Human-Centric Intelligent Systems (2022) 2:31–43
https://doi.org/10.1007/s44230-022-00002-2

REVIEW ARTICLE

Review on Pneumonia Image Detection: A Machine Learning Approach

Amer Kareem[1] · **Haiming Liu**[1] · **Paul Sant**[1]

Received: 25 November 2021 / Accepted: 5 February 2022 / Published online: 4 May 2022

Abstract

This paper surveys and examines how computer-aided techniques can be deployed in detecting pneumonia. It also suggests a hybrid model that can effectively detect pneumonia while using the real-time medical image data in a privacy-preserving manner. This paper will explore how various preprocessing techniques such as X-rays can detect and classify multiple diseases. The survey also examines how different machine learning technologies like convolution neural network (CNN), k-nearest neighbor (KNN), RESNET, CheXNet, DECNET and artificial neural network (ANN) can be used in detecting pneumonia disease. In this article, we have performed a comprehensive review of the literature to find how we can combine hospitals and medical institutions to train the machine learning models from their datasets so that the ML algorithms can detect disease more efficiently and correctly. We have proposed the future work of using transfer learning combined with federated knowledge that could help the medical institutions and hospitals form a combined approach of performing medical image detection using real-time datasets. We have also explored the scope, future work and limitations of the proposed solution.

Keywords Machine learning techniques · Pneumonia detection · CNN · Chest X-ray · Transfer learning · Federated learning · k-Nearest neighbors (KNN) · Artificial neural network (ANN) · DECnet

1 Introduction

The number of individuals suffering from pneumonia is approximately more than 450 million a year [1]. It is 7% of the overall population around the globe. Each year more than four million people die from Pneumonia [2]. Pneumonia disease is prevalent among young children below 5 years old [3]. According to the report released by "our World in data" [4], children below five have the highest death rate caused by pneumonia (Fig. 1). In 2017, 808,920 children died due to pneumonia, and this figure is 16 folds more than the deaths caused by cancer a year and ten folds higher than people who died from HIV.

According to the report released during World Pneumonia Day, it is estimated that more than 11 million infant children below the age of 5 years are likely to die from pneumonia by the year 2030 [5]. In the early nineteenth century, pneumonia was considered one of the significant causes of death amongst people.

In the past, medical doctors relied on several methods such as clinical examination, medical history, and chest X-rays to diagnose patients suffering from pneumonia. Nowadays, Chest-X-rays have become increasingly cheaper due to rapid advancements in technologies such as biomedical equipment. The Chest X-ray is commonly used in detecting pulmonary diseases like pneumonia. The problem of lack of experts can be addressed through the use of different computer-aided diagnosis techniques. Technological advancements in artificial intelligence (AI) have proven to be helpful in the diagnosis of disease. For instance, techniques like CNN are utilised for classifying Chest-X-rays in order to determine whether pneumonia is present. Some of the exciting research has been done in areas like abnormal-patterns detection [6–13], biometric recognition [14, 15], trauma seriousness valuation [16–19], accident prevention at the

✉ Amer Kareem
amer.kareem@study.beds.ac.uk

Haiming Liu
haiming.liu@beds.ac.uk

Paul Sant
paul.sant@beds.ac.uk

[1] University of Bedfordshire, Luton, UK

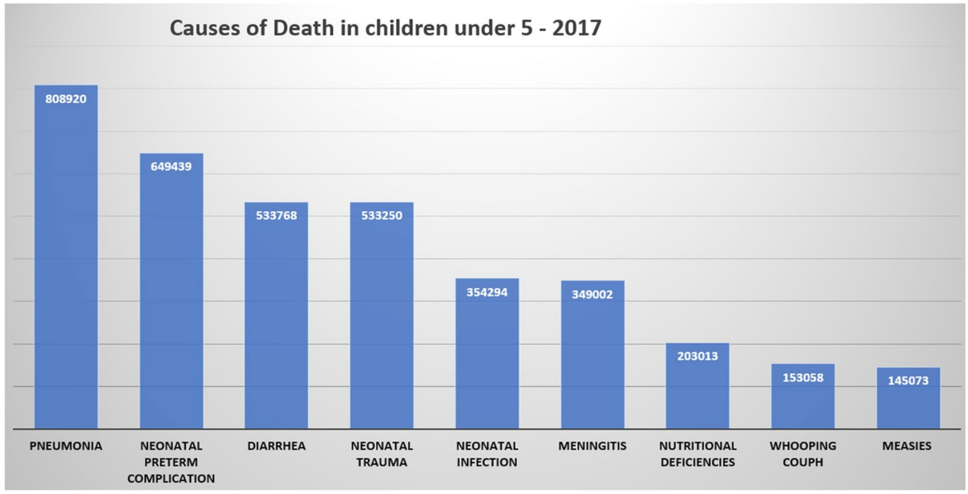

Fig. 1 This figure shows the number for causes of death in the Children under the age of 5 during 2017. It shows the pneumonia has the highest number of deaths in contrast to other common diseases

airport [20], predicting efficiency in information using ANN [21] and diagnoses of bone pathology [22]. However, the higher divergence in the image features impacts the retrieval accuracy [1].

1.1 Scope and Motivation

This review paper is inspired by machine learning methods that can promise an effective pneumonia image detection. While considering the ML techniques, the primary concern is the datasets. Lab-based data is always limited; therefore, there is a need to have realtime data that is always sufficient and remain updated with the ML for practical training. Hospitals and medical institutions are unable to share data due to GDPR [23]. In one of the reports by Digital health records, 24.3 million image data have been found compromised by cyber-attacks [24]. In this review paper, we have examined the number of ML techniques that have been used by the researcher as state of the art for effective medical image detection and image data security.

The review paper is based on the below problem statements:

- Medical images are complex and heterogeneous compared to standard images; therefore, it is challenging to propose an effective model with restricted data availability [25].
- The lab-based datasets are limited to training the effective ML model [26].
- The heterogeneous nature of the medical images makes it harder to train the ML model from lab-based datasets.
- Understanding the medical image patterns is challenging for researchers [27].
- The use of real-time data can help to improve the model. However, data sharing for hospitals and medical institutions is challenging [27]. phase.

2 Literature Review

Artificial intelligence techniques can be used to diagnose various diseases such as pneumonia [28]. Research has been done by using multiple methods of machine learning techniques for detecting medical diseases. In this section, we have illustrated the work done in the field of medical image detection. We have reviewed the finding based on strengths and limitations. Concerning medical image detection, various datasets have been used to build up an effective model.

2.1 Deep Learning Methods

Medical image detection is a complicated task; therefore, an effective approach is needed. Deep learning is one of the techniques that can be used for the training of medical image datasets. In the study, deep learning model of RestNet-101 and RestNet50 was used for pneumonia detection [29]. While considering these techniques, it has resulted in different results based on individual features. Therefore, to compensate this difference, an effective deep learner strategy was introduced that involves the combination of these techniques. In this study, dataset of 14,863 X-ray images was used and the achieved precision is 96%. Although the model output good precision, however it possesses limitations due to the complexity of combining the RestNet models that can effect the precision when larger dataset is considered in a real time scenario. The experiment was performed to demonstrate how deep learning models can diagnose diseases [3]. In this case, the deep neural network was used to aid in diagnosing 14 diseases. The ChestXray14 database was used and trained with DenseNet and reduced pairwise error to relate their outcomes in diagnosing diseases. The architecture was developed to help in detecting and classifying diseases using multilabels. In addition, the cascade network aided in making all possible predictions by comparing several previous levels, which are used as inputs in each successive level in the Cascade network. The level-6 cascading network was used in both PWE loss and cross-entropy. The study results indicated that the Cascade network helped in increasing the performance classifiers. The use of DenseNets has produced positive outcomes that include reducing the gradient problem, reinforcing the features propagation, and reducing the parameters. However, this model is not capable of modelling the inner class.

2.1.1 Artificial Neural Network

Artificial neural network (ANN) effectively detects and diagnoses various chest diseases like breast cancer, tuberculosis, and pneumonia infection [30]. Different preprocessing techniques was used to eliminate any irrelevant data. Strategies for enhancing the imaging process was used, including Equalisation of the histogram and image filtering. These techniques are crucial in reducing noises and bringing images into sharper focus, thus promoting easy detection of pneumonia. Lung segmentation is an important area of interest in diagnosing pneumonia infection. Various diagnostic features like perimeter, areas, irregularity index, equal diameter, and statical methods like standard deviation and entropy were extracted and used to classify the images obtained to help detect the presence of pneumonia. The neural network is used in categorising images to assist in detecting lung diseases. The dataset used in this study was obtained from 80 patients. The feed-forward neural network helped to attain an accuracy of 92%. However, if changes were made in the position and size of CXR, the accuracy of results obtained declined significantly. Although the study suggests the use of pattern recognition techniques works well in medical image detection that includes chest diseases, the proposed method has limitations that include the alteration in the size and positions of chest x-ray image, which results in ineffective detection. Therefore, while considering this drawback, it is essential to devise a neural network model capable of detecting any changes in the size and structure of the images.

2.1.2 Conventional Neural Network

Medical image classification follows complex patterns recognition; therefore, a highly effective ML model is needed. To make it possible, deep learning plays an important role, and among them, CNN is one of the effective approaches for pattern recognition due to its layering topology. CNN structure is highly dense that consists of a stack of layers with its heights, widths, and depths. The depth also allows sharing weight [31]. CNN is trained by providing the input, and the various parameter is learnt to define the individual output. The idea of using the CNN approach is to limit the network distinctions between the predicted and actual outcomes (Fig. 2). The below figure demonstrates the architecture of the CNN model.

The study [32] shows that X-ray images are very effective in detecting and identifying the presence of diseases such as lung cancer. The study was followed in two steps: sequencing an image processing method to remove noise and reduce the area of concern, one nodule suspected of occupying space of 65 × 65 square. The squared pixel obtained was considered as the device's data. The intensity of pixels obtained was collected in a file. The next stage involved training the system. The database was grouped into various distinct categories, and the information obtained was utilised in training and checking the process. Following the second step, the researchers used CNN to examine pixels and numerical feature-based inputs. When the pixel-based method was used, an accuracy rate of 96% was attained, and an 88% accuracy rate was achieved when the feature-based technique was considered. Although the use of pixel-based and feature-based techniques has produced effective results, however, when it comes to the ML model to deploy

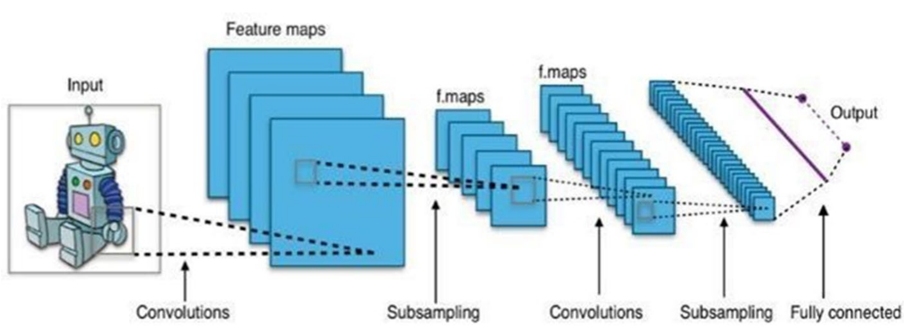

Fig. 2 This figure displays convolutional neural network (CNN) architecture [32]. The process of data input and output is followed across series of layering topology where each set of layers perform its task. Feature map layers perform feature engineering and filter map perform filtration of data

in real-time, it has limitations while considering that these methods have drawbacks of ignoring the feature dependencies. The primary reason for this involves the lack of interaction with the classifiers. Therefore, this method can cause feature selection issues in terms of ranking where it becomes harder to decide the exact selection of features and ignore the noise.

The CNN technique was applied for performing diagnostic of thorax X-rays [33].

Thorax is a type of disease that affects small localised areas. The poor alignment of CXR occurred due to the failure of network performance. The study proposed a three-branch AG-CNN framework that is crucial in avoiding noise and improving alignment from various regions infected by the disease. In addition, it integrates global branches to help in minimising local chapters in the lost discriminatory signs. The use of chestXray-14 datasets has enabled us to understand various regions of CNN. This method has produced the AUC of 0.87 while considering this dataset. However, this method has a limitation when it comes to parameter changes. It is not flexible to any parameter alterations that can prohibit the model from predicting the variety of data. The experiment was performed with the CheXNet algorithm with 121 layers of CNN and chest X-ray images as inputs in diagnosing and detecting the presence of pneumonia infection [34]. The dataset from various samples of patients was validated and tested using the training model. Then the images were compressed and resized to 224 × 224, normalised, and trained and augmented. It was combined with the modified alexnet framework (MAN), resulting in the model's adequate performance. However, this model has various lacking that includes the inability of the model to detect the subtypes of the lung disease, and instead it just detects the pneumonia disease. As the study was made on the classification of disease, the disease's segmentation is not identifiable.

The effectiveness of a CNN method was analysed in diagnosing tuberculosis disease by Chest X-rays, AlexNet as well as GoogleNet [35]. To carry out this experiment, two individual DCNN was utilised to help determine and detect the presence of respiratory conditions and other nutritious object. The untrained and trained network was utilised in determining the presence of pneumonia disease in ImageNet. Chest radiographs from various datasets were used to perform validation and testing processes. The Chest radiograph images were resized into 256 × 256 pixels and then converted into a portable network Graphic format that was then loaded into a computer learning machine with a Linux Operating system. The study suggested that the chest radiograph images were effective in detecting tuberculosis disease using 0.99 AUC. The pre-trained ImageNet DCNNs performed better when compared to the untrained networks with daily images. DCNN is effective in the detection of TB while considering other pulmonary diseases; it possesses limitations. These limitations are based on the fact that DCNN requires a higher number of parameters, and also, it is highly computationally intensive, which requires more research to make it adaptive to use effectively for detecting a variety of diseases.

A model was proposed based on the CNN approach for examining interstitial lung's disease [36] and other lung inflammatory disorders. The dataset of 14,696 image areas was obtained through 120 CT scans from various healthcare organisations, including pneumonia, cancer, and Tuberculosis images. In addition, a deep CNN model known as AlexNet was proposed. The model is comprised of five layers in conjunction with LeakyRevLu activations. It was also contrasted with several methods like VGG-Net and LeNet. The accuracy rate of 85.5% using the CNN Model was achieved. This model effectively detects the diseases; however, it has limitations that involve a higher number of parameters for training the model that could potentially result in overfitting the model. Therefore, a better approach is needed to avoid the requirements of higher parameters for model training.

The ResNet CNN template was used in differentiating between benign and malignant nodules when diagnosing lung cancer [1]. To carry out this experiment, the ResNet CNN template was used in identifying radiographs with a sensitivity of 92% to determine lung cancer in the nodules. The template also enabled to recognise general regions of the lung cancer nodules. However, they were not able to identify the specific positions of these nodules. The JSRT dataset was used for classifying radiographs by examining the accuracy of the dataset being tested. Determining the exact area of interest is lacking in this work.

The experiment was performed to examine ChestX-ray that could classify various diseases [37]. The results indicated that thoracic diseases could be identified using a unified multi-labelled image classification and infection locality procedure. These methods are widely used in determining thoracic diseases. These systems effectively perform the detection of several abnormalities and produce a boundary box. In addition, it can detect pathologies present in X-ray images, particularly in the DCNN system. The DCNN system was used in locating these pathologies in the body. The quantisation technique was used in the classification process while supervised learning like SVM can help achieve higher retrieval performance [31]. Higher training data and intensive GPU has been a limitation in the proposed work. The drawback of using this model involves the overfitting and spatial invariance of the input data.

It was examined how advanced calculation can address the programmed illustration of thoracic diseases using X-ray images [38]. Various approaches are used in the advancement of unused measure of illustrative programmed therapeutic images. The system incorporated four main steps; the image preprocessing step is used to classify and determine the accuracy of disease location in the lung. It also entails lung field division that allows the area of disease in the interior of lung borders and distinguishing each pathology depending on the changes in the organ shape. It also highlights the calculation performed on the therapeutic images. The classification method proved to be effective in diagnosing thorax diseases. The MIL-based approach helped in improving the preparation of classification algorithms. This strategy of following the four steps has resulted in the system's effective performance while considering feature engineering. However, it is ineffective in the process of detecting thoracic disease. Machine learning can be used to detect CAD in lungs using approach based on rules [39]. Typically, the rule-based approach uses a deep learning technique widely in image analysis and rib detection. This method is used primarily in establishing candidates using computer-assisted detecting systems like Deep learning technology. It is effective in overall image analyses; however, it has the limitation of identifying the certain image classes used in CT scans.

The experiment was performed on finding the effectiveness of CNN in detecting and distinguishing paediatric CXR, especially between bacterial and viral forms [40]. In this case, the visualisation techniques were used to locate various regions of interest, which was considered crucial in modelling predictions commonly used as inputs in the predicted classes. The visualisation method also helped in evaluating the quality of models used to carry out tasks statistically. The study results revealed that the VGG16 model effectively detected disease and differentiated between bacterial and viral pneumonia since it has a higher accuracy rate of 96.2%. The model is widely used in performance metrics; it effectively enhances the generalisation of results. However, this model is slower in performance while training the data and also, the overall architecture of the model is quite significant that constitute higher disk space and network bandwidth.

The two-step model in examining high-resolution medical images was used [41]. Medical images helped to exploit statistical dependency between various labels that are commonly essential in promoting the accuracy of disease detection. The LSTM and dataset of 14 chest x-ray was used in determining the trends and patterns in pathologies. The 2d convent was used in encoding and decoding with the aid of RNN-based activation function. In this work, an effective approach end of end neural network has been adopted; however, the proposed work is not quite capable because the experiment constitutes smaller datasets.

2.2 Privacy-Preserving Techniques for Image Detection

Training ML model demands large volume of data. Relying on lab-based data is ineffective as it has limited data, and also the medical image data is heterogeneous, and the ML model requires a continuous update for efficient training. The solution to solve this problem is to use real-time datasets from hospitals and medical institutions. However, maintaining privacy and confidentiality is challenging while following GDPR law [23]. Therefore, it is required to have a framework for using the real-time datasets while following GDPR rules and regulations.

Some of the work has been proposed that could remain intact the privacy of the data, including gossip learning, federated learning, and Blockchain technologies. Gossip learning is a decentralised method that can be used for data security. Experiments have been conducted to contrast the difference between gossip learning and federated learning [42]. The data was taken from the cell phones that including the network coverage and network distortions. Then the data was used to train the ML model in both frameworks of gossip learning and federated learning. It was analysed that federated learning performs better than gossip learning while considering privacy in terms of scalability, semi centralised nature and instant operation. However, on the other hand, gossip learning was slower in information exchange, and due to restricted messages size, scalability was an issue. For an effective medical image detection system for a real-time dataset, it is required to have a privacy-preserving framework that is highly scalable and faster.

The application of federated learning is limited because it has been recently introduced. Experiments have been conducted on using this technology on the electronic health record (EHR) on real-life medical data to predict disease and other research purposes [43]. In the experiment, the data was locally trained at each geographic location in hospitals and medical institutions. The model had an effective outcome when it was locally trained without sharing the data, and the trained model from various locations is aggregated together at a centralised location. This is an effective solution for training the data without sharing it. It is highly scalable, and the cyclic process helps to learn the new patterns as in the case of medical images where the heterogeneity is an issue, and the ML model requires updated data for effective training.

Blockchain is a decentralised technology that is used for data privacy utilising cryptographic elements. Experiments have been done on using blockchain technology for medical data and transactions information [44]. The results have shown effective results in retaining the privacy of the transactional details using cryptographic techniques of blockchain as the trial was performed on real-life datasets. The limitation of this research involves the slower process, scalability issues, processor-intensive, and while considering the real-time datasets, the slower process leads to the system's inefficiency [45]. Therefore, blockchain is not a considerable solution for keeping data private for considering medical image detection in real-time.

2.3 Other Techniques

The usefulness of computer-aided techniques was studied in detecting lung tuberculosis [46]. Examining various parameters like reducing patient waiting times was considered to obtain an X-ray and diagnosis lung tuberculosis. To perform diagnosis, the radiologists carried out a visual examination on textual features of thoracic X-ray images. They also used the principal component analysis method in measuring the outcomes of the study. It was identified, classified, and differentiated between TB and non-TB objects centered on various arithmetical feature from experiment. The challenge of considering the PCA includes the lower interpretability issues, and also data organisation is an essential requirement for PCA to work effectively. PCA finds linear correlation among the variable, which is not ideal in many cases.

The JSRT dataset was used in research, comprised of 247 X-ray images with various lung nodules to determine the presence of pneumonia infection. The finding from the JSRT dataset established that a small dataset was unbalanced since it was present and absent in some nodules. The dataset varied widely concerning the type of lung nodule, size, and distribution. Smaller datasets cause low precision and recall when used in realtime. The multiple JSRT datasets were extracted [28]. The bone shadow was removed to obtain the BSE-JSRT dataset as group 1. The JSRT dataset was segmented into various sets as group-2 and group-3. The dataset was segmented by removing the right and left lungs in the normal CXRs. The T-NSE was also released in outliers, including abnormal tiny lungs and other inclusions surrounding the heart regions provided in the JSRT as group-4. Then the datasets that were collected in performing the execution validation process was used. The most accurate dataset obtained was group-4, approximately 0.71, while the lowest was group-3, which stood at 0.56. The results of bone shadow exclusion in group 1 demonstrated a very little increment inaccuracy (0.65) compared to the original dataset. It can be observed from the effects that the smaller datasets (247 X-ray images) are not quite effective in achieving the higher accuracy; while performing the medical image detection process in the Real World, it is essential to consider the larger datasets, so the model is trained to detect the heterogeneity in the medial images in the production environment.

In the experiment [47], sound of the cough was used for diagnosing the pneumonia.

The sound of patient’s cough was taken by mobile device recorders, afterwards the wavelet sound decomposition was made using various arithmetical standards in classification pneumonia based on sound of the cough. The MATLAB R200a software was used to carry out programming. It was found that the signal analysis threshold effectively classified cough to determine whether pneumonia was present or not. The use of wavelet transform is computationally intensive, and also discretisation is the drawback that needs to be highlighted in the proposed work.

A hybrid technique was proposed for detecting and determining the pneumonia disease. In this case, the dataset that contained 20 frame, 19 non-cavity, and 110 standardised set cavities was considered. The hybrid result indicated an accuracy rate of 85.35% in detecting pneumonia in the lungs [48]. This method is effective in the detection of TB; however, it is ineffective in cavity detection.

2.4 Table of Comparison

Table 1 shows the list of the related work on the key review factors, namely disease, algorithm applied, evaluation method, dataset, pros and cons of the related work.

2.5 Evaluation Methods

The evaluation methods of the related work listed in Table 1. In an effort to compare the proposed method with the related work, the key evaluation method will be employed for the evaluation of the proposed method. This section will discuss the key evaluation methods, namely accuracy, precision, recall, F1 score, ROC (receiver operating characteristic) and AUC (area under curve).

2.5.1 Accuracy

Accuracy will justify the amount of predicted datapoints with respect to the rest of datapoints. It is used to identify the performance of the model with respect to all classes [51]. In regard to our proposed model, the performance evaluation of accuracy will help to determine the total number of accurate predictions among the total amount of predictions. It is represented as:

$$\text{Accuracy} = \text{number of correct predictions}/\\ \text{total number of predictions.}$$

Accuracy will help to determine the efficiency of the proposed model. It will help to analyse the effectiveness of the model with respect to other literatures in terms of correct predictions.

Table 1 Comparison of different machine learning algorithms for detecting diseases

References	Disease	Algo	Accuracy	Dataset	Pros	Cons
[28]	Lung cancer	CNN	78%	JSRT	Higher accuracy in large datasets	The process constitute smaller datasets
[30]	Pneumonia, TB, lung cancer	ANN	92%	Sasoo Hospital	Capable of detecting multiple chest diseases	Ineffective whenever there is a change in CXR image
[32]	Lung cancer	ANN	96% (Pixel) 88% (feature)	JSRT	Effective way of nodules detection	Nodule detection is difficult due to the surrounding vessel and rib
[33]	Thorax disease	CNN	AUC: 0.871	Chest X- ray-14	Effective results in detection	It is not quite flexible while considering the parameters change
[34]	Pneumonia	DCNN	AUC: 0.76	ChectXray 14	It can possibly detect the availability of 14 various pathological class	Very limited radiographs were available
[49]	Pneumonia	CNN (RestNet-50)	Internals AUC: 0.931 Externals AUC: 0.815	NIH IU MSH	Effective Accuracy	CNN is not as effective in external data in contrast to internal-data
[35]	TB	AlexNet/Goo gleNet	AUC: 0.99	1007-chest radiograph	Effective performance of ImageNet in contrast to untrained procedures	This procedure is limited to the identification of TB only
[5]	Pneumonia	Wavelet	99.70	COPD dataset	Effective and affordable	It has utilised relatively fewer inputs
[3]	Thoracic diseases	DenseNet161	(AUC) 0.7876	ChestXRay14	Effective performance	It is not capable of modelling the innerclass
[1]	Lungs cancer	CNN	68%	JSRT	Effective performance	It is ineffective in determining the exact location
[37]	Thoracic diseases	DCNN	81%	ChestXray-8	It can identify the lungs region and boundaries	The threshold was not effective in the experiments
[38]	Thoracic diseases	CAD	AUC 0.778	Chest X-ray 14	The higher performance of image features detection	The identification of thorax disease is ineffective
[39]	Lung Disease	Rule-based techniques	96.2%	JSRT	ConvNets is better in feature extraction	CT is not good
[40]	Pneumonia	VGG-16	96.2%	Chest X-ray 14	Effective accuracy	The datasets training was not correct
[46]	TB	PCA	95.7%	Sardjito Hospital	Preprocessed images can be utilised	Fewer datasets were considered
[36]	Lung Disease	CNN	85.5%	14,696 CT scans images	This technique can easily be used for training the extra textural-lung pattern	A higher number of parameters slowed the process
[48]	TB	Hybrid techniques	85.35%	Cavity set: 20 Non-cavity set: 19	Effective procedure for detection of TB	Cavity detection was failed
[43]	Various disease	Federated Learning	95.02%	EHR	Privacypreserving approach	System is heterogenous

Table 1 (continued)

References	Disease	Algo	Accuracy	Dataset	Pros	Cons
[44]	Various disease	Blockchain	95.08%	EHR	Privacypreserving approach	Time-consuming and privacy limitations
[50]	Various disease	Gossip Learning	95% LR 96% SVM	Spambase binary classification	Privacypreserving approach	Not effective in real-time datasets

2.5.2 Precision

Precision will involve the number of positive predictions of the model, that means precision is enhanced when the amount of correct positive predictions is higher and also the total number of incorrect positive predictions are fewer [52]. Precision is abbreviated as:

$$\text{Precision} = \text{true positive}/(\text{true positive} + \text{false positive}),$$

where true positive represents the correct prediction of positive class and false positive represents the correct prediction of negative class. In regard to the proposed model, the evaluation matrix of precision will help to compare the trustiness of model in terms of classifying the positive samples correctly with the other state of art.

2.5.3 Recall

Recall will compare the correct identification of positive sample with respect to all the available positive samples. It is involved in detecting the positive class and it is apart from the classification of negative samples [52]. Recall is represented as:

$$\text{Recall} = \text{true positive}/(\text{true positive} + \text{false negative}),$$

where true positive represents the correct prediction of positive class and false negative represents the incorrect prediction of negative class. The proposed model will compare the number of positive samples being correctly classified.

2.5.4 F1 score

Mean of the precision and recall will be compared by the F1 score and it is represented as:

$$\text{F1} := 2/(1/\text{precision} + 1/\text{recall}).$$

The evaluation metrics of F1 score is used to contrast the performance of 2 classifiers [53]. For example, if the classifier 1 has higher precision and classifier 2 has higher recall. In the proposed method, F1 score will be considered to understand the balance between precision and recall which will be compared with other work been done as an evaluation metrics.

2.5.5 ROC

ROC (receiver operating characteristic) is a graph representing the performance of possible classification. It is based on two factors [54].

The first one is true positive rate (TPR) and it can be calculated as:

$$\text{TPR} = \text{true positive}/(\text{true positive} + \text{false negative}),$$

where true positive represents the correct prediction of model for positive class and false negative as incorrect prediction of negative class.

The second one is false positive rate (FPR) and it can be calculated as:

$$\text{FPR} = \text{false positive}/(\text{false positive} + \text{ true negative}).$$

where false positive represents the incorrect prediction of positive class and true negative as correct prediction of negative class.

2.5.6 AUC

AUC (area under curve) can be used for calculating two-dimensional area under the ROC curve [54]. In other words, it can be used to contrast the classes and it represents the summary of the ROC curve. Higher AUC represents the better performance of the model in terms of comparison between the positive and negative classes. In the proposed model, area under the curve will help to determine the difference between normal images and pneumonia images which can be used to compare the with the aid of visual representation.

The above evaluation methods are applied widely in the related work, therefore, to compare the proposed model and the related work, these evaluation methods will be applied to the proposed model evaluation.

3 Proposed Model

We have analysed the various work done on medical image detection in the previous section. The experiments were performed based on available datasets. It has been observed that the machine learning models effectively detects medical images when the model is fed with a larger quantity of data. The use of ML algorithms has been proven effective in detection while compared to the traditional procedures mentioned in the literature review.

ML models need a higher volume of data for effective training capable of achieving higher accuracy in detection.

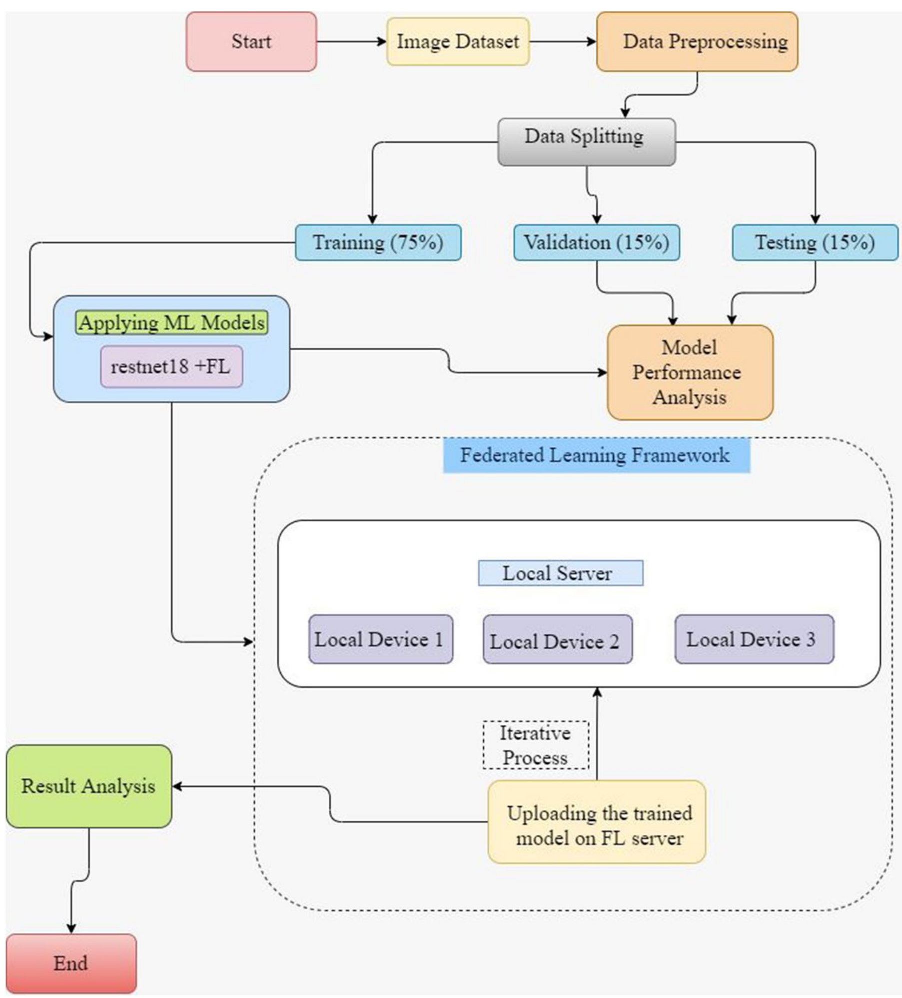

Fig. 3 This figure shows our Proposed Model. This model is followed across data splitting, training model, performing model performance analysis and using FL framework for training the model on local devices in a privacy-preserving manner

The lab-based datasets are always limited for effective ML model training. In the real-time medical image data, constant change in the feature variables determines the accuracy of ML models (Fig. 3). Therefore, we need a solution that can fulfil the datasets requirements for effective training of the model. The below image shows the flow chart of our proposed model.

The above figure shows our proposed model that follows series of steps from start to end. In the proposed model, image data is followed with data processing by splitting the data at the ratio of 75%, 15% and 15% into training, validation and testing respectively. After the model is trained by the training data, then the model is used for performance analysis by testing and validation data. In the proposed architecture, the training of the model is performed in a FL framework, where the restnet18 model is sent across local devices and model is trained on individual device data. After training it comes back to central server and the process carries on as iterative to get more updates from the local device. In this framework, data is not shared, instead only the trained model is shared to the central server (FL server), therefore the privacy of the data is promised. Eventually, a fully trained model can be effectively used for various purposes for example in detecting pneumonia.

Our proposed solution involves using privacy-preserving procedures that will allow using the real-time data, which fulfil the requirements of having massive data and variant patterns of medical images for ML model training. Privacy of the data is ensured in the proposed method that involves using the Federated Learning approach. The use of FL will involve the mutual collaboration of hospitals and medical institutes to train the ML model in their local servers, and the trained model from individual entities is shared centrally and aggregated together without sharing data. The central aggregation constitutes the trained model that repeats the cycle of training periodically, which helps to attain the higher efficiency of training the model for effective medical image detection. By using this approach, the privacy of the real-time data is ensured. Deep learning is one of the effective ML models that will be aggregated together with the FL, and it will ultimately help attain the maximum feature variables pattern to produce the effective outcome for medical image detection like pneumonia. The proposed method is unique because it will allow hospitals and medical institutions to collaborate to use real-time datasets in a privacy-preserving manner. The use of transfer learning will involve the training of the datasets locally, and ultimately it is aggregated together centrally to form an effective model that can be used to detect the medical image pattern efficiently. The collective use of federated learning technology with transfer learning will increase efficiency in training the data in a privacy-preserving manner. In the FL framework, data is not shared, instead only the trained model from the local host is shared to the client in an encrypted manner. The process of trained model transmission over the Internet is followed across the homomorphic encryption in FL that ensures the model is encrypted at all times while in transit. It promises the confidentiality, integrity, and accountability of the proposed model. In regard to the cost of the proposed model, the client's resources will be followed in accordance with the client permission. The proposed framework will allow client to manage the system with respect to their resource consumption. The model will only get trained if the certain time schedule is set by the client or it can follow the off-peak time of resource consumption to ensure the less load on the client resources.

4 Conclusion

This survey demonstrates the number of procedures been used in the past for the detection of lungs disease, especially pneumonia. Various tools and techniques have followed effective detection; however, it can be observed from the literature that the methods based on the ML are quite effective in the medical image detection from the image datasets. To make the ML model more productive, it is required to have a larger volume and variety of datasets to train the model. The lab-based datasets are limited to be used for effective training of ML model in a real-time scenario like in hospitals or medical institutions. Therefore, we need to have a solution of using real-time data to fulfil the requirements of having a more significant and variety of data. Our proposed model of using a federated learning strategy with deep learning can significantly enhance the capability of the ML model. FL will be responsible for ensuring data privacy, while deep understanding (neural networks) can be used to learn the image patterns effectively that will ultimately enhance the detection process. Our proposed work will give the researcher new dimensions in the field of medical image detection [28, 48, 55–61].

The proposed work fulfils all lack of using the current procedure of lab-based datasets, and it will allow using real-time datasets from hospitals and medical institutions that will ultimately result in achieving the effective system for medical image detection. However, the proposed work has certain limitations when it comes to the use of realtime datasets. Accepting this technology can be challenging as hospitals, and medical institutes are quite strict to their rules and regulations. Adapting the technology in the broader spectrum to form a central model must be reliable regarding security and data privacy. This technology involves trained model transportation over the internet medium, and here the security can be compromised. Also, adopting this technology will involve the extra cost of the client's resource consumptions. Therefore, these limitations should be addressed during the deployment of this technology in the real world.

Author contributions AK has made significant contribution to the conceptual aspects of the paper's design. HL and PS have assisted in report improvement and evaluation.

Funding Not applicable.

Availability of Data and Material Not applicable.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

1. Liu H, Song D, Rüger S, et al. Comparing dissimilarity measures for content-based image retrieval. Berlin: Springer; 2008.
2. Harsono IW, Liawatimena S, Cenggoro TW. Lung nodule detection and classification from Thorax CT-scan using RetinaNet with transfer learning. J King Saud Univ Comput Inf Sci. 2020. https://doi.org/10.1016/j.jksuci.2020.03.013.
3. Wang H, Jia H, Lu L, Xia Y. Thorax-Net: an attention regularized deep neural network for classification of thoracic diseases on chest radiography. IEEE J Biomed Health Inform. 2020;24:475–85. https://doi.org/10.1109/JBHI.2019.2928369.
4. (2021) Pneumonia—no child should die from a disease we can prevent. In: Our World in Data. https://ourworldindata.org/child-deaths-from-pneumonia
5. Naqvi SZH, Choudhry MA. An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis. Sensors. 2020;20:6512. https://doi.org/10.3390/s20226512.
6. Jakaite L, Schetinin V, Maple C. Bayesian assessment of newborn brain maturity from two-channel sleep electroencephalograms. Comput Math Methods Med. 2012. https://doi.org/10.1155/2012/629654.
7. Jakaite L, Schetinin V, Maple C, Schult J (2010) Bayesian decision trees forEEG assessment of newborn brain maturity. In: The 10th annual workshop on computational intelligence UKCI 2010. https://doi.org/10.1109/UKCI.2010.5625584
8. Jakaite L, Schetinin V, Schult J (2011) Feature extraction from electroencephalograms for Bayesian assessment of newborn brain maturity. In: Proceedings of the 24th IEEE international symposium on computer-based medical systems. https://doi.org/10.1109/CBMS.2011.5999109
9. Jakaite L, Schetinin V, Schult J (2011) Feature extraction from electroencephalograms for Bayesian assessment of newborn brain maturity. In: 24th international symposium on computer-based medical systems (CBMS), pp. 1–6. https://doi.org/10.1109/CBMS.2011.5999109
10. Nyah N, Jakaite L, Schetinin V, Sant P, Aggoun A (2016) Evolving polynomial neural networks for detecting abnormal patterns. In: 2016 IEEE 8th international conference on intelligent systems (IS), pp 74–80. https://doi.org/10.1109/IS.2016.7737403
11. Nyah, N., Jakaite, L., Schetinin, V., Sant, P., Aggoun, A.: Learning polynomial neural networks of a near-optimal connectivity for detecting abnormal patterns in biometric data. In: 2016 SAI Computing Conference (SAI), pp. 409–413 (2016). https://doi.org/10.1109/SAI.2016.7556014
12. Schetinin, V., Jakaite, L.: Classification of newborn EEG maturity with Bayesian averaging over decision trees. Expert Syst Appl 2012;39(10):9340–7. https://doi.org/10.1016/j.eswa.2012.02.184
13. Schetinin V, Jakaite L. Extraction of features from sleep EEG for Bayesian assessment of brain development. PLoS ONE. 2017;12(3):1–13. https://doi.org/10.1371/journal.pone.0174027.
14. Hassan MM, Billah MAM, Rahman MM, Zaman S, Shakil MMH, Angon JH (2021) Early Predictive Analytics in Healthcare for Diabetes Prediction Using Machine Learning Approach. In: 2021 12th international conference on computing communication and networking technologies (ICCCNT). IEEE, pp 01–05
15. Selitskaya N, Seliski S, Jakaite L, Schetinin V, Evance F, Conrad M, Sant P (2020) Deep learning for biometric face recognition: experimental study on benchmark data sets. In: Jiang R, Li C, Crookes D, Meng W, Rosenberger C (eds) Deep biometrics. Springer, pp 71–970. https://doi.org/10.1007/978-3-030-32583-1
16. Schetinin V, Jakaite L, Jakaitis J, Krzanowski W. Bayesian decision trees for predicting survival of patients: A study on the US national trauma data bank. Comput Methods Programs Biomed. 2013;111(3):602–12. https://doi.org/10.1016/j.cmpb.2013.05.015.
17. Schetinin V, Jakaite L, Krzanowski W. Bayesian averaging over decision tree models: an application for estimating uncertainty in trauma severity scoring. Int J Med Inform. 2018;112:6–14. https://doi.org/10.1016/j.ijmedinf.2018.01.009.
18. Kabiraj S, Akter L, Raihan M, Diba NJ, Podder E, Hassan MM (2020) Prediction of recurrence and non-recurrence events of breast cancer using bagging algorithm. In: 2020 11th international conference on computing, communication and networking technologies (ICCCNT). IEEE, pp 1–5
19. Schetinin V, Jakaite L, Krzanowski WJ. Prediction of survival probabilities with Bayesian decision trees. Expert Syst Appl. 2013;40(14):5466–76. https://doi.org/10.1016/j.eswa.2013.04.009.
20. Hassan MM, Peya ZJ, Mollick S, Billah MAM, Shakil MMH, Dulla AU (2021) Diabetes prediction in healthcare at early stage using machine learning approach. In: 2021 12th international conference on computing communication and networking technologies (ICCCNT). IEEE, pp 01–05
21. Rejwan Bin S, Schetinin V (2022) Deep neural-network prediction for study of informational efficiency. In: Arai K (eds) Intelligent systems and applications. IntelliSys 2021. Lecture notes in networks and systems, vol 295. Springer, Cham. https://doi.org/10.1007/978-3-030-82196-8_34
22. Jakaite L, Schetinin V, Hladuvka J, Minaev S, Ambia A, Krzanowski W. Deep learning for early detection of pathological changes in

x-ray bone microstructures: case of osteoarthritis. Sci Rep. 2021. https://doi.org/10.1038/s41598-021-81786-4.
23. (2021) Data Protection Act 2018. In: Legislation.gov.uk. https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted
24. (2021) Thousands of NHS medical images found 'unprotected' on web. In: Digital Health. https://www.digitalhealth.net/2019/09/thousands-nhs-medical-images-unprotected-web/. Accessed 25 Nov 2021
25. Kenny SPK. Optimizing space complexity using color spaces in CBIR systems for medical diagnosis. World News Nat Sci. 2020;9:96–103.
26. Abbas A, Abdelsamea MM, Gaber MM. DeTrac: transfer learning of class decomposed medical images in convolutional neural networks. IEEE Access. 2020;8:74901–13. https://doi.org/10.1109/ACCESS.2020.2989273.
27. Pandey P, Pallavi S, Pandey SC. Pragmatic medical image analysis and deep learning: an emerging trend. Singapore: Springer Singapore; 2019.
28. Khatri A, Jain R, Vashista H, et al. Pneumonia identification in chest X-ray images using EMD. Singapore: Springer; 2020.
29. Yang Z-Y, Zhao Q (2020) A multiple deep learner approach for X-ray image-based pneumonia detection. In: 2020 international conference on machine learning and cybernetics (ICMLC), pp 70–75. https://doi.org/10.1109/ICMLC51923.2020.9469043
30. Sarada N, Rao K. A neural network architecture using separable neural networks for the identification of "pneumonia" in digital chest radiographs. IjeC. 2021;17:89–100. https://doi.org/10.4018/IJeC.2021010106.
31. Artemi M, Liu H (2020) Image optimization using improved grayscale quantization for content based image retrieval. In: IEEE, pp 1–6
32. Deepal DAA, Fernando TGI. Convolutional neural network approach for the detection of lung cancers in chest X-ray images. Singapore: Springer Singapore; 2020.
33. Guan Q, Huang Y, Zhong Z, et al. Thorax disease classification with attention guided convolutional neural network. Pattern Recognit Lett. 2020;131:38–45. https://doi.org/10.1016/j.patrec.2019.11.040.
34. Bhandary A, Prabhu GA, Rajinikanth V, et al. Deep-learning framework to detect lung abnormality—a study with chest X-Ray and lung CT scan images. Pattern Recognit Lett. 2020;129:271–8. https://doi.org/10.1016/j.patrec.2019.11.013.
35. Lee S, Seo J, Yun J, et al. Deep learning applications in chest radiography and computed tomography: current state of the art. J Thoracic Imaging. 2019;34:75–85. https://doi.org/10.1097/RTI.0000000000000387.
36. Huang S, Lee F, Miao R, et al. A deep convolutional neural network architecture for interstitial lung disease pattern classification. Med Biol Eng Comput. 2020;58:725–37. https://doi.org/10.1007/s11517-019-02111-w.
37. Ye W, Yao J, Xue H, Li Y (2020) Weakly supervised lesion localisation with probabilistic-cam pooling. http://arxiv.org/abs/2005.14480
38. Huang X, Fang Y, Lu M, et al. Dual-ray net: automatic diagnosis of thoracic diseases using frontal and lateral chest X-rays. J Med Imaging Health Inform. 2020;10:348–55. https://doi.org/10.1166/jmihi.2020.2901.
39. Tilve A, Nayak S, Vernekar S et al. (2020) Pneumonia detection using deep learning approaches. IEEE, pp 1–8
40. Rajaraman S, Candemir S, Thoma G, Antani S (2019) Visualizing and explaining deep learning predictions for pneumonia detection in pediatric chest radiographs. In: SPIE, pp 109500S–109500S–12
41. Ge Z, Mahapatra D, Chang X, et al. Improving multi-label chest X-ray disease diagnosis by exploiting disease and health labels dependencies. Multimed Tools Appl. 2019;79(14889):14902. https://doi.org/10.1007/s11042019-08260-2.
42. Hegedűs I, Danner G, Jelasity M. Decentralized learning works: an empirical comparison of gossip learning and federated learning. J Parallel Distrib Comput. 2021;148:109–24. https://doi.org/10.1016/j.jpdc.2020.10.006.
43. Huang L, Shea AL, Qian H, et al. Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. J Biomed Inform. 2019;99: 103291. https://doi.org/10.1016/j.jbi.2019.103291.
44. Vyas J, Han M, Li L et al (2020) Integrating blockchain technology into healthcare. ACM, pp 197– 203
45. Maleh Y, Shojafar M, Alazab M, Romdhani I. Blockchain for Cybersecurity and Privacy. Milton: CRC Press; 2020.
46. Rohmah RN, Handaga B, Nurokhim N, Soesanti I. A statistical approach on pulmonary tuberculosis detection system based on X-ray image. Telkomnika. 2019;17:1474–82. https://doi.org/10.12928/telkomnika.v17i3.10546.
47. Yadav P, Menon N, Ravi V, Vishvanathan S. Lung-GANs: unsupervised representation learning for lung disease classification using chest CT and X-ray images. IEEE Trans Eng Manag. 2021. https://doi.org/10.1109/TEM.2021.3103334.
48. Xie Y, Wu Z, Han X, et al. Computer-aided system for the detection of multicategory pulmonary tuberculosis in radiographs. J Healthc Eng. 2020;2020:1–12. https://doi.org/10.1155/2020/9205082.
49. Yi P, Kim T, Lin C. Generalizability of deep learning tuberculosis classifier to COVID-19 chest radiographs: new tricks for an old algorithm? J Thoracic Imaging. 2020;35:W102–4. https://doi.org/10.1097/RTI.0000000000000532.
50. Hegedűs I, Danner G, Jelasity M. Decentralised learning works: an empirical comparison of gossip learning and federated learning. J Parallel Distrib Comput. 2021;148: 109124.
51. Pal K, Patel BV (2020) Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques. In: 2020 fourth international conference on computing methodologies and communication (ICCMC), pp 83–87. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00016
52. Oi H, Kawakami R, Nacmura T (2021) Analysis of evaluation metrics with the distance between positive pairs and negative pairs in deep metric learning, In: 2021 17th international conference on machine vision and applications (MVA), pp 1–5. https://doi.org/10.23919/MVA51890.2021.9511393
53. HossainMY, Sayeed A (2021) A comparative study of motor imagery (MI) detection in electroencephalogram (EEG) signals using different classification algorithms. In: 2021 international conference on automation, control and mechatronics for Industry 4.0 (ACMI), pp 1–6. https://doi.org/10.1109/ACMI53878.2021.9528276.
54. Singla J, Nikita K (2021) Comparing ROC curve based thresholding methods in online transactions fraud detection system using deep learning. In: 2021 international conference on computing, communication, and intelligent systems (ICCCIS), pp 9–12. https://doi.org/10.1109/ICCCIS51004.2021.9397167
55. Gang P, Zeng W, Gordienko Y, et al. Effect of data augmentation and lung mask segmentation for automated chest radiograph interpretation of some lung diseases. Cham: Springer International Publishing; 2019.
56. Chen K-C, Yu H-R, Chen W-S, et al. Diagnosis of common pulmonary diseases in children by X-ray images and deep learning. Sci Rep. 2020;10:17374. https://doi.org/10.1038/s41598-020-73831-5.
57. Datta S, Roberts K. A dataset of chest X-ray reports annotated with spatial role labeling annotations. Data Brief. 2020;32:106056. https://doi.org/10.1016/j.dib.2020.106056.

58. Ryoo S, Kim HJ. Activities of the Korean Institute of Tuberculosis. Osong Public Health Res Perspect. 2014;5:S43–9. https://doi.org/10.1016/j.phrp.2014.10.007.
59. Nam JG, Park S, Hwang EJ, et al. Development and validation of deep learning–based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. Radiology. 2019;290:218–28. https://doi.org/10.1148/radiol.2018180237.
60. Wang J, Li Z, Jiang R, Xie Z (2019) Instance segmentation of anatomical structures in chest radiographs. In: IEEE, pp 441–446
61. Liu Y, Liu G, Zhang Q. Deep learning and medical diagnosis. Lancet. 2019;394:1709–10. https://doi.org/10.1016/S0140-6736(19)32501-2.