Review

Artificial Intelligence-Based Software with CE Mark for Chest X-ray Interpretation: Opportunities and Challenges

Salvatore Claudio Fanni [1], Alessandro Marcucci [1], Federica Volpi [1], Salvatore Valentino [2], Emanuele Neri [1] and Chiara Romei [3,*]

[1] Department of Translational Research, Academic Radiology, University of Pisa, 56126 Pisa, Italy; s.fanni1@studenti.unipi.it (S.C.F.); a.marcucci2@studenti.unipi.it (A.M.)
[2] EBIT sr.l. Esaote Group, Via di Caciolle, 50127 Florence, Italy
[3] Department of Diagnostic Imaging, 2nd Radiology Unit, Pisa University-Hospital, Via Paradisa 2, 56124 Pisa, Italy
[*] Correspondence: ch.romei@ao-pisa.toscana.it or chiararomei@gmail.com

Abstract: Chest X-ray (CXR) is the most important technique for performing chest imaging, despite its well-known limitations in terms of scope and sensitivity. These intrinsic limitations of CXR have prompted the development of several artificial intelligence (AI)-based software packages dedicated to CXR interpretation. The online database "AI for radiology" was queried to identify CE-marked AI-based software available for CXR interpretation. The returned studies were divided according to the targeted disease. AI-powered computer-aided detection software is already widely adopted in screening and triage for pulmonary tuberculosis, especially in countries with few resources and suffering from high a burden of this disease. AI-based software has also been demonstrated to be valuable for the detection of lung nodules detection, automated flagging of positive cases, and post-processing through the development of digital bone suppression software able to produce digital bone suppressed images. Finally, the majority of available CE-marked software packages for CXR are designed to recognize several findings, with potential differences in sensitivity and specificity for each of the recognized findings.

Keywords: artificial intelligence; chest c-ray; CE-mark; deep learning; pulmonary tuberculosis; lung nodules

1. Introduction

Since the introduction of digital radiography in 1980, chest X-ray (CXR) has remained the most important imaging technique for chest imaging [1]. CXR is a cheap and easily accessible imaging technique with broad indications, characterized by short scan time and a lower dose compared to other imaging techniques, such as computed tomography, for example [2]. Nevertheless, CXR is limited in scope and sensitivity, and its interpretation is challenging due to the overlay of several different tissues belonging to a complex three-dimensional volume within a two-dimensional image [3].

The difference in density between pathological and healthy lung parenchyma may be subtle and challenging to detect, also because about 40% of lung parenchyma is covered by the ribs and mediastinum [4]. The experience of the radiologist and the cognitive process behind the interpretation of a CXR directly affect the clinical utility and effectiveness of CXR in patient management. To complicate matters even further, the rising demand for imaging exams, particularly in the emergency setting, may contribute to increasing numbers of errors [5].

The intrinsic limitations of CXR and the increasing number of examinations has prompted the proliferation of Artificial Intelligence (AI)-based software packages dedicated to CXR interpretation over the past decade [6].

AI-based image analysis, especially through deep learning, has become popular in recent years for classifying and segmenting the wealth of unlabeled data usually available in medical institutions in order to reduce, for example, the subjectivity and the time burden introduced by a manual labeling [7].

Still, to effectively tackle the above-mentioned issue, it is necessary to make these new technologies part of the daily routine. However, the gap between scientific research and clinical practice can be very significant. As a first essential step to achieving the clinical translation of all these research efforts, AI-based software needs to be certified for medical use. The providers affirm by means of CE marks the good conformity of the software with European health, safety, and environmental protection standards [8].

The primary aim of this review is to summarize the state-of-the-art of CE-marked AI-based software for CXR interpretation, focusing on their role, but also on the open issues of these applications and all the features of these software the reader must be aware of.

2. Materials and Methods

The online database "AI for radiology", a non-commercial project tracking all CE-certified AI products, was queried to identify CE-marked AI-based software available for CXR interpretation [9]. Filters were applied to include only software developed for the chest, as subspecialty, and for X-ray, as imaging modality. No filters were applied regarding the level of CE and FDA (Food and Drug Administration) class. We ran the product and company name of each of the AI-based software packages through the PubMed database. Original research published in English was included in this review. The last search was run on 13 September 2022.

After collecting the studies, we used a narrative approach to analyze the state of the art of AI-based interpretation of different disease on CXR.

3. Results

We found 26 different CE-marked AI-based software packages.

The characteristics of the included software are summarized in Table 1. The studies were divided into three categories according to the targeted disease: AI-based software for tuberculosis-related abnormalities; AI-based software for nodule and lung cancer detection; and AI-based software for multiclass findings and other clinical scenarios. However, several software packages were not developed for a single disease or finding, and a significant overlap among categories was found.

3.1. AI-Based Software for Tuberculosis-Related Abnormalities

Pulmonary tuberculosis (TB) diagnosis is based on clinical features, sputum/blood test examinations positive for tubercle bacilli, and CXR presenting confluent dense shadows or signs of consolidation [10]. However, CXR as a screening and triage tool presents some relevant drawbacks, such as high inter- and intra-reader variability, moderate specificity, and limited radiology availability, especially in regions of the world with higher burdens of pulmonary TB [11].

Artificial intelligence (AI)-powered computer-aided detection (CAD) is already a fundamental tool in screening and triaging pulmonary TB, especially in countries suffering from a high burden of this disease. In these countries, CAD software packages are already widely adopted in mobile unit machines, which can operate without access to an electrical power grid [12]. Indeed, in 2021, the WHO recommended the use of CAD software for TB screening and triage, and the minimum required values of sensitivity and specificity were defined as 0.90 and 0.70, respectively [13,14].

Table 1. Characteristics of the included Software (information provided from the online database "AI for radiology").

Software Name	Company	Disease Targeted	Population	Processing Time
ClearRead XRay	Riverain Technologies	Lung cancer	Adult, not pediatric	1–10 min
ChestLink	Oxipit	75 different pathologies	Older than 18 years old	10–60 s
CAD4TB	Thirona	Tuberculosis	All chest X-rays	3–10 s
AI-Rad Companion Chest X-ray	Siemens Healthineers	Pulmonary lesions, pleural effusion, pneumothorax, consolidation, atelectasis	All upright chest X-rays	1–10 min
SenseCare-Chest DR Pro	SenseTime	Pneumonia, tuberculosis, pneumothorax, pleural effusion, cardiomegaly, rib fractures	All chest X-rays	3–10 s
Annalise Enterprise CXR	annalise.ai	124 findings present in the emergent, urgent, and non-urgent care settings	Patients over 16 years of age	3–10 s
Chest ǀ MSK AI	Arterys	Fracture, dislocation, elbow joint effusion, pleural effusion, pulmonary nodule, pulmonary opacity, pneumothorax	Emergency department population	Not available
Chest X-ray Classifier	Quibim	14 different findings	All chest X-rays	10–60 s
qXR	Qure.ai	Tuberculosis, COVID-19, signs seen in Lung Parenchyma, Pleura, Mediastinum, Cardiac and bones	All chest X-rays	10–60 s
TIRESYA	Digitec	Unspecified	Unspecified	Not available
Milvue Suite	Milvue	Bone fracture, pleural effusion, lung opacity, elbow joint effusion, lung nodule, pneumothorax, dislocation	All chest X-rays	10–60 s
CheXVision	XVision	17 pathologies	All chest X-rays of patients above 16 years old	3–10 s
CAD4COVID-XRay	Thirona	COVID-19	All chest X-rays	3–10 s
Chest Solution	Nanox.AI	Pneumothorax, pleural effusion	Unspecified	Not available
VUNO Med®-Chest X-ray™	VUNO	Nodule/Mass, Consolidation, Interstitial Opacity, Pneumothorax, Pleural Effusion	All chest X-rays	<3 s
InferRead DR Tuberculosis	Infervision	Tuberculosis	Unspecified	Not available
InferRead DR Chest	Infervision	Lung cancer, pneumothorax, fracture, tuberculosis, lung infection, aortic calcification, cord imaging, heart shadow enlargement, pleural effusion.	Any	Not available

Table 1. *Cont.*

Software Name	Company	Disease Targeted	Population	Processing Time
X1	Visionairy Health	12 different findings	Unspecified	Not available
Auto Lung Nodule Detection	Samsung Electronics	Lung cancer	Unspecified	<3 s
Pneumothorax (Ptx)	Aidoc	Pneumothorax	Radiograph	1–10 min
Red Dot	Behold.ai	Pneumothorax	Unspecified	Not available
JLD-02K	JLK Inc.	Lung cancer	All Chest X-rays	3–10 s
Lunit INSIGHT CXR	Lunit	11 different findings	Patients aged 14 years or older	3–10 s
Critical Care Suite	GE Healthcare	Pneumothorax	Adult-size patients. Suspected of pneumothorax or intubated.	<3 s
ChestView	GLEAMER	Pneumothorax, pleural effusion, alveolar syndrome, nodule, mediastinal mass	All Chest X-rays	10–60 s
ChestEye Quality	Oxipit	75 different pathologies	Patients over 18 years old	3–10 s

However, the definition of software diagnostic accuracy is not that straightforward, as the performance may vary significantly according to several technical and clinical variables, see Figure 1. Furthermore, it is complicated to perform a direct comparison between the available software packages, or even between different versions (vrs) of the same software package. Only the newer vrs of the software CAD4TB (vrs 6 and vrs 7) and qXR (vrs 2 and vrs 3) utilize deep learning (DL) techniques and reach the required minimum values of sensitivity and specificity fixed by the WHO [13,14].
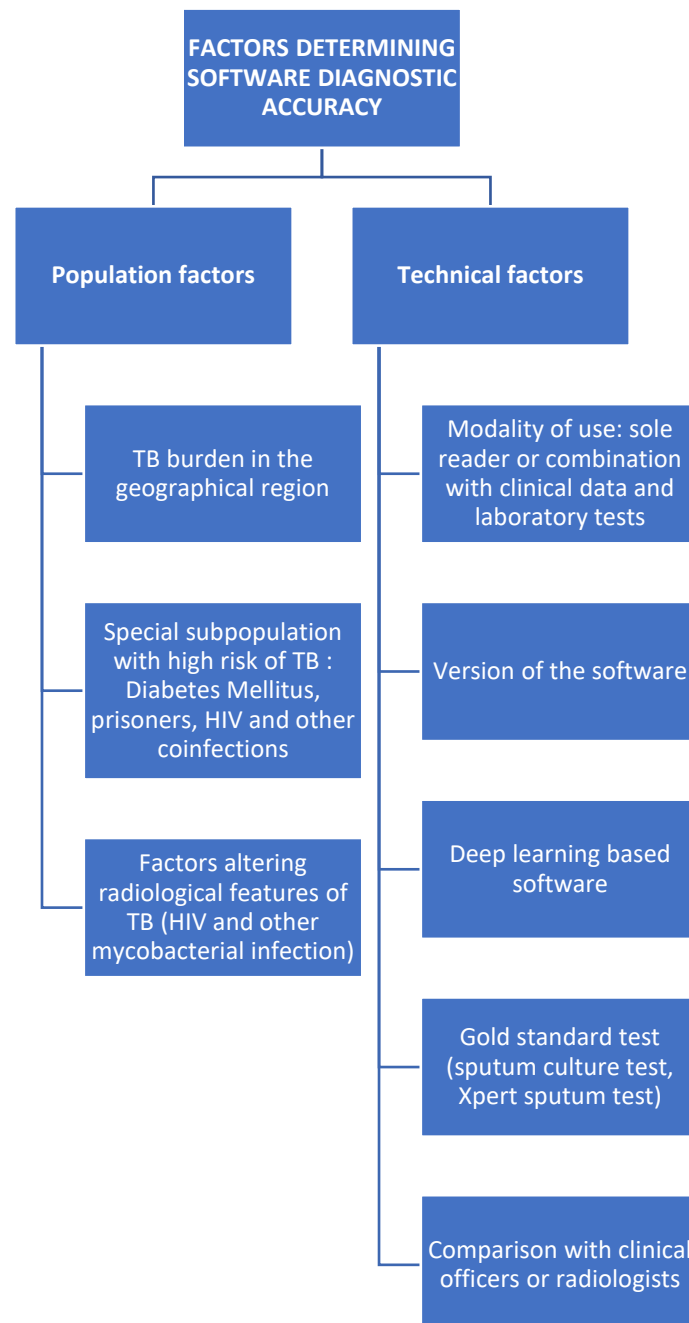

Figure 1. Factors determining software diagnostic accuracy.

CAD software also differs in image processing and output. CAD4TB produces an abnormality heatmap indicating suspicious regions and provides a likelihood score ranging from 0 (no TB) to 100 (TB positive). Conversely, qXR compares the computed abnormality score with a prespecified threshold and provides a binary classification (positive or negative

for pulmonary TB) [12,14]. This threshold must be tailored to achieve the optimal sensitivity and specificity according to the characteristics of the population studied, as suggested by Fehr et al., who achieved the best possible sensitivity for detecting pulmonary TB (84.8%), with a decisional threshold of 25 [15]. The proper tailoring of this threshold requires the consideration of clinical variables such as the TB burden in the studied population, diabetes, and HIV status [14].

According to the TB burden, we can differentiate high-burden and low-resource regions and low-burden high-resource regions. In high-burden low-resource regions (Pakistan, Nigeria, Brazil, etc.), the TB prevalence is almost 0.5%, but may reach as high as 2% when considering patients living with diabetes mellitus [16]. Mass screening surveys have been conducted in Karachi, Pakistan and Nigeria [17,18]. Due to the shortage of experienced CXR readers, in these regions, CAD could be used as the sole screening reader and may outperform the diagnostic accuracy of human reading [15]. In a resource-constrained setting, Philipsen et al. prospectively compared sputum testing alone and with the integration of CAD software for the selection of patients for sputum testing in terms of accuracy and cost. The results showed that the use of CAD software as a triage tool significantly reduced the cost, both per screened subject and per notified TB case, with only a few TB cases being missed [12,19]. Conversely, in low-burden high-resource countries, CAD software could be used to support triage or screening programs. Melendez et al. implemented CAD software in a screening program in London (UK) for a high-risk population including homeless people, prisoners and the drug and alcohol addicted [20].

Qin et al. compared the diagnostic accuracy of three DL algorithms (CAD4TB, Lunit and qXR), and concluded that the area under the curve (AUC) values for the three software packages were similar (0.92, 0.94, and 0.94, respectively), but most importantly, they pointed out the absence of a universal cut-off and the importance of identifying the proper threshold according to the TB prevalence in the population to achieve this level of performance [21].

However, even within the same region, there are different population subsets requiring specific adjustments. For instance, many studies have been conducted on prisoners [22,23]. The most significative study with respect to yield, efficiency and costs of mass screening algorithms for TB in prison was conducted in Brazil [24]. As a result, the integration of CAD4TB in this screening did not significantly change the cost per case diagnosed compared to other screening strategies without AI, and missed a higher number of cases. The most important drawback in this population subset was the high prevalence of HIV positivity and coinfection in the screened population [23,25].

Indeed, the detection of TB on CXR in HIV-positive subjects or patients with nontuberculous mycobacteria coinfection is quite challenging [19,26]. A strategy for avoiding this bias could be the stratification of the analysis according to the HIV status [19].

Two additional drawbacks to consider in the included studies are the significant heterogeneity regarding the gold standard and the methodology of comparison between algorithms and humans reader performance.

The majority of the included studies lack sputum culture testing for TB as a gold standard, usually because this test is expensive and requires several weeks to obtain a conclusive result, with rare exceptions [14,15,21,27]. As an alternative to sputum culture, the Xpert sputum test is usually adopted as a reference standard [12,26,28].

In regions where radiologists are not widely available for CXR reading, some studies have compared CAD software performance with the performance of clinical officers [29]. The difference between a clinical officer and an expert reader, such as a radiologist, was highlighted by Breuninger et al., who conducted a study with an older version of CAD4TB that outperformed only the clinical officers and not the experienced readers [26]. Moreover, only newer versions of CAD4TB and qXR achieved performance that was comparable to that of human radiologists in terms of triage ability [13,30]. A similar heterogeneity was reported by Qin et al., who assessed five AI algorithms during a screening campaign and also reviewed the literature related to CAD software for TB.

3.2. AI-Based Software for Nodule and Lung Cancer Detection

CXR remains, to date, the first line method for the diagnosis of lung nodules. Lung cancer benefits, in terms of survival, from early detection on CXR, but it has also been demonstrated that CXR fails to identify lung cancer in 18% of patients and that about 90% of mistakes in lung cancer diagnosis occur on CXR [31,32].

AI could assist radiologists in lung nodule diagnosis in several ways, such as detection, flagging of positive cases, and digital bone suppression.

A DL-based automatic detection algorithm has already been shown to be able to outperform the performance of physicians. However, in clinical practice, AI should be considered an aid to radiologists, and not an alternative [33].

Indeed, many papers have demonstrated that radiologists' performance is enhanced when using AI as a second reader [34,35]. This benefit has been proven for young radiologists as well as for senior radiologists, thus being independent of reader experience [36]. However, as seen in the previous section on TB, the nature of this benefit may be affected by several technical and clinical–radiological variables, such as the nodule characteristics or the adopted ground truth.

AI-based detection performance in terms of accuracy and robustness seems to depend on the size and conspicuousness of the nodule solid portion. Indeed, inconspicuous nodules may cause fluctuations in the model output, leading to misclassification of findings [37].

Additionally, the algorithm performance may vary significantly according to the different adopted methods of ground truth. In a recent paper by Kim et al., a DL algorithm showed an AUC of 0.771 and 0.839 when using CT and radiologist readings as ground truth, respectively [38].

Liang et al. used the QUIBIM CXR Classifier to detect pulmonary nodules/masses. The four different algorithms of the software were analyzed and validated, and each one of them was characterized by different pros, such as higher sensitivity for the nodule probability algorithm or higher specificity for the heat map algorithm. Thus, to obtain the maximum benefits and minimize both false positives and false negatives, the radiologist may combine different algorithm strategies [39]. The increase in sensitivity, specificity and diagnostic confidence could be even higher for trained radiographers, which is particularly helpful in countries facing both a shortage of radiologists and high medical imaging demand [40].

Another application of AI exploiting the synergy between AI and radiologists in lung cancer diagnosis on CXR is the automated flagging of positive cases. For this purpose, Tam et al. proposed a triage workflow and achieved a significant reduction in missed cancer of 60% by using AI to flag positive case [41].

Moving to post-processing, to overcome the obscuration of lung cancer by overlying bone structures, digital bone suppression software has been developed, which is able to produce bone-suppressed images (BSIs). Digital bone suppression does not require special dual-energy hardware or additional dosage, and it is not affected by motion artifacts [42].

In 2013, Schalekamp et al. showed that BSIs improved radiologists' detection performance for lung nodules compared to CXR alone (AUC = 0.883 vs. 0.855), particularly for nodules with moderate and subtle conspicuity [43].

However, BSIs have been shown to reduce specificity and lead to an increased overcall for focal lesions, which may result in unnecessary follow-up or further expansive investigations in normal patients [44,45]. It can be hypothesized that the lack of experience of the reader with BSIs may have caused this overcall, and it is likely that more practice will partially overcome this issue.

Another possible solution is to combine both detection and post-processing strategies by applying CAD software after digital bone suppression. As reported by Schalekamp et al., CAD improved radiologists' performance and detected 127 of 239 (53%) of the nodules that were initially missed by the radiologists. However, only 57 (45%) of these detections were accepted by the observer. According to the authors, the limited ability of the readers to reliably differentiate true-positive from false-positive CAD marks is currently one of the most important limits to CAD's beneficial effect [46]. To optimize CAD's beneficial

effect and address this issue, two alternative methods already implemented with success in mammograms have been proposed for CXR. First, in the interactive mode, a CAD mark is shown together with a score of suspicions only if the radiologist clicks on the mark's location. Secondly, the observer evaluates its score without viewing the CAD analysis, then a mathematical combination of the reader and the CAD score is computed. No improvement was seen with the interactive mode, while the mathematical combination significantly improved detection performance.

Thus, CAD provides information that might be even more useful when independently combined with a radiologist's evaluation, without interfering with the reading process itself [47]. However, this approach could raise other legal and ethical problems, as the radiologist would have no way of checking the software output.

3.3. AI-Based Software for Multiclass Findings and other Clinical Scenarios

However highly accurate, the clinical benefit of a software package limited to a single or a small number of findings is at least questionable. The use of different software packages for different findings is unrealistic in clinical practice, and software covering the full range of CXR findings could be more effectively integrated into clinical practice.

Indeed, the majority of available CE-marked software packages for CXR fulfill this need, and are designed to recognize multiple findings. Multiclass finding detection must address differences in size and extension, ranging from subtle nodular opacities to extensive pleural effusion or pneumothorax [48]. Thus, the software may exhibit different performances for each finding, and the radiologist must be aware of this.

One clinical scenario necessarily requiring multiclass detection is the post-traumatic CXR performed in the emergency department, where findings vary from subtle ones to time-critical pathology [49]. In this setting, AI may aid in timely identification and worklist prioritization. Gipson et al. externally validated the software Annalise.ai for the detection of traumatic injuries on supine CXR, using CT as ground truth. In this study, the software outperformed radiologists for detection of pneumothorax and rib fractures, two of the most common pathologies in thoracic trauma, but was found to be inferior to radiologists for clavicle, humerus and scapula fractures [50]. However, as already pointed out in the previous paragraphs, comparing the performance of radiologists and AI is not enough to objectively investigate the beneficial effect provided by AI on clinical practice. An initial step was taken by Hwuang et al., who did not limit themselves to comparing the standalone detection performance of Lunit software and radiology residents in the emergency setting, but also assessed the effect of AI on residents' performance. The results showed that when radiology residents used Lunit, the sensitivity improved from 65.6% to 73.4%, but, not surprisingly, a decrease in specificity was observed (98.1% vs. 94.3%) [51].

To evaluate the real-world usefulness of AI as a diagnostic assistance device for radiologists, Jones et al. set up a prospective multicenter study involving a group of radiologists using Annalise.ai in their daily reporting workflow [52]. Out of 2927 patients, 92 (3.1%) had significant reported changes, 43 (1.4%) had changed patient management, and 29 (1.0%) had further imaging recommendations. Moreover, radiologist attitudes towards the software were questioned, and almost all of them (9 out of 10) felt an improvement in their accuracy and a more positive attitude at the end of the study.

Another relevant clinical scenario to address is pediatric radiology. As CT cannot be freely performed in children, CXR plays an even more critical role in pediatric radiology.

Shin et al. evaluated whether Lunit, an AI-based detection software developed and approved for adult CXR, could be used for pediatric CXR as well. The software assessed the presence of nodules, consolidation, fibrosis, atelectasis, cardiomegaly, pleural effusion, pneumothorax and pneumoperitoneum. Diagnostic accuracies up to 96.9% were found when cardiomegaly and children 2 years old or younger were excluded [53]. To detect cardiomegaly, other strategies may be explored. Bercean et al. trained two U-Net segmentation algorithms to contour the heart and the lungs and automatically compute the cardiothoracic ratio. This kind of approach reduced the reading time from 22.54 s to 5.14 s, and the F1 score

for cardiomegaly detection was 0.85. However, AI may not necessarily reduce reading time, as showed by Kim et al., who reported a slight but significant increase in reading time from 14 s to 19 s when using AI [54].

Finally, COVID-19 pneumonia is a clinical scenario deserving of special mention, considering the ways in which it boosted radiology research, and particularly the AI field, and despite the SARS-CoV-2 pandemic is relatively recent, two different CE-certified deep-learning-based software packages are already available, qXR and CAD4COVID [55,56]. Mushtaq et al. investigated the prognostic utility of qXR on initial CXR. In this study, the software output was personalized to only report a score reflecting the percentage of pixels involved by opacity or consolidation, and a score ≥30 was shown to be an independent predictor for both mortality and severity of COVID-19 pneumonia ($p < 0.001$) [57]. The second software package is CAD4COVID, redesignated from CAD4TB vrs 6. CAD4COVID was efficiently integrated by Tovar et al. into a mobile TBC screening unit in Lima, Peru. Integrated TB and COVID-19 screening and testing services may ensure that TB case detection is maintained and not de-prioritized by health systems. The software produced a continuous abnormality score from 0 to 100, and all patients with a score >50 were tested for SARS-CoV-2 antibodies and by polymerase chain reaction (PCR) test, leading to a diagnosis of SARS-CoV-2 antibodies per three people screened and PCR-confirmed SARS-CoV-2 infection per eight people screened [58].

The radiologists' interpretation of CXR, regardless of the clinical scenario, may be hindered by bone overlay. Thus, as well as for lung nodules, digital bone suppression may be helpful. BSIs have been demonstrated to be useful for improving radiologists' detection performance for invasive pulmonary aspergillosis, single or multiple focal opacities, or signs of cardiac congestion [44,59]. These studies demonstrated that, without any significant negative effect on the interpretation of diffuse lung disease, the detection of focal chest abnormalities in CXRs, beyond just lung nodules, can be improved when their evaluation is aided by digital bone suppression.

3.4. Opportunities and Challenges

As highlighted in the previous paragraphs, the number of publications in the field of AI and of the available commercial applications for CXR interpretation is steadily increasing. Radiologists will definitely have to deal with these solutions, and therefore must be aware of both the opportunities they offer as well as their challenges.

Artificial intelligence could be the answer to the problem of an ever-increasing population faced by healthcare systems and the rising number of examinations referred to the radiology departments [60].

These software packages may help in finding and detecting various types of CXR abnormalities, saving work for the radiologists and time that can be dedicated to harder issues. This is possible today thanks to the availability of software covering the full range of CXR findings, from fractures to pleural effusions. In those regions in which expert radiologists may be lacking, such as developing countries, these software packages may really be game changers.

The integration of AI as a second reader is a topic of extreme interest and an emergent subject of study. AI has been demonstrated able to significantly increase radiologists' sensitivity; however, it is also proven to decrease specificity, especially in unexperienced radiologists. This dual effect makes necessary to emphasize that the responsibility of final diagnosis still lies with the radiologists, not AI systems, [61] and that radiologists must be aware of when to trust the AI output.

An additional point of complexity concerns the previously mentioned AI-based software for multiclass findings. The software may present different levels of sensitivity and specificity for each of the recognized findings, and the radiologist must be aware of this in order to trust the different outputs with more or less confidence.

Some issues still undermine the building of a trust relationship between radiologists and AI, significantly slowing down its integration into clinical practice. One of these issues

is the lack of explainability of the software and the lack of transparency of the available literature, particularly regarding algorithm and dataset characteristics and limitations [62].

However, promising progress has been achieved in recent years, such as the increasing number of papers reporting objective comparisons between different AI software packages performed with independent datasets.

Another relevant issue is the lack of large-scale multicentric datasets, which are pivotal for training the algorithms, and this is particularly true for pediatric CXR. Pediatric imaging requires expertise and also dedicated or adaptable equipment [63]. Similarly, AI-based software requires specific training to achieve acceptable performance for pediatric CXR interpretation, and the application of AI in this promising field is currently limited.

4. Conclusions

CXR is still the most important imaging technique for chest imaging, as evidenced by the number of published studies and software packages that have been developed for this technique. AI-based software packages that have obtained CE markings have been created for diverse clinical scenarios, including TB screening and nodule detection. AI was demonstrated to be able to improve the diagnostic accuracy of CXR in TB screening, with performance varying according to the software characteristics and the targeted population. In this setting, to be aware of this dependence is critical to properly tailoring the software according to the clinical context. AI-based software also affirmed its role as a second reader when performing nodule detection by CXR. Nodule features and combination BSIs seem to have more impact on diagnostic performance than the level of experience of the user. Finally, AI-based software for the detection of multiclass findings was developed to comprehensively address the complexity and wide range of potential findings.

In conclusion, the primary objective of this software is not to outperform the diagnostic performance of radiologists, but rather to enhance it. In order to achieve this objective, the radiologist must be aware not only of the benefits but, more significantly, of the limitations of this software.

Author Contributions: Conceptualization, C.R.; methodology, S.C.F., A.M. and S.V.; investigation, S.C.F. and A.M. data curation, A.M.; writing—original draft preparation, S.C.F., A.M. and F.V. writing—review and editing, E.N. and C.R. visualization, C.R.; supervision, C.R.; project administration, C.R. and E.N. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Puddy, E.; Hill, C. Interpretation of the chest radiograph. *Contin. Educ. Anaesth. Crit. Care Pain* **2007**, *7*, 71–75. [CrossRef]
2. Schalekamp, S.; Van Ginneken, B.; Karssemeijer, N.; Schaefer-Prokop, C.M. Chest Radiography: New Technological Developments and Their Applications. *Semin. Respir. Crit. Care Med.* **2014**, *35*, 3–16. [PubMed]
3. Wright, F.W. *Radiology of the Chest and Related Conditions: Together with an Extensive Illustrative Collection of Radiographs, Conventional and Computed Tomograms, Isotope Studies, MRs, etc.on CD-ROM. (Cross-Platform for Mac and Windows Version 3.1 or later)*, 1st ed.; CRC Press: London, UK, 2022.
4. Jones, C.M.; Buchlak, Q.D.; Oakden-Rayner, L.; Milne, M.; Seah, J.; Esmaili, N.; Hachey, B. Chest radiographs and machine learning-Past, present and future. *J. Med. Imaging Radiat. Oncol.* **2021**, *65*, 538–544. [CrossRef]
5. Bruls, R.J.M.; Kwee, R.M. Workload for radiologists during on-call hours: Dramatic increase in the past 15 years. *Insights Imaging* **2020**, *11*, 121. [CrossRef] [PubMed]
6. Winder, M.; Owczarek, A.J.; Chudek, J.; Pilch-Kowalczyk, J.; Baron, J. Are We Overdoing It? Changes in Diagnostic Imaging Workload during the Years 2010–2020 including the Impact of the SARS-CoV-2 Pandemic. *Healthcare* **2021**, *9*, 1557. [CrossRef]
7. Xiong, Z.; Stiles, M.K.; Yao, Y.; Shi, R.; Nalar, A.; Hawson, J.; Lee, G.; Zhao, J. Automatic 3D Surface Reconstruction of the Left Atrium From Clinically Mapped Point Clouds Using Convolutional Neural Networks. *Front. Physiol.* **2022**, *13*, 880260. [CrossRef] [PubMed]

8.   Commission Notice—The 'Blue Guide' on the implementation of EU products rules 2016 (Text with EEA relevance) *C/2016/1958 OJ C 272, 26.7.2016, p. 1–149 (BG, ES, CS, DA, DE, ET, EL, EN, FR, HR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV)*. Available online: https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=uriserv%3AOJ.C_.2016.272.01.0001.01.ENG&toc=OJ%3AC%3A2016%3A272%3ATOC (accessed on 5 March 2023).

9.   Van Leeuwen, K.G.; Schalekamp, S.; Rutten, M.J.; van Ginneken, B.; de Rooij, M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur. Radiol.* **2021**, *31*, 3797–3804. [CrossRef]

10.   Schwartz, W.S.; Moyer, R.E. The management of massive tuberculous pneumonia; a therapeutic review of seventy-five cases. *Am. Rev. Tuberc.* **1951**, *64*, 41–49.

11.   WHO. *Chest Radiography Intuberculosis Detection: Summary of Current WHO Recommendations and Guidance on Programmatic Approaches*; World Health Organization: Geneva, Switzerland, 2016.

12.   Murphy, K.; Habib, S.S.; Zaidi, S.M.A.; Khowaja, S.; Khan, A.; Melendez, J.; Scholten, E.T.; Amad, F.; Schalekamp, S.; Verhagen, M.; et al. Computer aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system. *Sci. Rep.* **2020**, *10*, 5492. [CrossRef]

13.   Qin, Z.Z.; Barrett, R.; Ahmed, S.; Sarker, M.S.; Paul, K.; Adel, A.S.S.; Banu, S.; Creswell, J. Comparing different versions of computer-aided detection products when reading chest X-rays for tuberculosis. *PLoS Digit. Health* **2022**, *1*, e0000067. [CrossRef]

14.   Khan, F.A.; Majidulla, A.; Tavaziva, G.; Nazish, A.; Abidi, S.K.; Benedetti, A.; Menzies, D.; Johnston, J.C.; Khan, A.J.; Saeed, S. Chest x-ray analysis with deep learning-based software as a triage test for pulmonary tuberculosis: A prospective study of diagnostic accuracy for culture-confirmed disease. *Lancet Digit. Health* **2020**, *2*, e573–e581. [CrossRef]

15.   Fehr, J.; Konigorski, S.; Olivier, S.; Gunda, R.; Surujdeen, A.; Gareta, D.; Smit, T.; Baisley, K.; Moodley, S.; Moosa, Y.; et al. Computer-aided interpretation of chest radiography reveals the spectrum of tuberculosis in rural South Africa. *NPJ Digit. Med.* **2021**, *4*, 106. [CrossRef]

16.   Koesoemadinata, R.C.; Kranzer, K.; Livia, R.; Susilawati, N.; Annisa, J.; Soetedjo, N.N.M.; Ruslami, R.; Philipsen, R.; van Ginneken, B.; Soetikno, R.D.; et al. Computer-assisted chest radiography reading for tuberculosis screening in people living with diabetes mellitus. *Int. J. Tuberc. Lung Dis.* **2018**, *22*, 1088–1094. [CrossRef]

17.   Wali, A.; Safdar, N.; Manair, R.; Khan, M.D.; Khan, A.; Kurd, S.A.; Khalil, L. Early TB case detection by community-based mobile X-ray screening and Xpert testing in Balochistan. *Public Health Action* **2021**, *11*, 174–179. [CrossRef]

18.   Odume, B.; Chukwu, E.; Fawole, T.; Nwokoye, N.; Ogbudebe, C.; Chukwuogo, O.; Useni, S.; Dim, C.; Ubochioma, E.; Nongo, D.; et al. Portable digital X-ray for TB pre-diagnosis screening in rural communities in Nigeria. *Public Health Action* **2022**, *12*, 85–89.

19.   Philipsen, R.H.H.M.; Sánchez, C.I.; Maduskar, P.; Melendez, J.; Peters-Bax, L.; Peter, J.G.; Dawson, R.; Theron, G.; Dheda, K.; van Ginneken, B. Automated chest-radiography as a triage for Xpert testing in resource-constrained settings: A prospective study of diagnostic accuracy and costs. *Sci. Rep.* **2015**, *5*, 12215. [CrossRef] [PubMed]

20.   Melendez, J.; Hogeweg, L.; Sánchez, C.I.; Philipsen, R.H.H.M.; Aldridge, R.; Hayward, A.C.; Abubakar, I.; van Ginneken, B.; Story, A. Accuracy of an automated system for tuberculosis detection on chest radiographs in high-risk screening. *Int. J. Tuberc. Lung Dis.* **2018**, *22*, 567–571. [CrossRef] [PubMed]

21.   Qin, Z.Z.; Sander, M.S.; Rai, B.; Titahong, C.N.; Sudrungrot, S.; Laah, S.N.; Adhikari, L.M.; Carter, E.J.; Puri, L.; Codlin, A.J.; et al. Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Sci. Rep.* **2019**, *9*, 15000. [CrossRef]

22.   Philipsen, R.H.H.M.; Sánchez, C.I.; Melendez, J.; Lew, W.J.; Van Ginneken, B. Automated chest X-ray reading for tuberculosis in the Philippines to improve case detection: A cohort study. *Int. J. Tuberc. Lung Dis.* **2019**, *23*, 805–810. [CrossRef]

23.   Steiner, A.; Mangu, C.; Hombergh, J.V.D.; van Deutekom, H.; van Ginneken, B.; Clowes, P.; Mhimbira, F.; Mfinanga, S.; Rachow, A.; Reither, K.; et al. Screening for pulmonary tuberculosis in a Tanzanian prison and computer-aided interpretation of chest X-rays. *Public Health Action* **2015**, *5*, 249–254. [CrossRef] [PubMed]

24.   Santos, A.D.S.; de Oliveira, R.D.; Lemos, E.F.; Lima, F.; Cohen, T.; Cords, O.; Martinez, L.; Gonçalves, C.; Ko, A.; Andrews, J.R.; et al. Yield, Efficiency, and Costs of Mass Screening Algorithms for Tuberculosis in Brazilian Prisons. *Clin. Infect. Dis.* **2021**, *72*, 771–777. [CrossRef] [PubMed]

25.   Wali, A.; Khan, D.; Safdar, N.; Shawani, Z.; Fatima, R.; Yaqoob, A.; Qadir, A.; Ahmed, S.; Rashid, H.; Ahmed, B.; et al. Prevalence of tuberculosis, HIV/AIDS, and hepatitis; in a prison of Balochistan: A cross-sectional survey. *BMC Public Health* **2019**, *19*, 1631. [CrossRef] [PubMed]

26.   Breuninger, M.; Van Ginneken, B.; Philipsen, R.H.H.M.; Mhimbira, F.; Hella, J.J.; Lwilla, F.; Hombergh, J.V.D.; Ross, A.; Jugheli, L.; Wagner, D.; et al. Diagnostic Accuracy of Computer-Aided Detection of Pulmonary Tuberculosis in Chest Radiographs: A Validation Study from Sub-Saharan Africa. *PLoS ONE* **2014**, *9*, e106381. [CrossRef]

27.   Qin, Z.Z.; Ahmed, S.; Sarker, M.S.; Paul, K.; Adel, A.S.S.; Naheyan, T.; Barrett, R.; Banu, S.; Creswell, J. Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: An evaluation of five artificial intelligence algorithms. *Lancet Digit. Health* **2021**, *3*, e543–e554. [CrossRef]

28.   Nishtar, T.; Burki, S.; Ahmad, F.S.; Ahmad, T. Diagnostic accuracy of computer aided reading of chest x-ray in screening for pulmonary tuberculosis in comparison with Gene-Xpert. *Pak. J. Med Sci.* **2021**, *38*, 62–68. [CrossRef]

29.   Maduskar, P.; Muyoyeta, M.; Ayles, H.; Hogeweg, L.; Peters-Bax, L.; Van Ginneken, B. Detection of tuberculosis using digital chest radiography: Automated reading vs. interpretation by clinical officers. *Int. J. Tuberc. Lung Dis.* **2013**, *17*, 1613–1620.

30. Melendez, J.; Philipsen, R.H.H.M.; Chanda-Kapata, P.; Sunkutu, V.; Kapata, N.; Van Ginneken, B. Automatic versus human reading of chest X-rays in the Zambia National Tuberculosis Prevalence Survey. *Int. J. Tuberc. Lung Dis.* **2017**, *21*, 880–886. [CrossRef] [PubMed]
31. Del Ciello, A.; Franchi, P.; Contegiacomo, A.; Cicchetti, G.; Bonomo, L.; Larici, A.R. Missed lung cancer: When, where, and why? *Diagn. Interv. Radiol.* **2017**, *23*, 118–126. [CrossRef]
32. Bradley, S.H.; Bhartia, B.S.; Callister, M.E.; Hamilton, W.T.; Hatton, N.L.F.; Kennedy, M.P.; Mounce, L.T.; Shinkins, B.; Wheatstone, P.; Neal, R.D. Chest X-ray sensitivity and lung cancer outcomes: A retrospective observational study. *Br. J. Gen. Pract.* **2021**, *71*, e862–e868. [CrossRef]
33. Gampala, S.; Vankeshwaram, V.; Gadula, S.S.P. Is Artificial Intelligence the New Friend for Radiologists? A Review Article. *Cureus* **2020**, *12*, e11137. [CrossRef]
34. Nam, J.G.; Park, S.; Hwang, E.J.; Lee, J.H.; Jin, K.-N.; Lim, K.Y.; Vu, T.H.; Sohn, J.H.; Hwang, S.; Goo, J.M.; et al. Development and Validation of Deep Learning–based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. *Radiology* **2019**, *290*, 218–228. [CrossRef] [PubMed]
35. Sim, Y.; Chung, M.J.; Kotter, E.; Yune, S.; Kim, M.; Do, S.; Han, K.; Kim, H.; Yang, S.; Lee, D.-J.; et al. Deep Convolutional Neural Network–based Software Improves Radiologist Detection of Malignant Lung Nodules on Chest Radiographs. *Radiology* **2020**, *294*, 199–209. [CrossRef] [PubMed]
36. Homayounieh, F.; Digumarthy, S.; Ebrahimian, S.; Rueckel, J.; Hoppe, B.F.; Sabel, B.O.; Conjeti, S.; Ridder, K.; Sistermanns, M.; Wang, L.; et al. An Artificial Intelligence–Based Chest X-ray Model on Human Nodule Detection Accuracy From a Multicenter Study. *JAMA Netw. Open* **2021**, *4*, e2141096. [CrossRef]
37. Kim, H.; Park, C.M.; Goo, J.M. Test-retest reproducibility of a deep learning–based automatic detection algorithm for the chest radiograph. *Eur. Radiol.* **2020**, *30*, 2346–2355. [CrossRef]
38. Kim, E.Y.; Kim, Y.J.; Choi, W.J.; Lee, G.P.; Choi, Y.R.; Jin, K.N.; Cho, Y.J. Performance of a deep-learning algorithm for referable thoracic abnormalities on chest radiographs: A multicenter study of a health screening cohort. *PLoS ONE* **2021**, *16*, e0246472.
39. Liang, C.H.; Liu, Y.C.; Wu, M.T.; Garcia-Castro, F.; Alberich-Bayarri, A.; Wu, F.Z. Identifying pulmonary nodules or masses on chest radiography using deep learning: External validation and strategies to improve clinical practice. *Clin. Radiol.* **2020**, *75*, 38–45. [CrossRef]
40. Teng, P.H.; Liang, C.H.; Lin, Y.; Alberich-Bayarri, A.; González, R.L.; Li, P.W.; Weng, Y.H.; Chen, Y.T.; Lin, C.H.; Chou, K.J.; et al. Performance and educational training of radiographers in lung nodule or mass detection: Retrospective comparison with different deep learning algorithms. *Medicine* **2021**, *100*, e26270. [CrossRef] [PubMed]
41. Tam, M.; Dyer, T.; Dissez, G.; Morgan, T.N.; Hughes, M.; Illes, J.; Rasalingham, R. Augmenting lung cancer diagnosis on chest radiographs: Positioning artificial intelligence to improve radiologist performance. *Clin. Radiol.* **2021**, *76*, 607–614. [CrossRef] [PubMed]
42. Shah, P.K.; Austin, J.H.M.; White, C.S.; Patel, P.; Haramati, L.B.; Pearson, G.D.N.; Shiau, M.C.; Berkmen, Y.M. Missed Non–Small Cell Lung Cancer: Radiographic Findings of Potentially Resectable Lesions Evident Only in Retrospect. *Radiology* **2003**, *226*, 235–241. [CrossRef]
43. Schalekamp, S.; van Ginneken, B.; Meiss, L.; Peters-Bax, L.; Quekel, L.G.; Snoeren, M.M.; Tiehuis, A.M.; Wittenberg, R.; Karssemeijer, N.; Schaefer-Prokop, C.M. Bone suppressed images improve radiologists' detection performance for pulmonary nodules in chest radiographs. *Eur. J. Radiol.* **2013**, *82*, 2399–2405. [CrossRef]
44. Schalekamp, S.; Karssemeijer, N.; Cats, A.M.; De Hoop, B.; Geurts, B.H.; Berger-Hartog, O.; van Ginneken, B.; Schaefer-Prokop, C.M. The Effect of Supplementary Bone-Suppressed Chest Radiographs on the Assessment of a Variety of Common Pulmonary Abnormalities: Results of an Observer Study. *J. Thorac. Imaging* **2016**, *31*, 119–125. [CrossRef]
45. Freedman, M.T.; Lo, S.-C.B.; Seibel, J.C.; Bromley, C.M. Lung Nodules: Improved Detection with Software That Suppresses the Rib and Clavicle on Chest Radiographs. *Radiology* **2011**, *260*, 265–273. [CrossRef]
46. Schalekamp, S.; Van Ginneken, B.; Koedam, E.; Snoeren, M.M.; Tiehuis, A.M.; Wittenberg, R.; Karssemeijer, N.; Schaefer-Prokop, C.M. Computer-aided Detection Improves Detection of Pulmonary Nodules in Chest Radiographs beyond the Support by Bone-suppressed Images. *Radiology* **2014**, *272*, 252–261. [CrossRef] [PubMed]
47. Schalekamp, S.; Van Ginneken, B.; Heggelman, B.; Imhof-Tas, M.; Somers, I.; Brink, M.; Spee, M.; Schaefer-Prokop, C.; Karssemeijer, N. New methods for using computer-aided detection information for the detection of lung nodules on chest radiographs. *Br. J. Radiol.* **2014**, *87*, 20140015. [CrossRef]
48. Park, S.; Lee, S.M.; Lee, K.H.; Jung, K.-H.; Bae, W.; Choe, J.; Seo, J.B. Deep learning-based detection system for multiclass lesions on chest radiographs: Comparison with observer readings. *Eur. Radiol.* **2019**, *30*, 1359–1368. [CrossRef]
49. Wisbach, G.G.; Sise, M.J.; Sack, D.I.; Swanson, S.M.; Sundquist, S.M.; Paci, G.M.; Kingdon, K.M.; Kaminski, S.S. What is the role of chest X-ray in the initial assessment of stable trauma patients? *J. Trauma* **2007**, *62*, 74–78; discussion 78–79. [CrossRef] [PubMed]
50. Gipson, J.; Tang, V.; Seah, J.; Kavnoudias, H.; Zia, A.; Lee, R.; Mitra, B.; Clements, W. Diagnostic accuracy of a commercially available deep-learning algorithm in supine chest radiographs following trauma. *Br. J. Radiol.* **2022**, *95*, 20210979. [CrossRef] [PubMed]
51. Hwang, E.J.; Nam, J.G.; Lim, W.H.; Park, S.J.; Jeong, Y.S.; Kang, J.H.; Hong, E.K.; Kim, T.M.; Goo, J.M.; Park, S.; et al. Deep Learning for Chest Radiograph Diagnosis in the Emergency Department. *Radiology* **2019**, *293*, 573–580. [CrossRef]
52. Jones, C.M.; Danaher, L.; Milne, M.R.; Tang, C.; Seah, J.; Oakden-Rayner, L.; Johnson, A.; Buchlak, Q.D.; Esmaili, N. Assessment of the effect of a comprehensive chest radiograph deep learning model on radiologist reports and patient outcomes: A real-world observational study. *BMJ Open* **2021**, *11*, e052902. [CrossRef]

53. Shin, H.J.; Son, N.-H.; Kim, M.J.; Kim, E.-K. Diagnostic performance of artificial intelligence approved for adults for the interpretation of pediatric chest radiographs. *Sci. Rep.* **2022**, *12*, 10215. [CrossRef]

54. Kim, E.Y.; Kim, Y.J.; Choi, W.-J.; Jeon, J.S.; Kim, M.Y.; Oh, D.H.; Jin, K.N.; Cho, Y.J. Concordance rate of radiologists and a commercialized deep-learning solution for chest X-ray: Real-world experience with a multicenter health screening cohort. *PLoS ONE* **2022**, *17*, e0264383. [CrossRef]

55. Wang, L.; Zhang, Y.; Wang, D.; Tong, X.; Liu, T.; Zhang, S.; Huang, J.; Zhang, L.; Chen, L.; Fan, H.; et al. Artificial Intelligence for COVID-19: A Systematic Review. *Front. Med.* **2021**, *8*, 704256. [CrossRef]

56. Khan, M.; Mehran, M.T.; Haq, Z.U.; Ullah, Z.; Naqvi, S.R.; Ihsan, M.; Abbass, H. Applications of artificial intelligence in COVID-19 pandemic: A comprehensive review. *Expert Syst. Appl.* **2021**, *185*, 115695. [CrossRef] [PubMed]

57. Mushtaq, J.; Pennella, R.; Lavalle, S.; Colarieti, A.; Steidler, S.; Martinenghi, C.M.A.; Palumbo, D.; Esposito, A.; Rovere-Querini, P.; Tresoldi, M.; et al. Initial chest radiographs and artificial intelligence (AI) predict clinical outcomes in COVID-19 patients: Analysis of 697 Italian patients. *Eur. Radiol.* **2020**, *31*, 1770–1779. [CrossRef]

58. Tovar, M.A.; Puma, D.; Palomino, S.; Peinado, J.; Llanos, F.; Martinelli, C.; Jimenez, J.; Calderon, R.; Yuen, C.M.; Lecca, L. Integrated screening and testing for TB and COVID-19 in Peru. *Public Health Action* **2022**, *12*, 7–9. [CrossRef] [PubMed]

59. Schalekamp, S.; Van Ginneken, B.; Berk, I.A.H.V.D.; Hartmann, I.J.C.; Snoeren, M.M.; Odink, A.E.; Van Lankeren, W.; Pegge, S.A.H.; Schijf, L.J.; Karssemeijer, N.; et al. Bone Suppression Increases the Visibility of Invasive Pulmonary Aspergillosis in Chest Radiographs. *PLoS ONE* **2014**, *9*, e108551. [CrossRef] [PubMed]

60. Fanni, S.C.; Shalaby, S.M.; Neri, E. The Role of Artificial Intelligence (AI) in the Management of Geriatric Patients. In *Imaging in Geriatrics*; Guglielmi, G., Maas, M., Eds.; Springer International Publishing: Cham, Switzerland, 2023; pp. 445–456.

61. Coppola, F.; Faggioni, L.; Gabelloni, M.; De Vietro, F.; Mendola, V.; Cattabriga, A.; Cocozza, M.A.; Vara, G.; Piccinino, A.; Lo Monaco, S.; et al. Human, All Too Human? An All-Around Appraisal of the "Artificial Intelligence Revolution" in Medical Imaging. *Front. Psychol.* **2021**, *12*, 710982.

62. Daneshjou, R.; Smith, M.P.; Sun, M.D.; Rotemberg, V.; Zou, J. Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review. *JAMA Derm.* **2021**, *157*, 1362–1369. [CrossRef]

63. Erondu, O.F. *Challenges and Peculiarities of Paediatric Imaging, in Medical Imaging in Clinical Practice*; IntechOpen: London, UK, 2013.