POLITECNICO DI TORINO
Repository ISTITUZIONALE

Data-driven classification and explainable-AI in the field of lung imaging

(Article begins on next page)

18 January 2025

Check for updates

Data-driven classification and explainable-AI in the field of lung imaging

Syed Taimoor Hussain Shah[1†], Syed Adil Hussain Shah[1,2†],
Iqra Iqbal Khan[3], Atif Imran[4], Syed Baqir Hussain Shah[5],
Atif Mehmood[6,7], Shahzad Ahmad Qureshi[8], Mudassar Raza[9,10],
Angelo Di Terlizzi[2], Marco Cavaglià[1] and Marco Agostino Deriu[1]*

[1]PolitoBIOMed Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino,
Turin, Italy, [2]Department of Research and Development (R&D), GPI SpA, Trento, Italy, [3]Department of
Computer Science, Bahauddin Zakariya University, Multan, Pakistan, [4]College of Electrical and
Mechanical Engineering, National University of Sciences and Technology, Rawalpindi, Pakistan,
[5]Department of Computer Science, Commission on Science and Technology for Sustainable
Development in the South (COMSATS) University Islamabad (CUI), Wah Campus, Wah, Pakistan,
[6]School of Computer Science and Technology, Zhejiang Normal University, Jinhua, China, [7]Zhejiang
Institute of Photoelectronics & Zhejiang Institute for Advanced Light Source, Zhejiang Normal
University, Jinhua, Zhejiang, China, [8]Department of Computer and Information Sciences, Pakistan
Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan, [9]Department of Computer
Science, Namal University Mianwali, Mianwali, Pakistan, [10]Department of Computer Science, Heavy
Industries Taxila Education City (HITEC), University of Taxila, Taxila, Pakistan

Detecting lung diseases in medical images can be quite challenging for radiologists. In some cases, even experienced experts may struggle with accurately diagnosing chest diseases, leading to potential inaccuracies due to complex or unseen biomarkers. This review paper delves into various datasets and machine learning techniques employed in recent research for lung disease classification, focusing on pneumonia analysis using chest X-ray images. We explore conventional machine learning methods, pretrained deep learning models, customized convolutional neural networks (CNNs), and ensemble methods. A comprehensive comparison of different classification approaches is presented, encompassing data acquisition, preprocessing, feature extraction, and classification using machine vision, machine and deep learning, and explainable-AI (XAI). Our analysis highlights the superior performance of transfer learning-based methods using CNNs and ensemble models/features for lung disease classification. In addition, our comprehensive review offers insights for researchers in other medical domains too who utilize radiological images. By providing a thorough overview of various techniques, our work enables the establishment of effective strategies and identification of suitable methods for a wide range of challenges. Currently, beyond traditional evaluation metrics, researchers emphasize the importance of XAI techniques in machine and deep learning models and their applications in classification tasks. This incorporation helps in gaining a deeper understanding of their decision-making processes, leading to improved trust, transparency, and overall clinical decision-making. Our comprehensive review serves as a valuable resource for researchers and practitioners seeking not only to advance the field of lung disease detection using machine learning and XAI but also from other diverse domains.

1 Introduction

In modern healthcare, radiology imaging has emerged as a primary tool for the early detection of various diseases (Beets-Tan et al., 2020). Multiple modalities of radiological imaging are employed, including X-ray, Computed Tomography (CT), Magnetic-Resonance Imaging (MRI), Positron Emission Tomography (PET), and Ultrasound (US). Among these, chest radiography is particularly notable for its widespread use in achieving precise and rapid diagnoses (Abhisheka et al., 2023; Hussain et al., 2022; Philip et al., 2023). Radiologists, in the context of chest diseases, predominantly rely on X-rays and CT scans (Hall and Giaccia, 2006).

In most radiological imaging chest diseases, Pneumonia is an acute infection affecting the alveoli and distal airways of one or both lungs. It causes narrowing or closing of the airways due to inflammatory cells and fluid deposition in the alveolar sack of the lungs. It is a common and potentially lethal illness, affecting more frequently susceptible individuals, especially children of <5 years of age, accounting for 15 % of all deaths caused in this year's groups. Babies born too early (prematurely) face higher health risks because their organs are still developing. The younger the baby is, the higher the chance of getting lung infection (Prodanovic et al., 2024). The infectious agents can be viruses, bacteria, or fungi with a great geographical prevalence and outcome variations, among developed and low-and-middle-income countries. Yearly over millions of people are being infected with this virus. Early diagnosis is crucial for the successful treatment process. Generally, the disease can be diagnosed from chest X-ray images, but it is challenging due to unclear appearances and confusion with other diseases. There is an increasing effort to reach the poor community, in term of helping strategies recognizing and managing pneumonia (Rudan et al., 2004; Torres et al., 2021; Bhutta, 2007; Nelson et al., 1995; UNICEF, 2022; Franquet, 2001). In Africa, a significant shortage of healthcare workers poses a critical challenge, with an estimated deficit of 2.4 million doctors and nurses. Numerous studies have highlighted the severe scarcity of medical professionals and other healthcare resources across the continent (Narasimhan et al., 2004; Naicker et al., 2009). Due to critical health issues, the early detection of this disease, especially in neonates, is very important as it interferes with brain growth development and leads to other visual and hearing impairments including heart diseases (Maeda et al., 2023; Yildirim and Canayaz, 2023; Prodanovic et al., 2024).

Considering AI as a promising tool in medical imaging, AI can improve pediatric pneumonia diagnosis in the early stages (Yoon and Kang, 2024; Bal et al., 2024). In this scenario, specific kinds of challenges arise when working with pediatric chest X-ray datasets. Firstly, the available data scarcity, as compared to adult datasets, can hinder the development of robust trained AI models. Secondly, pediatric X-ray images reveal diverse anatomical development structure as compared to adults. Thirdly, it is also very challenging to acquire high quality images from uncooperative young ages patients and will result in various noise factors and artifacts caused by positioning and breathing patterns. Due to this, it will lead to more complicated analysis. Finally, ethical rules and regulations come into play even more critically when dealing with pediatric patients as compared to adults. Strict anonymization protocols and parental consent procedures are very important to ensure data collection and further utilization (Singh, 2024; Candemir and Antani, 2019; Ciet et al., 2024). Considering all these challenges, the collaborative effort of data collection and the development of robust AI powered diagnostic tools require standard image normalization techniques and strict anonymity rules and regulations for early diagnostic of pneumonia in pediatric patients.

Over the last 20 years, it is noted that chest infection has decreased. New conjugate vaccines for diagnosing *Haemophilus influenzae* type b and *Streptococcus pneumonia* helped to decrease in radiologic, clinical and immense pneumonia cases have reduce hospitalization (Le Roux and Zar, 2017). In 2011, the western pacific region, recorded that 61,900 yearly deaths are due to pneumonia in <5-year-old children (Nguyen et al., 2017). According to world health organization (WHO), all these extraordinary reductions in deaths, pneumonia is one of biggest killer of children <5 years old (Chavez et al., 2015).

Today, chest X-ray is used to show the presence of different lung diseases in humans. Experienced radiologists can assume the probability of chest disease in humans after checking the x ray images. Notably, in 2019, hospitals in Italy and the UK primarily employed radiography imaging to diagnose coronavirus patients (Fields et al., 2020). In some regions of the world, there is a limited amount of medical equipment and a shortage of doctors and experienced professionals who can effectively interpret chest X-rays. The most often experts of radiology examine the two-dimension (2D) chest x-ray images to clarify the pneumonia bacteria or virus instead of three-dimension (3D) chest structure (Mahomed et al., 2020).

The ease of obtaining radiographic images, however, also poses a significant challenge, as a single radiologist may need to evaluate as many as 100 images daily through radiography imaging [3]. Therefore, medical image analysis is a time taking process to reach chest disease diagnosis. In this scenario, information technology (IT) has a cardinal importance and is assisting the radiologists. It has also demonstrated itself as a backbone in other medical fields too. Further, health IT is administering the information of different kinds of disease using computational knowledge and its progress. The promising capability of health IT in decision making is a lot more as compared to a human. In this way, health IT can provide more better assistance in diagnosing various diseases to all over the world's medical community such as for acute coronary syndrome (D'Ascenzo et al., 2021), cardio-vascular risk (Navarini et al., 2020), and so on. Moreover, the most tremendous things in IT is the available data on the internet and that anyone can access at any time (Bohr and Memarzadeh, 2020). In parallel, advancements in artificial intelligence (AI) has led the medical field to consider AI as a promising tool for tackling complex biomarkers and supporting medical researchers in comprehending the progression of various diseases (Xing et al., 2020).

In this comprehensive review, we have examined and compared 20 computer-aided systems designed to assist clinicians in diagnosing pneumonia. Our exploration began with an in-depth analysis of publicly available datasets. We also provided detailed insights into various pre-processing techniques to aid researchers related to the available data. Furthermore, we conducted an extensive discussion on the various machine learning and deep

learning algorithms that have been predominantly utilized in recent research efforts. Our aim is to highlight the best models for a reliable and versatile tool useful in the middle- and low-income communities.

The paper is divided into several sections. Section 2 focuses on the dataset and its modality. Section 3 describes the data preprocessing and balancing techniques, and Section 4 discusses the different performance measures used in the studies. Section 5 elaborates numerous techniques proposed in this domain and section 6 provides a detailed comparative analysis and discussion of all techniques used for chest diseases' classification. Section 7 discusses progress and challenges in research works related to chest diseases. Finally, Section 8 concludes the review work.

2 Available training datasets

In contemporary times, a variety of datasets for lung diseases, complete with labels, have become accessible for training diverse machine learning and deep learning algorithms. These datasets have consistently served as standard references in numerous research endeavors, facilitating the comparison of results achieved through different techniques. Considering the wealth of available datasets, it is evident that they provide ample resources for training, validating, and testing machine learning and deep learning algorithms, including the creation of holdout datasets for evaluation. In our study, we focused on several benchmark datasets containing chest disease data and reviewed pneumonia experiments on these datasets, as mentioned below.

2.1 OCT-CXR

This dataset (Kermany et al., 2018) comprises 5,863 chest X-ray (CXR) scans from children, organized into three partitions: training, validation, and testing. The CXR images were taken meticulously and carefully selected from pediatric patients treated at the China Medical Center. All X-rays were taken as part of routine checkups.

2.2 CHEST X-ray 14

This dataset (Wang et al., 2017) was collected from the clinical picture archiving and communication system (PACS) and consisted of several frontal chest X-ray images. The dataset contains 112,120 front-facing chest X-ray images, which were subdivided into 14 classes. Each class represents a disease that could be identified from radiology reports.

2.3 MIMIC-CXR-JPG

This dataset (Johnson et al., 2019) is the largest publicly available chest X-ray dataset. The dataset was gathered by the USA Medical Center from 2011 to 2016. More than 35,000 chest X-ray images that represent 14 different chest diseases, including pneumonia and normal cases, were obtained.

Additionally, Table 1 provides an overview of the chest X-ray datasets that were used in previous studies for pneumonia classification. Some datasets were custom designed by compiling for pneumonia classification tasks, while others included multiple chest diseases, including pneumonia by default. The general steps that were taken by most of the strategies to make the data ready for classification have been described previously Figure 1 and have also been employed by many related studies (Benhar et al., 2020). The dataset with the highest number of pneumonia cases is called CXR, but in terms of image count, the MIMIC-CXR-JPG dataset is larger due to its inclusion of multiple chest diseases.

TABLE 1 Comparison of datasets for pneumonia cases.

Dataset	No. of images	Description
OCT-CXR	5,856	Pneumonia binary cases either pneumonia or normal class
Chest X-ray14	108,948	1,500 pneumonia cases. The dataset includes 8 different chest diseases
MIMIC-CXR-JPG	377,110	Enormous dataset consisting of 65,079 number of pneumonia patients
PneumoCAD	156	Only 78 pneumonia cases exist
COVID-19	219	Comprises samples of three classes COVID-19 positive, normal, and viral pneumonia
PediCXR	9,125	Consists of 481 pneumonia cases and remaining dataset is divided into other 14 different labels
CheXpert	224,316	Having 14 different classes while pneumonia cases are 4,576
NIH Chest X-ray	112,120	Comprises 1,062 pneumonia images and remaining other pathologies are divided into 7 classes

2.4 PneumoCAD

This dataset (Oliveira et al., 2008) comprises pneumonia presence (PP) and pneumonia absence (PA) images based on children scans. All these findings were confirmed by two radiologists. These authors interpreted the chest radiography images according to the WHO instructions. These radiography images had a resolution of 1,024 x 768 pixels and an 8-bit grayscale. A total of 156 X-ray images were inspected by radiologists; 78 images were from pneumonia patients, and the remaining 78 were labeled as normal.

2.5 COVID-19

The COVID-19 dataset (Cohen et al., 2020) consists a of total 219 chest X-ray images. It comprises three classes COVID-19 positive, normal, and viral pneumonia. Further, each image varies in size, typically around 1,024 x 1,024 pixels, though this can vary due to different sources and preprocessing steps. The dataset is organized into three classes, each representing a different health condition, providing a balanced set of examples to train
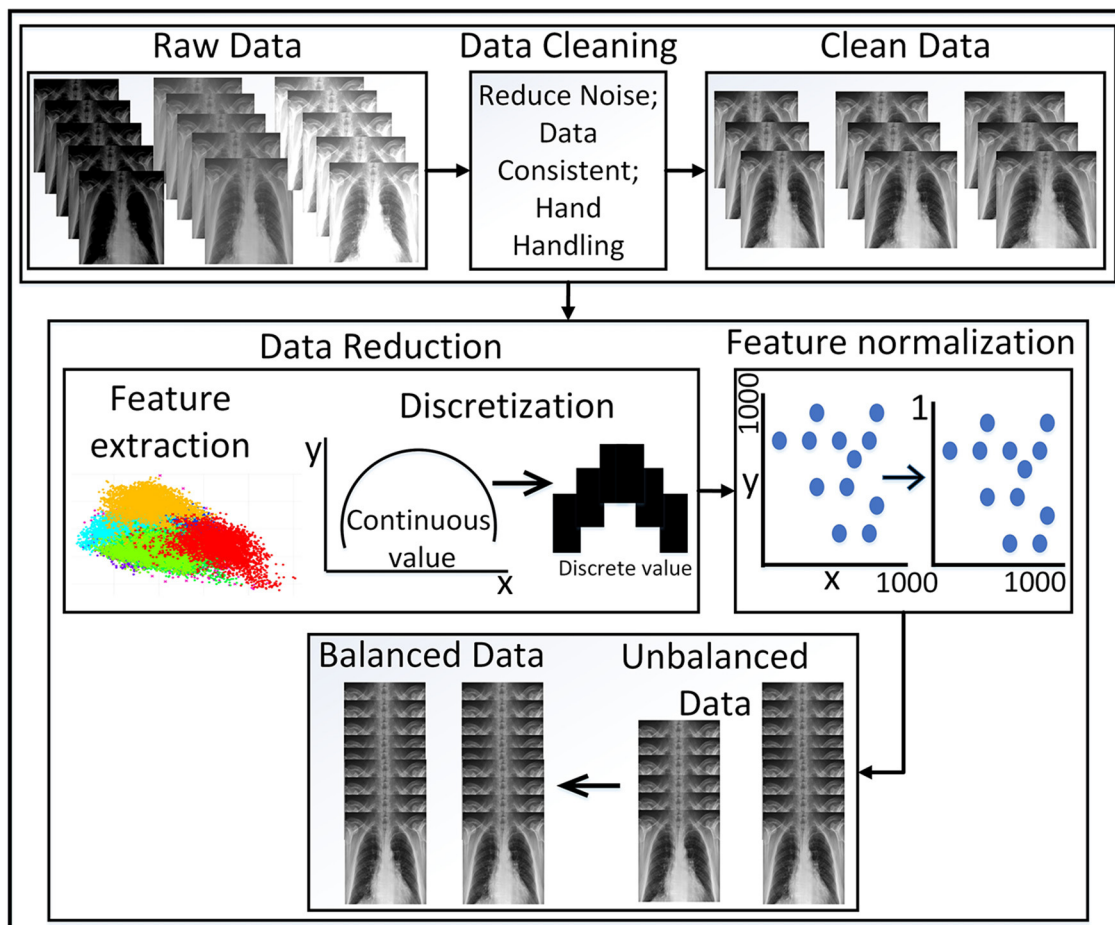
FIGURE 1
Illustration of the general preprocessing steps.

and evaluate machine learning models for accurate diagnosis and classification of COVID-19 and related respiratory conditions.

2.6 PediCXR

This dataset (Pham et al., 2023) is a very valuable resource for training and evaluating AI models for pediatric chest X-ray images. It consists of 9,125 X-ray scans and gathered at major pediatric hospital in Vietnam between 2020 and 2021. Each image possesses a different pixel range, but most cases have more than 1,000 pixels in both x and y dimensions. The dataset is enriched with 36 critical findings and 15 distinct disease categories. This comprehensive labeling scheme allows researchers to train AI models for various tasks, from pinpointing specific abnormalities to making broader diagnoses in pediatric chest X-rays.

2.7 CheXpert

CheXpert (Irvin et al., 2019) is a larger collection, containing 224,316 chest radiographs from adult patients. While image sizes aren't explicitly stated, they likely adhere to medical imaging standards, potentially ranging from 512 x 512 pixels to 1,024 x 1,024 pixels. In addition, CheXpert includes labels for 14 different pathologies.

2.8 NIH Chest X-ray

The NIH Chest X-ray dataset (Wang et al., 2017) is containing 112,120 frontal and lateral chest X-rays from adult patients. Image sizes are likely similar to CheXpert's. However, the dataset utilizes a simpler labeling scheme, focusing on the presence or absence of eight common findings, including pneumonia.

3 Preprocessing

The purpose of preprocessing is to improve the quality of images by reducing variations and eliminating any discrepancies that may cause complications during the processing stage, thereby reducing the consumption of resources (Tripathy and Swarnkar, 2020). Preprocessing achieves three main objectives, namely noise removal, contrast enhancement, and brightness

and color correction. To remove noise, filters such as mean and median filters, and Gaussian low-pass filtering are widely used. Morphological techniques are also used for image information enhancement purposes (Courtenay et al., 2020). Contrast stretching techniques and histogram equalization techniques have been widely used for contrast enrichment. For brightness, color correction, and color standardization techniques have been used also such as Gastal and Oliveira (2012).

Furthermore, standard x-ray images generally have a dimension of (3,000 x 2,000) pixels, which is quite large and contains unnecessary information. These large images require high storage space and powerful hardware for analysis. To save time and obtain better results, many publicly available datasets have resized their images from their original size. For example, the Chest X-ray 14 dataset has reduced the image sizes to 512 x 512 pixels, while the MIMIC-CXR-JPG and OCT-CXR datasets have image sizes of (2,048 x 2,048) and (1,024 x 1,024) pixels, respectively. In our studies, we have also emphasized the importance of a balanced experimental dataset for training, validation, test sets. An unbalanced data set can lead to biased classifications toward the majority image set. Therefore, data balancing techniques are used to prevent such biased classification results. Table 2 illustrates some traditional and deep learning data balancing techniques along with their descriptions.

TABLE 2 Data balancing techniques.

References	Technique	Description
Wang and Lu (2018)	Mean Square Error (MSE)	Sum square error of every sample class wise and then calculated their average
Wang et al. (2016)	Mean False Error (MFE)	Calculate the loss based on average error
Baltruschat et al. (2019)	Manually sampling	Samples randomly duplicates in minority class while samples randomly removed from majority class
Feng et al. (2019)	Synthetic Minority Oversampling Technique (SMOT)	Oversampling method

4 Performance metrics

The performance measures (Michelucci et al., 2021) for pneumonia detection and classification depend upon various factors. These factors are discussed below:

- True positives (TPs) that identified the correct pneumonia cases from the training data.
- True negatives (TNs) were those that identified normal cases from the training data.
- False positives (FPs) that incorrectly identified pneumonia cases.
- A false negative (FN) was used to identify the wrongly normal cases.

The following equations represent the performance metrics used to evaluate the robustness of the models: accuracy (1), sensitivity (2), specificity (3), and precision (4).

$$Accuracy\ (acc) = \frac{TP + TN}{TP + FN + TN + FP} \qquad (1)$$

$$Recall/Sensitivity\ (rec/sen) = \frac{TP}{TP + FN} \qquad (2)$$

$$Specificity\ (spe) = \frac{TN}{TN + FP} \qquad (3)$$

$$Precision\ (pre) = \frac{TP}{TP + FP} \qquad (4)$$

5 Classification methods

Over time, machine learning approaches, a general concept, have been shown Figure 2 to be implemented in numerous studies (Hameed et al., 2020; Javaid and Haleem, 2020), significantly assisting in the medical field. Even today, machine learning is useful in the medical field and helps researchers continue to address novel medical issues. A number of relevant studies on pneumonia have been published in the subsequent subsections based on CXR datasets. These used CXR datasets are frequently used in research as a standard dataset for comparison with the results of other studies. In our review, we adhered to the same framework.

5.1 Machine learning models

In traditional machine learning, researchers have applied feature extraction algorithms to images to address specific problems. In this context, datasets have been analyzed for only a single domain at a time. More complexity and more time are required to obtain results from these algorithm techniques.

Oliveira et al. (2008) proposed a machine learning technique that classifies X-ray images as pneumonia or normal cases. On the images, they employed the Haar wavelet transform feature extraction computer vision technique. The KNN algorithm was subsequently used, where K was 15, to train and classify the model on the extracted features. As a result, they reported an accuracy of 97% compared to that of other previously published feature extraction studies in this domain. The results showed that this method achieved good accuracy due to the application of relevant feature extraction techniques.

Sousa et al. (2013) compared different machine learning techniques and evaluated their results. In this work, three random sets of 40 images were used for training, and 15 images were randomly selected for the testing phase. For features, different texture-based feature extractors based on the Haar wavelet transform (Yu et al., 2023) were used to extract features for classification tasks (Choras, 2007; Guido, 2018; He et al., 1987). Three machine learning techniques, namely, SVM, KNN (K = 9), and naïve Bayes (Kumar et al., 2019; Zheng and Ding, 2020; Kaur and Oberoi, 2020) classifiers, were applied to these extracted features and achieved accuracies of 77, 70, and 68%, respectively.

Yao et al. (2011) proposed an automated intelligent system that identifies five different chest diseases, including pneumonia. They used 40 CXR images and applied machine learning to texture
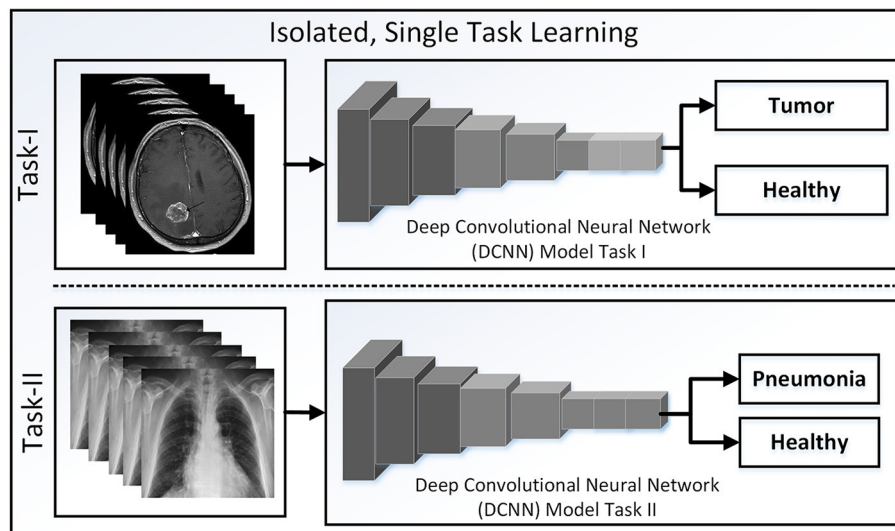
FIGURE 2
Basic concept of traditional learning where brain tumor images are taken from Chakrabarty (2019).

analysis. Their extracted feature vector contains 25 texture features (mean and variance features from histogram statistics, energy and correlation features from a co-occurrence matrix, etc.) for every image in the dataset. These extracted features were further passed to train and test the SVM model, which yielded 85% accuracy in the pneumonia detection task.

5.2 Deep learning models

DL has pushed research work in all fields by limiting the use of standard handcrafted mechanisms of detection and classification (Bhandary et al., 2020). Deep learning models are composed of many layers connected to each other and have designated functions. All these layers are used for their designated function in the respective successions. Deep learning works best on a large number of training sets that are not easy to obtain in the biomedical field because it is supervised (Suzuki, 2017; Khan et al., 2021).

The dependence on human operators for the diagnosis of pneumonia is not suitable because it requires expertise and experience. Therefore, the domain of deep learning is currently being adopted to make systems fully self-driven. This approach provides better results, fewer miscalculations and fewer chances of failure (Sun et al., 2018). Deep methodology is the latest innovation inside the domain of machine learning. It is constantly being pruned and enhanced daily, so much research is being conducted on this topic. We can expect more work in the field of cell diagnosis and disease classification soon, which will involve the use of deep networks.

CNNs have been widely used to improve classification and segmentation tasks. In CNN, the convolutional layer helps to extract multidimensional features from the input image. The applied weighted distributed technique helps to reduce the complexity of the network (Albawi et al., 2017). Architecturally,

CNNs are simply feedforward artificial neural networks. The general structure of the CNN model based on several blocks is shown in Figure 3. This strategy is widely employed in various studies, such as those of Yamashita et al. (2018) and Acharya et al. (2017), who adopted a similar approach to provide a comprehensive overview. In addition to classification tasks, explainable-AI (XAI) techniques are proposed for revealing black boxes, such as the following:

- **Local Interpretable Model-Agnostic Explanations (LIME):** Ribeiro et al. (2016) works by developing local interpretable models around a specific target, and the developed model is used to explain why that prediction was made.
- **Shapley Additive exPlanations (SHAP):** This technique (Lundberg and Lee, 2017) explains the black box model by assigning a shap value to each feature that tells the feature's contribution to the model prediction.
- **Saliency maps:** This technique (Zeiler and Fergus, 2014) unveils black boxes on images by highlighting the most important features in an input image.
- **Gradient-weighted Class Activation Mapping (Grad-CAM):** Grad-CAM (Selvaraju et al., 2016) works on the saliency map technique for CNNs. The gradient of the model's output is computed with respect to the input image, and gradient maps are developed.
- **Activation maps:** This technique (Zeiler and Fergus, 2014) produces images that possess the activations of neurons in a CNN. This technique is helpful for understanding the internal operation of CNNs.

In the upcoming subsections, we have mentioned custom-designed deep models and transfer learning-based models that researchers have employed in their work regarding imaging-based pneumonia disease detection.

FIGURE 3
Convolutional neural network (CNN) model with XAI techniques.
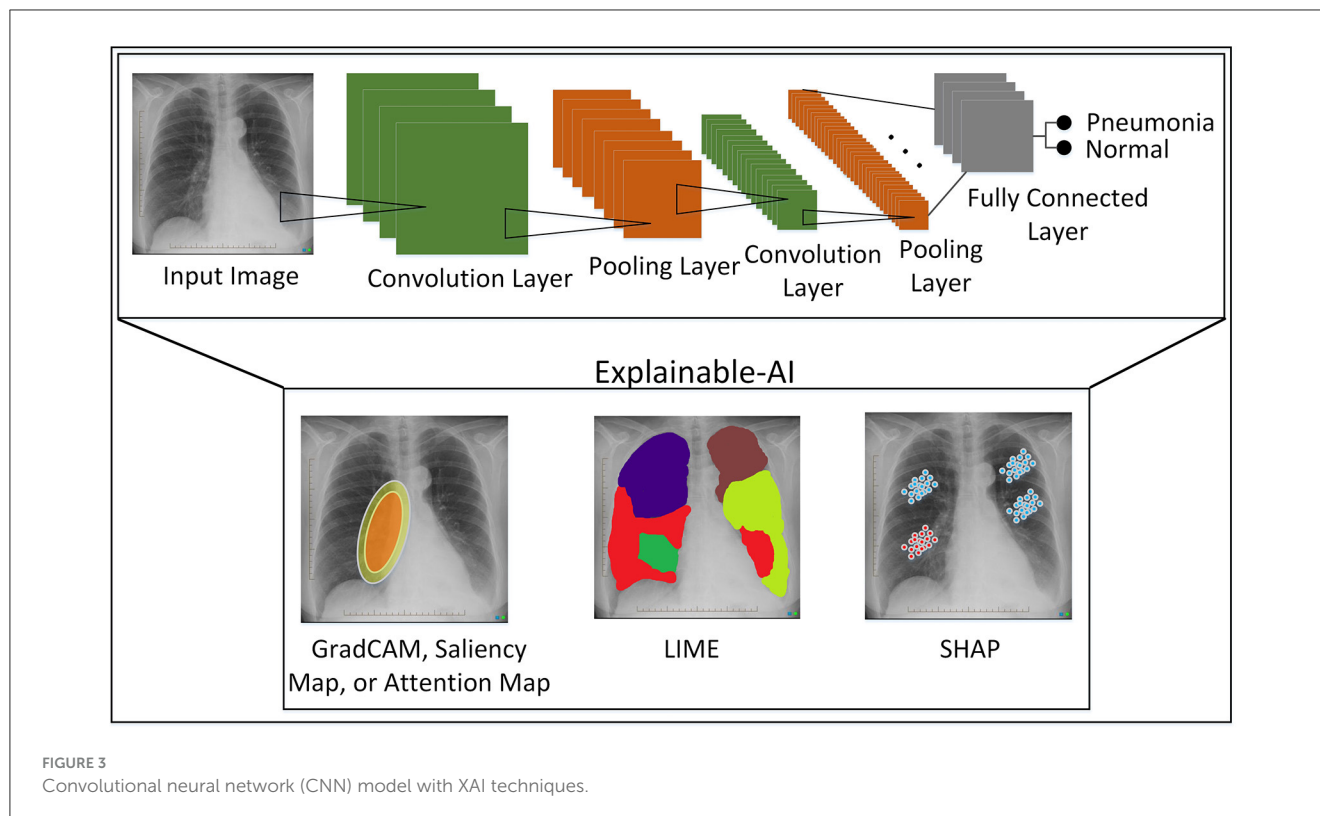
5.2.1 Customized CNN models

Stephen et al. (2019) proposed a CNN with four convolutional layers and two dense layers. They proposed a model trained on the OCT-CXR dataset. They split the dataset for training and testing on their own dividing threshold from the original dataset and used data augmentation techniques to increase the performance of the model up to 93.7% testing accuracy, as reported in their work.

Jain et al. (2020) worked to classify pneumonia and non-pneumonia chest X-ray images. The author used six models in this research to compare the performances of these models. The first 2 models were constructed by using several convolutional layers, and the other four models were pretrained (VGG-16, VGG-19, ResNet-50 and Inception-V3). All the models were trained and tested on the CXR dataset. The best classification result of the second model, which was based on three convolutional layers, achieved 92.31% accuracy compared to the other models.

Siddiqi (2019) proposed a CNN model that used an 18-layer convolutional neural network and trained it using the OCT-CXR dataset following a dataset with 80% disturbance in the training set and 20% disturbance in the testing set. The proposed technique achieved accuracy, specificity, and sensitivity of 94.3, 86, and 99%, respectively. This study showed that these methods improved the classification accuracy by 1.6%, but on the other hand, they had lower specificity than did the state-of-the-art methods.

In addition to the research in the same domain, Labhane et al. (2020) obtained better results from different neural network models using custom-CNN and transfer learning techniques on three renowned models, namely, Inception-V3, VGG-19, and VGG-16. In addition to the results of other studies in the same domain, better results were obtained from different neural network models

trained with custom-CNNs and transfer learning techniques on three renowned models, namely, Inception-V3, VGG-19, and VGG-16. The dataset (OCT-CXR) used in these models included 2,992 pneumonia images and 2,972 chest X-ray images. Using augmentation techniques, the training data increased, which helped in obtaining the most promising performances, ∼97% accuracy, for customized and all the renowned deep networks trained in this work.

Similarly, Liang and Zheng (2020) proposed a custom 49 convolutional layer residual model and renowned the VGG-16, DenseNet-121, Inception-V3, and Xception models for transfer learning. In this work, the OCT-CXR dataset was used to classify patients into binary response groups, such as normal vs. pneumonia. This work achieved 90.5, 74.2, 81.9, 85.3, and 87.8% accuracy, respectively, as mentioned previously. Liang and Zheng also concluded that their customized model performed better in classification than did transfer learning.

Saraiva et al. (2019) shows comparative research between conventional (multilayer perceptron) and custom-designed deep learning models on the OCT-CXR dataset considering binary classification. The experimental data had two classes consisting of 5,840 images with infected and non-infected traits. During training, a cross-validation technique was used to validate the models. Based on the performance, the deep learning models performed very well, with 94% accuracy, whereas the conventional neural network model achieved 92% accuracy.

Yi et al. (2023) designed an intelligent system for the detection of normal and pneumonia diseases using chest X-ray images. The main objective of this study was to illustrate the strength of deep neural models on image-based data. In the first step,

they acquired a publicly available OCT-CXR dataset. In the next step, they applied data augmentation and preprocessing steps such as rescaling, rotation, width-height shift, shearing, zooming, horizontal-flipping, and filling to prevent bias in the data sample class. For feature learning and extraction, they designed a proposed DCNN model with 42 conv layers and 2 dense layers. After the training step, they utilized the trained deep model as a feature extractor to extract robust features from the training and validation datasets and then applied a supervised machine learning classifier for pneumonia classification. For the performance evaluation of machine learning models, standard parameters such as accuracy, specificity, sensitivity, and the F1-measure were acquired. The proposed methodology achieved 98.02% training accuracy and 96.09% test accuracy scores.

Akbulut (2023) introduced an innovative and robust algorithm, the ACL model, which was designed as a customized deep learning architecture. This model was purposefully combined with attention and LSTM models and CNN frameworks to accurately classify patients into distinct categories: healthy, COVID-19, and pneumonia. To enhance the performance of this approach, crucial features and patterns present within chest X-ray images were emphasized. This was achieved through the application of the marker-controlled watershed (MCW) segmentation algorithm, which highlights essential stains and traces vital for accurate classification. Throughout the experimental phase, the ACL model underwent training across varying training–test ratios—specifically, 90–10, 80–20, and 70–30%. The attained accuracy scores demonstrated exceptional performance, recording a perfect accuracy of 100% for the 90–10% ratio while achieving an impressive 96% accuracy for both the 80–20 and 70–30% ratios. These outcomes underscore the adaptability and reliability of the ACL model across different training-test distributions, demonstrating its potential as an effective diagnostic tool for classifying health conditions based on CXR imaging data.

In another work, Kiliçarslan et al. (2023) developed a novel deep learning methodology aimed at pneumonia classification. This study introduced an innovative activation function termed Superior Exponential (SupEx) and conducted comparative evaluations against established activation functions such as ReLU, LReLU, Mish, Sigmoid, Swish, Logish, and Softplus. The study integrated the MNIST and CIFAR-10 datasets to substantiate the efficacy of the proposed SupEx activation function. The findings were subsequently applied to the CNN, which trained and tested pneumonia identification via chest X-ray images. Notably, classification accuracies were achieved, reaching 95.37% for pneumonia detection. This research signifies the introduction of a promising activation function for pneumonia detection, demonstrating its efficacy across both emerging and established benchmark datasets.

5.2.2 Transfer learning models

Transfer learning is the concept of overcoming individual learning complexities and reusing the knowledge that is obtained from any similar pretrained model. Figure 4 illustrates the fundamental principles of transfer learning, and numerous

investigations have explored its application in the classification of lung diseases, as demonstrated in Raghu et al. (2019).

Rajaraman et al. (2019) proposed a customized CNN model and compared their results with those of a VGG-16 retrained model on the OCT-CXR dataset. In this research, an atlas-based detection algorithm was used to detect lung tissue, and a DCNN model was used for classification. In comparison, VGG-16 achieved better performance than did the customized CNN. The author performed two experiments in terms of class distribution in this research. In these experiments, 96.2% accuracy for normal and pneumonia classes and 93.6% accuracy for viral and bacterial classes were the best performances. Finally, they utilized Grad-CAM to visualize the regions of the images that contributed most to the model's predictions.

Rahman et al. (2020) performed a transfer learning technique using four different pretrained convolutional neural network (CNN) models, such as AlexNet, ResNet-18, DenseNet-201, and SqueezeNet. All these trained models used an OCT-CXR dataset that included 5,237 chest X-ray images. In this research, the authors reported three different experiments in terms of dataset class combinations: case-1: normal and pneumonia; case-2: normal, bacterial pneumonia and viral pneumonia; and case-3: bacterial and viral pneumonia. The accuracies of these experiments were 98% (case-1: DenseNet), 95% (case-2: DenseNet), and 93.3% (case-3: DenseNet). Moreover, they also incorporated LIME to better explain the predictions of the model.

Hashmi et al. (2020) applied a transfer learning approach to classify binary classes as normal or pneumonia. Several data augmentation techniques simulate the regeneration of data images. The OCT-CXR dataset was utilized to retrain several experimental deep models (basic ResNet-18, Xception, Inception-V3, DenseNet-121, and MobileNet-V3 models and weighted mentioned renown models). This research concludes that renowned basic deep models with weightage perform better (accuracy: 98%) than other normal deep learning models (accuracy: 96–97%) by employing a data augmentation technique that reduces data overfitting. Additionally, they unveiled the model's reasoning by incorporating a technique called SHAP to explain the predictions of the model.

Ayan and Ünver (2019) performed an addition to this research to diagnose pneumonia by applying two CNN models, namely, VGG-16 and Xception, on the OCT-CXR dataset. The VGG-16 model achieved 87% accuracy, which was better than that of the Xception model (82%), and the confusion matrix showed that both models had their own significance against the data. The Xception network is more efficient at detecting pneumonia cases than is the normal network, while the Vgg-16 network focuses on normal patients. In addition, they elaborated upon the model and features by employing the saliency map technique to visualize the regions of the images that contributed most to the model's predictions.

Alshmrani et al. (2023) proposed a deep learning architecture for the identification and classification of multiple classes of lung diseases, such as pneumonia, lung cancer, and COVID-19. For multiclassification, they utilized large amounts of data for deep learning classification, including the OCT-CXR dataset. The proposed technique uses the ensembled methodology of the VGG-19 and custom-CNN models. This ensembled technique achieved the best performances—96.48, 97.56, and 93.75%—for
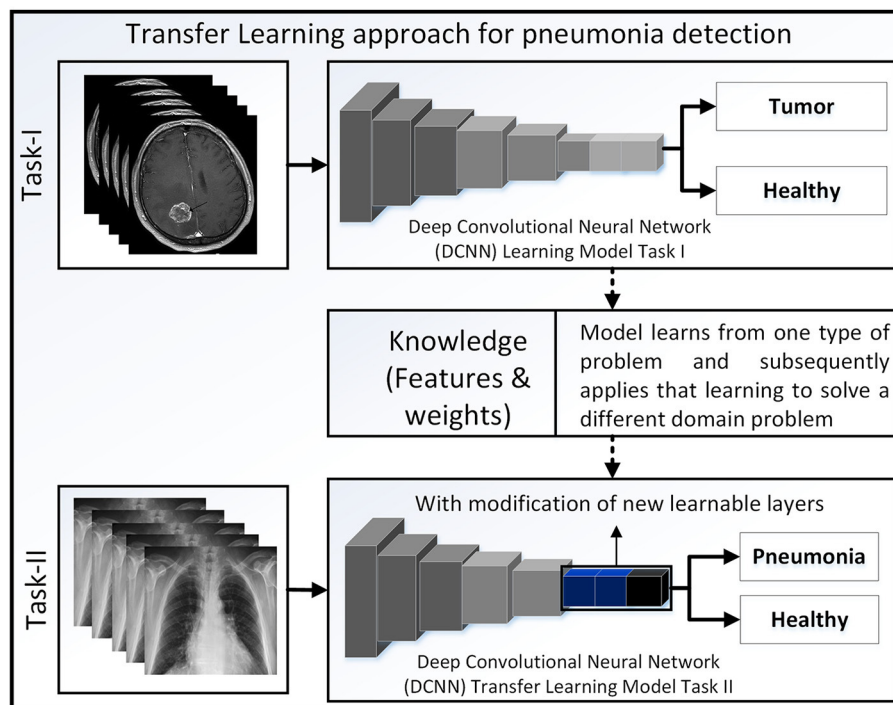
accuracy, precision and sensitivity, respectively. They also provided explanations by utilizing a technique called attention maps to visualize the regions of the images that the model is paying attention to.

5.2.3 Ensemble features/models

The ensemble approach is a technique used to ensemble multiple features from various trained models or deep models on the same dataset and provides better results than other techniques. In this domain, Chouhan et al. (2020) proposed a technique in which two or more deep learning algorithm models (AlexNet, ResNet-18, Inception-V3, DenseNet-121 and GoogleNet) were used for classification of images in the OCT-CXR dataset. After embedding, a voting classifier is employed to predict the best relevance class. The ensembled models outperform other state-of-the-art methods and achieve a 96% accuracy score. For explainability, they also employed a technique called integrated gradients to explain the predictions of the model.

Togaçar et al. (2020) applied three different pretrained deep models (AlexNet, VGG-16, and VGG-19). In the training process, every model obtained 1,000 features from every image in the dataset. Furthermore, the minimum redundancy maximum relevance algorithm was used to reduce the feature space to 100. On the basis of the combination of these features, a linear discriminant analysis classifier was trained and tested on a 99% binary classification accuracy on the OCT-CXR dataset. The authors also used a feature importance technique to explain the importance of different features in the model's predictions.

Ukwuoma et al. (2023) presented a hybrid workflow based on fused capabilities of convolutional networks and the transformer encoder (TE) mechanism. This ensembled learning technique employed to extract meaningful features from the dataset X-ray input images in two different ways: ensemble-A (i.e., GoogleNet, DenseNet201, and VGG16) and ensemble-B (i.e., Xception, DenseNet201, and InceptionResNetV2). Whereas, the TE is built following the scheme of self-attention structure considering multilayer perceptron (MLP) for accurate disease identification. The proposed whole designed pipeline underwent training in two fashions binary and multi-class classification. On test, end-to-end hybrid learning model resulted in 99.21% classification performance for both overall accuracy and F1-score in the binary classification task, while 98.19 and 97.29% scores for overall-accuracy and F1-score respectively have been expressed in the multi-classification task. In addition, they also showed the visual results by employing XAI, LIME and attention maps algorithms, to explain the predictions of the model.

6 Comparative analysis

In comparative analysis, we have evaluated the performance of various conventional machine learning and deep learning models, including customized CNN, transfer learning, and embedded models, utilizing different feature extraction techniques. For a more enhanced overview, different research works have been clustered into meaningful groups.

The performance metrics listed in this review are recommended methods for determining each model's performance. However, we have provided all the relevant measurements as shown in Table 3. To facilitate comparison,

TABLE 3 Summarizing related studies that employ machine and deep learning methods where metrics are accuracy (Acc), specificity (Spe), precision (Pre), and recall (Rec), all presented as percentages; additionally, "N" denotes not present, respectively.

References	Method	Features extractor	Classifier	Acc	Spe	Pre	Rec	XAI
Oliveira et al. (2008)	Machine Learning Techniques	Haar Wavelet Transforms	KNN	97	90	N	100	N
Sousa et al. (2013)		Coefficient of Variation, Correlation, Entropy, Standard Deviation, etc., based on Haar wavelet	SVM, KNN, and Naïve Bayes	77, 70, and 68	N	N	N	N
Yao et al. (2011)		Mean, variance, energy and correlation from correlation matrix, etc.	SVM	80	N	N	N	N
Stephen et al. (2019)	Customized CNN	4-layers CNN	SoftMax Layer	94	N	N	N	N
Jain et al. (2020)		CNN Model		92	N	N	98	N
Labhane et al. (2020)		Inception-V3, VGG-19, VGG-16, Customized-CNN		97	N	98	97	N
Siddiqi (2019)		18-layers Convolutional		94	86	92	99	N
Liang and Zheng (2020)		Custom 49- layers CNN, VGG-16, DenseNet-121, Inception-V3, and Xception		91, 74, 82, 85, and 88	N	89, 72, 79, 92, and 86	97, 95, 96, 84, and 97	N
Saraiva et al. (2019)		Custom CNN and MLP		92 and 94	92 and 94	92 and 94	92 and 94	N
Yi et al. (2023)		Custom CNN		96	99	N	94	N
Akbulut (2023)		LSTM + Attention-CNN in one model		96	98	94	94	N
Kiliçarslan et al. (2023)		Custom CNN		95	N	N	N	N
Rajaraman et al. (2019)	Transfer Learning	VGG-16		96	96	98	96	Grad-CAM
Rahman et al. (2020)		AlexNet, ResNet-18, DenseNet-201, and SqueezeNet		98, 95, and 93	97, 94, and 97	97, 95, and 93	99,96, and 93	LIME
Hashmi et al. (2020)		Original and Weighted (ResNet-18, Xception, Inception-V3, DenseNet-121, and MobileNet-V3)		96-97 and 98	N	98	99	SHAP
Ayan and Ünver (2019)		VGG-16 and Xception		87 and 82	91 and 76	87 and 84	82 and 85	Saliency maps
Alshmrani et al. (2023)		VGG-19 and Custom CNN		97	N	98	94	Attention maps
Chouhan et al. (2020)	Ensemble Model	AlexNet, ResNet-18, Inception-V3, DenseNet-121 and GoogleNet	Voting Classifier	96	N	93	100	Integrated gradients
Togaçar et al. (2020)		AlexNet, VGG-16 and VGG-19	Linear Discriminant Analysis Classification (Merged features)	99	99	99	100	Feature importance by model's predictions
Ukwuoma et al. (2023)		DenseNet201, VGG16, and GoogleNet	Multilayer Perceptron	97	97	97	97	LIME and attention maps

all the studies previously described were evaluated based on accuracy, with additional performance measures noted where available.

In the groups, ML group depicts that Oliveira et al. (2008) achieved the best 97% accuracy by employing the k-NN classifier on numerous extracted features. With respect to customized neural networks, Labhane et al. (2020) achieved ∼97% accuracy, with the best results obtained with all the customized and pretrained models. Using the pretrained transfer learning technique, Rahman et al. (2020) and Hashmi et al. (2020) achieved 98% accuracy by employing numerous pretrained models by comparing among them. In the last group, Togaçar et al. (2020) achieved 99% accuracy by employing transfer learning and the ensemble feature technique. Overall, we found that Togaçar et al. (2020) performed very well, with 99% accuracy, by using transfer learning and the ensemble feature technique. This shows that transfer learning with an ensemble method is an efficient and effective technique for learning and representing better visual features of images.

7 Progress and challenges in chest research applications

As AI continues to advance over time, it has become increasingly convenient to develop improved algorithms, reducing the reliance on manually crafted features for radiographic images. Researchers have made significant strides in creating numerous algorithms designed to automatically extract features and employ machine/deep learning techniques. The primary benefit of these methods lies in their enhanced performance, particularly in the detection and classification of various chest diseases using deep learning algorithms. Many recent studies have harnessed convolutional neural networks (CNNs) to augment early-stage detection by learning intricate patterns, thereby helping doctors comprehend these complex situations.

However, a notable challenge is the lack of practical applicability, which hinders the widespread use of these highly efficient algorithms. There is a pressing need to bridge the gap between related research and wider public and medical communities by creating user-friendly interfaces on well-performing models. Moreover, there is a deficiency in training resources because of the limited tools available, leaving many users uncertain about how to effectively utilize these tools in middle- and low-income regions for early diagnosis and mitigation of chest and related diseases.

In addition, while the number of studies on the detection of pneumonia using machine learning/deep learning is growing, the number of works that focus on explainability is still relatively limited. This is a significant gap in the literature, as explainability is crucial for understanding how models make decisions and for building trust in their results.

8 Conclusion

The detection of lung diseases from medical images poses significant challenges for radiologists, even for experienced professionals, due to the intricacies of these diseases and the time-intensive nature of the diagnostic process. However, artificial intelligence (AI) has emerged as a promising solution for addressing these challenges. Artificial intelligence (AI), particularly through machine learning and deep learning techniques, has revolutionized the field of medical research, replaced traditional handcrafted methods and significantly improved diagnostic accuracy. In this review, we have presented a vast amount of research on chest disease classification using four different standard radio imaging datasets of lungs. Among the datasets considered, the CXR dataset is the most commonly used dataset among the reviewed works compared to the ChestX-ray14, JSRT, and MC datasets. For classification, different image preprocessing and feature extraction techniques have been employed to train different machines and deep learning algorithms. Overall, it has been found that transfer learning techniques with ensemble models/features result very well-compared to machine learning, custom-designed deep models, and transfer learning models. Additionally, our work highlights the importance of XAI, and its usage in limited works can unveil hidden reasoning. Ultimately, this comprehensive review will not only benefit researchers related to lung diseases but also increase the interest of researchers working on images, machine learning, machine vision, deep learning, and other related areas.

Marie Sklodowska-Curie-Innovative Training Network 2020, Grant Agreement N° 956394 (https://parenth2020.com/). This work was supported by the Politecnico di Torino, Italy and GPI, Italy.

Conflict of interest

SAS and AD were employed by GPI SpA.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abhisheka, B., Biswas, S. K., Purkayastha, B., Das, D., and Escargueil, A. (2023). Recent trend in medical imaging modalities and their applications in disease diagnosis: a review. *Multimed. Tools Appl.* 23, 1-36. doi: 10.1007/s11042-023-17326-1

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., Gertych, A., et al. (2017). A deep convolutional neural network model to classify heartbeats. *Comput. Biol. Med.* 89, 389–396. doi: 10.1016/j.compbiomed.2017.08.022

Akbulut, Y. (2023). Automated pneumonia based lung diseases classification with robust technique based on a customized deep learning approach. *Diagnostics* 13:260. doi: 10.3390/diagnostics13020260

Albawi, S., Mohammed, T. A., and Al-Zawi, S. (2017). "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)* (Antalya: IEEE), 1–6.

Alshmrani, G. M. M., Ni, Q., Jiang, R., Pervaiz, H., and Elshennawy, N. M. (2023). A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images. *Alexandria Eng. J.* 64, 923–935. doi: 10.1016/j.aej.2022.10.053

Ayan, E., and Ünver, H. M. (2019). "Diagnosis of pneumonia from chest X-ray images using deep learning," in *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science (EBBT)* (Istanbul: IEEE), 1–5.

Bal, U., Bal, A., Moral, Ö. T., Düzgün, F., and Gürbüz, N. (2024). A deep learning feature extraction-based hybrid approach for detecting pediatric pneumonia in chest X-ray images. *Phys. Eng. Sci. Med.* 47, 109–117. doi: 10.1007/s13246-023-01347-z

Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T., and Saalbach, A. (2019). Comparison of deep learning approaches for multi-label chest X-ray classification. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-42294-8

Beets-Tan, R. G., Oyen, W. J., and Valentini, V. (2020). *Imaging and Interventional Radiology for Radiation Oncology*. Berlin: Springer.

Benhar, H., Idri, A., and Fernández-Alemán, J. (2020). Data preprocessing for heart disease classification: a systematic literature review. *Comput. Methods Progr. Biomed.* 2020:105635. doi: 10.1016/j.cmpb.2020.105635

Bhandary, A., Prabhu, G. A., Rajinikanth, V., Thanaraj, K. P., Satapathy, S. C., Robbins, D. E., et al. (2020). Deep-learning framework to detect lung abnormality—a study with chest X-Ray and lung CT scan images. *Patt. Recogn. Lett.* 129, 271–278. doi: 10.1016/j.patrec.2019.11.013

Bhutta, Z. A. (2007). Dealing with childhood pneumonia in developing countries: how can we make a difference? *Archiv. Dis. Childh.* 92, 286–288. doi: 10.1136/adc.2006.111849

Bohr, A., and Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. *Artif. Intell. Healthc.* 2020:2. doi: 10.1016/B978-0-12-818438-7.00002-2

Candemir, S., and Antani, S. (2019). A review on lung boundary detection in chest X-rays. *Int. J. Comput. Assist. Radiol. Surg.* 14, 563–576. doi: 10.1007/s11548-019-01917-1

Chakrabarty, N. (2019). *Brain MRI Images for Brain Tumor Detection*. Available at: https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection/data (accessed October 10, 2023).

Chavez, M. A., Naithani, N., Gilman, R. H., Tielsch, J. M., Khatry, S., Ellington, L. E., et al. (2015). Agreement between the World Health Organization algorithm and lung consolidation identified using point-of-care ultrasound for the diagnosis of childhood pneumonia by general practitioners. *Lung* 193, 531–538. doi: 10.1007/s00408-015-9730-x

Choras, R. S. (2007). Image feature extraction techniques and their applications for CBIR and biometrics systems. *Int. J. Biol. Biomed. Eng.* 1, 6–16.

Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., et al. (2020). A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Appl. Sci.* 10:559. doi: 10.3390/app10020559

Ciet, P., Eade, C., Ho, M.-L., Laborie, L. B., Mahomed, N., Naidoo, J., et al. (2024). The unintended consequences of artificial intelligence in paediatric radiology. *Pediatr. Radiol.* 54, 585–593. doi: 10.1007/s00247-023-05746-y

Cohen, J. P., Morrison, P., Dao, L., Roth, K., Duong, T. Q., and Ghassemi, M. (2020). COVID-19 image data collection: prospective predictions are the future. *arXiv preprint arXiv:2006.11988.* doi: 10.59275/j.melba.2020-48g7

Courtenay, L. A., Herranz-Rodrigo, D., Huguet, R., Maté-González, M. Á., González-Aguilera, D., and Yravedra, J. (2020). Obtaining new resolutions in carnivore tooth pit morphological analyses: a methodological update for digital taphonomy. *PLoS ONE* 15:e0240328. doi: 10.1371/journal.pone.0240328

D'Ascenzo, F., De Filippo, O., Gallone, G., Mittone, G., Deriu, M. A., Iannaccone, M., et al. (2021). Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. *Lancet* 397, 199–207. doi: 10.1016/S0140-6736(20)32519-8

Feng, W., Dauphin, G., Huang, W., Quan, Y., Bao, W., Wu, M., et al. (2019). Dynamic synthetic minority over-sampling technique-based rotation forest for the classification of imbalanced hyperspectral data. *IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens.* 12, 2159–2169. doi: 10.1109/JSTARS.2019.2922297

Fields, B. K., Demirjian, N. L., and Gholamrezanezhad, A. (2020). Coronavirus disease 2019 (COVID-19) diagnostic technologies: a country-based retrospective analysis of screening and containment procedures during the first wave of the pandemic. *Clin. Imag.* 67, 219–225. doi: 10.1016/j.clinimag.2020.08.014

Franquet, T. (2001). Imaging of pneumonia: trends and algorithms. *Eur. Respirat. J.* 18, 196–208. doi: 10.1183/09031936.01.00213501

Gastal, E. S., and Oliveira, M. M. (2012). Adaptive manifolds for real-time high-dimensional filtering. *ACM Trans. Graph.* 31, 1–13. doi: 10.1145/2185520.2185529

Guido, R. C. (2018). A tutorial review on entropy-based handcrafted feature extraction for information fusion. *Inform. Fus.* 41, 161–175. doi: 10.1016/j.inffus.2017.09.006

Hall, E. J., and Giaccia, A. J. (2006). *Radiobiology for the Radiologist*. Philadelphia, PA: Lippincott Williams & Wilkins.

Hameed, N., Shabut, A. M., Ghosh, M. K., and Hossain, M. A. (2020). Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques. *Expert Syst. Appl.* 141:112961. doi: 10.1016/j.eswa.2019.112961

Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., and Geem, Z. W. (2020). Efficient pneumonia detection in chest xray images using deep transfer learning. *Diagnostics* 10:417. doi: 10.3390/diagnostics10060417

He, D.-C., Wang, L., and Guibert, J. (1987). Texture feature extraction. *Pat. Recogn. Lett.* 6, 269–273. doi: 10.1016/0167-8655(87)90087-0

Hussain, S., Mubeen, I., Ullah, N., Shah, S., Khan, B. A., Zahoor, M., et al. (2022). Modern diagnostic imaging technique applications and risk factors in the medical field: a review. *Biomed. Res. Int.* 2022:5164970. doi: 10.1155/2022/5164970

Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., et al. (2019). "CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 590–597.

Jain, R., Nagrath, P., Kataria, G., Kaushik, V. S., and Hemanth, D. J. (2020). Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning. *Measurement* 165:108046. doi: 10.1016/j.measurement.2020.108046

Javaid, M., and Haleem, A. (2020). Virtual reality applications toward medical field. *Clin. Epidemiol. Glob. Health* 8, 600–605. doi: 10.1016/j.cegh.2019.12.010

Johnson, A. E., Pollard, T. J., Greenbaum, N. R., Lungren, M. P., Deng, C.-Y., Peng, Y., et al. (2019). MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. *arXiv preprint arXiv:1901.07042*. doi: 10.1038/s41597-019-0322-0

Kaur, G., and Oberoi, A. (2020). *Novel Approach for Brain Tumor Detection Based on Naïve Bayes Classification. Data Management, Analytics and Innovation.* Berlin: Springer.

Kermany, D., Zhang, K., and Goldbaum, M. (2018). Labeled optical coherence tomography (OCT) and Chest X-Ray images for classification. *Mendeley Data* 2.

Khan, W., Zaki, N., and Ali, L. (2021). Intelligent pneumonia identification from chest x-rays: a systematic literature review. *IEEE Access*. 9, 51747–51771. doi: 10.1109/ACCESS.2021.3069937

Kiliçarslan, S., Közkurt, C., Baş, S., and Elen, A. (2023). Detection and classification of pneumonia using novel Superior Exponential (SupEx) activation function in convolutional neural networks. *Expert Syst. Appl.* 217:119503. doi: 10.1016/j.eswa.2023.119503

Kumar, S., Singh, S., and Kumar, J. (2019). *Multiple Face Detection Using Hybrid Features With SVM Classifier. Data and Communication Networks.* Berlin: Springer.

Labhane, G., Pansare, R., Maheshwari, S., Tiwari, R., and Shukla, A. (2020). "Detection of pediatric pneumonia from chest X-ray images using CNN and transfer learning," in *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)* (Jaipur: IEEE), 85–92.

Le Roux, D. M., and Zar, H. J. (2017). Community-acquired pneumonia in children-a changing spectrum of disease. *Pediatr. Radiol.* 47, 1392–1398. doi: 10.1007/s00247-017-3827-8

Liang, G., and Zheng, L. (2020). A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Comput. Methods Progr. Biomed.* 187:104964. doi: 10.1016/j.cmpb.2019.06.023

Lundberg, S. M., and Lee, S-I. (2017). A unified approach to interpreting model predictions. *Adv. Neural Inform. Process. Syst.* 30, 4768–4777.

Maeda, R., Fujita, D., Tanaka, K., Ozawa, J., Haga, M., Miyahara, N., et al. (2023). *Predicting the Development of Chronic Lung Disease in Neonataes from Chest X-ray Images Using Deep Learning.* London: Nature Publishing Group.

Mahomed, N., Van Ginneken, B., Philipsen, R. H., Melendez, J., Moore, D. P., Moodley, H., et al. (2020). Computer-aided diagnosis for World Health Organization-defined chest radiograph primary-endpoint pneumonia in children. *Pediatr. Radiol.* 50, 482–491. doi: 10.1007/s00247-019-04593-0

Michelucci, U., Sperti, M., Piga, D., Venturini, F., and Deriu, M. A. (2021). A model-agnostic algorithm for bayes error determination in binary classification. *Algorithms* 14:301. doi: 10.3390/a14110301

Naicker, S., Plange-Rhule, J., Tutt, R. C., and Eastwood, J. B. (2009). Shortage of healthcare workers in developing countries–Africa. *Ethnicity Dis.* 19:60.

Narasimhan, V., Brown, H., Pablos-Mendez, A., Adams, O., Dussault, G., Elzinga, G., et al. (2004). Responding to the global human resources crisis. *Lancet* 363, 1469–1472. doi: 10.1016/S0140-6736(04)16108-4

Navarini, L., Sperti, M., Currado, D., Costa, L., Deriu, M. A., Margiotta, D. P. E., et al. (2020). A machine-learning approach to cardiovascular risk prediction in psoriatic arthritis. *Rheumatology* 59, 1767–1769. doi: 10.1093/rheumatology/kez677

Nelson, S., Mason, C. M., Kolls, J., and Summer, W. R. (1995). Pathophysiology of pneumonia. *Clin. Chest Med.* 16, 1–12. doi: 10.1016/S0272-5231(21)00975-8

Nguyen, T., Tran, T., Roberts, C., Graham, S., and Marais, B. (2017). Child pneumonia-focus on the Western Pacific Region. *Paediatr. Respirat. Rev.* 21, 102–110. doi: 10.1016/j.prrv.2016.07.004

Oliveira, L. L. G., Silva, E. S. A., Ribeiro, L. H. V., De Oliveira, R. M., Coelho, C. J., and Andrade, A. L. (2008). Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. *Int. J. Med. Informat.* 77, 555–564. doi: 10.1016/j.ijmedinf.2007.10.010

Pham, H. H., Nguyen, N. H., Tran, T. T., Nguyen, T. N., and Nguyen, H. Q. (2023). PediCXR: an open, large-scale chest radiograph dataset for interpretation of common thoracic diseases in children. *Sci. Data* 10:240. doi: 10.1038/s41597-023-02102-5

Philip, B., Jain, A., Wojtowicz, M., Khan, I., Voller, C., Patel, R. S., et al. (2023). Current investigative modalities for detecting and staging lung cancers: a comprehensive summary. *Ind. J. Thoracic Cardiovasc. Surg.* 39, 42–52. doi: 10.1007/s12055-022-01430-2

Prodanovic, T., Petrovic Savic, S., Prodanovic, N., Simovic, A., Zivojinovic, S., Djordjevic, J. C., et al. (2024). Advanced diagnostics of respiratory distress syndrome in premature infants treated with surfactant and budesonide through computer-assisted chest x-ray analysis. *Diagnostics* 14:214. doi: 10.3390/diagnostics14020214

Raghu, M., Zhang, C., Kleinberg, J., and Bengio, S. (2019). Transfusion: understanding transfer learning for medical imaging. *Adv. Neural Inform. Process. Syst.* 2019, 3347–3357. doi: 10.48550/arXiv.1902.07208

Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., et al. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest x-ray. *Appl. Sci.* 10:3233. doi: 10.3390/app10093233

Rajaraman, S., Candemir, S., Thoma, G., and Antani, S. (2019). Visualizing and explaining deep learning predictions for pneumonia detection in pediatric chest radiographs. Medical Imaging 2019: Computer-Aided Diagnosis. *Int. Soc. Opt. Photon.* 2019:109500S. doi: 10.1117/12.2512752

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). ""Why should i trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* [New York, NY: Association for Computing Machinery (ACM)], 1135–1144.

Rudan, I., Tomaskovic, L., Boschi-Pinto, C., and Campbell, H. (2004). Global estimate of the incidence of clinical pneumonia among children under five years of age. *Bullet. World Health Org.* 82, 895–903.

Saraiva, A. A., Santos, D., Costa, N. J. C., Sousa, J. V. M., Ferreira, N. M. F., Valente, A., et al. (2019). Models of learning to classify x-ray images for the detection of pneumonia using neural networks. *Bioimaging.* 2019, 76–83. doi: 10.5220/0007574900760083

Selvaraju, R. R., DAS, A., Vedantam, R., Cogswell, M., Parikh, D., and Batra, D. (2016). Grad-CAM: why did you say that? *arXiv preprint arXiv:1611.07450*. doi: 10.48550/arXiv.1611.07450

Siddiqi, R. (2019). "Automated pneumonia diagnosis using a customized sequential convolutional neural network," in *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies* [New York, NY: Association for Computing Machinery (ACM)], 64–70.

Singh, S. (2024). CoVScreen: Pitfalls and recommendations for screening COVID-19 using Chest X-rays. *arXiv preprint arXiv:2405.07674*. doi: 10.48550/arXiv.2405.07674

Sousa, R. T., Marques, O., Soares, F. A. A., Sene, J. R., I. I., De Oliveira, L. L., et al. (2013). Comparative performance analysis of machine learning classifiers in detection of childhood pneumonia using chest radiographs. *Proc. Comput. Sci.* 18, 2579–2582. doi: 10.1016/j.procs.2013.05.444

Stephen, O., Sain, M., Maduh, U. J., and Jeong, D.-U. (2019). An efficient deep learning approach to pneumonia classification in healthcare. *J. Healthc. Eng.* 2019:4180949. doi: 10.1155/2019/4180949

Sun, M., Tang, F., Yi, J., Wang, F., and Zhou, J. (2018). "Identify susceptible locations in medical records via adversarial attacks on deep predictive models," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 793–801.

Suzuki, K. (2017). Overview of deep learning in medical imaging. *Radiol. Phys. Technol.* 10, 257–273. doi: 10.1007/s12194-017-0406-5

Togaçar, M., Ergen, B., Cömert, Z., and Özyurt, F. (2020). A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. *IRBM* 41, 212–222. doi: 10.1016/j.irbm.2019.10.006

Torres, A., Cilloniz, C., Niederman, M. S., Menéndez, R., Chalmers, J. D., Wunderink, R. G., et al. (2021). Pneumonia. *Nat. Rev. Dis. Prim.* 7:25. doi: 10.1038/s41572-021-00259-0

Tripathy, S., and Swarnkar, T. (2020). A comparative analysis on filtering techniques used in preprocessing of mammogram image. *Adv. Comput. Intell. Eng.* 6:39. doi: 10.1007/978-981-15-1081-6_39

Ukwuoma, C. C., Qin, Z., Heyat, M. B. B., Akhtar, F., Bamisile, O., Muaad, A. Y., et al. (2023). A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images. *J. Adv. Res.* 48, 191–211. doi: 10.1016/j.jare.2022.08.021

UNICEF (2022). *Pneumonia in Children Statistics.* Available at: https://data.unicef.org/topic/child-health/pneumonia/ (accessed July 3, 2023).

Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., and Kennedy, P. J. (2016). "Training deep neural networks on imbalanced data sets," in *2016 International Joint Conference on Neural Networks (IJCNN)* (Vancouver, BC: IEEE), 4368–4374.

Wang, W., and Lu, Y. (2018). "Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model," in *IOP Conference Series: Materials Science and Engineering* (Bristol: IOP Publishing).

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., and Summers, R. M. (2017). "Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Piscataway, NJ: IEEE), 2097–2106.

Xing, L., Giger, M. L., and Min, J. K. (2020). *Artificial Intelligence in Medicine: Technical Basis and Clinical Applications.* Cambridge, MA: Academic Press.

Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imag.* 9, 611–629. doi: 10.1007/s13244-018-0639-9

Yao, J., Dwyer, A., Summers, R. M., and Mollura, D. J. (2011). Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification. *Acad. Radiol.* 18, 306–314. doi: 10.1016/j.acra.2010.11.013

Yi, R., Tang, L., Tian, Y., Liu, J., and Wu, Z. (2023). Identification and classification of pneumonia disease using a deep learning-based intelligent computational

framework. *Neural Comput. Appl.* 35, 14473–14486. doi: 10.1007/s00521-021-0 6102-7

Yildirim, A. E., and Canayaz, M. (2023). A novel deep learning-based approach for prediction of neonatal respiratory disorders from chest X-ray images. *Biocybernet. Biomed. Eng.* 43, 635–655. doi: 10.1016/j.bbe.2023.08.004

Yoon, T., and Kang, D. (2024). Enhancing pediatric pneumonia diagnosis through masked autoencoders. *Sci. Rep*. 14:6150. doi: 10.1038/s41598-024-56819-3

Yu, Y., Gao, J., Mu, X., and Wang, S. (2023). Adaptive LSB quantum image watermarking algorithm based on Haar wavelet transforms. *Quant. Inform. Process.* 22:180. doi: 10.1007/s11128-023-0 3926-1

Zeiler, M. D., and Fergus, R. (2014). "Visualizing and understanding convolutional networks," in *Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12. 2014, Proceedings, Part I. 13* (Berlin: Springer), 818–833.

Zheng, S., and Ding, C. (2020). A group lasso based sparse KNN classifier. *Pat. Recogn. Lett*. 131, 227–233. doi: 10.1016/j.patrec.2019. 12.020