although the different appearances on CXRs, interpretation of the aetiologies on CXR varies amongst physicians [3] as the opacification presented is often variable and irregular [8].

There are roughly two billion CXRs performed in the United States annually [9], with approximately two million of these being on paediatric patients [10]. The accumulation of imaging data and the increasing complexity of medical history pose new challenges in

modern medicine but also open up new opportunities in implementing artificial intelligence (AI) for effective detection and diagnosis [3].

AI is defined, by the father of AI, Marvin Minsky, as "the science of making machines do things that would require intelligence if done by men" [11]. Machine learning (ML) and deep learning (DL) are both under the umbrella of AI. In ML, the system identifies patterns by learning from data and makes decisions with minimal programming and human interventions [12,13]. On the other hand, DL involves multiple processing layers, with individual layers extracting a large number of features from unstructured data before progressing to the next layer [14]. Higher levels typically represent more abstract concepts [15] and can give a more comprehensive depiction or decision after passing through the entire network [14].

Computer-aided detection (CAD) was first introduced between 1963–1973 [16–21], and utilised in facets of radiology, predominantly the identification of lung, colorectal, breast, and prostate cancer, over the past 20 years [22–33]. CAD is trained in a regimental fashion and can only be improved by inputting more data, while AI has the autonomous learning element in which explicit programming and human instructions are not necessary for its improvement [34].

AI interpretation of medical images can also potentially be utilised in low- to middle-income countries, where the availability of radiologists is limited. Currently, interpretations rely on teleradiology [35], though it is not flawless. Cross-border teleradiology, in particular, is challenging due to liability in case of malpractice, healthcare professional registration restrictions, data protection, quality of the reporting, healthcare system, and cultural differences [36]. This is where AI can potentially come in to make the diagnostic and treatment pathway of patients smoother, especially in these developing nations [35].

Recent studies exhibited that AI algorithms outperformed radiologists in the detection of skin cancer [37], diabetic retinopathy [38], and haemorrhage identification [39], probably owing to the recent AI model advancement [40] and widening availability of electronic health record, providing more training materials. The winner of the Turning Prize said in 2016, "We should stop training radiologists now. It's just completely obvious that within five years, DL is going to do better than radiologists" [41].

These successes have sparked interest in the automated diagnosis of paediatric pneumonia in CXRs, reflected by the increased number of studies regarding the diagnostic accuracy of AI for paediatric pneumonia over the last number of years. When being asked about AI's role in diagnosing pneumonia, Andrew Ng, the co-founder and head of Google Brain, went further—"radiologists should be worried about their jobs" [42]. A recent study by Kermany et al. [43] showed that a customised AI model demonstrated a good level of classification of paediatric pneumonia on CXR. By systematically reviewing the current literature regarding AI and paediatric pneumonia, the current gap in literature may be filled and potentially result in a dramatic improvement of accuracy and efficiency in differentiating bacterial and viral pneumonia in paediatric CXRs in a clinical setting [44].

2. Materials and Methods

2.1. Literature Search Strategy

A comprehensive search of the literature was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [45] in February 2021. A number of pre-determined keywords were pooled for this systematic search, and subsequent medical subject headings (MeSH terms) were generated. The MeSH terms used for this systematic search include: ('artificial intelligence' OR 'deep learning' OR 'CNN' OR 'convolutional neural network' OR 'deep residual network'), ('paediatric pneumonia' OR 'pediatric pneumonia' OR 'child* pneumonia'), 'classification', ('chest xray' OR 'chest x ray' OR 'chest x-ray' OR 'CXR' OR 'chest radiograph'). The pre-determined MeSH terms were then linked using Boolean operators specific to each database (PubMed, Science Direct, Embase, ProQuest, and Scopus) to retrieve all relevant articles evaluating the diagnostic efficacy of AI models in the classification of paediatric pneumonia in CXRs. These databases were chosen as opposed to others due to their scientific reliability

and coverage. Google Scholar and the reference lists of relevant studies were also screened to ensure as much grey literature was captured as possible. In addition to these criteria, only peer-reviewed articles published in English were included in this review.

2.2. Study Selection and Eligibility Criteria

The initial selection involved screening the paper title and abstract. After gathering all of the potential papers, a full-text assessment was undertaken. Studies were included if they met the following inclusion criteria: only original cross-sectional studies, cohort or case-control studies, randomised control trials (RCTs), or diagnostic accuracy studies in the format of either journal articles, dissertations, conference proceedings, or grey literature, disseminated between 2015–2021, were included. In addition, the paper must analyse/evaluate the AI performances on the classification of pneumonia aetiology (bacterial or viral) using CXR test datasets of children under the age of 16 years in acute healthcare settings. These studies should also evaluate the AI performance using at least three of the following parameters: accuracy, sensitivity, specificity, and area under the curve (AUC). Studies which did not meet one of the inclusion criteria, or inaccessible full-text articles, were excluded.

2.3. Data Extraction

Data extraction was performed on included studies using the amended version of the Cochrane 'Data collection form for Intervention Reviews for RCTs and Non-RCTS— Template' [46]. A general overview of the data extracted is as follows: general information (report title, study ID, date form completed, etc.), study eligibility, characteristics of study (the aim of the study, design, etc.), participants (description of images from dataset, setting, age, etc.), AI model (AI type, mode, detail, etc.), outcomes (sensitivity, specificity, accuracy, AUC), and other information (key conclusions, future work, etc.). Each data extraction form was reviewed by two independent reviewers, and any discrepancies were resolved by consensus. The original data extraction form applied to each included study can be obtained from the author on request.

2.4. Quality Assessment and Risk of Bias

A quality assessment of each included study was examined using the Critical Appraisal Skills Programme (CASP) Diagnostic Study Checklist, which focuses on the result validity, measuring parameters and transparency, and generalisability [47]. A risk of bias (ROB) assessment was also completed for each included study. The tool utilised for the ROB was adjusted from Hung et al.'s QUADAS-2 tool, whose systematic review investigated the use and performance of AI applications in the maxillofacial and dental radiology [48]. This adapted ROB consists of four key domains: patient selection, index test, reference standard, and study flow and timing with regard to both applicability and general ROB. Each domain was assessed using a three-point scale, low (green), high (red), or unclear (yellow), to reflect the level of bias concerns accordingly. All CASP and ROB checklists were reviewed by two independent reviewers, and any discrepancies were resolved by consensus.

3. Results

3.1. Study Selection

A total of 114 papers were identified following the initial systematic search from the six databases previously mentioned, combined with subsequent manual searches of reference lists based on the relevance of their title to the research question. After the removal of duplicates, 82 papers were considered for abstract screening, and 24 out of 82 were considered suitable and underwent full-text assessment. Of these 24 papers, 19 did not meet the required inclusion criteria due to a broad spectrum of reasons, such as the study did not classify pneumonia subtypes, the study did not include the desired (number of) outcome measures, and the dataset used was inapplicable. Five papers were included at last. The PRISMA flowchart exhibiting the study eligibility and selection

process is presented in Figure 1. The two independent reviewers agreed with the initial search and study selection/eligibility process, with no discrepancies found.
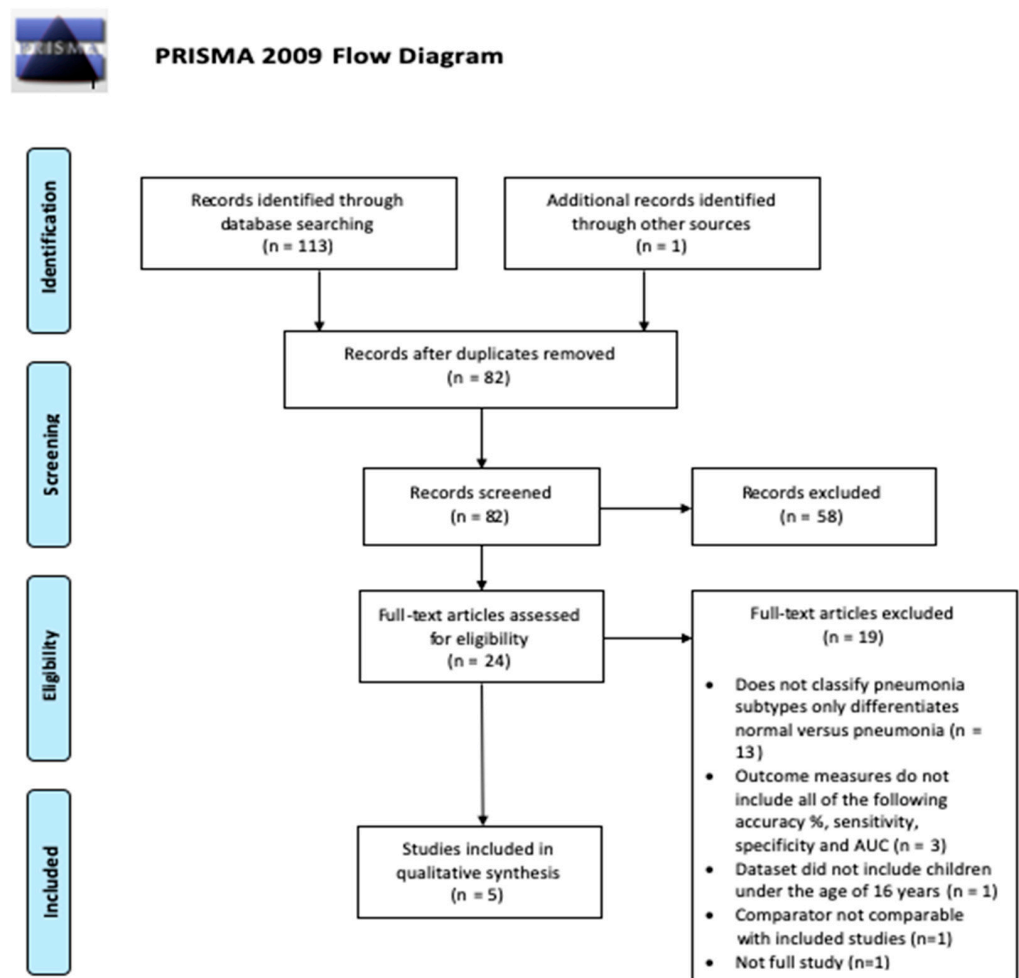

Figure 1. A PRISMA flow chart illustrating the filtering and gathering of eligible studies.

3.2. Study Quality Assessment

The quality assessment of each included study can be seen in Figure 2. All five papers were considered as having a "low" risk of concern in the workflow, and both reference standard domains. Gu et al.'s study [44] was regarded as a "high" risk in the subject selection domain because the authors omitted CXRs without the condition of interest. In the index test domain, Rajaraman et al.'s [5] and Karthikeyan's studies [49] were graded as "unclear", while Ferreira et al.'s study [50] was considered a "high" risk, in the same domain due to the fact that the performance of the AI model was not evaluated by an independent testing dataset that was excluded in the development of the AI model. Since the QUADAS-2 tool allows a study to contain one element ascertaining a high risk of bias without being eliminated [48], none of the selected studies were eliminated at this stage.

Figure 2. A risk of bias table of the five included studies from the adapted version of the QUADAS-2 ROB tool with a three-point scale indicating low, high and unclear concern [5,44,49–51].

3.3. Study Characteristics

Table 1 presents the characteristics of the included papers. All included papers utilised the same public dataset obtained from Guangzhou Women and Children's Medical Centre; however, each paper varied regarding the number of images used for training, testing and validation sets. All studies, apart from Gu et al.'s study [44], included 'normal' radiographs, as well as bacterial and viral pneumonia, labelled radiographs in their dataset. There were some similarities and overlaps with regard to pre-processing methods employed, though none of the studies utilised precisely the same pre-processing strategy.

Table 1. Characteristics of the AI algorithm employed, pre-processing methods utilised, as well as detailed descriptions of each dataset used in each of the included studies. Abbreviations: N/A (not applicable), ROI (region of interest), CLAHE (contrast limited adaptive histogram equalisation), CNN (convolutional neural network).

Author, Year	Algorithm Employed	Type of AI	Age of Participants (Years)	No. of Images (by Aetiology)				Total	No. of Images Used in			Pre-Processing Methods
				Normal	Pneumonia				Training	Validation	Testing	
					Bacterial	Viral	Total					
Gu et al., 2018 [44]	• AlexNet, • 3 Handcrafted Features (Gray level co-concurrence matrix-based feature extraction, Haar wavelet transform feature extraction, Histogram of oriented gradient-based feature extraction)	DL	5.5 ± 4.2	0	2665	1848	4513	4513	3211	802	500	• All input ROIs images were resized to 256 × 256 matrices. • An AlexNet-based fully convolutional networks model was applied for the segmentation of lung regions.
Ferreira et al., 2020 [50]	• VGG16, • Inception V3 architecture	DL	1–5	1349	2538	1345	3883	5232	5232	0	624	• CLAHE method • Chest cavity cropped from radiograph images • Combination of the above
Sousa et al., 2019 [51]	• CNN, • Inception V3 Architecture	DL	1–5	1349	2538	1345	3883	5232	624	0	624	• All of the images were re-dimensioned to a 300 × 300 pixels resolution and saved in a one-dimensional format
Rajaraman et al., 2018 [5]	• Sequential CNN, • Residual CNN, • Inception CNN, • Customised VGG16	DL	1–5	1349	2538	1345	3883	5232	5232	0	624	• Both baseline and cropped images are resampled to 1024 × 1024 pixel • Cropped using an algorithm based on anatomical atlases to detect the lung ROI automatically
Karthikeyan 2020 [49]	• AlexNet, • ResNet18, • DenseNet201, • SqueezeNet	DL	1–5	1341	2561	1345	3906	5247	4500	0	398	• The images were resized to 227 × 227 pixels

3.4. Diagnostic Accuracy of AI Algorithms in Distinguishing Viral Pneumonia from Bacterial Pneumonia

Table 2 presents the diagnostic accuracy measures (sensitivity, specificity, accuracy, and AUC) of each AI algorithm employed in respective papers. The ensemble set created achieved the highest sensitivity value of 96.3% [50] across all artificial algorithms examined (95% confidence intervals is not stated in the article). The DenseNet201 algorithm described by Karthikeyan reigned supreme with regard to both specificity and accuracy value obtaining 94% and 95%, respectively [49]. The AUC value closest to 1 was achieved by the VGG16 architecture (AUC = 0.9536) [5]. The baseline image, as well as the cropped ROI, scored 0.962 [5].

Table 2. The diagnostic accuracy of each AI algorithm employed in its respective paper in terms of sensitivity, specificity, accuracy, and AUC (area under the curve). The highest value for each outcome measure is highlighted in bold. $\pm$ indicates the standard deviations.

Author, Year	Algorithm	Sensitivity	Specificity	Accuracy	AUC
Gu et al., 2018 [44]	AlexNet (DCNN ONLY)	0.6322 ± 0.0023	0.7072 ± 0.0023	0.7360 ± 0.0023	0.7384 ± 0.0023
	GLCM Features	0.6378 ± 0.0058	0.8980 ± 0.0062	0.7060 ± 0.0672	0.7060
	Wavelet Features	0.5612 ± 0.0065	0.8779 ± 0.0205	0.6769 ± 0.0100	0.6769
	HOG Features	0.5714 ± 0.0617	0.8651 ± 0.0664	0.7511 ± 0.0127	0.6930
	All Handcrafted Features	0.6213 ± 0.0482	0.8848 ± 0.0387	0.7640 ± 0.0330	0.7200 ± 0.0060
	Fused Features (DCNN + all handcrafted features)	0.5567 ± 0.0379	0.9267 ± 0.0301	0.7692 ± 0.0122	0.8234 ± 0.0014
Ferreira et al., 2020 [50]	VGG16 and Baseline Set	*Not Stated*	*Not Stated*	*Not Stated*	0.85
	VGG16 and Set A	*Not Stated*	*Not Stated*	*Not Stated*	0.88
	VGG16 and Set B	*Not Stated*	*Not Stated*	*Not Stated*	0.83
	VGG16 and Set C (ensemble set)	**0.963**	0.851	0.921	0.91
	Inception V3 architecture	0.886	0.909	0.907	0.940
Sousa et al., 2019 [51]	'Best generated model'	0.913	0.696	0.831	0.831
	Inception V3 architecture	0.886	0.909	0.907	0.940
Rajaraman et al., 2018 [5]	Sequential CNN—Baseline	*Not specified*	0.838	0.928	0.954
	Residual CNN—Baseline	*Not specified*	0.784	0.897	0.921
	Inception CNN—Baseline	*Not specified*	0.714	0.854	0.901
	Customised VGG16—Baseline	*Not specified*	0.860	0.936	**0.962**
	Sequential CNN—Cropped	*Not specified*	0.838	0.928	0.956
	Residual CNN—Cropped	*Not specified*	0.798	0.908	0.933
	Inception CNN—Cropped	*Not specified*	0.730	0.872	0.919
	Customised VGG16—Cropped	*Not specified*	0.860	0.936	**0.962**
Karthikeyan, 2020 [49]	AlexNet	0.94	0.845	0.90	0.89
	ResNet18	0.92	0.82	0.87	0.87
	DenseNet201	0.96	**0.94**	**0.95**	0.952
	SqueezeNet	0.905	0.75	0.83	0.83

4. Discussion

CXR is routinely performed around the world to diagnose both subsets of pneumonia in paediatric patients [5]. Due to the complexity of lung diseases, the diagnosis of pneumonia on chest radiographs heavily relies on the eyes of a veteran radiologist. Therefore, there is a huge potential for AI algorithms to assist and further improve detection. This study aims to combine the results of all published literature focused on the classification of sub-types of paediatric pneumonia on chest X-rays using deep learning algorithms.

In this study, the efficacy of respective AI models in the classification of paediatric pneumonia on chest radiographs was evaluated by assessing accuracy, sensitivity, specificity, and AUC. Regarding diagnostic accuracy and specificity, the deep learning algorithm,

the DenseNet201 model utilised by Karthikeyan's study, performed far superiorly to the rest, yielding results of 95% and 94%, respectively [49]. AUCs measure the ability of a test (the individual algorithm in this case) to distinguish the presence or absence of a specific pathology [52] and take sensitivity and specificity into consideration. Thus, AUC can indicate how well a classifier is performing. The AI model, the customised VGG16 employed by Rajaraman et al. [5], achieved 96.2% in AUC for both of the baseline image sets (i.e., the original chest radiographs produced and cropped images), raising the question as to whether or not images that are cropped (just include the ROI) aid/improve AI models in the classification of paediatric pneumonia, or if it is simply one particular model being superior in performance to another. For example, both cropped and baseline images assessed by the customised VGG16 model achieved the same AUC result. However, the original VGG16 model utilised by Ferreira et al. [50] achieved an AUC of 85% with baseline images and 88% AUC with cropped images. Future studies should consider comparing cropped and uncropped images directly to make a definitive verdict. It would be unwise to conclude that AI models that are trained with the cropped ROI prior to testing with the dataset learn relevant feature representations toward classifying the task of interest without considering the probability of overfitting, reducing the generalisability of the results. In addition, cropping images before training involves more computational power and manpower.

When compared to the adapted Inception V3 architecture model employed by Kermany et al. [43], all four of the aforementioned outcome measures achieved greater results than the said platform in differentiating bacterial pneumonia from viral pneumonia on paediatric chest X-rays.

All five studies selected for this systematic review utilised the dataset obtained from Guangzhou Women and Children's Medical Centre, China. The selection of said dataset by all five groups of authors immediately excluded participation bias and allowed each AI model to be accurately compared. The included radiographs by Rajaraman et al. also included 'noisy images' to reduce bias and improve the model generalisation [5]. All authors also developed AI models with multi-level architectures, which, unlike the study by Kermany et al., avoided limited prediction accuracy. However, using the same dataset from Guangzhou place the studies' generalisability, as well as the algorithms' abilities in detecting pneumonia using noisy images since only one study included noisy images, into question. On the other hand, the upper age limit in Gu et al.'s study was 9.7 years old, diminishing its comparability to other studies included in this review.

At present, there is no sufficient guidance in critically appraising machine learning prediction models. Thus, one of the limitations of this study is that the information in this review may not be able to be combined or pooled together with other systematic reviews on the same subject matter and compare the data and information directly.

In order to improve and generalise results achieved by AI models in the classification of pneumonia on paediatric CXRs, further datasets should be developed worldwide. This is because pneumonia caused by different bacteria/viruses has different radiographical appearances [53], while the prevalence of different risk factors and types of paediatric pneumonia varies in different countries [54,55]. An algorithm that can effectively detect/rule out pneumonia and its ability to identify pneumonia aetiology can be the single best tool to be employed in the effort of reducing global paediatric pneumonia mortality. This is particularly crucial for resource-restrained countries with limited radiographic reporting capacity. The said datasets should include a variety of normal paediatric CXRs, as well as CXRs that belonged to those who had clinically confirmed having bacterial or viral pneumonia. Diagnosis should be confirmed by seasoned radiologists and other diagnostic test results, such as sputum cultures. This would improve the generalisation of AI models used to classify aetiologies. Finally, a major barrier impeding the translation of these results to a clinical setting is the comparison of said results to reporting clinicians' reports. Each of the included articles in this study assessed AI algorithms against one another rather than comparing outcome measures to that of consultant radiologists, and this would be

a great research question for future investigations. One study assessed the performance of an AI model to that of a number of radiologists in the classification of viral versus bacterial pneumonia in paediatric chest X-rays [56]. However, this study only assessed the differentiation of the two pneumonic aetiologies and not normal vs. pneumonic, so none of these algorithms could replace human radiologic interpretation currently [56]. In order for AI to translate into the clinical setting, future studies should compare different AI algorithms in differentiating normal versus bacterial versus viral pneumonia from that of human radiologists. AI algorithms are normally trained on one specific modality and on a specific pathology, while human radiologists have a basic and fundamental knowledge of all modalities and common pathologies then often specialise in one, sometimes multiple, organ systems.

There are only a few commercially available AI product that is capable of interpreting both chest and appendicular musculoskeletal X-ray images at the time of writing. This algorithm can detect seven different pathologies (fracture, pleural effusion, lung opacification, joint effusion, lung nodules, pneumothorax, and joint dislocation). A recent study which compared this AI algorithm's ability to human radiologists showed that the algorithm failed to pass the Fellowship of Royal College of Radiologists Rapid Reporting examinations [57]. This exam is normally taken between radiology speciality training years four and five in the UK [58].

This review has limitations. The age range used in these studies varied, making direct comparisons between the AI algorithms difficult. Secondly, only five studies are included in this review. Although all five studies used the same dataset from Guangzhou, the fact that some studies included noisy images and others included cropped images made it impossible to compare the algorithms on par with each other, let alone with other algorithms that are trained with a different dataset. Finally, the number of images that these algorithms were trained and validated on were small, compared to some of the commercially available algorithms [59–65] and some algorithms that are at the research stage [66].

Since the accuracy for all algorithms is so high, the amount of data is limited. Concerns about overfitting arise, which commonly occur due to the small training dataset [67,68]. Overfitting is a phenomenon where the algorithm fits in all the noises within the dataset, and the algorithm memorises all the peculiarities, finding the pattern best fit to the training data but not the general prediction trend, which is the goal of training [69]. Overfitted AI models are only applicable to the training dataset, but not the unseen dataset, losing the generalisability of the prediction [70–72]. Going forward, these algorithms should be further scrutinised. If overfitting did occur during training, the problem would need to be addressed in further studies before using these algorithms in clinical practice. Overfitting is an innate problem in using AI in radiology. First, there is no set threshold for a "sufficient" amount of data [70]. The more images and data fed into the algorithm for training, the less likely for overfitting to happen [49,73,74]. However, a large number of images can be very difficult to obtain in medicine due to confidentiality issues. Even if these data protection hurdles are overcome, the financial cost of acquiring medical images or data can be substantial [75,76]. Adding sub-optimal images into the training set, such as rotated or images with artefacts, can minimise the chance of overfitting [77,78] while increasing the overall image counts in the training set [78,79] can also mimic the situations in clinical practice.

This review gave the radiology community an insight into how AI can help us to reduce paediatric pneumothorax mortality rate and pose a potential solution to replace or to add on top of the current teleradiology system, especially in low- to middle-income or rural areas. This review took the essential step in starting the conversation in considering how radiology can utilise AI to improve workflow, but indeed, this is a start of a very long journey of research before AI can be used clinically. In this context, AI has the potential to assist in the classification of the aetiologies of pneumonia and, therefore, greatly increase treatment rates, potentially saving lives. The future landscape and the scope of practice of the radiology workforce are both going towards an exciting trajectory. "In the Age of the Algorithm, humans have never been more important" [80].

5. Conclusions

A number of the AI models achieved high accuracy in differentiating paediatric pneumonia. This showed potential in the automated classification of paediatric pneumonia on CXRs. Future studies should involve the comparison of AI models to that of a radiologist. Future research should focus on advancing the use of AI in identifying paediatric pneumonia in a clinical environment by including more sub-optimal images to ensure AI can correctly and accurately identify paediatric pneumonia in different circumstances and situations.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, B.; Kang, G.; Cheng, K.; Zhang, N. Attention-Guided Convolutional Neural Network for Detecting Pneumonia on Chest X-rays. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 4851–4854. Available online: https://ieeexplore.ieee.org/document/8857277/ (accessed on 28 September 2022).
2. Kassebaum, N.J.; Smith, A.G.C.; Bernabé, E.; Fleming, T.D.; Reynolds, A.E.; Vos, T.; Murray, C.J.L.; Marcenes, W.; GBD 2015 Oral Health Collaborators. Global, Regional, and National Prevalence, Incidence, and Disability-Adjusted Life Years for Oral Conditions for 195 Countries, 1990–2015: A Systematic Analysis for the Global Burden of Diseases, Injuries, and Risk Factors. *J. Dent. Res.* **2017**, *96*, 380–387. [CrossRef] [PubMed]
3. Chumbita, M.; Cillóniz, C.; Puerta-Alcalde, P.; Moreno-García, E.; Sanjuan, G.; Garcia-Pouton, N.; Soriano, A.; Torres, A.; Garcia-Vidal, C. Can Artificial Intelligence Improve the Management of Pneumonia. *J. Clin. Med.* **2020**, *9*, 248. [CrossRef] [PubMed]
4. Kurade, A.; Dhanawade, S.; Shetti, S. Induced Sputum as a Diagnostic Tool in Pneumonia in Under Five Children—A Hospital-based Study. *J. Trop. Pediatr.* **2018**, *64*, 510–515. [CrossRef] [PubMed]
5. Rajaraman, S.; Candemir, S.; Kim, I.; Thoma, G.; Antani, S. Visualization and Interpretation of Convolutional Neural Network Predictions in Detecting Pneumonia in Pediatric Chest Radiographs. *Appl. Sci.* **2018**, *8*, 1715. [CrossRef]
6. Franquet, T. Imaging of Community-acquired Pneumonia. *J. Thorac. Imaging* **2018**, *33*, 282–294. [CrossRef]
7. Kowalski, M.L.; Agache, I.; Bavbek, S.; Bakirtas, A.; Blanca, M.; Bochenek, G.; Bonini, M.; Heffler, E.; Klimek, L.; Laidlaw, T.M.; et al. Diagnosis and management of NSAID—Exacerbated Respiratory Disease (N-ERD)—A EAACI position paper. *Allergy* **2019**, *74*, 28–39. [CrossRef]
8. Longjiang, E.; Zhao, B.; Guo, Y.; Zheng, C.; Zhang, M.; Lin, J.; Luo, Y.; Cai, Y.; Song, X.; Liang, H. Using deep-learning techniques for pulmonary-thoracic segmentations and improvement of pneumonia diagnosis in pediatric chest radiographs. *Pediatr. Pulmonol.* **2019**, *54*, 1617–1626.
9. Hurt, B.; Yen, A.; Kligerman, S.; Hsiao, A. Augmenting Interpretation of Chest Radiographs With Deep Learning Probability Maps. *J. Thorac. Imaging* **2020**, *35*, 285–293. [CrossRef]
10. Smith-Bindman, R.; Lipson, J.; Marcus, R.; Kim, K.P.; Mahesh, M.; Gould, R.; Berrington de González, A.; Miglioretti, D.L. Radiation Dose Associated With Common Computed Tomography Examinations and the Associated Lifetime Attributable Risk of Cancer. *Arch. Intern. Med.* **2009**, *169*, 2078–2086. [CrossRef]
11. Bolter, J.D. *Turing's Man: Western Culture in the Computer Age [Internet]*; University of North Carolina Press: Chapel Hill, NC, USA, 1984; Available online: https://books.google.co.uk/books?id=4ZT6Th7CaDwC (accessed on 28 September 2022).
12. Chahal, A.; Gulia, P. Machine Learning and Deep Learning. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *8*, 4910–4914. [CrossRef]
13. Ongsulee, P. Artificial intelligence, machine learning and deep learning. In Proceedings of the 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE) [Internet], Bangkok, Thailand, 22–24 November 2017; pp. 1–6. Available online: http://ieeexplore.ieee.org/document/8259629/ (accessed on 28 September 2022).
14. Mathew, A.; Amudha, P.; Sivakumari, S. Deep Learning Techniques: An Overview. In *Advanced Machine Learning Technologies and Applications [Internet]*; Advances in Intelligent Systems and Computing; Hassanien, A.E., Bhatnagar, R., Darwish, A., Eds.; Springer: Singapore, 2021; Volume 1141, pp. 599–608. Available online: http://link.springer.com/10.1007/978-981-15-3383-9_54 (accessed on 28 September 2022).

15. Bengio, Y. Deep Learning of Representations: Looking Forward. In *Statistical Language and Speech Processing [Internet]*; Lecture Notes in Computer Science; Dediu, A.H., Martín-Vide, C., Mitkov, R., Truthe, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7978, pp. 1–37. Available online: http://link.springer.com/10.1007/978-3-642-39593-2_1 (accessed on 28 September 2022).

16. Lodwick, G.S.; Haun, C.L.; Smith, W.E.; Keller, R.F.; Robertson, E.D. Computer Diagnosis of Primary Bone Tumors: A Preliminary Report. *Radiology* **1963**, *80*, 273–275. [CrossRef]

17. Meyers, P.H.; Nice, C.M.; Becker, H.C.; Nettleton, W.J.; Sweeney, J.W.; Meckstroth, G.R. Automated Computer Analysis of Radiographic Images. *Radiology* **1964**, *83*, 1029–1034. [CrossRef]

18. Winsberg, F.; Elkin, M.; Macy, J.; Bordaz, V.; Weymouth, W. Detection of Radiographic Abnormalities in Mammograms by Means of Optical Scanning and Computer Analysis. *Radiology* **1967**, *89*, 211–215. [CrossRef]

19. Kruger, R.P.; Townes, J.R.; Hall, D.L.; Dwyer, S.J.; Lodwick, G.S. Automated Radiographic Diagnosis via Feature Extraction and Classification of Cardiac Size and Shape Descriptors. *IEEE Trans. Biomed. Eng.* **1972**, *19*, 174–186. [CrossRef]

20. Kruger, R.P.; Thompson, W.B.; Turner, A.F. Computer Diagnosis of Pneumoconiosis. *IEEE Trans. Syst. Man. Cybern.* **1974**, *4*, 40–49. [CrossRef]

21. Yamamoto, Y.; Nishiyama, Y.; Sota, T.; Miyai, M.; Uchibe, T.; Yada, N.; Nishiyama, Y.; Kitagaki, H. Improvement in the Quantification of Striatal Tracer Uptake in Single-photon Emission Computed Tomography With 123I-ioflupane Using a Cadmium-zinc-telluride Semiconductor Camera. *Shimane J. Med. Sci.* **2017**, *34*, 35–40.

22. Hadjiiski, L.; Sahiner, B.; Chan, H.-P. Advances in computer-aided diagnosis for breast cancer. *Curr. Opin. Obstet. Gynecol.* **2006**, *18*, 64–70. [CrossRef]

23. Hambrock, T.; Vos, P.C.; Hulsbergen–van de Kaa, C.A.; Barentsz, J.O.; Huisman, H.J. Prostate Cancer: Computer-aided Diagnosis with Multiparametric 3-T MR Imaging—Effect on Observer Performance. *Radiology* **2013**, *266*, 521–530. [CrossRef]

24. Dromain, C.; Boyer, B.; Ferré, R.; Canale, S.; Delaloge, S.; Balleyguier, C. Computed-aided diagnosis (CAD) in the detection of breast cancer. *Eur. J. Radiol.* **2013**, *82*, 417–423. [CrossRef]

25. Firmino, M.; Morais, A.H.; Mendoça, R.M.; Dantas, M.R.; Hekis, H.R.; Valentim, R. Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects. *Biomed. Eng. Online* **2014**, *13*, 41. [CrossRef]

26. Mosquera-Lopez, C.; Agaian, S.; Velez-Hoyos, A.; Thompson, I. Computer-aided prostate cancer diagnosis from digitized histopathology: A review on texture-based systems. *IEEE Rev. Biomed. Eng.* **2014**, *8*, 98–113. [CrossRef] [PubMed]

27. El Abbadi, N.K.; AL, J.; Taee, E. Breast Cancer Diagnosis by CAD. *Int. J. Comput. Appl.* **2014**, *100*, 25–29.

28. Brown, M.S.; Lo, P.; Goldin, J.; Barnoy, E.; Kim, G.H.J.; McNitt-Gray, M.; Aberle, D.R. Toward clinically usable CAD for lung cancer screening with computed tomography. *Eur. Radiol.* **2014**, *24*, 2719–2728. [CrossRef] [PubMed]

29. Mittal, A.; Kaur, M. Computer-aided-diagnosis in colorectal cancer: A survey of state of the art techniques. In Proceedings of the 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 26–27 August 2016.

30. Shariaty, F.; Mousavi, M. Application of CAD systems for the automatic detection of lung nodules. *Inform. Med. Unlocked* **2019**, *15*, 100173. [CrossRef]

31. Chan, H.P.; Samala, R.K.; Hadjiiski, L.M. CAD and AI for breast cancer—Recent development and challenges. *Br. J. Radiol.* **2020**, *93*, 20190580. [CrossRef]

32. Oliveira, S.P.; Neto, P.C.; Fraga, J.; Montezuma, D.; Monteiro, A.; Monteiro, J.; Ribeiro, L.; Gonçalves, S.; Pinto, I.M.; Cardoso, J.S. CAD systems for colorectal cancer from WSI are still not ready for clinical acceptance. *Sci. Rep.* **2021**, *11*, 14358. [CrossRef]

33. Porto-Álvarez, J.; Barnes, G.T.; Villanueva, A.; García-Figueiras, R.; Baleato-González, S.; Zapico, E.H.; Souto-Bayarri, M. Digital Medical X-ray Imaging, CAD in Lung Cancer and Radiomics in Colorectal Cancer: Past, Present and Future. *Appl. Sci.* **2023**, *13*, 2218. [CrossRef]

34. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence—An ESR white paper. *Insights Imaging* **2019**, *10*, 44. [CrossRef]

35. Frija, G.; Blažić, I.; Frush, D.P.; Hierath, M.; Kawooya, M.; Donoso-Bach, L.; Brkljačić, B. How to improve access to medical imaging in low- and middle-income countries ? *Eclinicalmedicine* **2021**, *38*, 101034. [CrossRef]

36. Legido-Quigley, H.; Doering, N.; McKee, M. Challenges facing teleradiology services across borders in the European union: A qualitative study. *Health Policy Technol.* **2014**, *3*, 160–166. [CrossRef]

37. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [CrossRef]

38. Son, J.; Shin, J.Y.; Kim, H.D.; Jung, K.-H.; Park, K.H.; Park, S.J. Development and Validation of Deep Learning Models for Screening Multiple Abnormal Findings in Retinal Fundus Images. *Ophthalmology* **2020**, *127*, 85–94. [CrossRef]

39. Grewal, M.; Srivastava, M.M.; Kumar, P.; Varadarajan, S. RADnet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 281–284. Available online: https://ieeexplore.ieee.org/document/8363574/ (accessed on 28 September 2022).

40. Stephen, O.; Sain, M.; Maduh, U.J.; Jeong, D.-U. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. *J. Healthc. Eng.* **2019**, *2019*, 1–7. [CrossRef]

41. Siddhartha, M.A.I. Versus M.D.—What happens when diagnosis is automated? *The New Yorker*, 27 March 2017; Annals of Medicine.

42. The Economist. Images Aren't Everything. Economists, 9 June 2018; Leaders Section. Available online: https://www.economist.com/leaders/2018/06/07/ai-radiology-and-the-future-of-work (accessed on 28 September 2022).

43. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122–1131.e9. [CrossRef]

44. Gu, X.; Pan, L.; Liang, H.; Yang, R. Classification of Bacterial and Viral Childhood Pneumonia Using Deep Learning in Chest Radiography. In Proceedings of the 3rd International Conference on Multimedia and Image Processing—ICMIP 2018 [Internet], Guiyang, China, 16–18 March 2018; ACM Press: New York, NY, USA, 2018. Available online: http://dl.acm.org/citation.cfm?doid=3195588.3195597 (accessed on 28 September 2022).

45. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef]

46. Chandler, J.; Cumpston, M.; Li, T.; Page, M.J.; Welch, V.A. (Eds.) *Cochrane Handbook for Systematic Reviews of Interventions [Internet], Version 6.3*; Cochrane, 2022; Available online: www.training.cochrane.org/handbook (accessed on 28 September 2022).

47. Critical Appraisal Skills Programme. CASP Diagnostic Study Checklist [Internet]. 2022. Available online: https://casp-uk.net/images/checklist/documents/CASP-Diagnostic-Study-Checklist/CASP-Diagnostic-Checklist-2018_fillable_form.pdf (accessed on 28 September 2022).

48. Hung, K.; Montalvao, C.; Tanaka, R.; Kawai, T.; Bornstein, M.M. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. *Dentomaxillofac. Radiol.* **2020**, *49*, 20190107. [CrossRef]

49. Karthikeyan, M.P. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. *Int. J. Adv. Res. Sci. Technol. IJARST* **2019**, *2019*, 4180949. Available online: https://ijarsct.co.in/Paper158.pdf (accessed on 28 September 2022).

50. Ferreira, J.R.; Armando Cardona Cardenas, D.; Moreno, R.A.; de Fatima de Sa Rebelo, M.; Krieger, J.E.; Antonio Gutierrez, M. Multi-View Ensemble Convolutional Neural Network to Improve Classification of Pneumonia in Low Contrast Chest X-ray Images. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) [Internet], Montreal, QC, Canada, 11 July 2020; pp. 1238–1241. Available online: https://ieeexplore.ieee.org/document/9176517/ (accessed on 28 September 2022).

51. Sousa, G.G.B.; Fernandes, V.R.M.; de Paiva, A.C. Optimized Deep Learning Architecture for the Diagnosis of Pneumonia Through Chest X-rays. In *Image Analysis and Recognition*; Lecture Notes in Computer Science; Karray, F., Campilho, A., Yu, A., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; Volume 11663, pp. 353–361. Available online: http://link.springer.com/10.1007/978-3-030-27272-2_31 (accessed on 28 September 2022).

52. Hoo, Z.H.; Candlish, J.; Teare, D. What is an ROC curve? *Emerg. Med. J.* **2017**, *34*, 357. [CrossRef]

53. O'Grady, K.A.F.; Torzillo, P.J.; Frawley, K.; Chang, A.B. The radiological diagnosis of pneumonia in children. *Pneumonia* **2014**, *5*, 38–51. [CrossRef]

54. McAllister, D.A.; Liu, L.; Shi, T.; Chu, Y.; Reed, C.; Burrows, J.; Adeloye, D.; Rudan, I.; Black, R.E.; Campbell, H.; et al. Global, regional, and national estimates of pneumonia morbidity and mortality in children younger than 5 years between 2000 and 2015: A systematic analysis. *Lancet Glob. Health* **2018**, *7*, e47–e57. [CrossRef]

55. GBD 2015. 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. *Lancet Infect. Dis.* **2017**, *17*, 1133–1161. [CrossRef]

56. Longjiang, E.; Zhao, B.; Liu, H.; Zheng, C.; Song, X.; Cai, Y.; Liang, H. Image-based deep learning in diagnosing the etiology of pneumonia on pediatric chest X-rays. *Pediatr. Pulmonol.* **2021**, *56*, 1036–1044.

57. Shelmerdine, S.C.; Martin, H.; Shirodkar, K.; Shamshuddin, S.; Weir-McCall, J.R. Can artificial intelligence pass the Fellowship of the Royal College of Radiologists examination? Multi-reader diagnostic accuracy study. *BMJ* **2022**, *379*, e072826. [CrossRef] [PubMed]

58. Ng, M.Y.; Ramamurthy, N. A career in radiology. *BMJ* **2012**, *345*, e8142. [CrossRef]

59. Gertych, A.; Zhang, A.; Sayre, J.; Pospiech-Kurkowska, S.; Huang, H. Bone age assessment of children using a digital hand atlas. *Comput. Med. Imaging Graph.* **2007**, *31*, 322–331. [CrossRef]

60. Rajpurkar, P.; Irvin, J.; Bagul, A.; Ding, D.; Duan, T.; Mehta, H.; Yang, B.; Zhu, K.; Laird, D.; Ball, L.R.; et al. Mura: Large dataset for abnormality detection in musculoskeletal radiographs. *arXiv* **2017**, arXiv:171206957.

61. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet], Honolulu, HI, USA, 21–26 July 2017; pp. 3462–3471. Available online: http://ieeexplore.ieee.org/document/8099852/ (accessed on 28 September 2022).

62. Irvin, J.; Rajpurkar, P.; Ko, M.; Yu, Y.; Ciurea-Ilcus, S.; Chute, C.; Marklund, H.; Haghgoo, B.; Ball, R.; Shpanskaya, K.; et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HA, USA, 27 January–1 February 2019; Volume 33, pp. 590–597.

63. Johnson, A.E.; Pollard, T.J.; Berkowitz, S.J.; Greenbaum, N.R.; Lungren, M.P.; Deng, C.Y.; Mark, R.G.; Horng, S. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci. Data* **2019**, *6*, 317. [CrossRef]

64. Halabi, S.S.; Prevedello, L.; Kalpathy-Cramer, J.; Mamonov, A.B.; Bilbily, A.; Cicero, M.; Pan, I.; Pereira, L.A.; Sousa, R.; Abdala, N.; et al. The RSNA pediatric bone age machine learning challenge. *Radiology* **2019**, *290*, 498–503. [CrossRef]

65. Bustos, A.; Pertusa, A.; Salinas, J.M.; de la Iglesia-Vayá, M. PadChest: A large chest X-ray image dataset with multi-label annotated reports. *Med. Image Anal.* **2020**, *66*, 101797. [CrossRef]

66. Selby, I. Automated Quality Control of Chest X-ray [Internet]. Oral Presentation Presented at: European Congress of Radiology 3 March 2023; Vienna. Available online: https://connect.myesr.org/?esrc_course=using-ai-for-quality-control-in-radiography (accessed on 28 September 2022).

67. Ying, X. An Overview of Overfitting and its Solutions. *J. Phys. Conf. Ser.* **2019**, *1168*, 022022. [CrossRef]
68. Whiteson, S.; Tanner, B.; Taylor, M.E.; Stone, P. Protecting against evaluation overfitting in empirical reinforcement learning. In Proceedings of the 2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL), Paris, France, 12–14 April 2011; pp. 120–127. Available online: http://ieeexplore.ieee.org/document/5967363/ (accessed on 28 September 2022).
69. Dietterich, T. Overfitting and undercomputing in machine learning. *ACM Comput. Surv. CSUR* **1995**, *27*, 326–327. [CrossRef]
70. Mutasa, S.; Sun, S.; Ha, R. Understanding artificial intelligence based radiology studies: What is overfitting? *Clin. Imaging* **2020**, *65*, 96–99. [CrossRef]
71. Anderson, D.; Burnham, K. *Model Selection and Multi-Model Inference*, 2nd ed.; Springer: New York, NY, USA, 2004; Volume 63, p. 10.
72. England, J.R.; Cheng, P.M. Artificial intelligence for medical image analysis: A guide for authors and reviewers. *Am. J. Roentgenol.* **2019**, *212*, 513–519. [CrossRef]
73. Sun, Y.; Wang, X.; Tang, X. Deep Learning Face Representation from Predicting 10,000 Classes. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1891–1898.
74. Gonçalves, I.; Silva, S.; Melo, J.B.; Carreiras, J.M.B. Random Sampling Technique for Overfitting Control in Genetic Programming. In *Genetic Programming*; Moraglio, A., Silva, S., Krawiec, K., Machado, P., Cotta, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 218–229.
75. Parmar, C.; Barry, J.D.; Hosny, A.; Quackenbush, J.; Aerts, H.J.W.L. Data Analysis Strategies in Medical Imaging. *Clin. Cancer Res.* **2018**, *24*, 3492–3499. [CrossRef]
76. Chartrand, G.; Cheng, P.M.; Vorontsov, E.; Drozdzal, M.; Turcotte, S.; Pal, C.J.; Kadoury, S.; Tang, A. Deep Learning: A Primer for Radiologists. *RadioGraphics* **2017**, *37*, 2113–2131. [CrossRef]
77. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Into Imaging* **2018**, *9*, 611–629. [CrossRef]
78. Kim, D.; MacKinnon, T. Artificial intelligence in fracture detection: Transfer learning from deep convolutional neural networks. *Clin. Radiol.* **2018**, *73*, 439–445. [CrossRef]
79. Ha, R.; Mutasa, S.; Karcich, J.; Gupta, N.; Pascual Van Sant, E.; Nemer, J.; Sun, M.; Chang, P.; Liu, M.Z.; Jambawalikar, S. Predicting Breast Cancer Molecular Subtype with MRI Dataset Utilizing Convolutional Neural Network Algorithm. *J. Digit. Imaging* **2019**, *32*, 276–282. [CrossRef]
80. Fry, H. *Hello World: How to Be Human in the Age of the Machine [Internet]*; Transworld, 2018; Available online: https://books.google.co.uk/books?id=72FCDwAAQBAJ (accessed on 28 September 2022).