



A systematic literature review on deep learning approaches for pneumonia detection using chest X-ray images

Shagun Sharma¹ · Kalpana Guleria¹

Received: 19 January 2023 / Revised: 22 May 2023 / Accepted: 24 July 2023 /

Published online: 9 August 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

As per World Health Organization, in 2019, 2.5 million deaths were reported due to pneumonia, of which 14% were observed among children between 0–5 years of age. Due to the increased mortality rate, it is essential to diagnose pneumonia to avoid the failure of the human body's functioning. Machine and deep learning techniques can be implemented for pneumonia prediction, but deep learning is preferred over machine learning due to its applicability of better performance outcomes along with an automatic feature extraction from the dataset. This systematic literature review meticulously discusses a wide range of techniques for detecting pneumonia using deep learning, including convolutional neural networks, pre-trained models, and ensemble models. The review provides an in-depth illustration of architecture and working process and evaluates the effectiveness of these models in solving various medical domain challenges. It presents a summarization and analytical discussion on convolutional neural networks-based, pre-trained, and ensemble models offering a deep insight into several factors, including performance measures, hyperparameters, and fine-tuning of the models. This meta-analysis also discusses the highly robust and outperforming ensemble pneumonia detection models. Furthermore, the review highlights various research gaps in the existing models, and probable solutions, enabling a deeper understanding of their performance and suitability for pneumonia detection tasks.

Keywords Pneumonia · Machine learning · Deep learning · Convolutional neural network · Pre-trained models · Ensemble models

1 Introduction

Researchers formally announced the term "artificial intelligence" in 1956, at Dartmouth University in a symposium, which introduced a new field of research for building machine intelligence to mimic human behavior [69]. In early 2016, a computer program called AlphaGo beat Lee Sedol, a top-level Go player in the Google DeepMind Challenge, which brought a revolution in the field of Artificial Intelligence (AI) all over the world. This

✉ Kalpana Guleria
guleria.kalpana@gmail.com

¹ Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

occurrence sparked widespread interest in AI, which has contributed to increasing the economic benefits for humanity and has assisted all aspects of society while also significantly promoting social development and leading the way into a new era of social improvement.

AI refers to a set of algorithms and tools having the goal of digitally simulating and replicating human intelligence [4, 56]. It is a broad term that trains machines to understand human actions, including judgment, learning, and decision-making by using computing powers and technology [69]. It is a learning initiative that accepts knowledge as its object, analyzes, investigates knowledge representation ways, and uses these techniques to simulate human cognitive operations. AI is a synthesis of logic, computer science, biology, philosophy, psychology, and various other disciplines that have yielded impressive breakthroughs in areas such as voice recognition, image classification, natural language processing, and intelligent robots. AI employs a variety of techniques drawn from the sub-disciplines of Machine Learning (ML) and Deep Learning (DL) to accelerate the automation and process the human skills, potentially resulting in a significant and tangible impact on healthcare [4, 6, 11, 19, 33, 41, 55, 61]. The medical application of AI has been recently expanded beyond clinical research and includes translational clinical and medical treatments for a variety of disorders, including cancers, covid-19, and many more [4, 40, 53, 65].

AI, big data, and telehealth predictive analytics are examples of digital health technologies that have the potential to reduce the consequences of various diseases due to pandemics and result in better public health [20]. Several of these digitalization have been tested and deployed to resolve the issues with the increased pandemic. The real-world creation and verification of such digitalization have been intriguing, and current studies have also emphasized major hurdles in installation and limits the clinical studies due to varying design and quality. Healthcare is an area that requires a huge amount of research to avoid mistakes and create an error-proof model [57].

Currently, various models are being utilized for healthcare data analysis and disease diagnosis. Convolution Neural Networks (CNN), pre-trained, and ensemble models are extensively used in the healthcare business to detect different diseases, including influenza, cancer, and pneumonia. Pneumonia is a respiratory infection caused by viruses, fungi, and bacteria [8, 14, 49, 60]. Due to the lack of early diagnosis of pneumonia, many people die every year [60]. It's a common infection that can be lethal. The frequency and the level of illness determine the severity of the disease and the level of medication required. Acute pneumonia is a severe condition that requires immediate medical care. In the past, the term "pneumonia" was used to describe a predicament that resulted in death and necessitated patients to be admitted to the Intensive Care Unit (ICU). Because the morbidity from pneumonia differs from the requirement for hospitalization, such interventions are increasingly seen as insufficient. The decision to move a patient to the hospital immediately is based on doctors' clinical judgment and their institutions' local practices, which could account for much of the difference in ICU initial treatment [15]. Identifying patients likely to have a major negative outcome is essential in minimizing the mortality rate due to pneumonia. As a result, some patients may not appear to be impacted in the beginning. The survival rates are connected to suboptimal care in the interval between hospitalization and ICU transfer. Accurate and early diagnosis of pneumonic patients or those with rapidly progressing pneumonia requires effective therapy for better recovery [15]. Pneumonia is identified and treated at emergency departments (EDs) in about 75% of incidents, making the facilities suitable for adopting a patient care approach for evaluating pneumonia. Identifying patients at higher risk for negative consequences can help improve evidence-based clinical judgments in the ED [15]. Hence, it is of utmost importance to design a system that can rapidly detect pneumonia in its initial stages so that the necessary measures can be taken to reduce the death rates.

Doctors utilize X-ray images to detect pneumonia. ML and DL techniques are extensively used in weather prediction and other sorts of forecasting, such as win-draw loss, earthquake prediction, and disease prediction [3, 8, 15, 21, 31, 46, 66]. Similarly, DL can be implemented on lung X-ray data to predict health concerns such as pneumonia and covid-19 [25, 44, 54]. The key reason for taking lung images is that pneumonia usually appears as an area of increased turbidity in the lung images [13]. Figure 1 illustrates pneumonia vs. normal lungs which highlights the region with increased turbidity with the red colour whereas the green colour highlights the normal lung [30, 60]. Moreover, many patients arrive in critical care having pre-existing lung disorders, making pneumonia diagnosis difficult. Fever, muscle aches, coughing, and difficulty in breathing are among the symptoms of pneumonia [30, 49].

Pneumonia disease has been documented throughout humanity's civilization, with references to the disease dating back to Greek civilization. Despite its lengthy background of sickness, pneumonia remains a primarily highly contagious infection for the global population, causing several severe morbidities and deaths [49]. As per the WHO, pneumonia claimed the lives of nearly 700,000 children under the age of 2 in 2016. Pneumonia is routinely evaluated with radiology, including CT scans and X-rays [2, 59]. This early examination is critical for determining the seriousness of the issue and developing a patient management solution. X-rays are a radiology process commonly used to assess a patient's illnesses, signs, and disorders. It is a rapid way of disease diagnosis in lung images [39]. It is usually used to diagnose disorders, but they've also proven valuable tools for monitoring cardiac disease. Radiologists frequently utilize radiographs to monitor how an illness spreads in patients' lungs. A detailed examination of radiographs reveals that it is used to assess where and how much recovery or deterioration has happened to establish a future treatment strategy. The proficiency of DL in image recognition and classification tasks enables its effective utilization for the early diagnosis of pneumonia. Researchers from diverse backgrounds in healthcare and technical fields have introduced a multitude of DL models aimed at achieving precise pneumonia prediction with enhanced outcomes.

This SLR provides an in-depth exploration of various existing DL pneumonia detection models and their performance analysis. An extensive study has been done by collecting articles from quality journals, along with their impact factor and h-index. The research process is carried out with the goal of achieving specific objectives. The following is an explanation of the planned research work's purpose and inspiration:

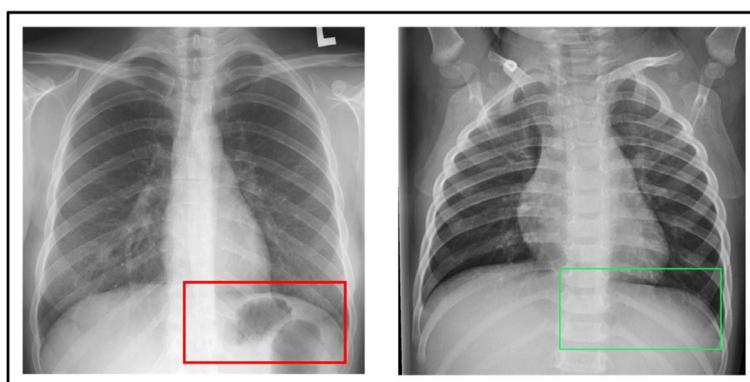


Fig. 1 Pneumonia Vs. normal lung

Pneumonia is associated with 14% of all fatalities in children younger than five years, killing 7,40,180 children in 2019 (World Health Organization, 2022). The major causes of pneumonia are bacteria, viruses, and fungi. Immunity vaccinations, a healthy diet, and mitigating environmental conditions can help to avoid pneumonia [48]. Medications can also treat pneumonia produced by bacteria, but barely one-third of children with acute pneumonia receive the antibiotics they require for the treatment. As a result, many patients die due to a late diagnosis of pneumonia and a lack of vaccines. Early detection leads to more effective treatment and a lower fatality rate. By examining the aforementioned factors, the article has contributed to identifying the empirical research for diagnosing pneumonia. This article put up an effort to bring out a guide to the researchers in terms of identifying the elementary material as well as principle factors that are required to develop a prediction model.

The contribution of this systematic review is as mentioned below:

- The presented work provides a taxonomy for DL-based pneumonia detection models, including CNN-based, pre-trained, and ensemble models. Further, it provides an in-depth illustration of architectures and working processes for the early detection of pneumonia.
- The article provides an analytical discussion on CNN-based, pre-trained, and ensemble models, which gives an insight into hyperparameter tuning, including optimizer, learning rate, epochs, batch size, and training/testing ratios. It also summarizes various factors, including dataset size, augmentation, and various performance parameters, including accuracy, sensitivity, precision, AUC, and recall.
- The work highlights various research gaps in the CNN-based, pre-trained, and ensemble models. Further, it also provides the probable solutions to overcome the challenges and limitations in this domain.

Table 1 provide the abbreviations used in the review.

The organization of this SLR for pneumonia detection is structured as follows: Section 1 introduces DL and pneumonia. Section 2 explores the related reviews carried out by various researchers for pneumonia detection. Section 3 elaborates on the methodology used for the research. Section 4 provides a detailed review of various pneumonia detection models. Section 5 provides the summary and analytical discussion of the existing works, along with performance measures and hyperparameter tuning. Lastly, the conclusion of this review is given in Section 6.

2 Related work

This section explores various systematic reviews in the domain of pneumonia classification methods. In [62], authors used the PRISMA technique to perform the SLR on pneumonia detection methods. The authors described various feature selection methods and performance estimation parameters for pneumonia detection. This SLR has considered only 16 articles to conduct a detailed study. Authors limited their review work to a small number of selected articles, and only three databases were considered, which include Scopus, Ovid, and Pubmed.

In [16], authors have worked to diagnose VAP accurately using ML algorithms. A SLR has been done to evaluate models to predict and detect VAP using patient reports,

Table 1 Abbreviation table

Abbreviation	Definition	Abbreviation	Definition
AI	Artificial Intelligence	MLP	Multi-Layer Perceptron
ANN	Artificial Neural Network	M R-CNN	Mask Recurrent-Convolutional Neural Network
BRISK	Binary Robust Invariant Scalable Key-points	NLP	Natural Language Processing
CAD	Computer Aided Diagnosis	NMS	Non Maximum Suppression
CGAN	Conditional Generative Adversarial Network	NN	Neural Network
CNN	Convolutional Neural Network	PCA	Principle Component Analysis
DBN	Deep Belief Network	ReLU	Rectified Linear Unit
DBNN	Deep Belief Neural Network	RF	Random Forest
DL	Deep Learning	RMSprop	Root Mean Squared propagation
DNN	Deep Neural Network	RNN	Recurrent Neural Network
EDCNN	Enhanced Deep Learning aided Convolutional Neural Network	RSNA	Radiological Society of North America
EDs	Emergency Departments	sCNN	Self Customized Simple CNN
FaNet	Fast Assessment Network	SCR	Segmentation in chest X-ray
FFNN	Feed Forward Neural Network	SGD	Stochastic Gradient Descent
GOM	Grande Ospedale Metropolitano	SLR	Systematic Literature Review
HIPAA	Health Insurance Portability and Accountability Act	SVM	Support Vector Machine
ICU	Intensive Care Unit	SVR	Support Vector Regression
KNN	K-Nearest Neighbor	VGG16	Visual Geometry Group-16
LSTM	Long Short Term Memory	VGG19	Visual Geometry Group-19
ML	Machine Learning	VAP	Ventilator Associated with Pneumonia

exhaled breath, clinical features, and demographic variables. This meta-analysis was tabulated by showing the results in performance outcomes, models, dataset size, and weights. The work is limited to only a few databases, which may not encompass a grey literature review. Furthermore, the work has focused only on the VAPs, which limits it to the generalization of other kinds of pneumonia.

In [52], authors have presented a review analysis of various DL models for identifying pneumonia in chest X-ray(CXR) images. The performance outcome, limitations, and effectiveness of the existing models have also been quantified. Although the work is a relevant and valuable resource on the application of DL for the classification and localization of pneumonia from covid-19, it is essential to note that this study has a narrow scope that encompasses only a few potential DL models in the context of pneumonia detection.

In [32], authors have proposed a meta-analysis of ML and DL models for predicting pneumonia existence in CXR images. The study is based on the security-based mechanism that can be used and implemented by various hospitals to extend this existing system to resolve various health challenges. This review has mentioned various existing ML and DL models, such as KNN, ANN, CNN, etc., for predicting pneumonia, along with the details of datasets, merits, and demerits of using these techniques. However, this SLR doesn't include an extensive quality assessment of pneumonia detection studies. Furthermore, this review also needs to provide the performance measure analysis of all the models presented in the literature, which limits the capability of the work to make meaningful comparisons of pneumonia detection techniques.

In [45], authors have depicted various publicly available pneumonia datasets and provided a survey of DL models which can be implemented on these datasets. However, this article has reviewed minimal studies on pneumonia detection models and lacks a comparative performance analysis and research gaps in the current work.

In [35], authors have performed a review analysis to identify the best-performing DL models for pneumonia prediction. The authors have tabulated the details of the architecture, the dataset used, and the images count, along with the train and test split. However, the review is limited to only 15 articles; hence, this lacks recent research in this domain.

Though there exist research articles in this domain, however, as per the related work, there are only a few articles that have included the research gaps and shortcomings of the existing models. These review articles are limited to a few CNN-based DL models; however, there is a strong need to introduce the current research methods in the domain of pneumonia detection using deep learning and also to put forward the challenges and probable solutions for the same. This article is an effort to explore the aforementioned criteria. A SLR of various DL models for pneumonia detection has been made, which includes CNN-based, pre-trained, and ensemble models. The available datasets, augmentation methods, DL model architectures, and the analytical summary of the existing work have been done. This article also addresses shortcomings, hyperparameters setting, and performance outcome of the existing methods, which can be helpful for the researchers in building a better pneumonia detection model.

3 Research methodology

To comprehend the DL approach for pneumonia detection, this section implies the PRISMA approach for SLR. It systematically assesses various publications and research on pneumonia detection using DL approaches. In addition, the effectiveness of the research

selection process is examined. The search procedure is outlined in depth in the subsequent sections, which include article selection, research gaps, and dataset selection.

The key aim of this study is to differentiate, review, classify, and assess all relevant publications found in DL applications for pneumonia prediction. The SLR can be utilized to investigate the key characteristics of the methodologies to attain the objectives (early pneumonia prediction).

3.1 Article selection process

This subsection provides information about the selection process of the articles in this complete literature survey.

The systematic literature search methodology for this analysis is divided into four steps, as represented in Fig. 2.

Table 2 shows the keywords used to review the publications. To choose articles for this SLR, a keyword search of pertinent databases has been done. This entails recognizing relevant keywords or phrases and utilizing them to explore databases like Scopus, PubMed, Web of Science, IEEE Xplore, Google Scholar, Springer Link, Elsevier, ACM, MDPI, Emerald Insight, Wiley, and Taylor & Francis. Furthermore, citation searching was utilized in which the identification of important articles related to the research questions and using their references to find additional relevant articles has also been made. In addition, the snowballing process was also used, which involves the identification of relevant articles by using their citations and reference list. The articles identified include conference papers, books, journals, technical analysis, book chapters, special issues, and notes. In the first stage, a total of 144 papers were gathered.

In the first stage, the publications are evaluated using the parameters shown in Fig. 3. The 96 papers have been left at this stage. The duplicate articles and various survey papers were removed from the list, leaving 87 articles in stage 3. The majority of publications for the study were determined in stage 2. Finally, in stage 4 and stage 5, the relevant papers are kept and selected on the basis of their H-index, impact factor, and quality, which resulted in 57 relevant articles for review. The criteria for the selection of relevant and related papers are impact factor, results, H-index, etc. There are various reasons for selecting only 57 articles, including their relevance, quality, results, and the DL techniques. The specifications of various selected articles are shown in Table 3. This specification includes various

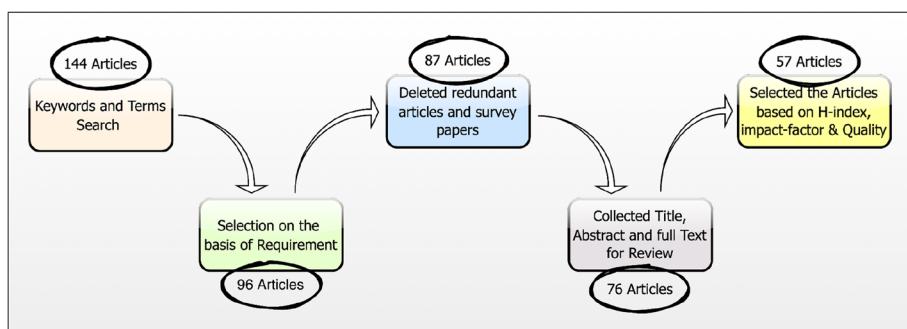
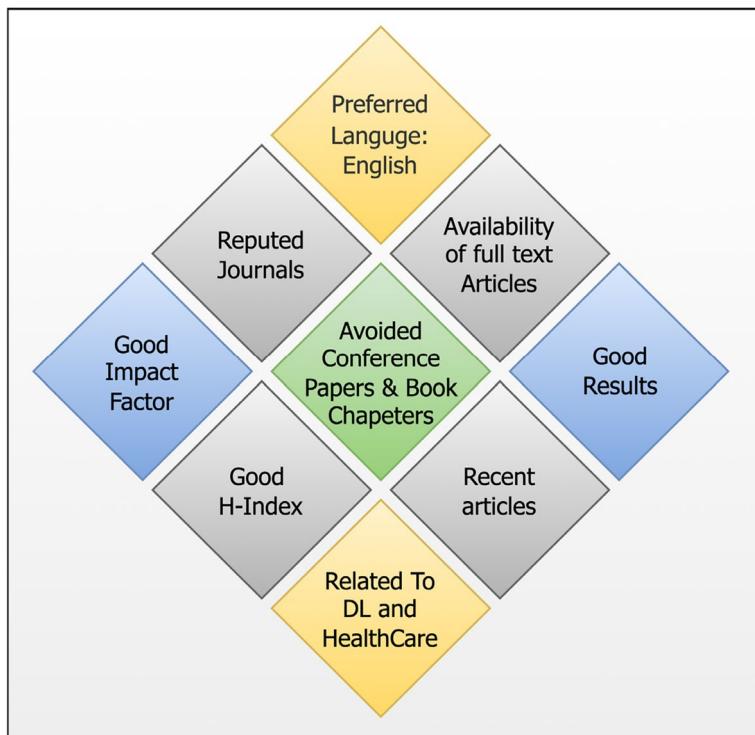


Fig. 2 Selection of Articles using Systematic Literature Survey Methodology

Table 2 Keywords and phrases used for systematic literature review

Search#	Keywords and Phrases
Search1	“Deep Learning” and “healthcare”
Search2	“Machine Learning” in “pneumonia detection”
Search3	“Deep Learning Applications” in “disease diagnosis”
Search4	“Machine Learning” and “pneumonia”
Search5	“Deep Learning” and “pneumonia”
Search6	“Deep Learning” in “disease prediction”
Search7	“AI algorithms” for “pneumonia detection/prediction”
Search8	“Transfer Learning methods” and “medical issues”
Search9	“Deep Learning” and “pneumonia dataset”
Search10	“Machine Learning and Deep Learning methods” for “healthcare sector”
Search11	“Transfer Learning” for “medical challenges detection and prediction”
Search12	“Deep Learning models” and “covid-19 pneumonia detection”
Search13	“Deep Learning and Machine Learning” applied to “chest X-ray images for pneumonia prediction”

**Fig. 3** Article Selection Criteria

articles used in this SLR, along with their publishers, citation, SJR ranking, journal name, h-index, and impact factor.

The distribution of the articles and their publisher have been illustrated in Fig. 4.

Figure 5 shows the distribution of the articles used in the SLR based on the publisher. This illustrated distribution shows that the maximum number of articles have been published under Elsevier and Springer.

3.2 Research questions identification and formalization

The objective of creating these questions is to inform the readers of the article about the scope of this SLR. The arrangement of the research questions is designed in such a way that all essential concepts of pneumonia, along with DL and ML approaches, are completely covered. The major purpose of this SLR is to examine the present state-of-the-art relevant to the early diagnosis of pneumonia. In this SLR, the major objective is to seek, investigate and integrate the existing resources about the examined research types, research subjects, research methods, research gaps, and contribution in the domain of early pneumonia prediction and classification.

Question 1. Why early diagnosis of pneumonia is an important field of research?

Pneumonia has been found as the leading cause of death due to late diagnosis and limited resources. Each year, a number of patients die, which results in an increased mortality rate. This research question aims to identify the technology that can be implemented for the accurate and early prediction of pneumonia to improve the increased mortality rate. Furthermore, the solution to this research question has been answered in Section 1 by introducing a broader perspective for effective and early pneumonia prediction technology.

Question 2. What are the related reviews available in this domain? What shortcomings have been analyzed in the existing SLRs?

To answer this research question, related studies conducted in this domain have been presented in Section 2. Additionally, this section also provides a summary of shortcomings along with the topic coverage issues.

Question 3. What is the best suitable technique for conducting an SLR for pneumonia diagnosis?

The purpose of this research question is to discuss the step-by-step procedure for conducting the SLR by using the PRISMA technique. In addition, the keyword searching, publication details, along with article selection criteria have been discussed in Section 3.

Question 4. What are the DL techniques and their frameworks for the early diagnosis of pneumonia?

The goal of this research question is to present the pneumonia diagnosis models along with their architectures. The architectural frameworks and detailed elaboration of each DL-based pneumonia diagnosis model have been presented in Section 4.

Question 5. What are the research gaps, performance outcomes, and pneumonia datasets in the existing studies?

The objective of this research question is to summarize the existing CNN-based, pre-trained, and ensemble models by incorporating their techniques, research gaps, dataset details, augmentation methods, comparative performance analysis, and shortcomings. Furthermore, the probable solutions for improving the performance of existing models and the most suitable techniques for pneumonia detection and classification have also been discussed in the summary presented in Section 5.

Table 3 Specifications of selected articles

Ref	Publisher	Citation	SJR	Scopus/SCI	Journal Name	H-Index	Impact Factor
[67]	PUBLIC LIBRARY SCIENCE	602	4.85/Q1	SCIE	PLoS medicine	228	11.07
[60]	PERGAMON-ELSEVIER SCIENCE LTD	102	0.63/Q1	SCIE	Computers & Electrical Engineering	64	4.143
[64]	HINDAWI LTD	282	0.51/Q2	SCIE	Journal of Healthcare Engineering	29	3.06
[24]	MDPI	95	0.622/Q3	SCIE	diagnostics	19	3.706
[63]	ELSEVIER SCIENCE INC	118	0.35/Q3	SCIE	IRBM	29	2.30
[47]	Elsevier Ltd	306	0.44/Q3	SCIE	Informatics in Medicine Unlocked	21	3.37
[28]	Elsevier	2	1.66/Q1	SCIE	Neurocomputing	157	5.719
[68]	LIPPINCOTT WILLIAMS & WILKINS	11	0.47/Q3	SCIE	Medicine	155	1.889
[34]	PUBLIC LIBRARY SCIENCE	11	0.85/Q1	SCIE	PLOS One	367	3.24
[27]	Springer	74	1.35/Q1	SCIE	Cognitive Computation	56	5.418
[22]	Springer	90	1.13/Q1	SCIE	Journal of Medical Systems	89	4.46
[23]	Hindawi Publishing Corporation	17	0.33/Q2	SCIE	Mathematical Problems in Engineering	68	1.305
[5]	Springer International Publishing AG	4	0.5/Q3	SCIE	Journal of Medical and Biological Engineering	39	2.602
[26]	Springer Netherlands	7	1.21/Q2	SCIE	Applied Intelligence	72	5.086
[49]	Multidisciplinary Digital Publishing Institute (MDPI)	7	0.8/Q1	SCIE	Sensors	196	3.576
[2]	Frontiers Media S.A	-	0.81/Q1	SCIE	International Journal of Environmental Research and Public Health	138	3.39
[18]	Elsevier Ltd	5	1.31/Q1	SCIE	Computers in Biology and Medicine	102	4.589
[9]	Elsevier BV	16	1.21/Q1	SCIE	Biomedical Signal Processing and Control	84	3.88
[59]	Springer Netherlands	1	0.72/Q1	SCIE	Multimedia Tools and Applications	80	2.757
[29]	Taylor and Francis Ltd	244	1.03/Q1	SCIE	Measurement	18	3.927
[17]	Elsevier Ltd	2	1.31/Q1	SCIE	Computers in Biology and Medicine	102	4.589
[58]	Elsevier BV	-	0.53/Q2	SCIE	Healthcare	23	1.76

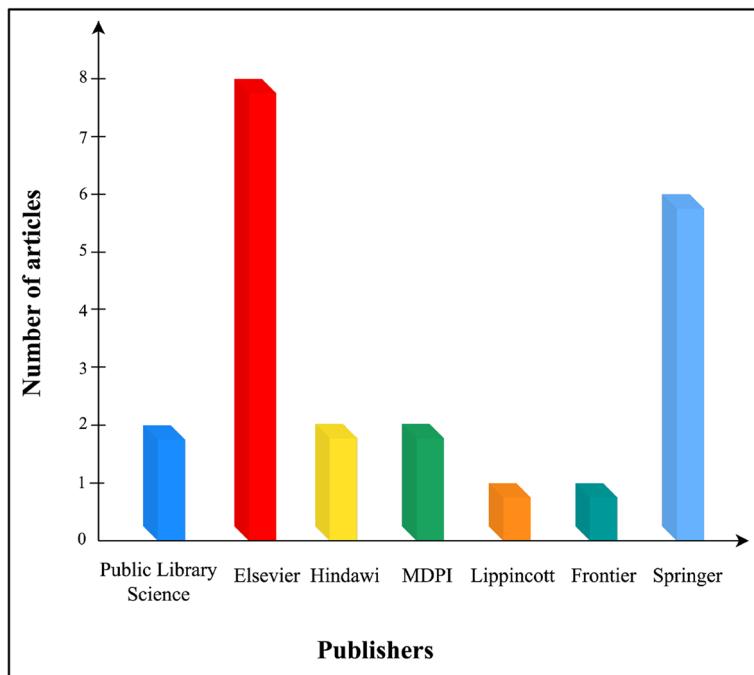


Fig. 4 Number of articles published by various publishers used in SLR

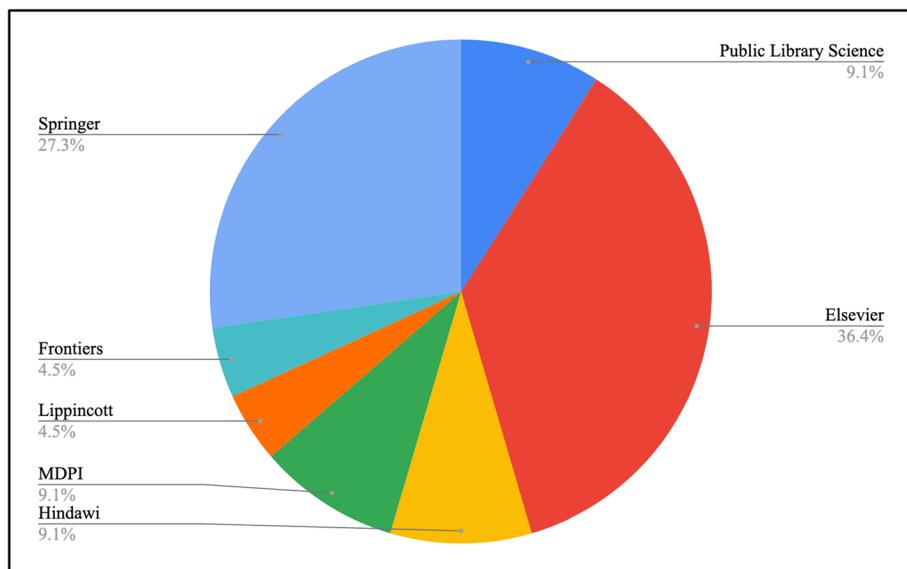


Fig. 5 Distribution of articles based on publisher

4 Literature review on deep learning-based models for pneumonia detection

The presence of pneumonia can be recognized by identifying the symptoms such as yellow, greenish, or crimson mucus produced by coughing, fever, perspiration, shivering, difficulty in breathing, chest pain, low energy, appetite loss, and lethargy. It is important to predict pneumonia at an early stage so that early medication and preventive measures can be taken to save precious human lives. Many researchers devised various ways to diagnose pneumonia disorders using ML and DL techniques. There is significant research made in this domain by various researchers, which includes a variety of models with varied degrees of prediction success. Lung disease detection from X-ray images has proven to be successful with DL and ML models. It's being utilized in healthcare practices to help diagnose diseases. Further, DL-based systems have also been created to automatically diagnose pneumonia in the covid-19 patients using radiology images [18].

This section provides a taxonomy of various pneumonia detection models, which include CNN-based, pre-trained, and ensemble models. CNN models are the DL models which often used in various computer vision applications. The main objective is to use image data to recognize and extract hierarchical features [36] consequently. The CNN architecture consists of convolutional, pooling, and FC layers. These layers work together to identify localized patterns, high-level representations, and spatial relationships that appear in images. CNNs enable efficient analysis and understanding of graphical information by utilizing these layers. The training methodology and application focus of CNN models set them apart from pre-trained models. CNN models are specialized architectures that are primarily made for performing image classification. On the other hand, the pre-trained models have undergone intensive training on a massive dataset by subject-matter experts [38]. The models are exposed to a lot of labeled data during training, including images or text, to help them recognize and extract significant features. The primary advantage of pre-trained models is to transfer learned skills from one task to another. These models gain a thorough understanding of the key trends and concepts underneath the data by experimenting with a variety of large-scale datasets. As a result, this learned information can be successfully applied to new, unforeseen tasks to identify efficient performance. Furthermore, the ensemble models combine various individual DL models to generate predictions [37]. The fundamental principle of ensemble modeling is that accuracy and performance can be enhanced over the use of a single model by combining predictions from various models. The ensemble models can be either heterogeneous or homogeneous types. Homogeneous models are those in which each model is of the identical kind, whereas, in heterogeneous ensemble models, each model is different. Every model in the ensemble architecture is independently trained using either a different algorithm or a subset of the training data. This strategy encourages a deeper investigation of the data patterns and adds diversity to their predictions.

There are several criteria for selecting articles within each of these three categories. The CNN-based models include techniques that have been built from scratch and trained on large datasets, whereas the pre-trained category has models which can be fine-tuned on specialized datasets to recognize a diverse set of features in images, and ensemble models include the ensemble of different types of DL techniques to enhance their performance. However, all models have the same ultimate goal of identifying pneumonia in CXR images, whereas their construction and operation procedures are different. Consequently, these models have been classified into three distinct categories. Figure 6 illustrates various pneumonia detection models under three major categories, which include CNN-based, pre-trained, and ensemble models.

4.1 Review of CNN-based models for pneumonia diagnosis

This section provides an in-depth elaboration of CNN-based models for pneumonia detection.

4.1.1 FaNet

In [26], authors have proposed a DL-based FaNet dual-task-based network model that quickly identifies and predicts the symptoms covid-19 by utilizing combined clinical symptoms and a 3D CT imaging dataset. This model has been applied to CT images containing information on covid-19 symptoms. This model has been implemented using three convolutional layers, and a batch normalization is performed after the second convolutional layer, as shown in Fig. 7. The input image is taken of size 512*512*160. The kernel size was taken as 1*1*64*32, 1*1*32*2, and 3*3*32*3, respectively.

The authors have proved that the FaNet model's accuracy for prediction and severity has been found as 98.28% and 94.83%, respectively.

4.1.2 Quaternion CNN

In [59], authors used a publicly accessible large chest radiographs dataset collected from Kaggle to train a Quaternion residual network algorithm and achieved F1-score and classification accuracy of 0.94 and 93.75%, respectively. The model has been divided into various steps, such as batch normalization, quaternion convolution, residual block, pooling, and activation. The quaternion convolution does the feature extraction, and normalizations

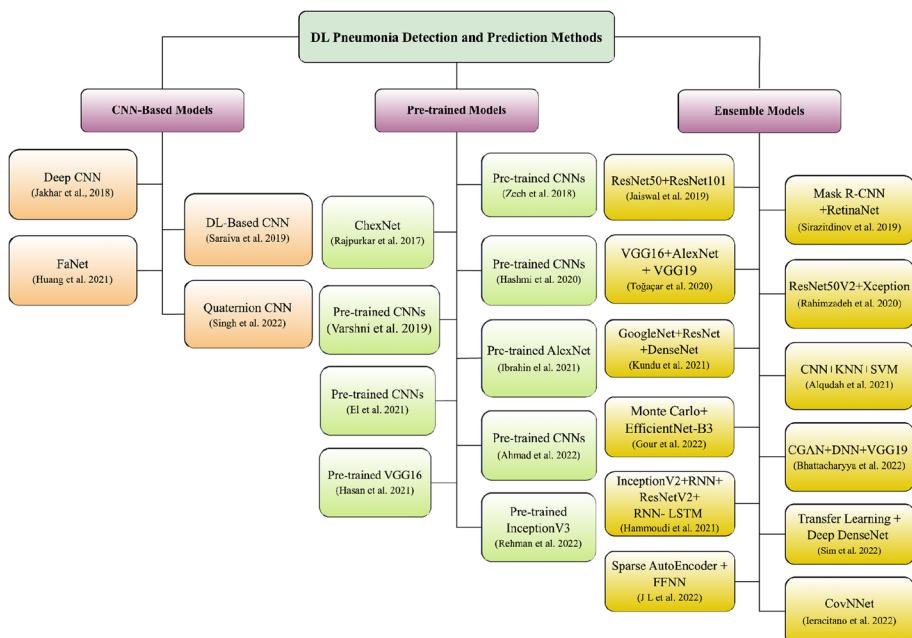


Fig. 6 Taxonomy of deep learning methods for pneumonia diagnosis and Prediction

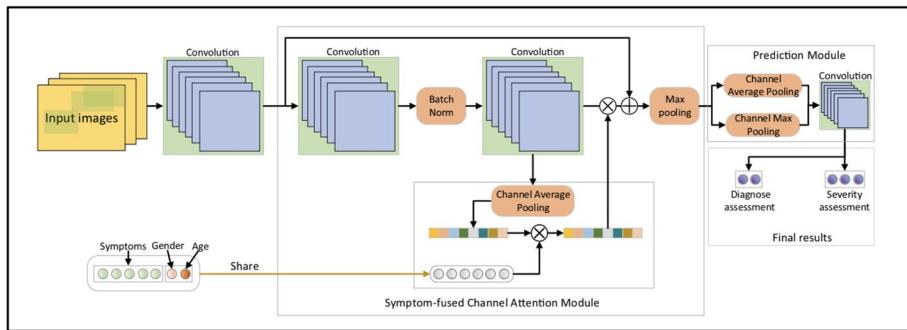


Fig. 7 FaNet framework proposed in [26]

pace up the training process while pooling, and ReLU(the range is shown in Eq. 1) is utilized for dimensionality reduction and avoiding the issue of vanishing gradient, respectively. Finally, a comparison between the results of the proposed model and existing CNN models was made, which has shown the Quaternion residual network as the outperforming algorithm as compared to others. The workflow of the model is shown in Fig. 8.

$$f(x) = \max(0, x) \quad (1)$$

4.1.3 Deep CNN

In [30], authors have proposed a Deep CNN model, as shown in Fig. 9, for the accurate prediction of pneumonia, wherein the model has massive data for training with numerous deep layers. The data quality has been improved by applying various pre-processing and data-cleaning operations. After extracting characteristics from high-quality radiology imaging data, the proposed implementations were carried out and achieved 84% prediction accuracy.

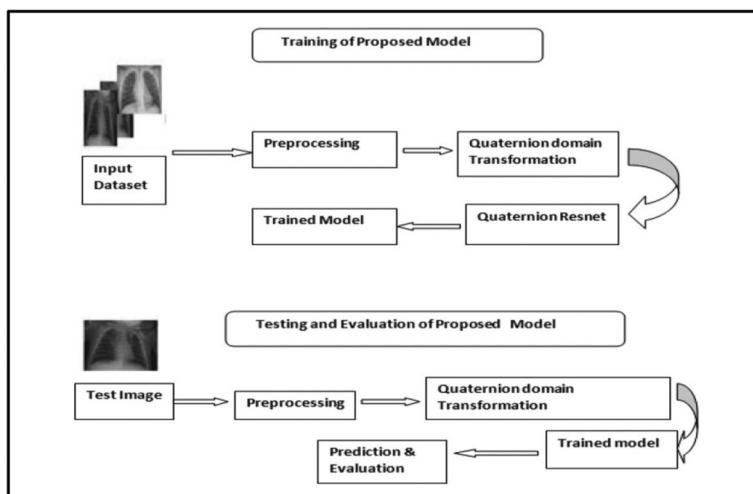


Fig. 8 Pneumonia diagnosis model proposed in [59]

4.1.4 DL-based CNN

In [51], authors used a DL-based CNN model, as shown in Fig. 10, to analyze 5,863 CXR. The CNN model contained 10 layers, among which the last three layers include softmax function without convolutions, while the other seven include convolutional layers. The complete model has been fed by 300*300 input images. The Adam optimizer and ReLU have been used in each convolution to increase the performance results of the model. After setting the parameters, the method is evaluated using k-fold cross-validation. The results were presented as a confusion matrix, with the proposed model's accuracy determined as 95.3%.

4.2 Review of pre-trained models for pneumonia diagnosis

Pre-trained models are the networks that have undergone extensive training on massive datasets, generally for huge image categorization and prediction tasks. In this process, either a transfer learning model or a pre-trained model can be adapted to perform a specific task. It has been proved through extensive research that for image identification and categorization tasks, utilizing a pre-trained classifier is advantageous for several reasons because

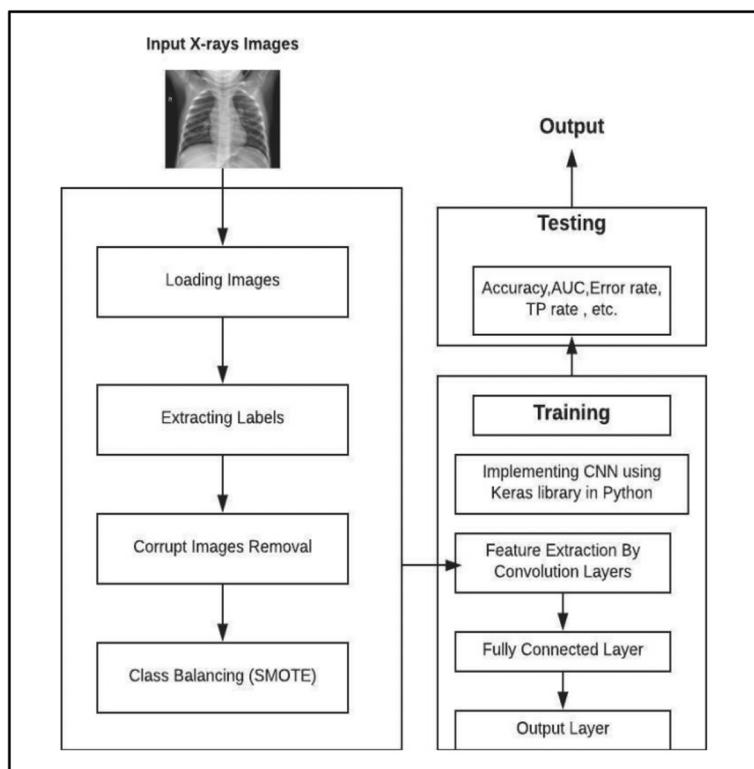


Fig. 9 Deep CNN model proposed in [30]

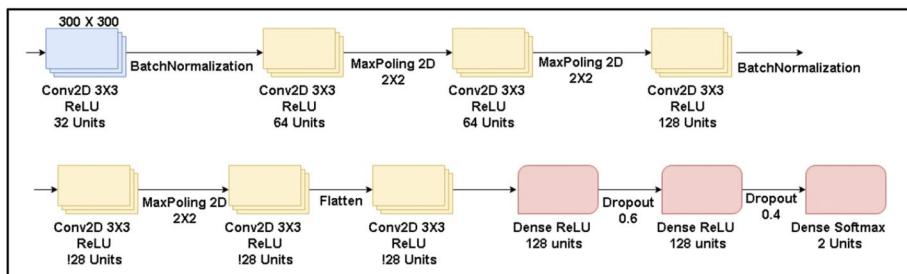


Fig. 10 Pneumonia diagnosis using CNN network presented in [51]

a pre-trained system reduces time utilization and effort by requiring minimal training and model design development. In this SLR, various pre-trained models have been identified for the prediction of pneumonia, as introduced below.

4.2.1 CheXNet

CheXNet is a DenseNet-based model consisting of 121 dense layers. In [48], authors have trained the model on the chestX-ray14 dataset. DenseNet models are used to enhance the gradients and data flow inside the prediction models and increase the optimization. This algorithm consists of a 121-layer CNN trained on the chestX-ray14 dataset, consisting of 100,000 frontal view radiology images with 14 diseases and presently found as the largest available CXR dataset to the public. On a test set annotated by four academic radiologists, a comparison among the performance of the CheXNet model and other models has been made. The outcome resulted in CheXNet as the best model by predicting the highest accuracy and effective performance in the form of the F1 metric. In chestX-ray14, an expansion of CheXNet was proposed to detect all 14 diseases, and the model used a sigmoid activation function at the fully connected(FC) layer with a single output neuron. Sigmoid activation function has been shown in Eq. (2).

$$S(x) = \frac{1}{1 + e^{-1}} \quad (2)$$

The input images for the training and testing of the model are taken of the size 224*224. Each layer in the model is connected with every other layer in the network for an enhanced architecture to reduce the error and increase the performance, as shown in Fig. 11.

4.2.2 Pre-trained CNN Models

In [67], authors have proposed two CNN models using pre-trained DenseNet121 and ResNet50. The models were trained on pooled datasets collected from various websites with varying pneumonia occurrences. The performance of the collected test data has been established, which resulted in the proposed model as the accurate method to identify pneumonia in the X-ray scans with a significant variance value. In this work, the different types of cross-sectional frameworks have been used to identify the generalizability of extrinsic sites by utilizing a validation which is named "test-split." The dataset has a size of 1,58,323 CXR, among which the mean age of the patients have been identified as 63.2 years, 46.9 years, and 49.6 years for Mount Sanai, NIH, and Indiana University patient care

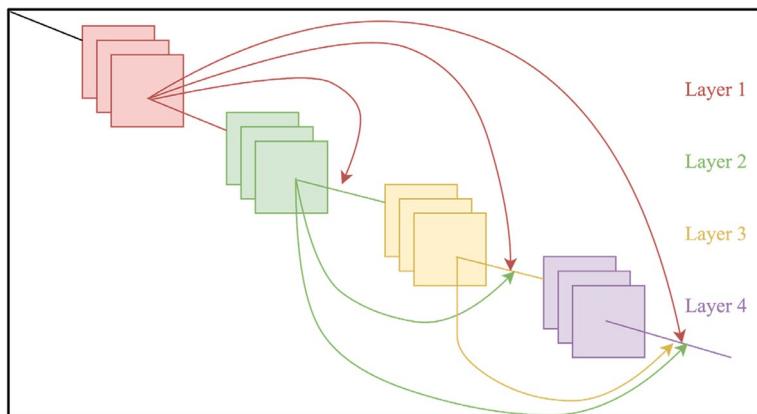


Fig. 11 DenseNet framework[64]

dataset with a percentage rate of female patients as 44.8%, 43.5%, and 57. 3%, respectively. The model's results have been analyzed in the AUC metric and finally compared with the existing models using Delong's method. This method is widely used for comparing the AUC results of different ML and DL techniques to identify the best-performing model.

4.2.3 Pre-trained CNNs

In [64], authors have used pre-trained CNN models (VGG16, ResNet50, DenseNet121, VGG19, and DenseNet169) and supervised ML (SVM, KNN, naive bayes, and random forest) as the feature extractors and classifier, respectively. The feature extraction from the dataset has been done using all the pre-trained models, whereas the classification has been performed with supervised ML to evaluate whether a person is infected with pneumonia. The proposed method differs from existing models as it does not rely solely on transfer learning to get classification outcomes. The results have identified that the feature extraction performed with DenseNet169 and classification with the SVM classifier achieved better results, showing an AUC of 0.8002. This model consists of 169 deep layers in which each layer has been taken as a feature extractor, and the output of the previous layer is the input to the next layer. Furthermore, the model has been developed by the implementation of four different dense blocks, whereas each block has two convolutional layers (1*1 size for the first layer and 3*3 for the second layer). The final classification layer used an average pooling of 7*7 size followed by the FC layer, which utilized softmax activation for resulting better performance outcomes. The framework for this model is shown in Fig. 12.

4.2.4 Pre-trained CNNs

In [24], authors have proposed a framework for pneumonia disease detection using chest radiographs, as shown in Fig. 13. The DL framework consists of a pre-processing and a data augmentation block wherein various augmentation operations have been considered to enhance the dataset which includes rotation, shift, shear, flip, crop, and pad. Further, the DL system has a fine-tuning block constructed using pre-trained models,

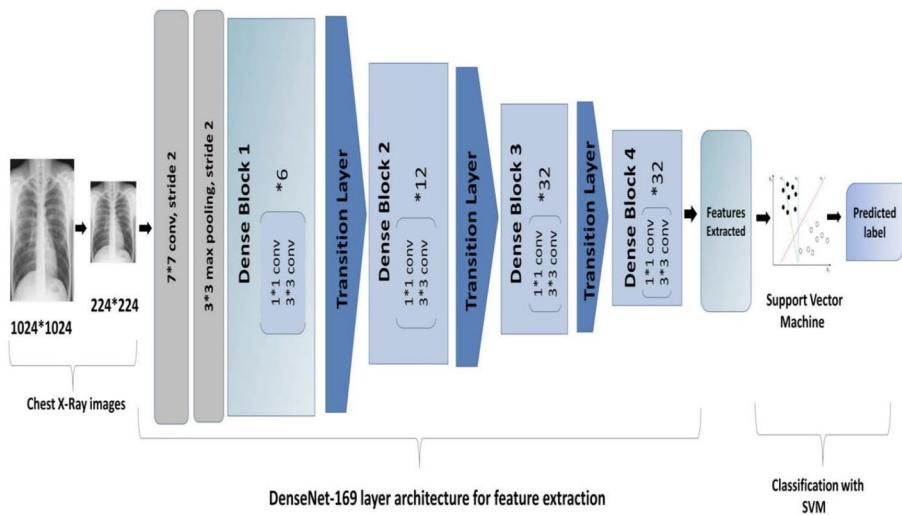


Fig. 12 Pre-trained CNNs proposed in [64]

MobileNetv2, ResNet18, InceptionV3, DenseNet121, and Xception. These DL models predict pneumonia by aggregating the outcome of each model. This framework uses different pre-trained architectures, which results in different input image size requirements. However, to achieve better results with these pre-trained models, the authors first resized the images to 224*224 and applied them to the models, including DenseNet121, MobileNetv2, and ResNet18, while the remaining models were fed with 299*299 input image size. The dataset used for feeding the model contained only 5,156 images; however, to solve the challenge of overfitting, the dataset has been augmented, which increased the number of images to 7,022. Additionally, the SGD optimizer, learning rate as 0.001, and momentum and epochs have been kept as 0.9 and 25, respectively, for the successful implementation and development of the pneumonia prediction model.

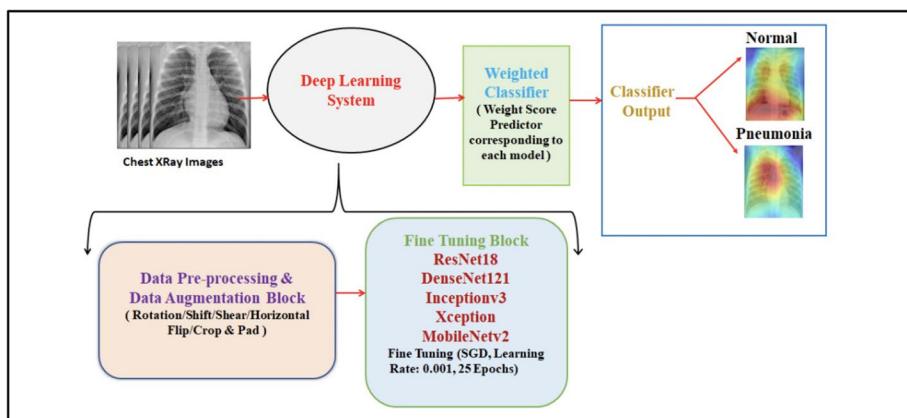


Fig. 13 Pneumonia detection framework using pre-trained models [24]

4.2.5 Pre-trained CNNs

In [7], authors have compared contemporary Deep CNN frameworks for the binary classification of pneumonia CXR using fine-tuned variants of CNN which includes InceptionResNetV2, VGG19, VGG16, InceptionV3, DenseNet201, MobileNetv2, Xception, and ResNet50. This method was evaluated using a CT dataset (1583 normal and 4273 pneumonia) and 5,856 CXR images, wherein data augmentation was applied to enhance the dataset for better model training. Figure 14 illustrates the framework of the proposed model. The collected dataset has been pre-processed by resizing them to 224*224 pixels for InceptionResNetV2, VGG19, VGG16, DenseNet201, MobileNetv2, Xception, and ResNet50 models, whereas for InceptionV3, all the images have been resized to 229*229. The proposed model is trained using various hyperparameters such as the batch size, epochs, optimizer, and learning rate have been taken as 32, 300, Adam, and 0.00001, respectively. Additionally, the results of each model used in the experiment have been analyzed, and it is found that ResNet50 has shown the highest accuracy of 96.61% whereas InceptionResNetV2, VGG19, VGG16, DenseNet201, MobileNetv2, Xception, and InceptionV3 have shown 96.09%, 85.94%, 86.26%, 93.66%, 96.27%, 83.14%, and 94.59%, respectively.

4.2.6 Pneumonia detection model using pre-trained AlexNet

In [27], authors have proposed a model based on pre-trained AlexNet for binary and multiclass disease classification. In the binary classification, authors classified between covid-19 and normal lungs, bacterial pneumonia and normal lungs, viral pneumonia and normal lungs, covid-19 and bacterial lungs, whereas, for multiclass, the classification was made among covid-19, bacterial pneumonia, viral pneumonia, and normal lungs. This pre-trained AlexNet model contained five convolutional layers, each having a filter of size 3*3 and padding of 2*2 for the max pooling layer. The AlexNet model has been fed with the input image size of 227*227*3, and five convolutional layers have been used with an input size of 55*55*96, 27*27*256, 13*13*384, 13*13*384, and 13*13*256 for each layer, respectively. In this model, the first two convolutional layers are passed through the max pooling and a normalization function to get the desired

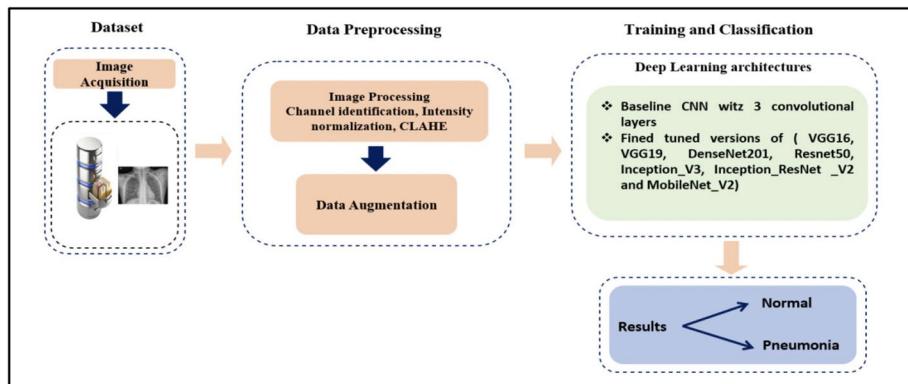


Fig. 14 Framework for pneumonia detection using various fine-tuned pre-trained models proposed in [7]

output of $27 \times 27 \times 256$ and $13 \times 13 \times 256$, as shown in Fig. 15. Furthermore, the next two convolutional layers have not been passed through max pooling layers which provide the outcome as $13 \times 13 \times 384$ in each one of them. Further, the output of the previous layer has been passed through one convolutional layer followed by max pooling resulting in $6 \times 6 \times 256$ outcome. Lastly, the model has used two FC layers and one output layer for the pneumonia classification.

4.2.7 Pneumonia detection model using pre-trained VGG16

In [23], authors used a pre-trained VGG16 model to detect pneumonia in CXR images collected from Kaggle. The data pre-processing and augmentation have been performed to remove noise and increase the number of images for better outcomes. The dataset has been divided into 80:20 ratios for training and testing of the model. The proposed VGG16 architecture consists of three FC layers followed by convolutional layers. During the training phase of this model, the learning rate and batch size were set as 0.001 and 16, respectively. Image flattening has been applied to convert the n-dimensional array to a 1-dimensional array. In the end, the classification layer has been added for predicting the outcome. Additionally, the ML-based LabelBinarizer tool has also been used to conduct an encoding on labeled CXR and convert them to categorical form, as shown in Fig. 16. The results were identified in the form of accuracy and sensitivity as 91.69% and 95.92%, respectively.

4.2.8 Pneumonia disease progression model with sequence learning using pre-trained CNNs

In [2], authors have proposed pneumonia detection model using various pre-trained CNNs which include InceptionV3, ResNet50, EfficientNetB0, EfficientNetB2, ChexNet, and InceptionResNetV2 on lung disease datasets. The authors devised a method to identify disease progression using sequence modeling, which extracts coarse and fine-grained characteristics. The proposed model quantified disease progression in positive and negative categories, which were paired with age-related risk parameters. To assure sequence learning for disease progression, the dataset considers images of patients with multiple scans, whereas images with single scan have been eliminated. This model is shown in Fig. 17, wherein sequence modeling is utilized for segregating images into different types, namely lung-based, zone-based, and image-based. Further, it has also been identified that disease progression using ChexNet with zone-based feature extraction has shown the highest AUC of 0.98 whereas image-based and lung-based strategies have achieved an AUC of 0.92 and 0.96, respectively.

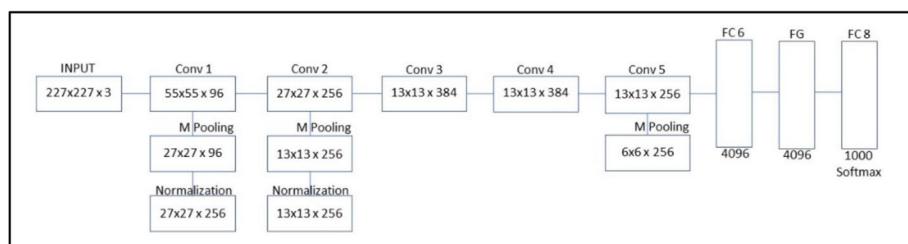


Fig. 15 Pneumonia detection model using pre-trained AlexNet proposed in [27]

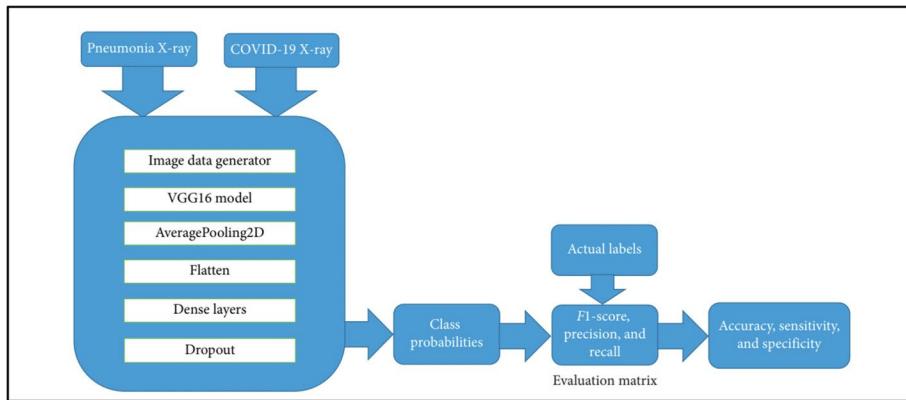


Fig. 16 Pneumonia detection model using pre-trained VGG16 proposed in [23]

4.2.9 Pneumonia prediction using pre-trained InceptionV3

In [49], authors have proposed a pneumonia detection model using pre-trained InceptionV3 on CXR images dataset. The method extracts various physiological characteristics like fever, chest pain, cough, low energy, flu, breathing issues, sweating, loss of appetite, fatigue, and headache from CXR images. The InceptionV3 model has been fine-tuned by adding two pooling layers to the architecture, followed by two convolutional layers. Afterward, a flattening layer was introduced to convert the N-dimensional array into a 1-dimensional array. Further, the classification has been performed, which identified that the pre-trained InceptionV3 achieves 97% accuracy. Additionally, the authors also implemented various ML-based models, which include RF, SVMs, naive bayes, and decision tree, and concluded that the proposed InceptionV3 performs better among all. The proposed model for pneumonia detection using InceptionV3 is shown in Fig. 18.

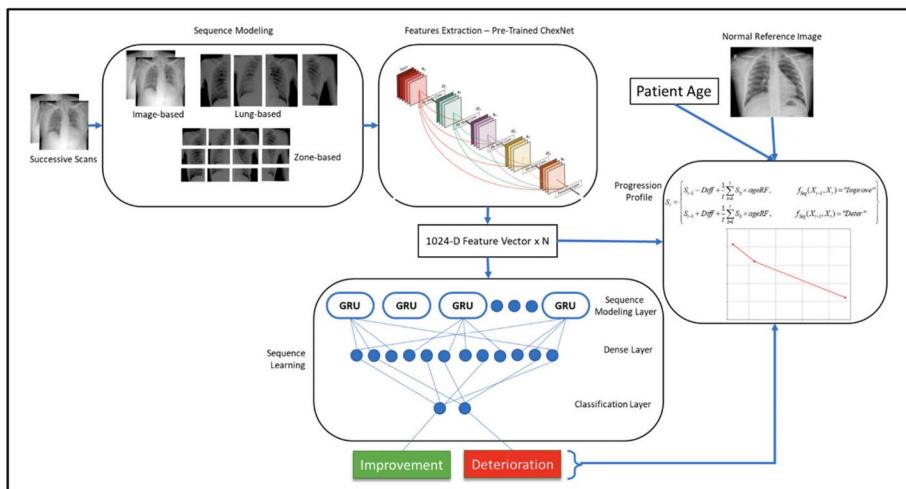


Fig. 17 Pneumonia disease progression model with sequence learning using pre-trained CNNs proposed in [2]

4.3 Review on ensemble models for pneumonia diagnosis

CNN classifiers that are uniform and coherent are combined to create ensemble DL models. Transfer learning approaches such as GoogleNet, AlexNet, VGG16, and VGG19 can be integrated to form an ensemble model. These models are created to increase prediction accuracy and to improve model performance. The ensemble DL approach integrates the predictions using various CNN models to decrease the generalization error in the prediction results. This section explores various ensemble models for the early prediction of pneumonia.

4.3.1 ResNet50 + ResNet101

In [29], authors have proposed an ensemble model using ResNet50 and ResNet101. This model has also been named Mask R-CNN. It is proposed by integrating two ResNet models consisting of 50 and 101 convolutional layers for identifying pneumonia in CXR. The CXR images contained the information of two stages: stage 1 and stage 2. Further, pixel-wise segmentation was done by identifying local and global features in the images. The input to the model is taken in the form of images with a size of 512*512. This model has integrated local and global data, as shown in Fig. 19, and achieved robustness by incorporating a novel post-processing step that combines bounding boxes from various models by enhancing the training procedure. The Mask R-CNN outperforms fast R-CNN by achieving a better threshold value of 0.218051 for pneumonia prediction.

4.3.2 Mask R-CNN + RetinaNet

In [60], authors have proposed an ensemble of two models, RetinaNet and Mask R-CNN, for pneumonia classification and identification. The input image has been taken of size 512*512

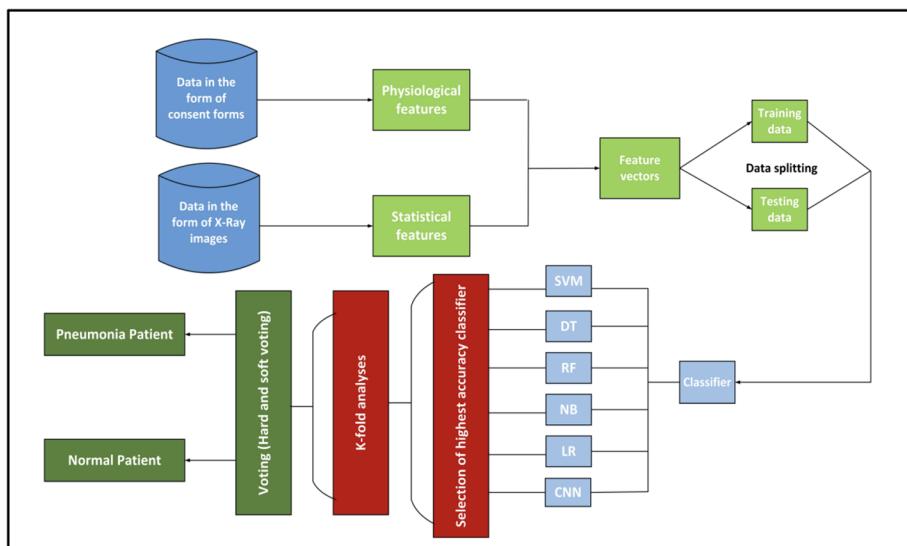


Fig. 18 Pneumonia prediction using pre-trained InceptionV3 [49]

for both models. The ResNet50 and ResNet101 were taken as the base models for RetinaNet and Mask R-CNN models, respectively. The Adam optimizer has been used with a learning rate of 0.0001 for RetinaNet and 0.001 for Mask R-CNN. Furthermore, for testing and validating the model against pneumonia, a publicly available dataset of 26,684 images collected from the Kaggle has been used. This ensemble model has been illustrated in Fig. 20, in which, after the successful extraction of the object detectors, the overlapping bounding boxes have been removed by using the NMS algorithm. The proposed model has been tested and validated for pneumonia detection using a publically available dataset. A comparison among proposed ensemble model, DenseNet121, and ResNet50 has been made. The results exhibit that the precision, recall, and F1-score of DensNet-121 as 0.883, 0.652, 0.731, ResNet50 is 0.855, 0.569, 0.683, and for the proposed ensemble model is 0.758, 0.793 and 0.775, respectively. Conclusively, it has been found that the combination of RetinaNet and Mask R-CNN shows better recall measures in comparison to DenseNet121 and ResNet50.

4.3.3 VGG16 + AlexNet + VGG19

In [63], authors have proposed an ensemble model consisting of VGG16, AlexNet, and VGG19. The proposed model is based on the traditional CNN model in which the input images were separately used for each model. The size of the input image and filter was taken as 227*227 and 3*3/5*5 for AlexNet model, whereas the input size and filter of 224*224 and 3*3 were substituted for both VGG16 and VGG19 models, respectively. The feature extraction was done using VGG16, AlexNet, and VGG19, wherein each model extracted 1000 features from the dataset. Thereafter, a complete set of features from each variant was reduced to 100 from 1000 by utilizing the low redundancy maximum relevance technique (mRMR). The proposed ensemble model performs classification using various ML classifiers, namely KNN, decision tree, linear regression, linear discriminant analysis, and SVM. The results exhibit that the linear discriminant analysis outperformed among all the models having an accuracy of 99.41%. The proposed architecture of the ensemble model is shown in Fig. 21.

4.3.4 ResNet50V2 + Xception

In [47], authors have proposed a pneumonia detection model wherein the extracted features of ResNet50V2 and Xception models have been concatenated to develop an ensemble framework. The concatenated features of size 10*10*4096 have been given as input to

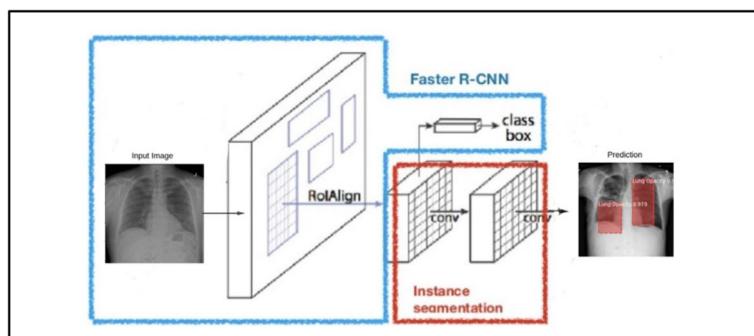


Fig. 19 Mask R-CNN proposed in [29]

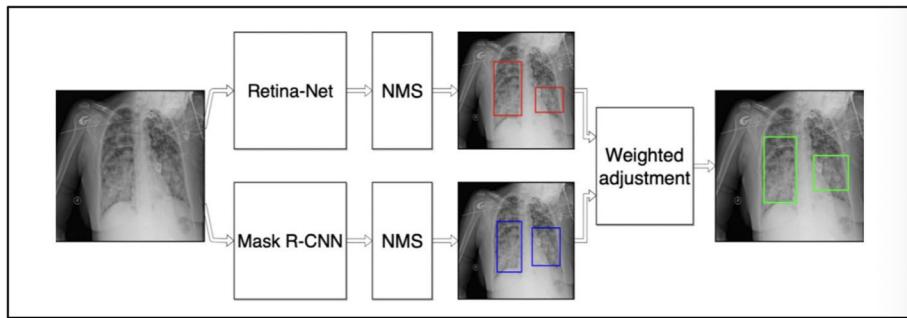


Fig. 20 Mask R-CNN + RetinaNet framework proposed in [60]

the convolutional layer. A total of 1024 filters, kernel size of 1×1 , and Nadam optimizer have been used to perform convolutions. Thereafter, the extracted features have been flattened and converted into a 1-dimensional array of 102,400 sizes. Lastly, a dropout layer is used to predict the output using a softmax activation function. The model performs binary as well as multiclass classification and categorizes the output as pneumonia, normal, and covid-19. The model exhibits an accuracy of 99.5% for binary classification of covid-19 and pneumonia, whereas the overall accuracy for multiclass classification has resulted as 91.4%. Figure 22 illustrates the proposed framework using ResNet50V2 + Xception model.

4.3.5 GoogleNet + ResNet + DenseNet

In [34], the authors have integrated three models GoogleNet, DenseNet, and ResNet, consisting of 22, 121, and 18 layers, respectively, to construct an ensemble framework for pneumonia prediction. The model is built using the RSNA and Kermany's dataset. The sizes of images in these datasets were different and varies from $127 \times 384 \times 3$ – $2713 \times 2517 \times 3$. The preprocessing was applied to reduce the input image size to $224 \times 224 \times 3$. The proposed framework utilizes the inception module from the GoogleNet model for performance enhancement, and the residual block from the ResNet model reduces the time complexity

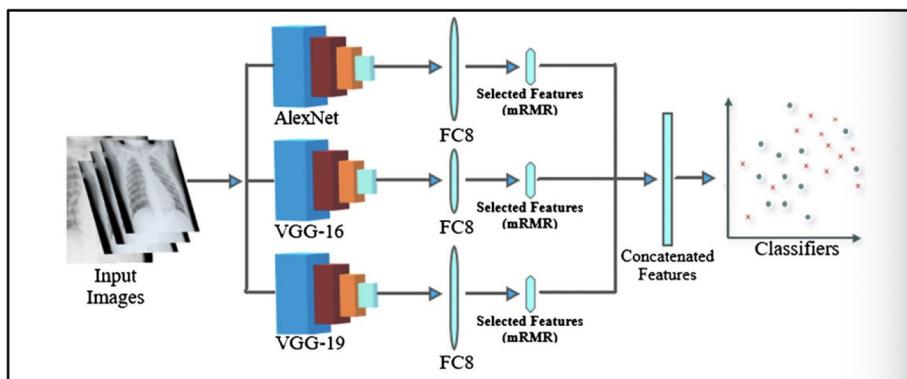


Fig. 21 VGG16 + AlexNet + VGG19 framework proposed in [63]

by skipping connections. Further, DenseNet extracts the most prominent features to enhance the efficiency of the ensemble architecture. The proposed model exhibits an accuracy of 98.81%, a sensitivity of 98.80%, and a precision of 98.82% for Kermany's dataset, whereas the respective scores for RSNA dataset were 86.85%, 87.02%, and 86.89%.

Figure 23 illustrates the inception module from GoogleNet architecture, Fig. 24 shows a residual block of the ResNet model, and Fig. 11 depicts the DenseNet model utilized in this ensemble architecture.

4.3.6 CNN + KNN + SVM

In [5], authors have proposed an AI-based framework for predicting and classifying pneumonia using different-sized CXR images. The model performs feature extraction using CNN, and it uses two classifiers for the classification of the disease, namely KNN and SVM. The dataset has 5,852 radiographs divided for training, testing, and validation of the proposed model. The proposed model exhibits an accuracy of 94% using the SVM classifier, whereas an accuracy of 93.9% has been achieved with the KNN classifier for multiclass classification. Figure 25 illustrates the CNN architecture used in the proposed ensemble model.

4.3.7 Monte Carlo + EfficientNet-B3

In [18], authors have suggested an ensemble method for classifying CXR by making use of the Monte Carlo dropout (MC-dropout) and EfficientNet-B3 model. The EfficientNet-B3 model is a CNN-based pre-trained model used to fine-tune the images. The model utilizes the MC technique to remove overfitting and to achieve an enhanced network generalization. Generative predictive distribution was calculated to obtain the model's uncertainty and mean prediction score. The MC-dropout has been implemented on EfficientNet-B3 to make it a reliable EfficientNet-B3 bayesian model. During the preprocessing step, the images were resized to 224*224*3 to feed the model, whereas, for normalization, the pixels were ranged between 0–1. Figure 26 shows ensemble model architecture using Monte Carlo + EfficientNet-B3.

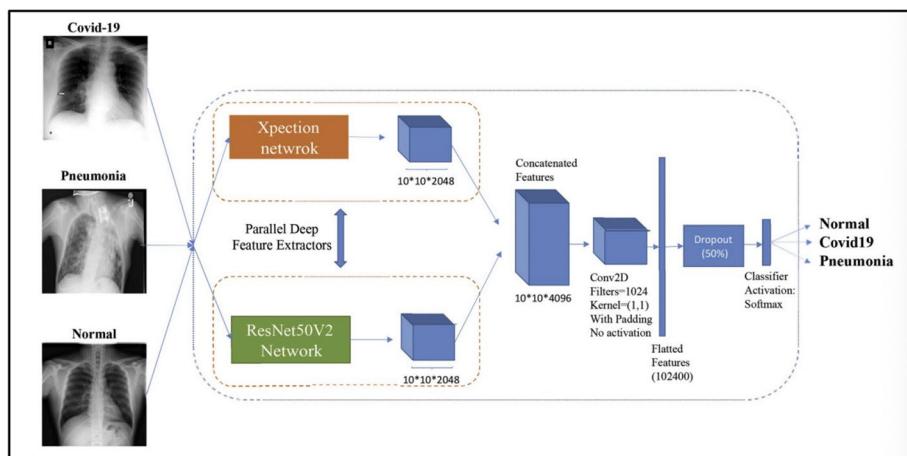


Fig. 22 ResNet50V2 + Xception framework for pneumonia detection proposed in [47]

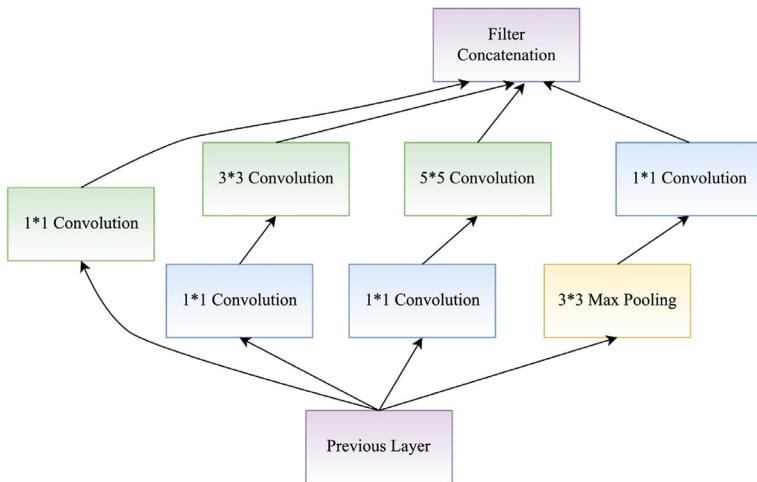
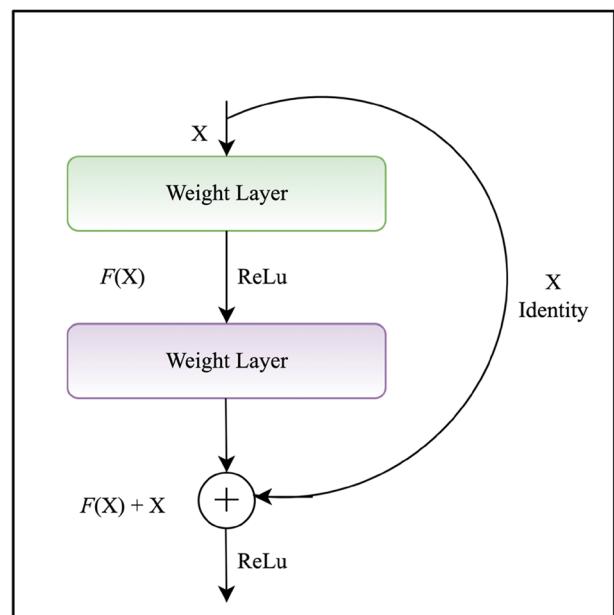


Fig. 23 Inception module to decrease the complexity of GoogleNet architecture [34]

Fig. 24 Residual block in ResNet model [34]



4.3.8 CGAN + DNN + VGG19

In [9], authors have presented an ensemble approach for classifying normal lungs, pneumonia, and covid-19 using CXR images. The first step includes a CGAN to segment the CXR data. In the second step, segmented lung radiographs were fed to a trained DNN for discriminatory feature extraction. VGG19 model has been used to perform multi-class classification. The model utilizes the Adam optimizer, categorical loss function, and

learning rate as 0.01 and 100 epochs. The proposed ensemble model exhibits an accuracy of 96.6% for multiclass classification. Figure 27 illustrates the architecture of the CGAN + DNN + VGG19 ensemble model.

4.3.9 InceptionV2 + RNN + ResNetV2 + RNN-LSTM

In [22], authors have constructed an ensemble of InceptionV2, ResNetV2, RNN, and RNN-LSTM to identify pneumonia infection. The integration of InceptionV2, ResNetV2, and RNN models was used to classify the input categories, whereas the RNN-LSTM model was used to predict the final outcome. The input images in the dataset had different sizes, so the image size was reduced to 310*310 to train and test the model. An effort is made based on realistic settings to provide convenient health indicators to monitor the infection rate and risk of covid-19 pneumonia. The model has been implemented using three different types of CXR: bacterial, viral, and normal. A dual method has also been developed for the characterization of the split of input images, as shown in Fig. 28.

4.3.10 Transfer learning + Deep DenseNet

In [58], authors proposed the DenseNet121 to detect pneumonia using the NIH chest-14 open-source dataset containing 4000 radiograph images. Initially, the model performs feature extraction, and thereafter, except for the last layer's weights, all other weights have been used for initializing the DenseNet121 model. Fine-tuning has been done on the Tan Tock Seng Hospital (TTSRH) dataset. The proposed model shows an AUC, sensitivity, specificity, and F1-score of 0.95, 79%, 97%, and 0.9120, respectively.

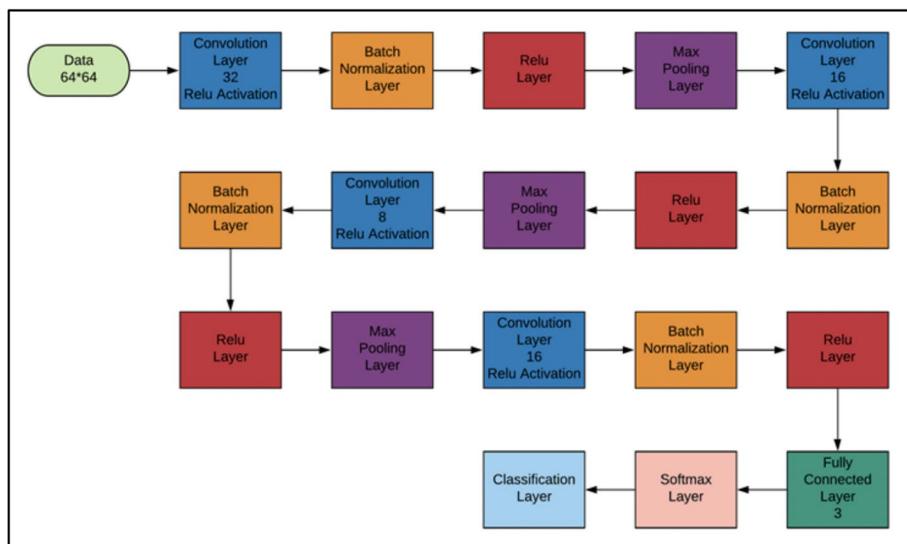


Fig. 25 CNN architecture used in ensemble model for pneumonia prediction presented in [5]

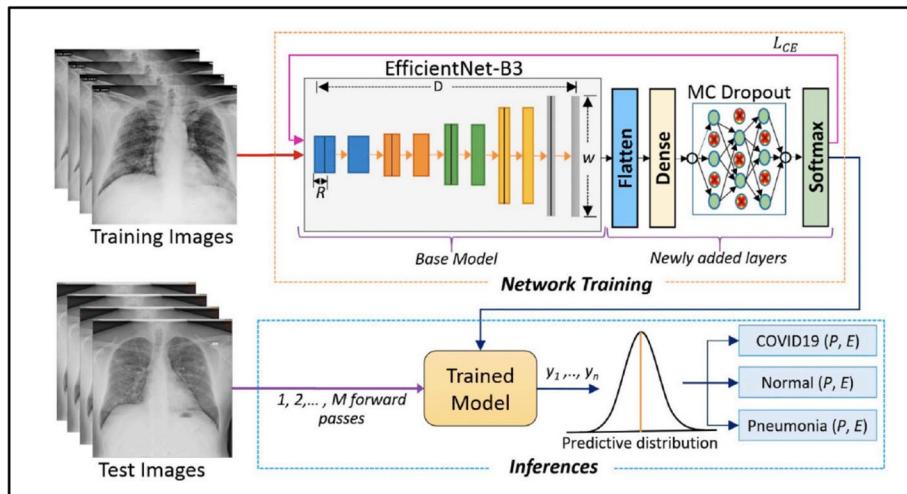


Fig. 26 Monte Carlo + EfficientNet-B3 ensemble model for pneumonia prediction proposed in [18]

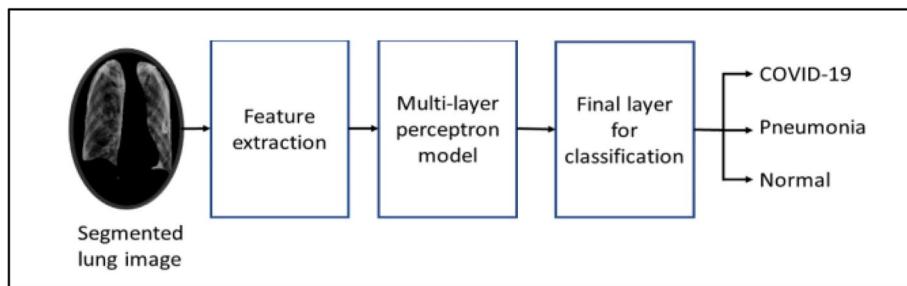


Fig. 27 CGAN + DNN + VGG19 ensemble model architecture for pneumonia detection proposed in [9]

4.3.11 Sparse Autoencoder + FFNN

In [17], authors have used the FFNN and Sparse Autoencoder to construct a CAD model for diagnosing covid-19 pneumonia using chest radiographs, as shown in Fig. 29. In the feature extraction, various pre-trained networks have been concatenated. The Sparse Autoencoder has been used to enhance the accuracy. This model contains three different phases; in the first phase, the CNN model is applied to extract features from the dataset. Sparse Autoencoder has been used in the second phase for dimensionality reduction from the images to avoid features that may decrease the performance outcome. Lastly, in phase three, the image categorization has been made using FFNN for binary classification in covid-19 and non-covid-19 pneumonia lungs.

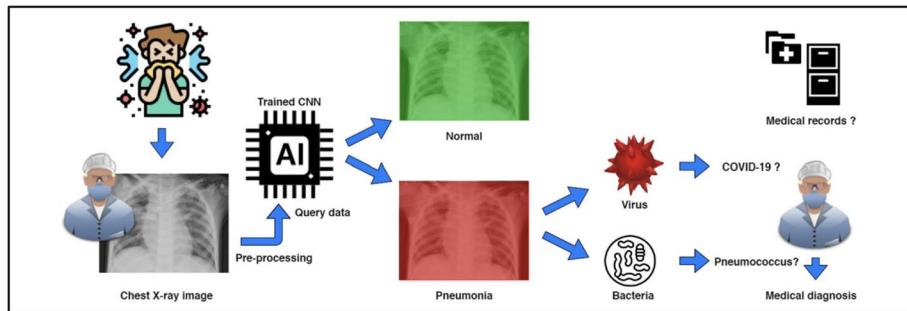


Fig. 28 An ensemble model using InceptionV2 + RNN + ResNetV2 + RNN-LSTM to classify pneumonia proposed in [22]

4.3.12 CovNNet-based ensemble model for pneumonia detection

In [28], authors have constructed a CovNNet model and ensemble model based on CovNNet framework for pneumonia classification. The CovNNet model is architected using ReLU activation function, three convolutions, pooling layers, and two FC and one softmax layer. It is capable of predicting the presence of pneumonia in a covid-19 CXR images. It is a fuzzy logic-based DL technique for distinguishing the CXR of patients with interstitial pneumonia and covid-19 pneumonia, as shown in Fig. 30. In the CovNNet model, the inputs of CXR and images developed by the fuzzy algorithm are combined and substituted for feature extraction while the classification task has been performed by MLP.

In the CovNNet-based ensemble model, the CXR and images developed by the fuzzy model have been separately substituted to two different CovNNet models for feature extraction, which are further combined and substituted to MLP for classifying the images. Further, the CovNNet model has achieved an accuracy of 67.2% for the first model, and the CovNNet-based ensemble model has achieved an accuracy of 81%. Fig. 31 illustrates the architecture of the CovNNet-based ensemble model for pneumonia prediction.

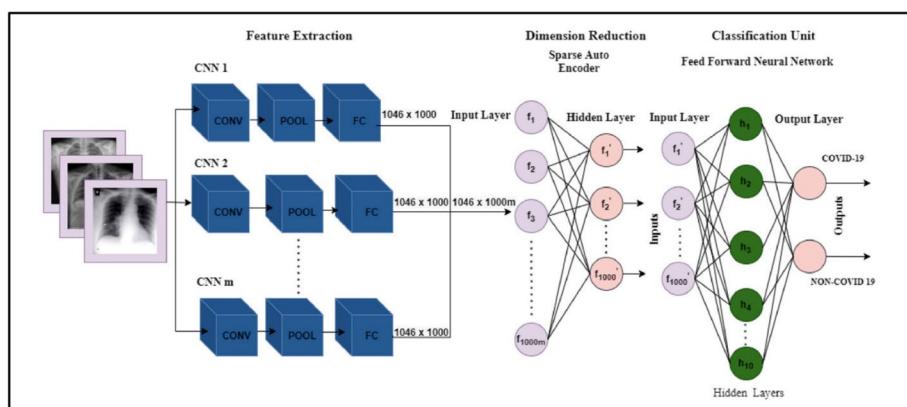


Fig. 29 An ensemble model using using Sparse Autoencoder+FFNN for the covid-19 and non covid-19 pneumonia prediction proposed in [17]

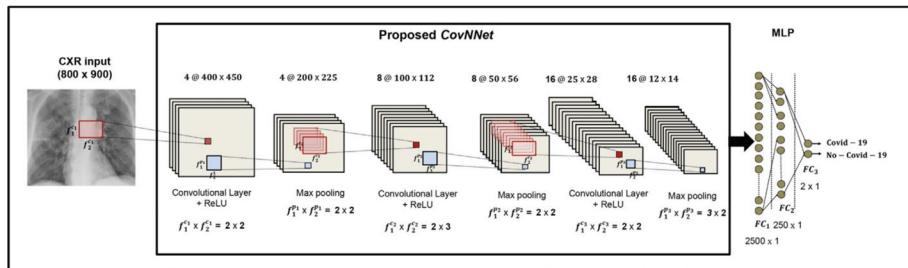


Fig. 30 CovNNet model for covid-19 and non covid-19 pneumonia prediction presented in [28]

In the recent past, research in healthcare has been increasing day by day due to the increased mortality rate caused by pandemics and acute diseases. In this SLR, researchers have contributed various models for pneumonia prediction, including CNN-based, pre-trained, and ensemble models. Therefore, the list of journals from various continents has also been identified to provide a broader perspective in this domain. Geo maps of various continents, Europe, Africa, North America, and the world, have been shown in Figs. 32, 33, 34, and 35, respectively.

5 Summary of literature review on pneumonia detection based on CNN, pre-trained and ensemble models

This section presents an analytical discussion and summarizes various pneumonia detection techniques. This work highlights the comparative summarization of research work done by various researchers in the domain of pneumonia prediction. The summarization includes pneumonia detection techniques, its objectives, dataset details, performance measures, and research gaps. In the performance comparison, few models have shown underperformance, and a few have been identified as superior models for pneumonia detection. The models have been classified into three categories: CNN-based, pre-trained, and ensemble

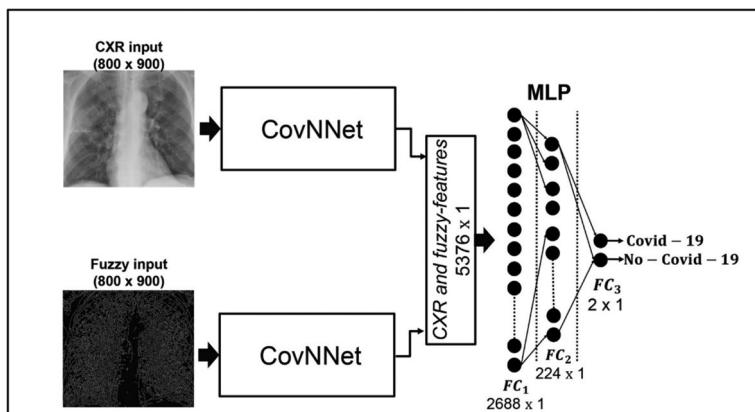


Fig. 31 Ensemble CovNNet model proposed in [28]

models. The performance analysis has shown that the ensemble models have shown better performance in contrast to CNN-based and pre-trained models. In this research, the accuracy measure has been used as the most significant parameter for identifying the correctness of any DL model.

Furthermore, other parameters such as AUC, precision, and recall also reflect the performance outcome of DL models. Specifically, precision focuses on the ratio of predicted true positive cases out of all the true positive and false positive pneumonia cases, whereas recall identifies the ratio of the true positive cases and the addition of true positives and false negatives. Further, the AUC measure has also been utilized as it provides the ability to distinguish between negative and positive classes.

The overview of all available literature reviews on pneumonia detection based on CNN, pre-trained, and ensemble models for the research are shown in Table 4.

In addition to Table 4, the existing research has identified several gaps, including the need to continuously improve the efficiency of DL models, which often require large datasets to achieve high-performance levels. Some researchers have used very limited datasets and have not applied sufficient data augmentation techniques, leading to the overfitting of the models. To enhance the quantity and diversity of data utilized for training the DL models, various kinds of data augmentation can be implemented, ultimately resulting in improved model performance and generalization. These techniques include various operations such as flipping, rotation, cropping, scaling, translation, and elastic deformation. The flipping operation entails creating a similar image of the original by flipping it horizontally/vertically. In contrast, rotation involves changing the orientation of an image by rotating it by a specific degree to produce a modified version of the actual image. Further, the cropping operation involves keeping the relevant part of the images to focus on the diseased area. Scaling is used to resize the actual images, and the translation technique moves the image horizontally/vertically by a certain distance

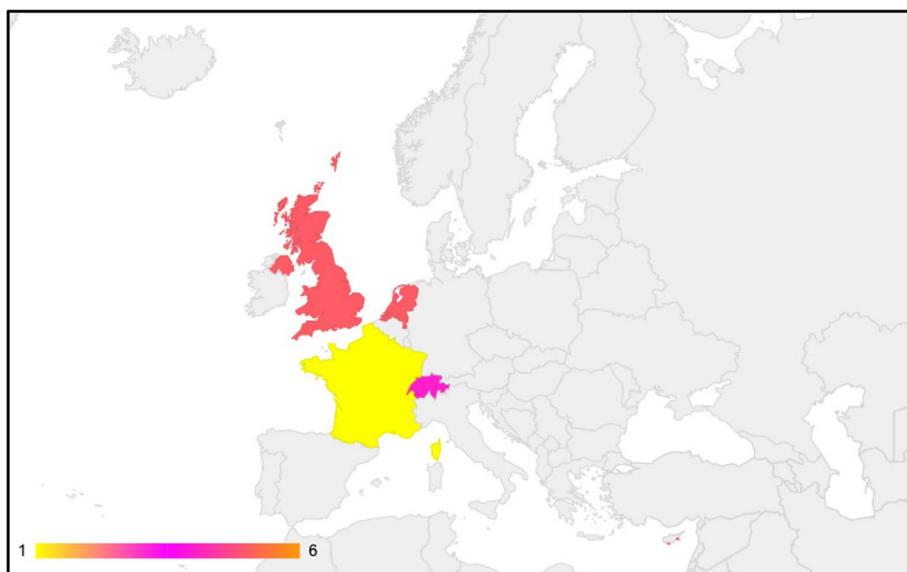


Fig. 32 Geo map highlighting the European countries publishing the research in the domain of pneumonia prediction

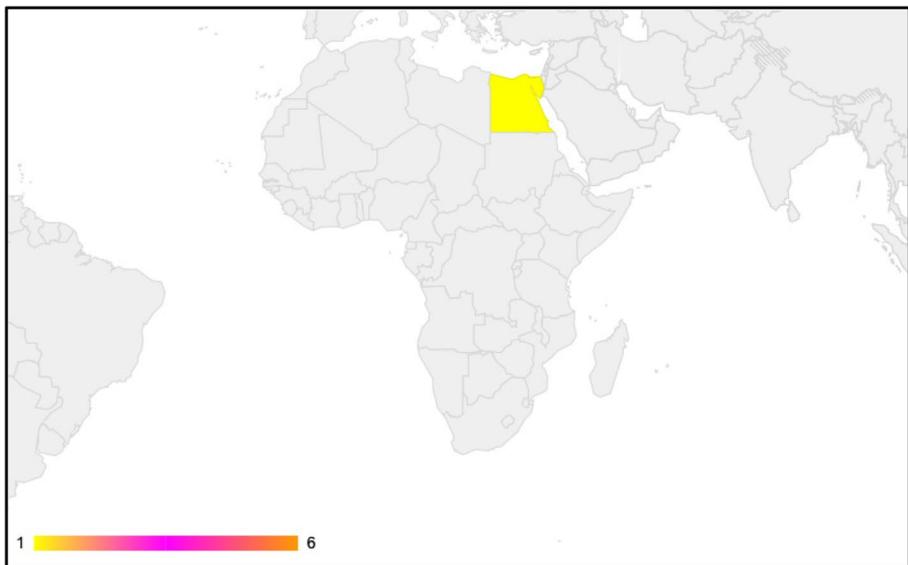


Fig. 33 Geo map highlighting the African countries publishing the research in the domain of pneumonia prediction

to create an updated image of the original image. These augmentation operations can be used for dataset enhancement which may lead to improved performance outcomes of pneumonia prediction models.

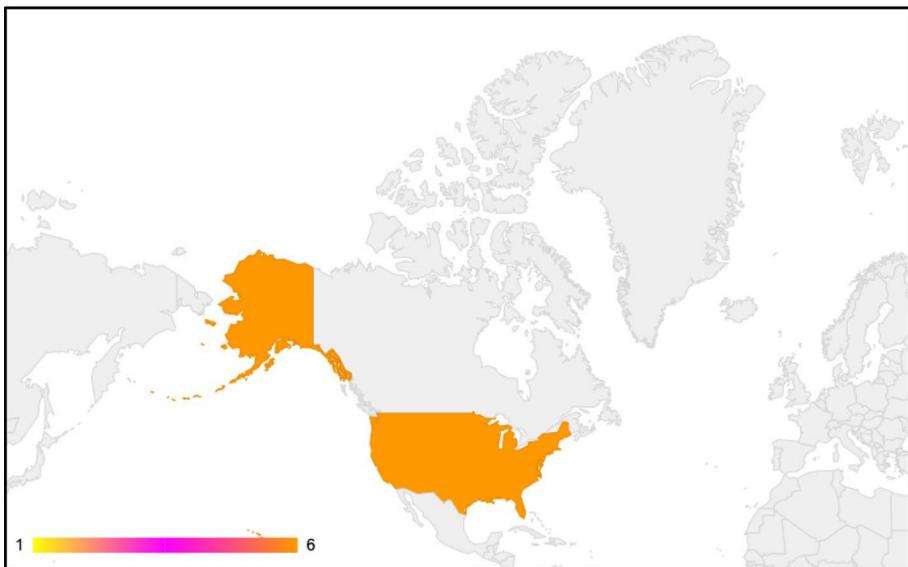


Fig. 34 Geo map highlighting the north American countries publishing the research in the domain of pneumonia prediction

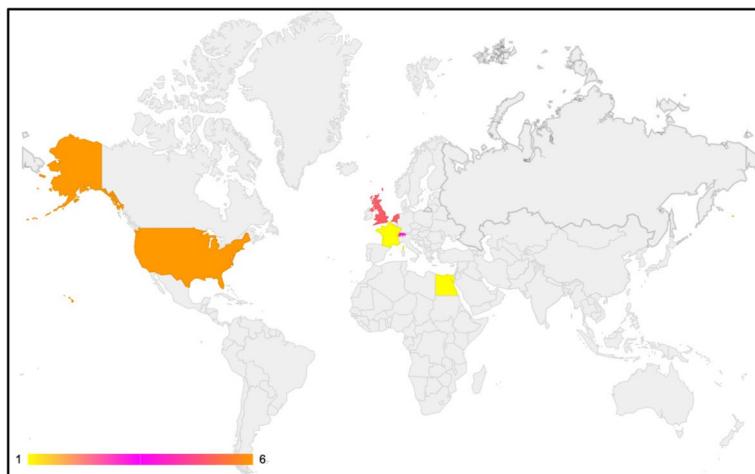


Fig. 35 Geo map highlighting various continents on the world map publishing the research in the domain of pneumonia prediction

Moreover, with the deep analysis of the proposed study, it has been identified that the current studies have explored few pre-trained models for pneumonia classification. In contrast, other models such as Xception, RetinaNet, GoogleNet, and MobileNet can also be explored more for enhanced results. These models offer a rich source of training data to help develop the DL models to recognize a wide range of images. Therefore, there is a possibility of improvement in pneumonia detection methods by incorporating these models. Additionally, while ensemble models have been used only to classify pneumonia cases, few studies have utilized pre-trained feature extraction models such as ResNet, DenseNet, MobileNet, and Inception to improve the feature extraction process. Combining these well-known feature extraction models into a single ensemble model can potentially improve feature extraction and classification results for pneumonia detection.

5.1 Summary and comparative analysis on CNN-based models

Various researchers have used CNN-based models for early pneumonia prediction by using the CXR dataset. The outcome of the research based on CNN models primarily identifies that the dataset utilized in the models was small, which directly impacts the performance outcomes of the DL models. A large amount of data is required in DL models to achieve better outcomes. Similarly, the CNN models are data-hungry, requiring hundreds of thousands of data samples. Therefore, to enhance the performance results, a huge dataset is required for training the model. Hence, the data can be augmented well to fulfill the requirements of CNN models. There are various operations, such as flipping, zooming, scaling, padding, and cropping, which can be applied to the images to enhance their quantity. The augmentation may help the DL model to train with different samples of data leading to improved results. Furthermore, it has also been noted that many authors who have worked on CNN-based models have used similar data for training; hence, it is required to enhance the dataset through augmentation, or dataset pooling may be applied.

Table 4 Summarization of literature review on pneumonia detection based on CNN, pre-trained and ensemble models

Ref	Technique	Dataset	Performance Estimates	Research gaps
[26]	FaNet model, Optimizer—Adam, Learning rate—0.001, Epochs—1000,	Covid- CT dataset collected from GitHub	Accuracy—98.28%	In the future, more CT scan datasets can be collected to increase the efficiency of the model
[59]	An extended version of traditional CNN is developed namely Quaternion CNN for dealing with RGB images, Epochs—20,30,40 and 60, Learning rate—0.001, Optimizer—Adam	CXR dataset containing 5,856 images collected from Kaggle	Accuracy at 20, 30,40 and 30 epochs—63.67%, 91.01%, 91.40%, 93.75% F1-Score at 20, 30, 40 and 60 epochs—0.5755, 0.9222, 0.9222, and 0.9640. Sensitivity at 20, 30, 40 and 60 epochs—0.6690, 0.9396, 0.9417 and 0.9560, Specificity at 20, 30, 40 and 60 epochs—0.9850, 0.7391, 0.7746 and 0.8311, AUC at 20, 30, 40 and 60 epochs—0.95, 0.94, 0.97 and 0.97	In this work, the ResNet model has been extended to Quaternion CNN which can further be extended for the octonion domain
[30]	Deep CNN	CXR dataset containing 5,863 images collected from a literature survey	Accuracy—84%	The dataset taken by the authors is very small for DL model, however the augmentation to enhance the size of the dataset can be explored further to magnify the performance outcome
[51]	DL-based CNN	CXR images named Kermany's dataset	Accuracy—95.3%	In this work, the authors have not done any fine-tuning of the CNN model and have implemented the same existing architecture presented in the state-of-the-art work. Further, the dataset can be increased for the better prediction outcome

Table 4 (continued)

Ref	Technique	Dataset	Performance Estimates	Research gaps
[48]	121-layer CheXNet pre-trained model Adam optimizer Batch size—16 Initial Learning rate—0.001	Chest X-ray14 Dataset containing 1,12,120 images of 30,805 patients	F1 Score—0.435	The authors only used only the frontal view of CXR images has been used, There is a requirement of working on side view CXR images to increase training and testing of the model
[67]	Pre-trained CNN model Stochastic gradient descent optimizer Learning rate—0.01	CXR images containing 1,58,323 images collected from the National Institutes of Health Clinical Center (1,12,120 images of 30,805 patients), Mount Sinai Hospital Newyork (42,396 images of 12,904 patients), and Indiana University Network for Patient Care (3,807 images of 3,683 patients)	NA ^a	The model has not performed better in the external data, hence the model is required to be trained on the data which has been collected from external sites
[64]	Pre-trained CNN model	CXR dataset consisting of 5,856 images of retrospective pediatric patients varying from 1–5 age	Training loss—0.1288, Validation loss—0.1835. Training accuracy—0.9531, and Validation accuracy—0.9373	As per the study, it has been found that it is difficult to distinguish between lung cancer and pneumonia, so, in the future, the study can be extended to identify the existence of pneumonia/ lung cancer
[24]	Pre-trained CNN model For fine-tuning, the model is developed by combining ResNet18, DenseNet21, InceptionV3, Xception and MobileNetv2 Epochs—25, Optimizer—Stochastic Gradient Descent(SGD), Learning rate—0.001	CXR dataset consisting of 5,836 images Guangzhou Women and Children's Medical Center pneumonia dataset	Accuracy—98.43%, F1 score—98.63, and AUC score—99.76	Shortage of dataset, due to which it is found quite complex and expensive to achieve accurate results

Table 4 (continued)

Ref	Technique	Dataset	Performance Estimates	Research gaps
[7]	Pre-trained CNN model, Optimizer—Adam, Batch size—32, Epochs—300, Learning rate—0.00001	Consists of X-ray and CT Scan images having 5658 images in JPEG format	Accuracy—>96%	In future, the dataset can be increased and more intellectual feature extraction methods can be applied for better results
[27]	Pre-trained AlexNet model, Optimizer—SGD, Learning rate—0.0001, Epochs—20	CXR dataset containing covid-19, viral pneumonia and normal chest images	Accuracy—94.43%, Sensitivity—98.19% and specificity—95.78%	In the future, more datasets can be used and trained on ResNet and GoogleNet to get better results The literature has also found that ensemble models can perform better than a single model to get better results
[23]	LabelBinarizer model and pre-trained VGG16	CXR images dataset containing 3500 images collected from Kaggle	Accuracy—91.69%, Sensitivity—95.92%, Specificity—100%	In the future the transfer learning models and hyperparameters can be adjusted to increase the prediction rate
[2]	Pre-trained CNN model Optimizer—Adam, Epochs—25, Learning rate—0.001, Batch size—16	CXR dataset containing 2,465 images collected from Valencian Region Medical Image Bank	Area Under the Curve (AUC)—0.98	Multipurpose DL models can be used for improving the quality of pneumonia detection
[49]	Pre-trained InceptionV3, Optimizer—SGD, Epochs—25	CXR images dataset	Accuracy—97%	There is a need to develop ensemble ensemble models for better prediction of pneumonia
[29]	Ensemble Mask R-CNN model of ResNet50+ResNet101, Epochs—20, Optimizer—SGD, Batch Size—16	NIH CXR14 dataset subset of RSNA pneumonia CXR dataset containing 54,368 images of stage 1 and stage 2 pneumonia including test dataset collected from GitHub	Loss—0.218051	Data augmentation techniques may be applied to increase the size of the dataset for getting better performance results

Table 4 (continued)

Ref	Technique	Dataset	Performance Estimates	Research gaps
[60]	Ensemble of Mask R-CNN and RetinaNet Adam optimizer, Batch Size—8 Learning rate—0.0001 and 0.001 for RetinaNet and Mask R-CNN respectively	Radiological Society of North America (RSNA) dataset consisting of 26,684CXR images	Precision—0.288, Recall—0.284, and F1-score—0.286	To acquire a more thorough view of the lungs and guarantee an accurate diagnosis of patients with problems, a frontal view of chest radiographs should be supplemented with lateral computed tomography and CXR images
[63]	Ensemble model consisting of VGG16, AlexNet, and VGG19 pre-trained models Optimizer—SGD, Batch Size—16, Learning rate—0.0001	CXR dataset of 5,849 images	Accuracy—99.41%, Sensitivity—99.61%, and Specificity—99.22%	As the used dataset is very small, in the future the dataset can be increased and the proposed method can be applied to identify the pneumonia prediction performance
[47]	Ensemble model having ResNet50V2 and Xception Optimizer—Nadam, Learning rate—0.0001, Batch size—30	Two CXR and scan datasets: 180 images from GitHub and 14,863 from Kaggle	Accuracy—91.4%, Sensitivity—80.53%	For the future more datasets can be collected to increase the performance of the proposed model
[34]	Ensemble model consisting of GoogLeNet, ResNet18, and DenseNet121 architecture Optimizer—Adam, Epochs—30	Two publicly available datasets consisting of CXR images namely RSNA and the second dataset named Germany's dataset collected from Kaggle	Accuracy and sensitivity for RSNA—86.85% and 87.02% Accuracy and sensitivity for Germany's dataset—98.81% and 98.80	There is a requirement of increasing the dataset by enhancing the contrast Along with it, pre-processing can be done to increase the image quality for better results
[5]	Ensemble model containing two steps Feature extraction- CNN, Classification- KNN and SVM,	CXR dataset containing 5,852 images collected from Kaggle	Accuracy—94% and 93.95 for, Sensitivity—93.33% and 93.19%, Specificity—96.68% and 96.60% SVM and KNN respectively	The proposed model can be enhanced to detect other X-ray images related to diseases

Table 4 (continued)

Ref	Technique	Dataset	Performance Estimates	Research gaps
[18]	Ensemble of Monte Carlo and Efficient-Net-B3 model, Optimizer—RMSprop(Root Mean Squared propagation)	Three covid-19 datasets namely X-ray images containing 127 X-ray images of covid-19 and 500 images of normal patients, COVID19xCxr(combining Figure-1 covid-19 CXR dataset initiative, covid-19 image data collection and Mendeley Data containing 3,040 images) and Kaggle-based dataset (containing 2,905 images)	Geometric mean—99.61%, Sensitivity—99.30%	The model can be updated by UA-ConvNet to increase the performance of covid-19 detection
[9]	An ensemble model containing CGAN, DNN, and VGG19 integrated with BRISK Optimizer—Adam, Learning rate—0.01, Epochs—100	Segmentation dataset—Publicly available CXR images dataset containing 247 images divided into 13 images for testing and the remaining 234 images for training the model	Accuracy of VGG19+ binary robust invariant scalable key points—96.6%	The dataset may be increased for training purposes to achieve better results, Transfer learning methods are required to be explored
[22]	A ensemble model consisting of InceptionV2, RNN, ResNetV2, and RNN-LSTM have been used	Classification Dataset—Publicly available dataset collected from [12] containing 940 X-ray images	Accuracy—95.72%	The dataset used for the implementation has been identified as very small. The augmentation can be utilized to increase the number of images. Further, the research can be extended by measuring the performance variations by setting the different learning rate values
[58]	Ensemble model containing transfer learning for feature extraction and Deep DenseNet for pneumonia CXR classification	Open source NIH chest-L4 open source dataset containing 9431 images	AUC—0.95, Specificity—97%	The model is required to be trained using an augmented dataset

Table 4 (continued)

Ref	Technique	Dataset	Performance Estimates	Research gaps
[17]	Ensemble model where dimensionality reduction using Sparse AutoEncoder and FFNN for covid-19 pneumonia detection has been done Epochs—100	Integrated CXR dataset containing 1046 images having 504 pneumonia positive and 542 pneumonia-negative patients	Accuracy—95.78% AUC—0.9821	For future purposes, other kinds of images like CT scans and ultrasound can be used rather than X-ray images
[28]	CovINet—An ensemble model used for feature extraction and classification of covid-19 and non covid-19 pneumonia Optimizer—Adam, Learning rate—0.001	CXR images were collected from the Advanced Diagnostic and Therapeutic Technology Department of the Frande Ospedale Metropolitano of Reggio Calabria, Italy	Accuracy for the first model—67.2% Accuracy for the second model—81%	The size of the dataset can be increased for improved performance results

^a Not Available

Another aspect that has also been noticed in all the CNN-based models is that the models have experimented with a single Adam optimizer, while the performance of the models can be explored by applying SGD, Nadam, Adamax, RMSprop, Adadelta, etc. The results with the Adam optimizer are pretty high, whereas SGD, Nadam, etc., may be explored to enhance the performance outcomes.

The batch size also plays an essential role while training the DL models, which shows the number of sample images substituted to train the model in a single iteration. With a large batch size value, the results can get impacted and produce a bad generalization or get stuck in local minima. In contrast, a smaller value of batch size can train the model well but has a high computational time and complexity. Therefore, the models must be experimented with the different values of batch size to avoid the challenges of local minima and high-time computation.

5.2 Summary and comparative analysis on pre-trained models

DL is used in various sectors, including healthcare, security, agriculture, and robotics. It includes the usage of image processing to solve the various challenges which can not be handled by humans in a short duration of time. The operation which can be performed in image processing is colorization, segmentation, and detection, along with classification. These image-processing operations require enough data and high intra-class similarity to provide better results.

In this work, the main focus is on pneumonia identification methods in healthcare. In real-time, this domain is very vigilant and requires to be handled with expertise due to high intra-class similarities in the CXR images. The researchers have implemented image processing methods for accurately classifying the disease into “pneumonia” and “normal” lungs. It is important to mention that the performance of the DL models degrades if enough data is not provided for training and may provide the transcend results due to overfitting and data leakage issues.

As per the study, various researchers have utilized pre-existing datasets and experimented with pre-trained models. These models have been rigorously trained with huge datasets. The usage of such models can provide better results if there are large datasets available for the problem domain. The pre-trained models are data-hungry models and can magnify the results if supplied with large and augmented data. The fine-tuning of the pre-trained models with various hyperparameters may also lead to better results.

The study provided by [30] is limited to only the frontal view of chest images, due to which the model may underperform if it has been tested with the top/down/left/right chest images. Since this research has yet to include images other than the frontal chest, the model can be added up and implemented with other types of images to improve the outcome tested with the different datasets.

The learning rate plays an essential role in predicting an accurate outcome. Another shortcoming in the existing pre-trained models has been identified in which many authors [23, 24, 27] have used the constant learning rate value of 0.001 and 0.0001 for training the pneumonia detection models. Hence, there is a need for more research that can be empirically experimented and explored with the variants of learning rate values. This identifies if the models can be experimented with by setting the variants of the learning rate, the performance outcomes can be compared, and may fall under the outperforming values for pneumonia detection challenges.

The current study relies on working with a limited set of pre-trained models such as InceptionV3, DenseNet, ResNet, VGG19, CheXNet, AlexNet, and VGG16. This necessitates performing more implementation contributions to the pre-trained models for better training and performance results.

5.3 Summary and comparative analysis on ensemble models

Ensemble models are techniques developed by integrating a wide range of DL models or datasets. The intermediate results of each model involved in this aggregation are merged to form a single prediction outcome.

This article reviews various works on ensemble models provided by different authors for pneumonia detection. The literature review has attempted to study the frameworks for ensemble models such as R-CNN, AlexNet, Xception, etc. As per the detailed review, it has been identified that there are still various limitations that are required to be addressed in this domain.

The first shortcoming is that various researchers have used limited datasets for ensemble model construction. It is important to mention that the ensemble model is a complex model which requires a large dataset for training and feature extraction of various integrated DL models. In the reviewed literature, ensemble models are required to be trained with huge datasets. Hence, to fulfill the requirement of large datasets, augmentation/pooling techniques need to be explored further. Although many authors have applied augmentation but a limited number of techniques have been explored. The existing studies can be extended by considering more augmentation and pooling techniques for improving the performance results.

The variance, noise, and bias are the major factors that have been used as the primary measures of error identification. Another finding is that the ensemble models are more prone to errors as compared to the CNN-based and pre-trained models only if the dataset is not large enough or the hyperparameters have not been set properly for the training. Hence, for developing effective models, the researchers can perform primary research to enhance the datasets and hyperparameter tuning, which can provide a way forward for better research outcomes in this domain with a lower percentage of errors and higher accuracy outcomes.

In addition to the aforementioned limitations, the ensemble models proposed in [9, 18, 34, 63] have been identified as the best-performing models for pneumonia classification. For instance, an ensemble model consisting of VGG16, AlexNet, and VGG19 has achieved an accuracy of 99.41% [63]. Similarly, in [34], the ensemble models have achieved an accuracy of 98.81% for identifying pneumonia in CXR images using Kermany's dataset. In [18], a fusion of Monte Carlo and EfficientNet-B3 models has resulted in an accuracy of 99.61% for accurate pneumonia prediction. Lastly, in [9], the ensemble model achieved a high-performance accuracy of 96.6% by integrating CGAN, DNN, and VGG19 models.

Table 5 tabulates the extracted information from various articles describing the names of datasets used, data augmentation, pre-processing information, image size, train/test split, and data dimensionality.

Conclusively, the summary and comparative analysis of ensemble pneumonia detection models sums up that the various techniques of data augmentation can be used to improve the outcome. Furthermore, the pre-trained models can be updated by fine-tuning the layered architectures of the models, and the performance can be accessed. Additionally, for real-time and actual data establishment, primary research can be performed. Lastly, the models can be

Table 5 Comparative evaluation of pneumonia detection based on CNN, pre-trained and ensemble models

Ref	Model	Dataset Name (Platform)	Pre-processing	Number of Images/Format	Data Augmentation	Access Publicly available/Not	Input Size	Train/ Test ratio	Type of Data (X-ray/CT Scan)
[26]	FaNet	Covid- CT dataset	No	416 images	Yes, Vertical/ horizontal flip and rotation by 90, 180 and 270 degrees	Open Access from GitHub	512*512	72:28	CT scan and CXR images
[59]	Quaternion CNN	CXR dataset	Yes	5856 images	No	Open Access collected from Kaggle	50*50	90:10	X-ray images
[48]	CheXNet pre-trained model	Chest X-ray14	No	1,12,120	Yes using Random Horizontal flipping	Open Access	224*224	Training: testing ratio—70:20, Validation set—10%	X-ray images
[67]	CNN	NIH chest X-ray14	Yes	1,58,323	No	Open Access	224*224	Training: testing ratio—70:20, Tune set—10%	X-ray images (png)
[64]	Pre-trained CNN model	CXR dataset	Yes	5,856	Yes, height/width shift, shear/zoom range change, rescaling, rotation(40 degrees), and horizontal shift	NM ^a	200*200	3722 was assigned to training and 2134 was assigned to testing the model	X-ray images
[24]	Pre-trained CNN model	CXR dataset	Yes	5,836	Yes, Flip, cropping, shift, horizontal, padding, rotation and shear	NM ^a	224*224 for ResNet18, DenseNet and MobileNetV2, 229*229 for InceptionV3 and Xception	70:30	X-ray images
[27]	Pre-trained AlexNet	CXR dataset	No	12,124 images	Yes	Open Access	227*227 for AlexNet model	70:30	CXR and CT Scan Images
[23]	LabelBinarizer and pre-trained VGG 16 model	Kaggle dataset	Yes	3500 images	Yes	Open Access	NM ^a	80:20	CXR images

Table 5 (continued)

Ref	Model	Dataset Name (Platform)	Pre-processing	Number of Images/ Format	Data Augmentation	Access Publicly available/Not	Input Size	Train/ Test ratio	Type of Data (X-ray/CT Scan)
[2]	Pre-trained CNN model	(BIMCV covid-19)	No	2465 images	No	Not open access but can be asked from Valencian Region Medical Image Bank	NM ^a	70:30	X-ray images
[49]	Pre-trained CNN model	X-ray images	No	NM ^a	No	NM ^a	NM ^a	X-ray images(jpeg)	X-ray images(png)
[29]	ResNet50 + ResNet101	NIH CXR14 dataset	Yes	54,368 images	Yes, rotation, flip, crop, scale, translate and noise insertion	Open Access	512*512	95:5	X-ray images(DICOM)
[60]	Ensemble of Mask R-CNN and RetinaNet	RSNA dataset	No	26,684	Yes, Horizontal/ vertical flip, rotation on random degree, Brightness change, Noise insertion, gamma transformation, and blurring	Open Access through Kaggle	512*512	Training—90;10 Validation—90:10	X-ray images(DICOM)
[63]	Ensemble of VGG16, AlexNet, and VGG19	CXR dataset	Yes	5,849	Yes, Cutting, changing height, zooming, brightness increase, rotation, filling operations and horizontal flip Before augmentation images—1583, after augmentation—4266	Publicly available dataset	AlexNet—227*227, VGG16 and VGG19—224*224	70:30	X-ray images(jpeg)

Table 5 (continued)

Ref	Model	Dataset Name (Platform)	Pre-processing	Number of Images/ Format	Data Augmentation	Access Publicly available/Not	Input Size	Train/ Test ratio	Type of Data (X-ray/CT Scan)
[47]	Ensemble of ResNet50V2 and Xception	CXR and scan dataset	Yes	180 images from GitHub and 14,863 from Kaggle images	Yes, Zoom, shift, rotation, width/height change, and re-scaling	Open source dataset	NM ^a	-	X-ray images and CT Scan
[34]	Ensemble of GoogLeNet+ResNet18+DenseNet121	Kermany dataset and RSNA dataset	Yes	5856 and 26,601 images	Yes	Publicly available on Kaggle	224*224 after resizing	NM ^a	X-ray images (jpeg in Kermany's and DICOM in RSNA dataset)
[5]	Ensemble of CNN, KNN and SVM models	Kermany's dataset	No	5852 images	No	Open Access	32*32, 64*64, 128*128, 256*256	70:30	X-ray images(jpeg)
[18]	Ensemble of Monte Carlo and EfficientNet-B3 model	covid-19CXR dataset, Kaggle Dataset and x-ray dataset	Yes	X-ray data-set—627 images, Kaggle data-set—2905 images, COVID 19CXR—3040 images	Yes, rotation, zoom, flip, shear and shifting	Kaggle database—Open Access on Kaggle, COVIDCXR dataset	224*224	70:20:10:20—Training/test: validation	X-ray images
[9]	Ensemble of CGAN, DNN, VGG19 integrated with BRISK	Combination of two datasets i.e. Segmentation in chest X-ray (SCR)	Yes	247 images	Yes, rotation and shearing	Open Access collected from [12]	256*256	90:10	X-ray images
[22]	Tailored model consisting InceptionV2, ResNetV2, RNN and RNN-LSTM	Kermany's dataset	No	5863 images	No	Open Access	310*310	90:10	CXR images(jpeg)

Table 5 (continued)

Ref	Model	Dataset Name (Patient form)	Pre-processing	Number of Images/Format	Data Augmentation	Access Publicly available/Not	Input Size	Train/Test ratio	Type of Data (X-ray/CT Scan)
[58]	Ensemble of transfer learning and deep DenseNet	NIH chest-14 dataset	Yes	9431 images	Yes	Open Access	224*224	NM ^a	CXR images (png)
[17]	Ensemble of Sparse AutoEncoder + FFNN	Combination of two datasets	No	1046 images: 504 from first and 542 from second dataset	No	First dataset is open access collected from [12] and Second dataset is also open access can be collected from Kaggle	ResNet101, EfficientNetB0—224*224, Xception, Inception-ResNetV2—299*299, DarkNet53—256*256	90:10	X-ray and CT Scan images
[28]	CovNNet	CXR images	Yes	121 images	No	Not publicly available but can be got from Advanced Diagnostic and Therapeutic Technology Department of the Grande Ospedale Metropolitano (GOM) of Reggio Calabria, Italy	800*900	70:30	CXR images
[68]	InfNet model	X-ray covid-19 Dataset	Yes	129 images	No	Publicly available	NM ^a	NM ^a	X-ray
[7]	Pre-trained CNN	KernGermany dataset	Yes	5658 images	Yes, Rescale, Rotation, vertical and horizontal shift, shear, zoom, fill and horizontal flip	Open Access	224*224 for VGG16, VGG19, ResNet 50, DenseNet201, and MobileNetV2 and 299*299 for Inception_V3, Xception	60:40	X-ray and CT scan images (jpeg format)

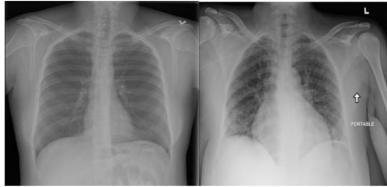
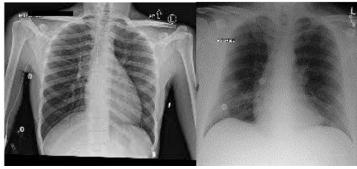
^aNot Mentioned

Table 6 Summary and details of open-source pneumonia datasets

Sr. No	Dataset name	Ref	Summary
1	Kermany's Dataset	[42]	The dataset contains pneumonia and normal CXRs with a total number of 5,856 images in three directories namely, train, test and validation containing 4216, 624 and 16 images, respectively
2	RSNA Dataset	[50]	This dataset is in the form of CSV file containing images of pneumonia and normal lungs. Each of the image has an extension of DICOM
3	ChestX-ray8 Dataset	[43]	The chestX-ray8 is in the form of CSV format. The complete dataset contains three.CSV files for training, testing and validating the model
4	NIH CXR Dataset	[1]	The NIH dataset contains a number of 1,12,120 X-ray images of 30,805 patients. There are a number of 14 diseases in this dataset and the 15 th class contains the X-ray images which have “No Disease”
5	CXR Dataset	[12]	This dataset contains the images of covid-19 CXR. These images were in the.png and.jpeg format
6	NIH chest X-ray14 Dataset	[10]	NM ^a

^aNot Mentioned

Table 7 Sample images in the datasets

Sr. No.	Dataset	Image Samples
1.	Kermany's Dataset [42]	
2.	RSNA Dataset [50]	
3.	ChestX-ray8 Dataset [43]	
4.	NIH chest X-ray Dataset [1]	
5.	CXR Dataset [12]	
6.	NIH chest X-ray14 Dataset[10]	

trained with varied values of learning rates, epochs, and different optimizers, and their performance can be compared to identify the best suitable pneumonia detection models.

Table 6 shows the available open-source pneumonia dataset names along with the summary and image details in each dataset.

Table 7 tabulates the sample images of each dataset mentioned in Table 6.

6 Conclusion

In the recent past, a lot of research has been done in the field of healthcare. Although various researchers have proposed several DL models for the early prediction of pneumonia, there still exists potential for improvement in this domain. This SLR presents a taxonomy for pneumonia detection models based on DL techniques, including CNN-based, pre-trained, and ensemble models. In addition to categorization, the review also provides a deep dive into the architectures and processes involved in developing pneumonia prediction models. It further includes detailed hyperparameter information, including optimizers, learning rates, epochs, batch sizes, and training, testing ratios, offering a deeper understanding of the model's configuration and training procedure. The review also incorporates the research gaps in CNN-based, pre-trained, and ensemble DL models. Furthermore, the work discusses the potential solutions for the challenges faced by researchers while developing these models. This SLR summarizes the existing work by encompassing various pneumonia detection techniques, frameworks, dataset details, and performance measures. This review has identified that improvement in the performance outcomes of the DL models can be made by enhancing the dataset. However, healthcare data collection is challenging due to securing patient privacy. With the sensitive nature of healthcare data, which includes a medical history and personally identifiable information; hence, healthcare departments follow strict privacy protocols to protect patient confidentiality and comply with regulations like HIPAA. In the future, to ensure data privacy, federated learning-based DL models can be explored to identify pneumonia in CXR images.

Author's contributions Conceptualization: Shagun Sharma, Kalpana Guleria; Methodology: Shagun Sharma, Kalpana Guleria; Literature Search: Shagun Sharma; Data Analysis: Shagun Sharma, Kalpana Guleria; Manuscript Writing: Shagun Sharma and Kalpana Guleria; Review and editing: Kalpana Guleria;

Data availability Data sharing not applicable to this article as no datasets were generated or used to carry out experiments since it is a systematic literature review.

Declarations

Informed consent All authors have read and agreed to this version of the manuscript.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

1. (2018) NIH Chest X-ray multi-classification. <https://www.kaggle.com/code/adamjgoren/nih-chest-x-ray-multi-classification/data>. Accessed 15 Jan 2023
2. Ahmad J, Saudagar AKJ, Malik KM et al (2022) Disease progression detection via deep sequence learning of successive radiographic scans. Int J Environ Res Public Health 19. <https://doi.org/10.3390/ijerph19010480>
3. Alhasan M, Hasaneen M (2021) Digital imaging, technologies and artificial intelligence applications during COVID-19 pandemic. Comput Med Imaging Graph 91:101933. <https://doi.org/10.1016/j.compmedimag.2021.101933>
4. Allegra A, Tonacci A, Sciacchitta R et al (2022) Machine learning and deep learning applications in multiple myeloma diagnosis, prognosis, and treatment selection. Cancers 14. <https://doi.org/10.3390/cancers14030606>

5. Alqudah AM, Qazan S, Masad IS (2021) Artificial Intelligence Framework for Efficient Detection and Classification of Pneumonia Using Chest Radiography Images. *J Med Biol Eng* 41:599–609. <https://doi.org/10.1007/s40846-021-00631-1>
6. Alyasseri ZAA, Al-Betar MA, Doush IA et al (2022) Review on COVID-19 diagnosis models based on machine learning and deep learning approaches. *Expert Syst* 39:e12759. <https://doi.org/10.1111/exsy.12759>
7. Asnaoui El K, Chawki Y, Idri A (2021) Automated Methods for Detection and Classification Pneumonia Based on X-Ray Images Using Deep Learning. In: Maleh Y, Baddi Y, Alazab M et al (eds) Artificial Intelligence and Blockchain for Future Cybersecurity Applications. Springer International Publishing, Cham, pp 257–284
8. Basheera S, SatyaSai Ram M (2020) A novel CNN based Alzheimer's disease classification using hybrid enhanced ICA segmented gray matter of MRI. *Comput Med Imaging Graph* 81:101713
9. Bhattacharyya A, Bhaik D, Kumar S et al (2022) A deep learning based approach for automatic detection of COVID-19 cases using chest X-ray images. *Biomed Signal Process Control* 71:103182. <https://doi.org/10.1016/j.bspc.2021.103182>
10. Box. <https://nihcc.app.box.com/v/ChestXray-NIHCC/folder/37178474737>. Accessed 15 Jan 2023
11. Chen Z, Lalande A, Salomon M et al (2022) Automatic deep learning-based myocardial infarction segmentation from delayed enhancement MRI. *Comput Med Imaging Graph* 95:102014. <https://doi.org/10.1016/j.compmedimag.2021.102014>
12. Cohen JP images at master ieee8023/covid-chestxray-dataset. Github. Accessed 15 Jan 2023
13. Diehr P, Wood RW, Bushyhead J et al (1984) Prediction of pneumonia in outpatients with acute cough—A statistical approach. *J Chronic Dis* 37:215–225. [https://doi.org/10.1016/0021-9681\(84\)90149-8](https://doi.org/10.1016/0021-9681(84)90149-8)
14. Ding K, Zhou M, Wang Z et al (2022) Graph Convolutional Networks for Multi-modality Medical Imaging: Methods, Architectures, and Clinical Applications. *arXiv [eess.IV]*
15. España PP, Capelastegui A, Gorordo I et al (2006) Development and validation of a clinical prediction rule for severe community-acquired pneumonia. *Am J Respir Crit Care Med* 174:1249–1256. <https://doi.org/10.1164/rccm.200602-177OC>
16. Frondelius T, Atkova I, Miettunen J et al (2022) Diagnostic and prognostic prediction models in ventilator-associated pneumonia: Systematic review and meta-analysis of prediction modelling studies. *J Crit Care* 67:44–56. <https://doi.org/10.1016/j.jcrc.2021.10.001>
17. Gayathri JL, Abraham B, Sujarani MS, Nair MS (2022) A computer-aided diagnosis system for the classification of COVID-19 and non-COVID-19 pneumonia on chest X-ray images by integrating CNN with sparse autoencoder and feed forward neural network. *Comput Biol Med* 141:105134. <https://doi.org/10.1016/j.combiomed.2021.105134>
18. Gour M, Jain S (2021) Uncertainty-aware convolutional neural network for COVID-19 X-ray images classification. *Comput Biol Med* 140:105047. <https://doi.org/10.1016/j.combiomed.2021.105047>
19. Guleria K, Sharma S, Kumar S, Tiwari S (2022) Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning. *Meas Sensors* 24:100482. <https://doi.org/10.1016/j.measen.2022.100482>
20. Gunasekeran DV, Tseng RMWW, Tham Y-C, Wong TY (2021) Applications of digital health for public health responses to COVID-19: a systematic scoping review of artificial intelligence, telehealth and related technologies. *NPJ Digit Med* 4:40. <https://doi.org/10.1038/s41746-021-00412-9>
21. Haghifar A, Majdabadi MM, Choi Y et al (2022) COVID-CXNet: Detecting COVID-19 in frontal chest X-ray images using deep learning. *Multimed Tools Appl* 81:30615–30645. <https://doi.org/10.1007/s11042-022-12156-z>
22. Hammoudi K, Benhabiles H, Melkemi M et al (2021) Deep Learning on Chest X-ray Images to Detect and Evaluate Pneumonia Cases at the Era of COVID-19. *J Med Syst* 45:75. <https://doi.org/10.1007/s10916-021-01745-4>
23. Hasan MDK, Ahmed S, Abdullah ZME et al (2021) Deep learning approaches for detecting pneumonia in COVID-19 patients by analyzing chest X-Ray images. *Math Probl Eng* 2021. <https://doi.org/10.1155/2021/9929274>
24. Hashmi MF, Katiyar S, Keskar AG et al (2020) Efficient Pneumonia Detection in Chest Xray Images Using Deep Transfer Learning. *Diagnostics* 10:417
25. Hasoon JN, Fadel AH, Hameed RS et al (2021) COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray images. *Results Phys* 31:105045. <https://doi.org/10.1016/j.rinp.2021.105045>
26. Huang Z, Liu X, Wang R et al (2021) FaNet: fast assessment network for the novel coronavirus (COVID-19) pneumonia based on 3D CT imaging and clinical symptoms. *Appl Intell (Dordr)* 51:2838–2849. <https://doi.org/10.1007/s10489-020-01965-0>
27. Ibrahim AU, Ozsoz M, Serte S et al (2021) Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19. *Cognit Comput* 1–13. <https://doi.org/10.1007/s12559-020-09787-5>

28. Ieracitano C, Mammone N, Versaci M et al (2022) A fuzzy-enhanced deep learning approach for early detection of Covid-19 pneumonia from portable chest X-ray images. *Neurocomputing* 481:202–215. <https://doi.org/10.1016/j.neucom.2022.01.055>
29. Jaiswal AK, Tiwari P, Kumar S et al (2019) Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement* 145:511–518. <https://doi.org/10.1016/j.measurement.2019.05.076>
30. Jakhar K, Hooda N (2018) Big Data Deep Learning Framework using Keras: A Case Study of Pneumonia Prediction. In: 2018 4th International Conference on Computing Communication and Automation (ICCCA). pp 1–5
31. Juneja M, Singh S, Agarwal N et al (2020) Automated detection of Glaucoma using deep learning convolution network (G-net). *Multimed Tools Appl* 79:15531–15553. <https://doi.org/10.1007/s11042-019-7460-4>
32. Kareem A, Liu H, Sant P (2022) Review on Pneumonia Image Detection: A Machine Learning Approach. *Human-Centric Intelligent Systems* 2:31–43. <https://doi.org/10.1007/s44230-022-00002-2>
33. Kumar N, Gupta M, Gupta D, Tiwari S (2023) Novel deep transfer learning model for COVID-19 patient detection using X-ray chest images. *J Ambient Intell Humaniz Comput* 14:469–478. <https://doi.org/10.1007/s12652-021-03306-6>
34. Kundu R, Das R, Geem ZW et al (2021) Pneumonia detection in chest X-ray images using an ensemble of deep learning models. *PLoS One* 16:e0256630. <https://doi.org/10.1371/journal.pone.0256630>
35. Li Y, Zhang Z, Dai C et al (2020) Accuracy of deep learning for automated detection of pneumonia using chest X-Ray images: A systematic review and meta-analysis. *Comput Biol Med* 123:103898. <https://doi.org/10.1016/j.combiomed.2020.103898>
36. Mahmoudi N, Ahadi SM, Rahmati M (2019) Multi-target tracking using CNN-based features: CNN-MTT. *Multimed Tools Appl* 78:7077–7096. <https://doi.org/10.1007/s11042-018-6467-6>
37. Majeed A, Beg MO, Arshad U, Mujtaba H (2022) Deep-EmoRU: mining emotions from roman urdu text using deep learning ensemble. *Multimed Tools Appl* 81:43163–43188. <https://doi.org/10.1007/s11042-022-13147-w>
38. Mehmood A, Khan MA, Sharif M et al (2020) Prosperous Human Gait Recognition: an end-to-end system based on pre-trained CNN features selection. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-020-08928-0>
39. Mijwil MM (2021) Implementation of machine learning techniques for the classification of lung X-ray images used to detect COVID-19 in humans. *Iraqi J Sci* 2099–2109. <https://doi.org/10.24996/ij.s.2021.62.6.35>
40. Mijwil MM, Abtan RA, Alkhazraji A (2022) Artificial intelligence for COVID-19: A short article. *Artificial Intelligence* 10:1–6. <https://doi.org/10.24203/ajpnms.v10i1.6961>
41. Mohammed MA, Al-Khateeb B, Yousif M et al (2022) Novel Crow Swarm Optimization Algorithm and Selection Approach for Optimal Deep Learning COVID-19 Diagnostic Model. *Comput Intell Neurosci* 2022:1307944. <https://doi.org/10.1155/2022/1307944>
42. Mooney P (2018) Chest X-Ray Images (Pneumonia). <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>. Accessed 15 Jan 2023
43. Mrad H (2021) ChestX-ray8 dataset. <https://www.kaggle.com/datasets/habibmrad1983/chestxray8-dataset>. Accessed 15 Jan 2023
44. Nagi AT, Awan MJ, Mohammed MA et al (2022) Performance Analysis for COVID-19 Diagnosis Using Custom and State-of-the-Art Deep Learning Models. *NATO Adv Sci Inst Ser E Appl Sci* 12:6364. <https://doi.org/10.3390/app12136364>
45. Padash S, Mohebbian MR, Adams SJ et al (2022) Pediatric chest radiograph interpretation: how far has artificial intelligence come? A systematic literature review. *Pediatr Radiol* 52:1568–1580. <https://doi.org/10.1007/s00247-022-05368-w>
46. Rahim T, Usman MA, Shin SY (2020) A survey on contemporary computer-aided tumor, polyp, and ulcer detection methods in wireless capsule endoscopy imaging. *Comput Med Imaging Graph* 85:101767. <https://doi.org/10.1016/j.compmedimag.2020.101767>
47. Rahimzadeh M, Attar A (2020) A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. *Inform Med Unlocked* 19:100360. <https://doi.org/10.1016/j.imu.2020.100360>
48. Rajpurkar P, Irvin J, Zhu K et al (2017) CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv [cs.CV]*
49. Rehman MU, Shafique A, Khan KH et al (2022) Novel privacy preserving non-invasive sensing-based diagnoses of pneumonia disease leveraging deep network model. *Sensors* 22. <https://doi.org/10.3390/s22020461>
50. RSNA pneumonia detection challenge. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>. Accessed 15 Jan 2023

51. Saraiva A, Ferreira N, de Sousa LL et al (2019) Classification of Images of Childhood Pneumonia using Convolutional Neural Networks. Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies
52. Shah A, Shah M (2022) Advancement of deep learning in pneumonia/Covid-19 classification and localization: A systematic review with qualitative and quantitative analysis. Chronic Dis Transl Med 8:154–171. <https://doi.org/10.1002/cdt3.17>
53. Sharma S, Guleria K (2022) Deep Learning Models for Image Classification: Comparison and Applications. In: 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). pp 1733–1738
54. Sharma S, Guleria K (2023) A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks. Procedia Comput Sci 218:357–366. <https://doi.org/10.1016/j.procs.2023.01.018>
55. Sharma R, Kukreja V, Kadyan V (2021) Hispa Rice Disease Classification using Convolutional Neural Network. In: 2021 3rd International Conference on Signal Processing and Communication (ICPSC). pp 377–381
56. Sharma S, Khanra P, Ramkumar KR (2021) Performance Analysis of Biomass Energy using Machine and Deep Learning Approaches. J Phys Conf Ser 2089:012003. <https://doi.org/10.1088/1742-6596/2089/1/012003>
57. Sharma S, Guleria K, Tiwari S, Kumar S (2022) A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans. Meas Sensors 24:100506. <https://doi.org/10.1016/j.measen.2022.100506>
58. Sim JZT, Ting Y-H, Tang Y et al (2022) Diagnostic performance of a deep learning model deployed at a national COVID-19 screening facility for detection of pneumonia on frontal chest radiographs. Healthcare (Basel) 10. <https://doi.org/10.3390/healthcare10010175>
59. Singh S, Tripathi BK (2022) Pneumonia classification using quaternion deep learning. Multimed Tools Appl 81:1743–1764. <https://doi.org/10.1007/s11042-021-11409-7>
60. Sirazitdinov I, Kholiavchenko M, Mustafaev T et al (2019) Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. Comput Electr Eng 78:388–399. <https://doi.org/10.1016/j.compeleceng.2019.08.004>
61. Sivaranjini S, Sujatha CM (2020) Deep learning based diagnosis of Parkinson's disease using convolutional neural network. Multimed Tools Appl 79:15467–15479. <https://doi.org/10.1007/s11042-019-7469-8>
62. Stokes K, Castaldo R, Federici C et al (2022) The use of artificial intelligence systems in diagnosis of pneumonia via signs and symptoms: A systematic review. Biomed Signal Process Control 72:103325. <https://doi.org/10.1016/j.bspc.2021.103325>
63. Toğacıar M, Ergen B, Cömert Z, Özürt 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. <https://doi.org/10.1016/j.irbm.2019.10.006>
64. Varshni D, Thakral K, Agarwal L et al (2019) Pneumonia Detection Using CNN based Feature Extraction. In: 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT). pp 1–7
65. Wu N, Xie Y (2022) A Survey of Machine Learning for Computer Architecture and Systems. ACM Comput Surv 55:1–39. <https://doi.org/10.1145/3494523>
66. Yasar H, Ceylan M (2021) A novel comparative study for detection of Covid-19 on CT lung images using texture analysis, machine learning, and deep learning methods. Multimed Tools Appl 80:5423–5447. <https://doi.org/10.1007/s11042-020-09894-3>
67. Zech JR, Badgeley MA, Liu M et al (2018) Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLoS Med 15:e1002683. <https://doi.org/10.1371/journal.pmed.1002683>
68. Zhang F (2021) Application of machine learning in CT images and X-rays of COVID-19 pneumonia. Medicine 100:e26855. <https://doi.org/10.1097/MD.00000000000026855>
69. Zhang C, Lu Y (2021) Study on artificial intelligence: The state of the art and future prospects. J Ind Inf Integr 23:100224. <https://doi.org/10.1016/j.jii.2021.100224>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.