Received: 8 August 2021 | Accepted: 20 January 2022

DOI: 10.1002/cdt3.17

REVIEW

Chronic Diseases® and Translational Medicine

Advancement of deep learning in pneumonia/Covid-19 classification and localization: A systematic review with qualitative and quantitative analysis

Aakash Shah[1] | Manan Shah[2]

[1]Department of Computer Science & Engineering, Institute of Technology, Nirma University, Ahmedabad, India

[2]Department of Chemical Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar, India

Correspondence
Manan Shah, Department of Chemical Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat 382007, India.
Email: manan.shah@spt.pdpu.ac.in

Edited by Yi Cui

Abstract

Around 450 million people are affected by pneumonia every year, which results in 2.5 million deaths. Coronavirus disease 2019 (Covid-19) has also affected 181 million people, which led to 3.92 million casualties. The chances of death in both of these diseases can be significantly reduced if they are diagnosed early. However, the current methods of diagnosing pneumonia (complaints + chest X-ray) and Covid-19 (real-time polymerase chain reaction) require the presence of expert radiologists and time, respectively. With the help of deep learning models, pneumonia and Covid-19 can be detected instantly from chest X-rays or computerized tomography (CT) scans. The process of diagnosing pneumonia/Covid-19 can become faster and more widespread. In this paper, we aimed to elicit, explain, and evaluate qualitatively and quantitatively all advancements in deep learning methods aimed at detecting community-acquired pneumonia, viral pneumonia, and Covid-19 from images of chest X-rays and CT scans. Being a systematic review, the focus of this paper lies in explaining various deep learning model architectures, which have either been modified or created from scratch for the task at hand. For each model, this paper answers the question of why the model is designed the way it is, the challenges that a particular model overcomes, and the tradeoffs that come with modifying a model to the required specifications. A grouped quantitative analysis of all models described in the paper is also provided to quantify the effectiveness of different models with a similar goal. Some tradeoffs cannot be quantified and, hence, they are mentioned explicitly in the qualitative analysis, which is done throughout the paper. By compiling and analyzing a large quantum of research details in one place with all the data sets, model architectures, and results, we aimed to provide a one-stop solution to beginners and current researchers interested in this field.

KEYWORDS

classification, Covid-19, deep learning, localization, pneumonia

1 | INTRODUCTION

Pneumonia is a respiratory disease responsible for significant morbidity all over the world. It causes a lower respiratory tract infection, leading to inflammation in the lungs' air sacs known as the alveoli. The infected alveoli are filled with fluid, which makes breathing difficult. Pneumonia, a contagious disease, is classified into two main types (hospital-acquired pneumonia and community-acquired pneumonia [CAP]) based on where it is acquired. The majority of pneumonia cases fall under the category of CAP (all cases of pneumonia that are not acquired from the hospital). If CAP is diagnosed early, the chances of 100% recovery are high, with little chances of reinfection. For a complete diagnosis of pneumonia, a combination of clinical awareness, specific microbiological tests, and radiographical studies are necessary. However, plain chest radiography alone can rapidly demonstrate the presence of pulmonary abnormalities in most cases.[1] Unfortunately, pneumonia is only one of many pulmonary abnormalities and, hence, radiographical findings often fail to lead to a definitive diagnosis of pneumonia. Consequently, the distinction of pneumonia from other pulmonary diseases cannot be made with certainty on radiological grounds with the current technology.

One of the significant problems of radiographical findings is that the distinction of pneumonia from other pulmonary diseases cannot be made with certainty on radiological grounds alone. Moreover, this is not the only problem with the current procedure of pneumonia diagnosis. A considerable number of medical images are produced in hospitals and medical centers daily. Consequently, radiologists are inundated with a large number of images that they have to analyze manually. In these cases, tried and tested deep learning algorithms might be helpful in assisting doctors by marking the part of the lungs where pneumonia/coronavirus disease 2019 (Covid-19) is present.

Many automated technologies related to medical imaging have shown promising results over the past few years, but deep learning has quickly gained prominence among them. Researchers have extensively exploited deep learning methods for detecting diseases in various body parts such as the eye, brain,[2,3] and skin.[4,5] In some medical imaging cases, it was shown that the classification performance of a deep learning model was better than that of medical specialists.[6] Since the proposal of AlexNet[7] in 2012, deep learning models have improved significantly in image classification tasks. Recent architectures such as ResNet and variations of ResNet have also provided a solid base for accurate object detection and localization. Although single-shot detectors such as Yolo[8] and RetinaNet[9] provide speedy detections useful in real time, generative adversarial networks (GANs)[10] have played an essential role in unsupervised learning and domain adaption whenever training images have been scarce. Hence, automated deep learning solutions can solve both problems mentioned above. Deep learning models for pneumonia classification and detection can automatically learn complex features from radiographs that may not be visible to the naked eye. This was proved in 2017 when Rajpurkar et al.[6] proposed CheXNet, a deep learning model, which achieved better results than radiologists on pneumonia detection and other pulmonary disease detection tasks.

The fact that deep learning models succeeded not only in the task of pneumonia detection but also in other pulmonary abnormality detection tasks was leveraged by many other researchers to detect other anomalies from the same models or training data. This case could prove useful, especially in recent situations (in 2021) such as the outbreak of Covid-19 because of the following reasons. Even though real-time polymerase chain reaction (RT-PCR) is the accepted as standard in the diagnosis of Covid-19, its sensitivity and specificity are not optimal.[11] Other than that, many countries or regions cannot conduct sufficient RT-PCR testing for thousands of subjects in a small span of time because of the lack of people who can perform these tests. In these cases, deep learning algorithms might help if the country has enough imaging machines but fewer people who can perform the test. RT-PCR testing may also be delayed in cases of newly evolved coronavirus, because detection of a newly evolved virus requires the extraction of the new DNA sequence.[11] In contrast, deep learning models with anomaly detection capabilities can detect the clustering effect of viral pneumonia occurrences such as Middle East respiratory syndrome (MERS),[12] severe acute respiratory syndrome (SARS),[13] and Covid-19 as proved by Zhang et al.[11] Thus, deep learning models provide a vital technique that might help in diagnosing pneumonia better and faster.

In this paper, we aimed to elicit, explain, and evaluate qualitatively and quantitatively all advancements in deep learning methods aimed at detecting bacterial or viral pneumonia from radiographical images. Since chest X-rays and computerized tomography (CT) scans are the most common radiographical tools doctors use today, we have covered deep learning methods that use chest X-rays, CT scans, or both as input images. As the quantitative results of these models depend on the data sets used, we group these models according to data sets, to perform a fair and uniform quantitative analysis. Although standard data sets are available for bacterial/viral pneumonia detection tasks, the same is not applicable for Covid-19 data sets due to the disease's novelty (in 2021). However, the models that leverage these data sets have been grouped by the amount and quality of images used for training and testing. This being said, it is not uncommon to find deep learning models that fail to perform well in the real world after being trained on data sets with specific sources. The poor performance in the real world is mainly because of the data set shift between training images and the images used in other hospitals. A significant amount of variability in individual hospital images also accounts for the poor performance of these models. To address this problem, we also evaluate and compare the features learned by various models to predict how well they would perform in the real world. The reason for comprehensively compiling all significant research in deep learning for pneumonia detection is to compare different models used in each scenario and identify the best deep learning architectures for each of those scenarios. Although similar work was performed by Li et al.,[14] we provide a significantly more comprehensive overview of models by including research with CT scans, localization tasks, and Covid-19 classification.

2 | METHODOLOGY

This review is based upon the qualitative and quantitative analysis of studies in the field of pneumonia/Covid-19 detection via chest X-rays and CT scans. The method for collecting relevant papers for this study was as follows. Platforms such as Elsevier, Google Scholar, IEEE Xplore, and Springer were searched with the keywords: "pneumonia detection with deep learning," "Covid-19 detection with deep learning," "pneumonia localization with deep learning," "Covid-19 localization with deep learning," "pneumonia detection with Chest X-rays," "pneumonia localization with chest X-rays," "Covid-19 detection with chest X-rays," and "Covid-19 localization with chest X-rays." Papers were excluded from the study as follows: all papers not related to deep learning, pneumonia, or Covid-19 were excluded. After the first exclusion process, all remaining papers were included in the final review according to the following criteria. As the main focus of this review is on the generalizability of models, all studies that made an explicit effort to make their model generalizable were included. Different studies used various metrics for accuracy, so there was no hard limit of accuracy (performance in general) for a paper to be included in this study. After that, studies were included with the goal of covering as much breadth in deep learning methods as possible. This was done because different deep learning methods often solve different problems (improper images, training data shortage, and insufficient training data variety). Furthermore, if a similar method was followed by more than one paper, then the most generalizable and the paper with the best performance was chosen.

On the medical front, pneumonia is mainly divided into two types: bacterial pneumonia and viral pneumonia. Although bacterial pneumonia does not have any subcategories worth discussing here, viral pneumonia is often subcategorized according to the virus responsible for causing viral pneumonia. The most recent example of viral pneumonia and of concern to us is Covid-19. Owing to these types and subtypes, researchers broadly classify input images into the following: (1) pneumonia/no-pneumonia, (2) bacterial pneumonia/viral pneumonia/no-pneumonia, and (3) Covid-19/all other pneumonia/no-pneumonia. Although most research papers fall into one of these three categories, some models do not consider no-pneumonia.

Radiologists use either chest X-rays or CT scans for diagnosing a patient. Both of these modes have their pros and cons. Although X-ray machines are portable and enable faster diagnosis, CT scans provide finer detail of the lungs that may be more difficult to see in a plain X-ray. Similarly, some deep learning models use X-rays as input images, whereas others use CT scans. This paper gives equal weightage to both models mentioned above but discusses them separately in Sections 3 and 4, respectively.

Other than classification, a significant task taken up by some deep learning models is that of detecting and localizing the region where pneumonia is present in the lungs. It is worth noting that some classification models also perform grad-cam analysis to analyze which features are being used to perform classification. These models, even after localizing features, are not considered localization/segmentation models. Localization/segmentation models provide bounding boxes/semantic segmentation in input images around the part of the chest affected by pneumonia. We will include these models in our discussion too. However, their comparison shall only be made with other localization models.

Data sets play one of the most prominent roles in the success or failure of deep learning models. The details of the three most frequently used data sets are shown in Table 1. The National Institutes of Health (NIH) data set consists of 15 classes, out of which one is pneumonia, one is no pulmonary disease, and the remaining 13 are other pulmonary diseases. It is worth noting that "other pulmonary diseases" may have any number of classes ranging from 0 to 13. This way, if it has 0 classes, the classification task simplifies to pneumonia/no-pneumonia (one sigmoid neuron or two softmax neurons in the output layer). On the other hand, if it has 13 classes, the model will classify a chest X-ray into pneumonia, no-pneumonia, or any one of the 13 pulmonary diseases (15 softmax neurons in the output layer). The classes of the Radiological Society of North America (RSNA) data set are normal, lung opacity, and no lung opacity-not normal, which can be explained as no pneumonia, pneumonia with visible lung opacity, and some pulmonary disease without visible damage to the lungs. Lastly, the classes of the Kaggle data set are divided as normal, bacterial-pneumonia, and viral-pneumonia, which need no further explanation.

TABLE 1 Data set for pneumonia detection

Data set	Images	Classes	Bounding boxes
NIH chest X-rays	1,12,120	14	985
RSNA chest X-rays	26,684	3	9555
Kaggle chest X-rays	5856	3	0
CheXpert	2,24,316	14	0
MIMIC-CXR	3,71,920	14	0

2.1 | Detection of pneumonia and its classification among other pulmonary diseases

Rajpurkar et al.[6] developed a deep learning model that could achieve radiologist-level accuracy on pneumonia detection from chest X-rays. They used the NIH data set, which consists of 112,120 chest X-ray images from 30,805 patients. This data set was first presented and used by Wang et al.[15] for the same task. However, the model was the first one that attained radiologist-level accuracy and it also served as a base for many future models. First, the entire data set is split into training and test sets such that no patients are repeated in the respective sets. The images are converted to size 224 × 224 and normalized by the ImageNet[16] training data set metrics. For training, these images are fed into the CheXNet model that uses a 121 layered dense convolution neural network (CNN) known as DenseNet.[17] DenseNet improves information flow and back-propagation through the network, which makes the optimization process easier. Hence, the entire model was used as it is, except for the output/classification layer. This layer was replaced by a single sigmoid neuron because the classification task was pneumonia/no-pneumonia. As the NIH data set consists of 15 classes, the classes pneumonia and no-pneumonia (14 classes including other pulmonary diseases) were highly imbalanced. To get rid of this problem, a weighted loss function is used while training the model. Finally, the model achieved an F1 score of 0.435 and an area under the receiver operating characteristic (AUROC) of 0.76 when tested with 420 images. The data set was randomly split into training (28,744 patients and 98,637 images), validation (1672 patients and 6351 images), and test (389 patients and 420 images). There was no patient overlap between the sets.

Zech et al.[18] demonstrated that deep learning pneumonia classifiers trained on two different hospital systems predicted results by learning the origin of those hospitals instead of learning relevant features that cause pneumonia. To address this problem, Janizek et al.[19] developed an adversarial training-based approach. They found that the occurrence of pneumonia in posterior–anterior (PA) chest X-rays was twice as much as that of pneumonia in anterior–posterior (AP) images (PA images are the ones in which X-rays enter from the back of the body, whereas AP is vice versa). They also found out that pneumonia detection classifiers as in Rajpurkar et al.[6] learned to distinguish between the two views (AP and PA) and leveraged that information to classify pneumonia. Their approach was different from standard adversarial approaches, where the classifier learns domain-invariant features. In their case, the classifier could not learn domain-invariant features, because they had no images from the target domain. In their adversarial approach, Janizek et al.[19] tried to train a classifier in which the final output score of the classifier would be invariant of the view (AP or PA). Although the training and architecture for their classifier were the same as that of Rajpurkar et al.,[6] they also added and trained an adversary network. This adversary network took the output score of the classifier as input and outputted a prediction of the view. The adversary network is a standard 3 layered feedforward network of 32 neurons, each with rectified linear unit (ReLU) activations. The classifiers' objective was to predict output scores such that the adversary could not predict the view of the input image from the output score. In contrast, the adversarial network's objective was to predict the output score's view (AP or PA). Both the classifier and the adversary network were trained alternatively for optimizing their respective objectives. To test their approach, Janizek et al.[19] tested their model on the CheXpert data set (source domain) and Massachusetts Institute of Technology MIMIC-CXR data set (target domain). Although the standard model (without the adversary network) achieved an AUROC of 0.79 on the source domain, it could only achieve an AUROC of 0.703 on the target domain. Alternatively, the adversarially trained model achieved almost similar AUROC's of 0.747 and 0.739 on the source and target domains.

In April 2020, Lu et al.[20] presented the MUXConv, a CNN layer specially designed to increase the flow of information by multiplexing channels and spatial input through the network. They also presented a multi-objective algorithm to automatically optimize hyperparameters while training. Although the MUXConv was not specially designed for pneumonia classification, it could achieve an AUROC of 84.1% on the same data set used by Rajpurkar et al.[6] while using 3× fewer parameters, being 14× more efficient than DenseNet-121 and without any manual hyperparameter optimizations. This result shows the scope of improvement in the accuracy of pneumonia detection through better deep learning architectures alone, i.e., without considering any medical knowledge. In September 2020, the same team presented the NSGANetV1, another multiobjective evolutionary algorithm. NSGANetV1 learns the designs of various architectures through the recombination and generation of multiple architectural components. NSGANetV1 makes its efficiency better by exploiting various patterns used in successful architectures by estimating their distributions with the help of a Bayesian model. Although made for general-purpose image classification, this model achieved an AUROC of 84.6% on the NIH data set without modifications or hyperparameter tuning. Moreover, the class activation mean average map) of NSGANetV1 showed that the model learns relevant features, which can also be used to pinpoint the region where pneumonia is present.

Using architectures such as DenseNet-121 in the pneumonia detection task is possible because of large data sets such as NIH or CheXpert. If such architectures are used with smaller data sets such as that of Kaggle, there is a considerable chance of overfitting. Li et al.[21] presented the PNet, an efficient yet effective architecture for pneumonia detection using a significantly smaller number of images. They collected their own data set from Shenzhen No.2 People's Hospital, consisting of 6339 X-rays labeled pneumonia and 4445 X-rays labeled normal. The architecture of PNet is straightforward, consisting of only five convolution blocks, each followed by a max-pooling layer. This small architecture allows PNet to be 25 times as efficient as AlexNet and about 50 times as efficient as visual geometric group (VGG) detection task with an accuracy of 92.79% and an F1 score of 0.93. Even though PNet has a smaller number of parameters, it outperforms both the AlexNet and VGG 16 in the pneumonia are many customized architectures such as PNet, which also get equivalent accuracy. However, only PNet was included in our research because of its excellent results on feature analysis. While analyzing the features of all models, it was found that VGG 16 focuses on the entire lung region instead of focusing on the pneumonia-affected region and AlexNet wanders off to the wrong regions. On the other hand, PNet focuses on only those features that correspond to the pneumonia-affected region in most cases. Hence, PNet is not only good at detecting pneumonia but it can also help doctors by highlighting the pneumonia-affected area. The detailed results were true positive/false positive/true negative/false negative: 617/86/360/19 with a sensitivity of 0.9701 and specificity of 0.8072.

Dong et al.[22] presented a network architecture that achieved high classification accuracy in pneumonia detection. They used an improved quantum neural network and trained this model on the Kaggle chest X-ray data set containing 5232 training images. This model was tested using 624 separate images in the test set and achieved an accuracy of 96.07%. They also trained AlexNet, ResNet, and InceptionV3 on the same data, giving 85.30%, 86.38%, and 95.53% accuracy, respectively. Although the authors do not conduct a feature analysis in their paper, chances are few that a quantum neural network would give such high accuracy while learning wrong or irrelevant features. The data set that these authors used was published by the University of California, San Diego. The sensitivity and specificity were 0.9756 and 0.9460, respectively.

Diving deeper into pneumonia detection with small data sets, most intuitively, we come across a solution based on GAN. Khalifa et al.[23] used a GAN with various deep learning models to generate more images and use those images to train the deep learning models. They took only 10% images from the Kaggle chest X-ray data set and generated the remaining 90% with the GAN for training purposes. These images were then used for training by AlexNet, SqueezeNet, GoogleNet, and ResNet with 8, 18, 12, and 18 layers, respectively. ResNet performed best with a testing accuracy of 99.0% and a recall of 0.9897. The catch, however, is that they used 624 images to train the GAN, which is the same number of images provided in the testing data set. Although the authors have mentioned that three separate trials were conducted with a different 10% of the data set, using test images in even one of the four trials would drastically change the average accuracy. Nonetheless, the idea of using GAN's to generate new data can certainly be applied when there is a dearth of training images.

Dey et al.[24] developed a model with an Ensemble Feature Scheme (EFS) for pneumonia detection. Their EFS combines handcrafted features and automatically extracted features from a deep learning model to classify an image into pneumonia or normal. Extraction of hand-crafted features is again completed by combining continuous wavelet transform, discrete wavelet transform, and gray level co-occurrence matrix (GLCM). The deep learning features are extracted using the standard VGG-19 architecture. The combined handcrafted features are then concatenated with features extracted using VGG-19 through PCA and serial feature concatenation. After concatenation, these features are given as an input to a random forest classifier for final classification. This model was trained using 5500 images from the NIH data set and achieved 97% accuracy when tested against 1650 separate images from the NIH data set. Similar to other models mentioned in this paper, the feature activations of this model also point to relevant regions in the lung where pneumonia is present. The detailed metrics were true positive rate/false positive rate/true negative rate/false negative rate: 0.9756/0.0244/0.9808/0.0192 with a sensitivity of 0.9807 and specificity of 0.9757 (Table 2).

2.2 | Detection of Covid-19 and classification of viral pneumonia from bacterial pneumonia

Capturing a chest X-ray is one of the primary methods of screening the occurrence of Covid-19. However, there is a general dearth of doctors even at places where equipment to capture such X-rays is available. To tackle this problem, a lot of research has been done to detect Covid-19 from chest X-rays automatically. Cases of Covid-19 emerged in the entire world in 2019, but a lot of research in pneumonia detection from chest X-rays had already been done before. Hence, much research on the detection of Covid-19 from chest X-rays is built upon the base provided by previous research into pneumonia detection. Due to the novelty of Covid-19 (in 2020–2021), no standardized databases are available and almost every research work uses a different database. Hence, the details of all databases and comments on their quality are given while explaining the research work rather than giving an overview of all databases beforehand.

TABLE 2 A comprehensive study on pneumonia detection and classification

Author	Model	Data set	AUROC	Accuracy
Rajpurkar et al.[6]	CheXNet (DenseNet-121)	NIH	0.760	NA
Janizek et al.[19]	CheXNet (DenseNet + Adversarial)	NIH + MIMIC	0.747	NA
Lu et al.[20]	MUXConv (multiplexed convolutions)	NIH	0.841	NA
Lu et al.[20]	NSGANetV1	NIH	0.846	NA
Li et al.[21]	P-Net (customized CNN)	Custom (10,784)	NA	92.79%
Dong et al.[22]	Quantum neural network	Kaggle	NA	96.07%
Khalifa et al.[23]	GAN (semi-supervised)	Kaggle (624)	NA	99.00%
Dey et al.[24]	EFS (CWT + DWT + GLCM)	NIH (5550)	NA	97.00%

Abbreviations: AUROC, area under the receiver operating characteristic; CNN, convolution neural network; CWT, continuous wavelet transform; DWT, discrete wavelet transform; GAN, generative adversarial network; GLCM, gray level co-occurrence matrix; NIH, National Institutes of Health.

Haghanifar et al.[25] made a hierarchical deep learning model for detecting Covid-19. In the first level, images of chest X-rays are classified into normal and pneumonia. In the second level, images classified as pneumonia are further classified into covid positive (CP) or CAP. The data set used by the authors contains 780 Covid-19-positive X-rays, 4600 X-rays having CAP, and 5000 normal X-rays. The approach taken by Haghanifar et al.[25] was very similar to that of Rajpurkar et al.[6] The key difference was that Haghanifar et al.[25] first segmented the lungs from chest X-ray and then they only used the part surrounding those lungs for classification. This approach, to a significant extent, solved the issue of "learning the wrong features to reach the right answer," because then, the model was forced to learn only from the lung region rather than learning from the entire X-ray, which usually contains a lot of regions other than the lungs. U-Net was used for segmentation of the lung region and then they performed dilation on the segmented lungs to cover some lung areas that the U-Net did not segment. After segmentation, they cropped the chest X-ray image such that only the segmented area was covered. This cropped image was then fed into the DenseNet-121 model given by Rajpurkar et al.[6] This model achieved an accuracy of 81.04% and *f*-scores of 0.85 and 0.76 for CP and CAP classes, respectively. Although the accuracy of this model is 0.4% less than that of CheXNet,[6] it is more robust than CheXNet on unseen data because of the cropped images. The precision and recall for (normal/pneumonia/Covid-19) were P: (0.8251/0.9340/0.9420) and R: (0.9516/0.7797/0.9420), respectively.

While on the topic of lung segmentation, we cover another research work,[26] which uses lung segmentation to classify a chest X-ray into bacterial pneumonia or viral pneumonia. The data set used by them consists of 241 X-ray images where lungs have been separated manually. The rest of the data set consists of 4513 pediatric chest X-ray images, out of which 2665 are bacterial pneumonia and 1848 are viral pneumonia. The entire model is divided into three parts. The first part is where the lung region is segmented from the chest X-ray by an eight-layer fully convolution network (FCN).[27] The FCN model was trained using the 241 segmented images from the Japanese Society of Radiological Technology data set and used pretrained weights from the Pascal visual object class[28] segmentation data set. The second part consists of feature extraction, where features are extracted using three different methods. The first method uses a deep CNN (DCNN), the second method uses a mixture of GLCM-based (Gray-Level Co-occurrence Matrix) texture features and histogram of oriented gradients-based shape features, whereas the third method uses HAAR wavelet texture features. The third part of the model uses a simple support vector machine (SVM) classifier to classify a given image into bacterial pneumonia or viral pneumonia. This particular approach achieved an accuracy of 76.92% with an area under curve (AUC) of 82.34%. At this point, it is imperative to reiterate that metrics like accuracy, F-scores, and AUC should not be the only parameters to judge the performance of a deep learning (DL) Model. In fact, in most cases, perfect or close to perfect metrics suggest the opposite of sound, because in most cases, the underlying model is overfitted, not because of the complexity of the model or the lack of data, but because of learning irrelevant features that are specific to the source of train data. The model achieved a sensitivity of 0.5567 and specificity of 0.9267.

Covid-19 is a type of viral pneumonia, but it is not the only type of viral pneumonia. Several different respiratory diseases such as MERS and SARS fall into the category of viral pneumonia. Moreover, the occurrence of clusters of viral pneumonia cases over a short period can be a signal of an upcoming outbreak or a pandemic. Keeping this in mind, Zhang et al.[11] developed a Confidence Aware Anomaly Detection (CAAD) model to detect the occurrence of viral pneumonia from chest X-rays. To train their model, they used two in-house data sets named X-Viral and X-Covid. The X-Viral data set contains 5977 viral pneumonia images, 18,619 nonviral pneumonia images, and 18,774 normal images. The X-Covid data set contains 106 CP images and 107 normal images. They also used the Open-Covid data set containing 493 CP images. The CAAD model has three main parts. A feature extractor, an anomaly detector, and a confidence predictor. Before we go any further, it is essential to clarify that the "anomaly" we are trying to predict is viral pneumonia and all other classes (pneumonia and normal) are considered normal. Moving back to the model, after passing an image to the feature extractor, the features are passed simultaneously into the anomaly detector and the confidence predictor. If the anomaly detector predicts the image as an anomaly or the confidence predictor predicts our model's confidence below a particular threshold, the image is considered an anomaly, i.e., viral pneumonia. The feature extractor is made up of EfficientNet B0.[29] The authors designed the anomaly predictor and the confidence detector, and they are not as common as other ones mentioned in this review, so they deserve an explanation. However, the explanation is too involved and out of the scope of this review, so readers are requested to read the original paper for an explanation of those modules. Coming to the results of this approach, it achieved 80.33% accuracy on the X-viral data set with training and 78.57% accuracy on the X-Covid and Open-Covid data sets combined without any training. This shows us that the model could categorize Covid-19 cases as viral pneumonia without any specific training on Covid-19 images, which shows that this model can be useful in predicting upcoming cases and different mutations of viral pneumonia. The sensitivity and specificity on various data sets for viral and normal classes were: (X-Viral: 85.88/79.44), (X-Covid: 71.70/73.83), (Open-Covid: 100/100), (X-Covid + Open-Covid: (77.13/78.97)).

Another instance of a region-based discriminator for Covid-19 was given by Wang et al.[30] in August 2021. They used the Covid-CXR data set consisting of 204 CP X-rays and the RSNA pneumonia detection data set for 2004 CAP and 1314 normal chest X-rays to train their model. The authors proposed a Discrimination-DL and a Localization-DL, but their approach was completely different. They divided all chest X-ray images into superpixels first and then they ran a proposal of lung (POL) regressor over those superpixels. This approach is very similar to that of YOLO,[8] with a critical difference that only the outer boundaries of all superpixels inside the POL-proposed rectangles are used to extract two lungs. After both lung regions are extracted, they are passed into the Discrimination-DL, which comprises a ResNet and a feature pyramid network over the ResNet to rebuild the image after feature extraction. Focal loss is then measured against the rebuilt image and the original lung region is passed into the Discrimination-DL. This method helps the Discriminator-DL in learning optimal features. If the Discriminator-DL classifies the image into CP, both the softmax score and original image are passed into the Localization-DL. The Localization-DL only gives one out of three results, that is, it classifies the Covid-19 as either present in the left lung or the right lung or both lungs. The name Localization-DL might thus seem to be misleading, because it is more of a classifier. Nevertheless, the Localization-DL uses a residual attention mechanism to determine the occurrence of Covid-19 in both lungs. The residual attention mechanism looks at the features extracted by the feature extractor to determine where the attention of the classifier lies. For a deeper analysis of the residual attention mechanism, the reader is referred to the original paper.[31] Coming to the accuracy of this model, it achieves 99%, 90%, and 93% accuracy on CP, CAP, and normal classes, respectively.

Arias-Londono et al.[32] presented a thoughtful evaluation approach for DL networks that detect Covid-19. Not only that, but they also compiled the most extensive known data set of 8573 unique Covid-19 chest X-rays. The entire data set consisted of 49,000 normal, 2400 CAP, and 8573 Covid-19-positive images. They used the same deep learning model used in Covid-Net[33] and ran three different experiments on this data set and model. The first experiment used raw images as an input, with the only preprocessing being histogram equalization. In the second experiment, they used U-Net to segment the lung region and cropped the image so that only the region encompassing the two lung regions remained. In the third experiment, the same segmentation approach was used, but this time they only kept the segmented lung part while the remaining region was filled with a black mask. Upon Grad-Cam analysis, it was found that only experiment three learned relevant features even if the accuracy was lower than that of the other two experiments. They also showed that the accuracies of the AP X-ray projection were significantly higher than that of the PA projection. The showings of this study take us to an important point worth noticing. As shown below, metrics such as accuracy and F-scores can be bolstered if the deep learning model is not extracting the right features. However, models made in such a manner may be poor at generalizing to new data from a new source. Hence, Grad-Cam analysis is crucial to determine whether a given model will be able to perform well in the real world, and one should not judge a model solely based on its metrics, especially if the train/test data is less or if the train/test data belong to the same source.

Before we continue with our quest for the best deep learning models for Covid-19 detection and classification of viral pneumonia from bacteria pneumonia, we should make a note. The constructions of all models discussed above show an explicit effort to make the model perform well in the real world. These efforts are shown in the form of Grad-Cam evaluations or segmenting the lungs so that the models learn only relevant features. The models described below this point, however, do not showcase any effort of such kind. Hence, even though the accuracies and other metrics of the models below this point might seem significantly higher than those mentioned above, the reader should keep in mind that they are not proven to generalize well in the real world.

To overcome the problem of a significantly smaller number of Covid-19 images as compared with normal and CAP images, Sakib et al.[34] used a custom GAN to generate more Covid-19 images for training. The data set used by them consisted of 27,228 normal, 5794 CAP, and 209 Covid-19 images. On analysis, they found that generating precisely 100%, that is, 209 new Covid-19 images by GAN, led to the highest classification accuracy. On top of GAN, they used a customized CNN with exponential linear unit activation and Adagrad optimizer. The idea of using a customized and lean CNN works well in cases where data used for training is less. In such cases, even if the metrics are not necessarily excellent, we can be assured that the model will not overfit our small data set, ensuring good generalizability. Talking about the results, this model achieved 93.94%, 88.52%, and 95.91% accuracy on CP, CAP, and normal cases, respectively.

Ali et al.[35] proposed a dual attention module to classify viral pneumonia and bacterial pneumonia. For training, they used the popular data set available on Kaggle, which consists of 5856 chest X-rays. The dual attention module consists of a spatial attention module and a channel attention module. For readers that do not know what "attention" is, attention was primarily used fornatural language processing (NLP) in recurrent neural networks to allow the network to remember the relevant parts of a sentence. Later on, it was adopted into computer vision to determine the relevance of each feature with respect to the output. After that, each feature is multiplied by its weight to give importance to those features that contribute more to the output. The channel attention modules measure the importance of each channel with regard to other channels, whereas the spatial attention module measures the importance of each feature in a channel with regard to other features in the same channel. This model achieved an accuracy of 97.82%.

Ohata et al.[36] used MobileNet to classify chest X-rays with Covid-19 and normal chest X-rays. The data set used consisted of 194 Covid-19 images and the normal images were collected from Kaggle and NIH data sets. They used MobileNet for feature extraction and tried six different classifiers for classification purposes. In the end, they decided to use linear SVM for classification purposes, which gave an accuracy of 98.62%. Lastly, Chowdhary et al.[37] tried using various models such as SqueezeNet, MobileNet, InceptionV3, ResNet18, ResNet101, CheXNet, DenseNet201, and VGG19 on 423 Covid-19 images, 1579 normal images, and 1485 CAP images. They concluded that DenseNet and CheXNet perform best (99.70% accuracy) in two-class classification, that is, Covid-19 and other, whereas DenseNet performs best (97.94% accuracy) in three-class classification problems, that is, Covid-19, CAP, and normal. The sensitivity and specificity were 0.979 and 0.988, respectively (Table 3).

2.3 | Localization of pneumonia in chest X-rays

Although we have already covered some research that localized the entire lung region with the help of segmentation models such as *U-Net* or *a-YoLo-like-lung-regressor*, it is worth noting that the research covered previously only localized the entire lung regions and not pneumonia-affected regions. Localization of pneumonia-affected regions in a chest X-ray can be beneficial in two ways. Mainly, it can assist radiologists in giving a quicker

TABLE 3 A systematic study on detection of Covid-19 and classification of viral pneumonia from bacterial pneumonia

Author	Model	Task	Data set	Accuracy
Haghanifar et al.[25]	U-Net + DenseNet-121	CP/N/CAP	780/4600/5000	81.06%
Gu et al.[26]	FCN + (DCNN)	Bacterial/viral	2655/1848	76.92%
Zhang et al.[11]	ResNet + AD + CoP	CP/N/CAP	5977/18619/18774	80.33%
Wang et al.[30]	POL + ResNet	CP/N/CAP	204/2004/1314	99%/90%/93%
Arias-Londoño et al.[32]	U-Net + Covid-Net	CP/N/CAP	8573/400/49000	91.53%
Sakib et al.[34]	GAN + Custom CNN	CP/N/CAP	209/5794/27228	94%/88.5%/96%
Ali et al.[35]	ResNet + Attention	Bacterial/Viral	Kaggle	97.82%
Ohata et al.[36]	MobileNet	CP/CN	194/NIH-RSNA	97.00%
Chowdhury et al.[37]	Multiple	CP/N/CAP	423/1485/1579	97.94%

Abbreviations: CAP, community-acquired pneumonia; CN, covid negative; CNN, convolution neural network; Covid-19, coronavirus disease 2019; CP, covid positive; DCNN, deep convolution neural network; FCN, fully convolution network; GAN, generative adversarial network; N, normal; NIH, National Institutes of Health; RSNA, Radiological Society of North America.

and more accurate diagnosis. Not only that, but localization also solves a significant problem of generalizability that we have encountered so far. If the primary goal of our deep learning model is to localize pneumonia-affected regions, we can be assured that the model is not looking at the wrong features to arrive at the right decision. As far as data sets are concerned, only one data set (RSNA) has enough images with bounding boxes to train a DL that localizes well. Thus, it will be easy to compare all research work in this section based on metrics alone.

We start by explaining the approach[38] because they won the RSNA Pneumonia Detection Challenge hosted by Kaggle. The authors used an ensemble of five models to localize pneumonia in chest X-rays. These five models were divided into two groups. The output regions from the first group (three models) were ensembled into one region. Similarly, the output regions from the second group (two models) were separately ensembled into a single region. Finally, the output regions from the two groups are ensembled into one output region using appropriate thresholds. The first group is made up of one Deformable Object Relation Network and two Deformable region-based FCNs (R-FCNs). Here, the prefix *Deformable* simply suggests the use of deformable convolutions in the respective architectures. Deformable convolutions are different from regular convolutions in that every pixel/feature is offset by a certain amount in a certain direction. In this way, the shape of the receptive field of the convolution becomes free and is not limited to a rectangle. The offsets are learnable and thus play an essential role in correctly locating the entire object.

The object relation network is not used very commonly and thus deserves some explanation. The object relation module is an adapted version of a basic attention module used in NLP. Although the primitive elements of an NLP attention module are words, the primitive elements of an object relation module are objects. As objects have a two-dimensional spatial arrangement and vary in terms of scale/shape, their locations and geometrical features are much more complex than the positions of words in a single sentence. Hence, the object relation module has an added geometric weight other than the original weight commonly found in NLP attention modules. The geometric weight considers the relative geometry of objects and models spatial relationships between them.

The second type of module used in the first group is deformable R-FCN, which is just R-FCN with deformable convolutions. R-FCN is explained during the discussion of GeminiNet in this section itself.

Moving on, the second group is made up of two RetinaNets. The difference between these two RetinaNets is not in their architectures but in the type of input images used for training. The first RetinaNet, also called the ConcatRetinaNet, uses concatenated images for training. Each concatenated image is made by concatenating a pneumonia-negative image with a pneumonia-positive one. This way, the RetinaNet improves its distinguishing capacity, while distinguishing between lung opacity with pneumonia and lung opacity without pneumonia. Images of 10 different sizes are given as input to all five models. Hierarchical ensembles are then formed from the two main groups and, finally, the bounding boxes from both models are ensembled according to different thresholds.

Li et al.[39] used 30,000 images to train their model and the rest of the images from the RSNA data set were used for testing. Before using the raw images for input, they segmented the lung region from the original image using U-Net, much similar to Haghanifar et al.[25] After segmenting the lung region, they combined the segmented and raw images to make a final data set for training their model. They used the SE-ResNet34 for localizing regions containing pneumonia.[40] SE-ResNet is short for squeeze-and-excitation ResNet, which is basically an encoder–decoder model that serves multiple purposes. The SE-ResNet acts as a feature extractor and its side branch can automatically learn weights to assign importance to each channel. Moreover, the model can learn smoothly even over significantly deep layers without risk of degradation because of the residual blocks. Hence, the model works as a channel attention module over a ResNet34. For the final output, each pixel in the output channel represents the probability of that pixel belonging to the pneumonia class. The regions can then be extracted by applying thresholds to those probabilities. Coming to the results of this model, it was able to achieve an mean average precision (mAP) score of 0.262. The mAP was calculated under intersection over union (IoU) thresholds of 0.3, 0.4, 0.5, 0.6, and 0.7.

Dimitrov's team placed second in the RSNA pneumonia detection challenge hosted by Kaggle. The paper (including Poplavskiy) describes their model and approach in detail.[41] For their model, they used RetinaNet, which is a single-shot detector. For the base of RetinaNet, they decided to use the encoder part of SE-ResNext-101. This particular design was chosen to accommodate both the speed of a single shot detector and the accuracy of a deep model such as ResNext-101. Using this approach, they were able to achieve an mAP score of 0.26097. The official score on the leaderboard was 0.24781, but they optimized the model with heavy augmentations and zero rotation after the competition was over. A lot more trial and error went into making this model, mainly because it was made as a part of a competition. Almost all hyperparameters in this model are optimized and with good reasons, which are provided in their paper.

Up until now, we have talked about research that uses single-shot detectors for the localization of pneumonia-affected regions. However, two-stage detectors have a significant advantage over single-shot detectors in terms of accuracy. There is, of course, a time tradeoff involved while using two-stage detectors, but the question to ask is: how much does the detection time matter? At testing time, the difference between single-shot and two-stage detectors is not big enough to make any significant difference, because real-time detection is not required for any use case of pneumonia localization.

Keeping this in mind, Yao et al.[42] presented the GeminiNet in March 2020. Before we begin with the explanation of this study, there is a note worth taking. Some terminology in the following four or five sentences might sound new to beginners, but all of it is elaborated upon in considerable detail in the two successive paragraphs. Continuing with GeminiNet, it is a two-stage detector that builds upon the concept of R-FCN.[43] The difference between R-FCN and GeminiNet is that the latter uses RFB[44] blocks instead of simple convolution blocks for multiscale context information. Moreover, they changed the base model used for feature extraction. Instead of using ResNet-50 they used DetNet59, because it yielded better performance metrics. This model (DetNet59 + GeminiNet) presented by the authors achieved an mAP score of 0.3259 at IOU thresholds 0.4, 0.5, 0.6, and 0.7.

Now onto the elaboration, the RFB block is much like an InceptionV1 block, except it has an extra shortcut such as residual blocks in a ResNet. RFB blocks are especially useful in object detection scenarios, because they have variable receptive fields (e.g., inception) and they can handle deep models smoothly (e.g., ResNet). Moreover, instead of simple convolutions, the authors used dilated convolutions in the RFB block.[45] Dilated convolutions convolve upon a larger size (say 5 × 5 instead of 3 × 3) but select only a few features (3 × 3 = 9) from the big block (5 × 5), thereby keeping the number of parameters small but increasing the receptive field (Figure 1).

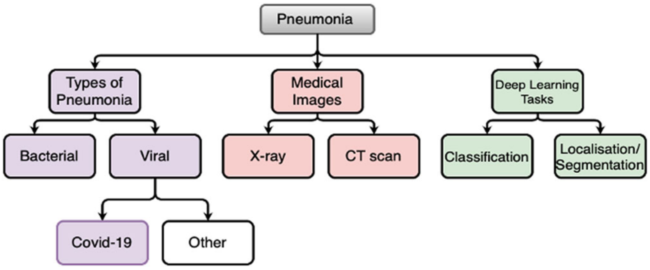

FIGURE 1 Classifications of pneumonia and its detection techniques. Covid-19, coronavirus disease 2019; CT, computerized tomography

Although RFB blocks are important in GeminiNet, its heart is the R-FCN. R-FCN is used as a substitute for Fast-residual convolutional neural network) and Faster R-CNN. Fast R-CNN improves upon the speed of R-CNN by calculating the feature map of the entire image at once and uses that feature map to derive region of interest (ROIs) directly. Feature maps do not have to be calculated for different ROI's separately. R-FCN works by simultaneously generating ROIs and region-based feature maps, thus saving a lot of time. After that step, for all regions generated in the ROI step, region-based feature maps are checked to vote for the probability of a particular ROI containing a particular part of the entire object. The final vote array (consisting of probabilities from all ROIs) is averaged to determine which object is present in the image. This process of calculating probabilities for all ROIs and storing them in a vote array is called position-sensitive ROI (PS-ROI) pooling. GeminiNet does not use R-FCN as it is. The changes are as shown in Figure 2.

While on the topic of R-FCN, the approach of the DeepRadiology Team[46] is worth mentioning. They used a modified version of R-FCN called CoupleNet.[47] CoupleNet adds a second branch to R-FCN for processing global features. This way, the resulting architecture learns features from a larger area through the global branch by adding extra ROI features and local features learn from the local branch by using PS-ROI features. The DeepRadiology Team used an ensemble of four models having the same architecture. All four of these models gave unique outputs, which were used for generating the final regions. First, all bounding boxes that had a confidence score < 0.5 were eliminated. After that, bounding boxes from all four groups, which had an IOU > 0.25 were grouped together. Lastly, the coordinates of all bounding boxes in one group were used to derive a final bounding box. This model was able to achieve an mAP of 0.23089 and placed seventh in the competition.

Next, we move on to models that use a combination of single-shot detector and two-stage detectors. Sirazitdinov et al.[48] presented a model that used a combination of RetinaNet (single-shot detector) and Mask R-CNN (two-stage detector). RetinaNet worked as the main unit, whereas Mask R-CNN was used as an

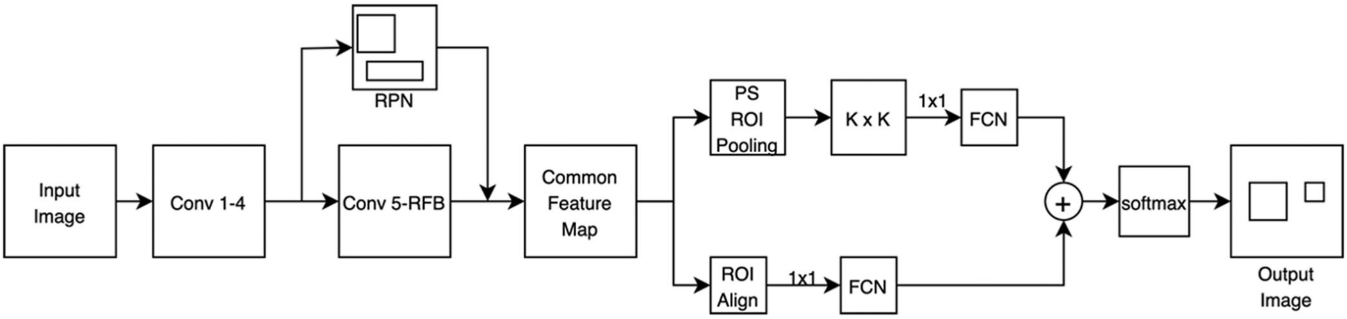

FIGURE 2 Architecture of GeminiNet. FCN, fully convolution network; PS ROI, position-sensitive ROI

auxiliary unit to adjust the regions of RetinaNet. The working of the entire model is straightforward. Both the RetinaNet and the Mask R-CNN models work separately and predict bounding boxes with corresponding classes. After applying non-max suppression in both models, a weighted average of predictions from both models is calculated where the weight of RetinaNet: Mask R-CNN predictions is 3:1. This ratio was calculated by an iterative grid search over many such ratios ranging from 1:1 to 4:1.

Another research work explored the combination of RetinaNet and Mask R-CNN for pneumonia detection.[49] They tried various ensembles of RetinaNet and Mask R-CNN with different sizes and different weights. Finally, a model with RetinaNet 178, RetinaNet 184, RetinaNet 201, Mask R-CNN 150, and Mask R-CNN 162 in the ratio 2:2:3:2:3 was used for detection. This model achieved an mAP of 0.21746, which could be placed at the 21st place in the competition approximately (Table 4).

2.4 | Classification of Covid-19 and CAP via CT scans

Harmon et al.[50] made a deep learning model detect Covid-19 from CT scans using multinational data sets. Their data set consisted of CP scans from China (369), Japan (100), and Italy (57). In total, 1059 scans were used for training and 1397 separate scans were used for testing. Their deep learning model consists of a lung segmentation module and a classifier module. The lung segmentation module segments the lung region from the entire CT scan. After the lung region is segmented, the segmented region is given as an input to the classifier, which classifies the input into CP or covid negative. For the lung segmentation module, the AH-net[51] architecture is used. AH-Net is an encoder–decoder-based segmentation module used for three-dimensional (3D) segmentation and it mostly works similar to U-Net. The segmented regions used while training had a mean dice score of 0.95. Dice scores are similar to IOU scores and are used widely as a metric in segmentation tasks. Moving on, the classification module is made up of the DenseNet-121 architecture just like CheXNet and takes a fixed input of size 192 × 192 × 64. Finally, this model achieved an accuracy of 89.6% with an AUC score of 0.941 on independent testing sets. Although the architecture of the classifier in this model is the same as CheXNet, the number of training images is significantly fewer. Nevertheless, Grad-CAM evaluations of this model show that the model can learn correct features to arrive at the right decision. Hence, the segmentation module that precedes the classification module plays a vital role in the generalizability of this model. The sensitivity and specificity of this model were 0.840 and 0.930, respectively.

Ouyang et al.[52] presented a deep learning model with dual sampling and an online, trainable class activation mapping (CAM) module to ensure that the model learned important features. The training data set used for this model contains 2186 images, of which 1092 are CP and 1094 are CAP. The data set used for testing is also quite large, with 2796 images, of which 2295 are CP and 501 are CAP. The authors also use a standard lung segmentation module called the VB-Net toolkit[53] for lung segmentation. Feature extraction is then done using a ResNet34. After segmentation, the entire data set is sampled in two ways. The first one is uniform sampling, where each minibatch contains images in the same ratio as the entire data set. The second method is size-balanced sampling. Size-balanced sampling is required, because the data set has only a small number of Covid-19 images with a small infection area. Similarly, only a few images with a large area of infections are available in the CAP category. Hence, size-balanced sampling is applied such that the ratio of: CAP images with large infection; CAP images with small infection; covid images with large infection; and covid images with small infection remain approximately the same in each minibatch. This ratio is maintained by oversampling. However, oversampling poses another challenge of overfitting. This challenge is resolved by using the first of its kind, online CAM module. The online CAM module

TABLE 4 A comprehensive review on localization of pneumonia in chest X-rays

Author	Model	Type	IOU thresholds	mAP
Li et al.[40]	U-Net (SE-ResNet34)	SSD	0.3–0.7 (0.1)	0.262
Gabruseva et al.[41]	RetinaNet (SE-ResNext101)	SSD	0.4–0.75 (0.05)	0.260
Yao et al.[42]	GeminiNet (modified R-FCN)	TSD	0.4–0.7 (0.1)	0.326
The DeepRadiology Team[46]	CoupleNet (modified R-FCN)	TSD	0.4–0.75 (0.05)	0.231
Sirazitdinov et al.[48]	RetinaNet + Mask R-CNN (3:1)	SSD + TSD	0.4–0.75 (0.05)	0.204
Ko et al.[49]	RetinaNet + Mask R-CNN (7:5)	SSD + TSD	0.4–0.75 (0.05)	0.217
Pan et al.[38]	R-FCN + RelNet + RetinaNet	SSD + TSD	0.4–0.75 (0.05)	0.255

Abbreviations: mAP, R-FCN, region-based fully convolution network; SSD, single shot detector; TSD, two stage detector.

is generated by applying a 1 × 1 × 1 convolution to the weights of the fully connected layer and then convolving that layer over the feature map. A ReLU operation is applied at last to get the final activation map. This model achieved 95.4% accuracy with an AUC of 0.988. The sensitivity and specificity of this model were 0.872 and 0.907, respectively.

The work of Wang et al.[54] is yet another example of a deep learning model that consists of a lung segmentation module followed by a classifier with attention. Their data set consists of 4657 scans where 936 are Normal, 2406 are CAP, and 1315 are CP. For segmentation, the authors used the 3D-UNet[55] models. After lung lobe segmentation, the images are cropped into a size of 96 × 96 × 96 and passed into the classifier. The classifier consists of two parts, the pneumonia detector and the pneumonia classifier. If an image is detected to have pneumonia by the pneumonia detector, it is passed to the pneumonia classifier, which classifies the image into interstitial lung disease (ILD) or Covid-19. The fact that the pneumonia classifier only comes into action after the pneumonia detector has performed its job was leveraged into using a prior attention residual block. As shown in Figure 3, the prior attention residual block has one additional input other than the regular residual block, which is borrowed from the weights of the final layer of the pneumonia detection module. The prior attention residual block can get the attention weights before backpropagation takes place and they can be used to train the pneumonia classifier simultaneously. This method ensures that the classifier is trained on the right features. This model achieved an accuracy of 93.3% on the Covid-19 class, 89.4% on the ILD class, and 91.5% on the normal class. The sensitivity and specificity for normal/viral/covid-19 classes were (91.5/89.4/93.3) and (93.5/90.6/95.5), respectively.

Lai et al.[56] proposed the novel coronavirus-infected pneumonia (NCIP)-Net for the detection of Covid-19 from CT scans. The authors of NCIP-Net used a multitask DCNN for determining the presence of Covid-19 based on the entire image, Segmentation of Covid-19 lesions from the entire CT scan and determining the probability of Covid-19 from the segmented lesions. The data set used for training this model consists of 323 Covid-19-positive CT scans and 501 normal scans. Before providing the images to the model as an input, all images went through a lung lobe segmentation process where the lung region was separated from the entire image. The model is constructed like a normal encoder–decoder, but the encoder is connected to three branches. Out of those three branches, one is the decoder, which is used for lesion segmentation. The second branch from the encoder is used for the prediction of Covid-19 directly from the image. The third branch is used for determining the probability of Covid-19 based on the ROI with lesions. The training is divided into two stages. In the first stage, the second branch from the encoder is connected to three convolution layers with a residual block concatenated with a softmax function to determine the probability of Covid-19 from the image directly. Still, in the first training stage, the features encoded by the encoder are passed on to the decoder for lesion segmentation based on dice loss. In the second stage of training, CT volume patches are used as an input and the third branch extended from the encoder (C-Net) is used to identify a maximum of 10 proposals with the likelihood of lesions to predict the presence of Covid-19. The encoder can predict the proposals with the likelihood of lesions, because it was previously trained to segment lesions from the CT scan. This model achieved an accuracy of 74.4% in Covid-19/normal and 82.9% in Covid-19/other lung diseases.

Looking at all this study work, some patterns clearly stand out. The first and the most important one is to segment the lung region from the entire CT scan. This way, a lot of computation time is saved, and the model is

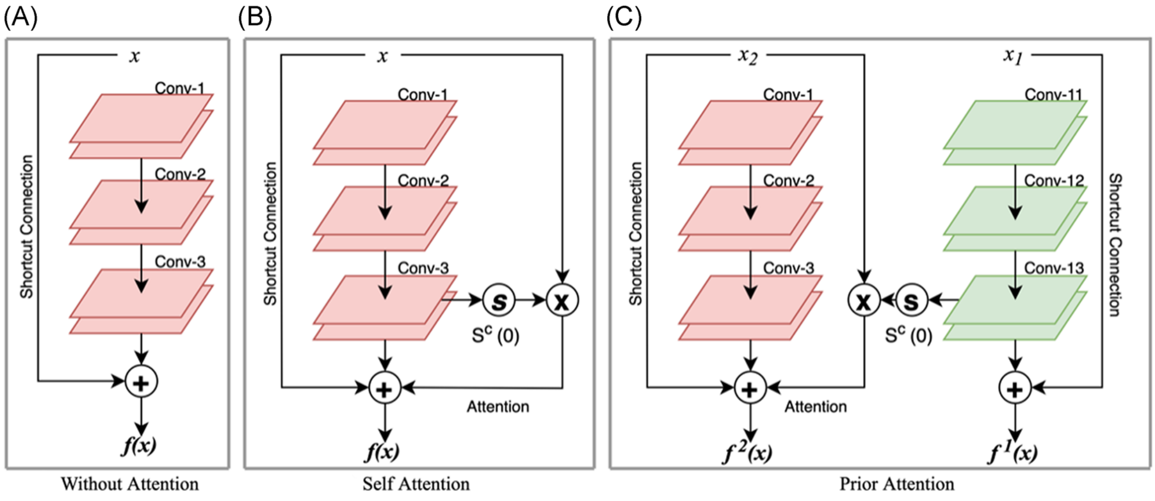

FIGURE 3 Comparison of different attention mechanisms

forced to learn features from the right region. However, the model can still learn the wrong features from the lung region. To overcome this problem, some kind of attention mechanism, online or offline, is used in all models that are proven to generalize well. Next, we move on to some research work that distinguishes pneumonia from normal cases and does not include Covid-19 cases. A separate section was not created to include the detection of pneumonia via CT scans, because not enough research has been carried on that topic. This is because detection of pneumonia is usually done with X-rays rather than with CT scans.

Wang et al.[57] proposed a multichannel multimodal deep regression framework for the screening of pneumonia from CT scans. For their model, they used 450 pneumonia-positive CT scans and 450 normal CT scans. Not only that, but they also used the complaints of those patients and their demographic information to improve the performance of their model. The entire model is divided into three parts that process demographic information, complaint information, and CT scans, respectively. Intuitively, the demographic information and the complaint information are processed with the help of an LSTM (long short term memory). The CT scans, however, are processed differently. First, three slices from the CT, namely the lung window, high attenuation, and low attenuation (LA), are extracted and concatenated into a three-channel image. This three-channel image is then passed onto an R-CNN with a base of ResNet-50. The R-CNN is an object detection module, so it detects the region of the CT scan where pneumonia is present. The features extracted from the region detected by the R-CNN are then passed on to an LSTM network. The features extracted by the R-CNN were passed on to the LSTM for two reasons. First, the authors wanted to use the three channels as a sequence of video frames that were dependent on each other. The second reason is that an LSTM was the only feasible way to concatenate the demographic and complaint information with the spatial information of CT scans. Finally, all three LSTMs are concatenated and used for pneumonia detection. This model achieved an accuracy of 94.6% in the pneumonia detection task. The sensitivity and specificity of this model were 0.933 and 0.922, respectively (Table 5).

2.5 | Localization of Covid-19 in CT scans

Wang et al.[58] presented the COPLE-Net, a noise-robust model for segmentation of Covid-19 lesions from CT images. To train their model, they used 558 CP CT images. The architecture of COPLE-Net was based on U-Net with some modifications. First, instead of using only max-pooling or average pooling for downsampling, the authors concatenated both methods, and it gave better results. Second, they modified the skip connections of U-Net by adding another layer of convolution between the encoder and the decoder. This additional layer contains half as many channels as the encoder. This layer was added to alleviate the semantic gap between the decoder's high-level features and the encoder's low-level features by forcing the encoder features to a lower dimension (half channels). Third, the authors added an atrous spatial pyramid pooling (ASPP)[44] layer at the end of the encoder. An ASPP layer contains four parallel layers of dilated convolutions with different dilation rates. This way, multiscale features can be extracted for small and large lesion segmentation.

COPLE-Net was trained using an adaptive self-ensembling technique with a noise-robust dice loss. The noise robustness in dice loss was achieved by using an mean absolute error analogous dice loss instead of the usual mean squared error analogous dice loss. To understand the self-ensembling, we must first understand which models were ensembled. The authors trained two COPLE-Nets via a teacher-student mechanism. The teacher model was an exponential moving average of the student model and was thus more stable than the student model. However, the weights of the moving average were not fixed from the beginning. If the loss of the student model was more than a defined threshold, the student model was not used to update the teacher model at all. Otherwise, the weight of the student model considered to update the teacher model was defined as a function of the loss constant (difference between the

TABLE 5 A detailed study on classification of Covid-19 and CAP via CT scans

Author	Model	Task	Data set	Accuracy
Harmon et al.[50]	AH-Net + CheXNet	CP/N/CAP	1059	89.6%
Ouyang et al.[52]	VB-Net + ResNet34	CP/CAP	1092/1094	95.4%
Wang et al.[54]	3D U-Net + ResNet	CP/N/CAP	1315/936/2406	93.3%/91.5%/89.4%
Lai et al.[56]	NCIP-Net	CP/N	323/501	74.4%
Wang et al.[57]	ResNet + LSTM	N/CAP	450/450	94.6%

Abbreviations: CAP, community-acquired pneumonia; CN, covid negative; Covid-19, coronavirus disease 2019; CP, covid positive; CT, computerized tomography; N, normal; 3D, three-dimensional.

losses) of the said models. This model was able to achieve a dice score of 0.8072% or 80.72%.

Gao et al.[59] presented a dual-branch combination network (DCN) for performing lesion segmentation and classification at once. Their data set consisted of 1918 CT scans from 1202 subjects across two hospitals. Before feeding the CT image slices into the DCN, the images underwent lung segmentation through a U-Net. These segmented lungs with a dice score coefficient of 0.99 were then used as an input to the DCN model. The model comprises two main parts, one for classification and another one for segmentation. The segmentation model is an encoder-decoder model analogous to a U-Net model. The classification model uses ResNet-50 as a backbone with lesion attention modules, as shown by brown color in Figure 4. The LA module is a combination of (the original CT slice)/(ResNet-50 downsampled slice) and the feature extracted slice of the corresponding size from the decoder of the segmentation module. A slice from the decoder module is chosen, because the decoder has more relevant features which correspond to Covid-19 lesions. Hence, the ResNet-50 classification module is forced to pay attention to features that contain Covid-19 lesions. This model was able to achieve a dice score of 0.8351% or 83.51%. The classification accuracy for internal validation (CT images from the same hospital that the model was trained on) was 96.74%, with an AUC of 0.9864, whereas the accuracy on external validation (CT images from a different hospital) was 92.87% with an AUC of 0.9771.

Zhou et al.[60] presented a three-way segmentation technique for segmentation of Covid-19 infected regions from a CT scan. The data set used by them consisted of CT scans of 120 patients. The total number of unique CT scans used is not disclosed in their paper. The authors, however, used a unique data augmentation technique to generate 200 CT scans from each unique patient. The detailed augmentation technique has not been disclosed in the paper, but the principles upon which the augmentation was based were delineated. Hence, the data set consists of about 24,000 CT scans. The authors used three-way segmentation in that they extracted *x–y*, *y–z*, and *x–z* slices from the CT scan and trained three different segmentation models to segment Covid-19 lesions from these models. This technique is analogous to how radiologists diagnose Covid-19 lesions. If a particular voxel cannot be clearly predicted as lesion or normal, radiologists often look at voxels surrounding that voxel. Similarly, if we have two-dimensional segmentations from all three axes (*x–y*, *y–z*, and *x–z*), our model can classify a voxel into lesion or normal by looking at surrounding voxels without being limited to that particular plane. This model was able to achieve a dice score of 0.783.

Fan et al.[61] presented the Inf-Net, a semisupervised deep learning model for the segmentation of Covid-19 lesions from CT scans. Their data set consisted of 50 CT scans, which aptly justifies the semisupervised learning. The architecture of Inf-Net begins with two convolution layers into which a CT scan slice is fed. The first two convolution layers extract the low-level features. In general, low-level features are known to detect edges in computer vision, so these features are passed through a simple convolution layer and compared against the ground truth segmented region to determine the edge loss. As shown in Figure 5, this edge loss is back-propagated to f2 so that f2 can learn correct edge features. Next, the features of convolution layers 3, 4, and 5 are passed on to a partial decoder, which yields a coarse global map of the region to be segmented. Only high-level features are used as an input to the partial decoder because Wu et al.[62] pointed out that low-level features are computationally intensive as compared to high-level

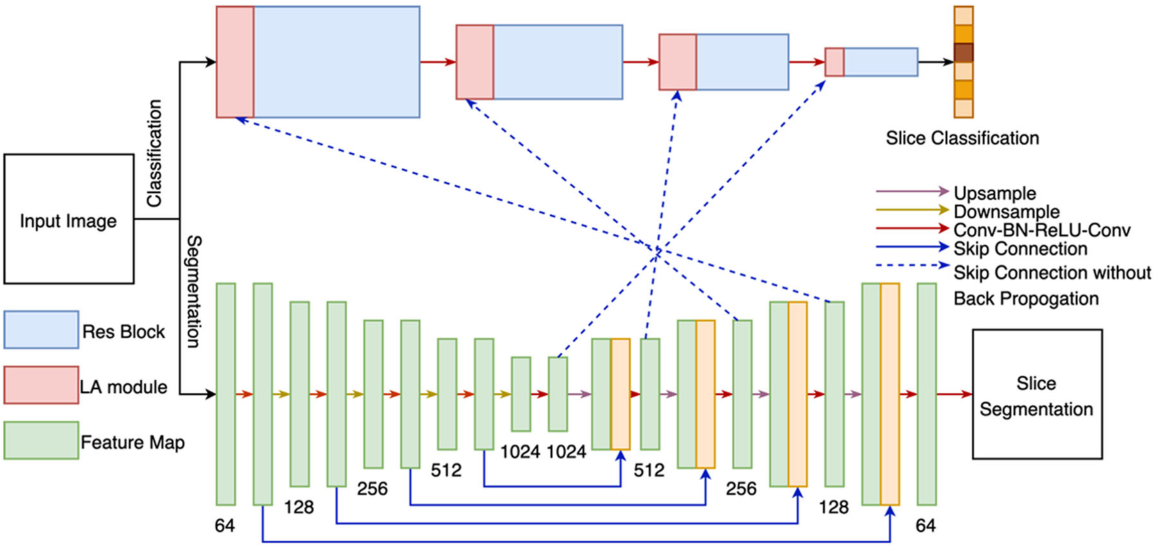

FIGURE 4 Architecture of dual-branch combination network

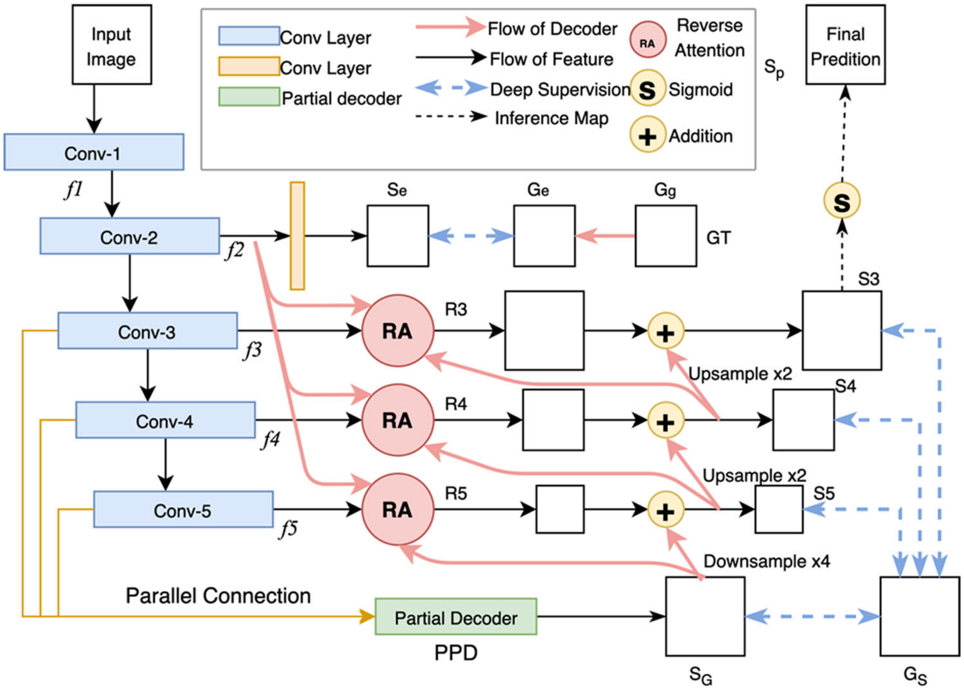

FIGURE 5 A detailed architecture of Inf-Net

features and contribute little to the process of segmentation. The global map provided by the partial decoder is labeled as coarse in that it contains an extra segmentation region that needs to be removed. Hence, a reverse attention module is used to erase the extra region from the coarse global map. The removal of this extra region is done with the help of edge features from the second convolution layer so that only the region inside the edge is preserved. Therefore, the reverse attention module takes input from both f2 and the global coarse map. Three such reverse attention modules, R3, R4, and R5, are stacked in a cascade manner such that the output of R5 is used as an input for the reverse attention module of R4 and so on. Finally, the output of R3 is followed by a sigmoid function to give the completely segmented infected region. The semisupervised learning approach of Inf-net is progressively enlarging the data set. This process is performed by predicting some labels from the limited training data and then using the predicted labels as the training data and the original training data. This process is repeated for a while until enough training data is gathered. The Inf-Net achieved a dice score of 0.739 on their data set and a dice score of 0.597 on a different data set.

Yang et al.[63] presented a unique approach for the localization of Covid-19 lesions in CT slices. The idea was to train a Generator Network, which would output normal (without Covid-19) slices even if the corresponding input slice had Covid-19 lesions. Afterward, the output slices could be subtracted from the input slices to localize the regions where Covid-19 lesions were present. The generator model was trained against a discriminator model, which tried to distinguish between real and generated normal pneumonia images. Moreover, a ResNet-18 was also trained on Covid-19-positive images so that the ResNet could grasp the low-level features and concatenate those features with the encoder of the generator network. This was done because the generator network itself was not powerful enough to grasp the low-level features of a CT slice. Finally, both normal and Covid-19-positive CT slices are provided to the generator model, but the loss is only calculated against normal images. In this way, the generator is forced to generate normal CT slices even from the Covid-19 lesion containing CT slices. This is analogous to a denoising autoencoder where noisy images are passed into the auto-encoder, but the loss is calculated against noise-less images. A major benefit of using this model is that it is weakly supervised. Hence, while training the generator, labeled image pairs are not necessarily required. This model achieved a dice score of 0.575, which is very competitive for weakly supervised models. However, fully supervised models have a much higher dice score (Table 6).

3 | CHALLENGES AND FUTURE SCOPE

The end goal of all research into automatic pneumonia/Covid-19 detection and localization is to have a model that can be used in (hospitals)/(chest X-ray centers)/(CT

TABLE 6 A study on localization of Covid-19 in CT scans

Author	Model	Type	Data set	DSC
Wang et al.[58]	COPLE-Net (Modified U-Net + ASE)	Fully supervised	558 Scans	0.8072
Gao et al.[59]	DCN (Modified U-Net + LA + ResNet)	Fully supervised	1918 Scans	0.8351
Zhou et al.[60]	U-Net (X–Y, Y–Z, X–Z axes segmentation)	Semisupervised	120 Patients	0.783
Fan et al.[61]	Inf-Net (Custom CNN + RA + PD)	Semisupervised	50 Scans	0.594
Yang et al.[63]	GAN + ResNet	Semisupervised	1252 Scans	0.575

Abbreviations: ASE, adaptive self-ensembling; Covid-19, coronavirus disease 2019; CT, computerized tomography; DSC, dice score coefficient; LA, lesion attention; PD, partial decoder; RA, reverse attention.

scan centers) on an everyday basis. For a single model to be used in different centers worldwide, the model should be able to generalize well to different CT scan/X-ray machines and different demographics.

This poses the problem of collecting a data set that contains such a wide variety of data. Although the problem of overfitting to a particular data set has been mitigated by attention mechanisms, Grad-CAM analysis, adversarial training, and segmentation-before-classification, this kind of work needs to be applied to a more distributed data set so that it can learn correct features from any chest X-ray/CT scan around the world without the need of tedious preprocessing. Hence, the first future scope would be to collect a data set with a wide variety of chest X-rays/CT scans, especially for Covid-19 classification.

Preprocessing an image of a chest-X-ray/CT-scan before using it as an input for a deep learning model poses another challenge. As most image preprocessing is dependent on the type of image. For example, chest X-rays taken on machine A would require a different kind of image preprocessing mechanism than a chest X-ray taken on machine B. Hence, another future scope would be creating deep learning models, which require little to no data-dependent preprocessing.

In this study, a lot of different research that tackles different problems has been illustrated. Although no single work tackles all challenges, a smart combination of some practices used in the mentioned research might yield a truly generalizable model. Furthermore, several small, custom data sets were compiled by different authors for their research. Combining those data sets or even using semisupervised domain adversarial training with different data sets would generalize the corresponding deep learning model better.

Practical application of research in such deep learning models might be restricted to assisting doctors in making a better diagnosis instead of working in complete autonomy. Keeping such applications in mind, deep learning models can be modified to output a prediction highlighting the most important features based on which the prediction was made. This way, doctors might get help if they miss some features in the image which are not apparent to the naked eye.

4 | CONCLUSION

The process for automating the detection of pneumonia from chest X-rays and CT scans has evolved a lot over the past few years, especially with the advent of deep learning methods. Looking back at the past 4 years, base deep learning model architectures have evolved a lot. However, base model architectures are not the most effective solutions for the specific task of pneumonia detection. The pioneering models that achieved good metrics on pneumonia detection tasks tweaked the architectures of base models so that the tweaked models were a better fit for the task of pneumonia detection. The models that followed these pioneering models were focused on generalizing the model architecture. This generalization was achieved through techniques such as adversarial training, Grad-CAM analysis, attention mechanisms, and many more.

The task of classifying Covid-19 from chest X-rays and CT scans is not very different from the pneumonia detection task. However, research into Covid-19 detection through deep learning models is relatively new, because Covid-19 is a relatively new disease (as of 2021). Because of the time gap, the models made for detecting Covid-19 from pneumonia use better base model architectures than those initially used in pneumonia detection. However, the techniques used to make the base models more effective toward the specific task of Covid-19 detection are similar to the techniques used for the pneumonia detection task, both for higher metrics and better generalization. This observation leads us to an important inference. The inference would be that those techniques, which make base model architectures more effective or more generalizable for a specific task (pneumonia detection) are at least as important if not more important than the base models.

Even as base model architectures keep improving, the techniques discussed in this paper can always be applied to the improved base models to further improve the base models' generalizability and effectiveness. With that thought, many different techniques and architecture tweaks, along with their merits, demerits, and tradeoffs, have been explained in this paper. A quantitative analysis table that corresponds to each section of the paper is also provided so that the readers can corelate between the qualitative and quantitative results of different models and techniques. With both qualitative and quantitative analysis, this paper can be a one-stop solution for aspiring researchers who want to study the field of pneumonia/Covid-19 detection in depth. Lastly, this paper serves as a means of initiating and propagating new research in the field of automatic pneumonia/Covid-19 detection and localization by providing a wide breadth of techniques along with enough depth in every technique so as to guide aspiring researchers in the right direction for their specific purpose.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

REFERENCES

1. Franquet T. Imaging of pneumonia: trends and algorithms. *Eur Respir J*. 2001;18:196-208. doi:10.1183/09031936.01.00213501
2. Kaymak S, Serener A. Automated age-related macular degeneration and diabetic macular edema detection on OCT images using deep learning: IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE. 2018:265-269.
3. Shi J, Zheng X, Li Y, Zhang Q, Ying S. Multimodal neuroimaging feature learning with multi-modal stacked deep polynomial networks for diagnosis of Alzheimer's disease. *IEEE J Biomed Heal Informatics*. 2018;22:173-183.
4. Kaymak S, Esmaili P, Serener A. Deep learning for two-step classification of malignant pigmented skin lesions: 14th Symposium on Neural Networks and Applications (NEURAL). IEEE. 2018:1-6.
5. Serte S, Serener A. A generalized deep learning model for glaucoma detection: 3rd International Symposium on Multidis-Ciplinary Studies and Innovative Technologies (ISMSIT). IEEE. 2019:1-5.
6. Rajpurkar P, Irvin J, Zhu K, et al. *CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning*. 2017. http://arxiv.org/abs/1711.05225
7. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017;60:84-90. doi:10.1145/3065386
8. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2016. doi:10.1109/CVPR.2016.91
9. Lin T-Y, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE; 2017. doi:10.1109/ICCV.2017.324
10. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Commun ACM*. 2020;63(11). doi:10.1145/3422622
11. Zhang J, Xie Y, Pang G, et al. *Viral pneumonia screening on chest X-ray images using confidence-aware anomaly detection*. 2020. http://arxiv.org/abs/2003.12338
12. Drosten C, Kellam P, Memish ZA. Evidence for camel-to-human transmission of MERS coronavirus. *N Engl J Med*. 2014;371:1359-1360. doi:10.1056/NEJMc1409847
13. Li W, Moore MJ, Vasilieva N, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*. 2003;426:450-454.
14. Li Y, Zhang Z, Dai C, Dong Q, Badrigilan S. Accuracy of deep learning for automated detection of pneumonia using chest X-ray images: a systematic review and meta-analysis. *Comput Biol Med*. 2020;123:103898. doi:10.1016/j.compbiomed.2020.103898
15. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chest X-Ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2017. doi:10.1109/CVPR.2017.369
16. Deng J, Dong W, Socher R, Li LJ, Kai L, Li FF. ImageNet: a large-scale hierarchical image database: IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2009. doi:10.1109/CVPR.2009.5206848
17. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2017. doi:10.1109/CVPR.2017.243
18. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med*. 2018;15:e1002683. doi:10.1371/journal.pmed.1002683
19. Janizek JD, Erion G, DeGrave AJ, Lee S-I. *An adversarial approach for the robust classification of pneumonia from chest radiographs*. ACM CHIL 2020 - Proc 2020 ACM Conf Heal Inference, Learn. 2020:69-79. doi:10.1145/3368555.3384458
20. Liang C, Li Y, Luo J. Multiobjective evolutionary design of deep convolutional neural networks for image classification. *IEEE Trans Evol Comput*. 2021;25:277-291. doi:10.1109/TEVC.2020.3024708
21. Li Z, Yu J, Li X, et al. PNet: an efficient network for pneumonia detection: 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE; 2019. doi:10.1109/CISP-BMEI48845.2019.8965660
22. Dong Y, Wu M, Zhang J. Recognition of pneumonia image based on improved quantum neural network. *IEEE Access*. 2020;8:224500-224512. doi:10.1109/ACCESS.2020.3044697
23. Khalifa NEM, Taha MHN, Hassanien AE, Elghamrawy S. *Detection of coronavirus (COVID-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest X-ray dataset*. 2020:1-15. http://arxiv.org/abs/2004.01184
24. Dey N, Zhang YD, Rajinikanth V, Pugalenthi R, Raja NSM. Customized VGG19 architecture for pneumonia detection in chest X-rays. *Pattern Recognit Lett*. 2021;143:67-74. doi:10.1016/j.patrec.2020.12.010
25. Haghanifar A, Majdabadi MM, Choi Y, Deivalakshmi S, Ko S. *COVID-CXNet: detecting COVID-19 in frontal chest X-ray images using deep learning*. 2020. http://arxiv.org/abs/2006.13807
26. Gu X, Pan L, Liang H, Yang R. Classification of bacterial and viral childhood pneumonia using deep learning in chest radiography. *ACM Int Conf Proc Ser*. 2018:88-93. doi:10.1145/3195588.3195597
27. Long J, Shelhamer E, Darrell T Fully convolutional networks for semantic segmentation: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2015. doi:10.1109/CVPR.2015.7298965
28. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal visual object classes (VOC) challenge. *Int J Comput Vis*. 2010;88:303-338. doi:10.1007/s11263-009-0275-4
29. Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. *36th Int Conf Mach Learn ICML*. 2019:10691-10700.
30. Wang Z, Xiao Y, Li Y, et al. Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays. *Pattern Recognit*. 2021;110:107613. doi:10.1016/j.patcog.2020.107613
31. Wang F, Jiang M, Qian C, et al. Residual attention network for image classification: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2017. doi:10.1109/CVPR.2017.683

32. Arias-Londono JD, Gomez-Garcia JA, Moro-Velazquez L, Godino-Llorente JI. Artificial intelligence applied to chest X-ray images for the automatic detection of COVID-19. A thoughtful evaluation approach. *IEEE Access*. 2020;8:226811-226827. doi:10.1109/ACCESS.2020.3044858

33. Wang L, Lin ZQ, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Sci Rep*. 2020;10:19549. doi:10.1038/s41598-020-76550-z

34. Sakib S, Tazrin T, Fouda MM, Fadlullah ZM, Guizani M. DL-CRC: deep learning-based chest radiograph classification for Covid-19 detection: a novel approach. *IEEE Access*. 2020;8:171575-171589. doi:10.1109/ACCESS.2020.3025010

35. Ali G, Shahin A, Elhadidi M, Elattar M. Convolutional neural network with attention modules for pneumonia detection. *2020 Int Conf Innov Intell Informatics, Comput Technol 3ICT 2020*. 2020;13:0-5. doi:10.1109/3ICT51146.2020.9311985

36. Ohata EF, Bezerra GM, Souza das Chagas JV, et al. Automatic detection of COVID-19 infection using chest X-ray images through transfer learning. *IEEE/CAA J Autom Sin*. 2021;8:239-248. doi:10.1109/JAS.2020.1003393

37. Chowdhury M, Rahman T, Khandakar A, et al. Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access*. 2020;8:132665-132676. doi:10.1109/ACCESS.2020.3010287

38. Pan I, Cadrin-chênevert A, Cheng PM. Tackling the radiological society of North America pneumonia detection challenge. *AJR Am J Roentgenol*. 2019;213:568-574.

39. Li B, Kang G, Cheng K, Zhang N. Attention-guided convolutional neural network for detecting pneumonia on chest X-rays. *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS*. 2019;2019:4851-4854. doi:10.1109/EMBC.2019.8857277

40. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell*. 2020;42:2011-2023. doi:10.1109/TPAMI.2019.2913372

41. Gabruseva T, Poplavskiy D, Kalinin A. Deep learning for automatic pneumonia detection. *IEEE Comput Soc Conf Comput Vis Pattern Recognit Work*. 2020;2020:1436-1443. doi:10.1109/CVPRW50498.2020.00183

42. Yao S, Chen Y, Tian X, Jiang R. GeminiNet: combine fully convolution network with structure of receptive fields for object detection. *IEEE Access*. 2020;8:60305-60313. doi:10.1109/ACCESS.2020.2982939

43. Dai J, Li Y, He K, Sun J. R-FCN: Object detection via region-based fully convolutional networks. *Adv Neural Inf Process Syst*. 2016:379-387.

44. Liu S, Huang D, Wang Y. Receptive field block net for accurate and fast object detection. *Lect Notes Comput Sci*. 2018;45(11215):404-419. doi:10.1007/978-3-030-01252-6_24

45. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell*. 2018;40:834-848. doi:10.1109/TPAMI.2017.2699184

46. The DeepRadiology Team. 2018. *Pneumonia detection in chest radiographs*. http://arxiv.org/abs/1811.08939

47. Zhu Y, Zhao C, Wang J, Zhao X, Wu Y, Lu H. CoupleNet: coupling global structure with local parts for object detection: IEEE International Conference on Computer Vision (ICCV). IEEE; 2017. doi:10.1109/ICCV.2017.444

48. Sirazitdinov I, Kholiavchenko M, Mustafaev T, Yixuan Y, Kuleev R, Ibragimov B. Deep neural network ensemble for pneumonia localization from a large-scale chest X-ray database. *Comput Electr Eng*. 2019;78:388-399. doi:10.1016/j.compeleceng.2019.08.004

49. Ko H, Ha H, Cho H, Seo K, Lee J. Pneumonia detection with weighted voting ensemble of CNN models 2019: 2nd Int Conf Artif Intell Big Data, ICAIBD 2019. 2019;306-310. doi:10.1109/ICAIBD.2019.8837042

50. Harmon SA, Sanford TH, Xu S, et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat Commun*. 2020;11:1-7. doi:10.1038/s41467-020-17971-2

51. Liu S, Xu D, Zhou SK, et al. 3D anisotropic hybrid network: transferring convolutional features from 2D images to 3D anisotropic volumes. *Lect Notes Comput Sci*. 2018;11071:851-858. doi:10.1007/978-3-030-00934-2_94

52. Ouyang X, Huo J, Xia L, et al. Dual-sampling attention network for diagnosis of COVID-19 from community-acquired pneumonia. *IEEE Trans Med Imaging*. 2020;39:2595-2605. doi:10.1109/TMI.2020.2995508

53. Shan F, Gao Y, Wang J, et al. Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction. *Med Phys*. 2021;48:1633-1645. doi:10.1002/mp.14609

54. Wang J, Bao Y, Wen Y, et al. Prior-attention residual learning for more discriminative COVID-19 screening in CT images. *IEEE Trans Med Imaging*. 2020;39:2572-2583. doi:10.1109/TMI.2020.2994908

55. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W, eds. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016*, 2016:424-432. doi:10.1007/978-3-319-46723-8_49

56. Lai Y, Li G, Wu D, et al. 2019 novel coronavirus-infected pneumonia on CT: a feasibility study of few-shot learning for computerized diagnosis of emergency diseases. *IEEE Access*. 8, 2020:194158-194165. doi:10.1109/ACCESS.2020.3033069

57. Wang Q, Yang D, Li Z, Zhang X, Liu C. Deep regression via multi-channel multi-modal learning for pneumonia screening. *IEEE Access*. 2020;8:78530-78541. doi:10.1109/ACCESS.2020.2990423

58. Wang G, Liu X, Li C, et al. A Noise-Robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images. *IEEE Trans Med Imaging*. 2020;39:2653-2663. doi:10.1109/TMI.2020.3000314

59. Gao K, Su J, Jiang Z, et al. Dual-branch combination network (DCN): towards accurate diagnosis and lesion segmentation of COVID-19 using CT images. *Med Image Anal*. 2021;67:101836. doi:10.1016/j.media.2020.101836

60. Zhou L, Li Z, Zhou J, et al. A rapid, accurate and machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis. *IEEE Trans Med Imaging*. 2020;39:2638-2652. doi:10.1109/TMI.2020.3001810

61. Fan DP, Zhou T, Ji GP, et al. Inf-Net: automatic COVID-19 lung infection segmentation from CT images. *IEEE Trans Med Imaging*. 2020;39:2626-2637. doi:10.1109/TMI.2020.2996645

62. Wu Z, Su L, Huang Q. Cascaded partial decoder for fast and accurate salient object detection. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2019;2019:3902-3911. doi:10.1109/CVPR.2019.00403

63. Yang Z, Zhao L, Wu S, Chen CYC. Lung lesion localization of COVID-19 from chest CT image: a novel weakly supervised learning method. *IEEE J Biomed Heal Informatics*. 2021;25:1864-1872. doi:10.1109/JBHI.2021.3067465

How to cite this article: Shah A, Shah M. Advancement of deep learning in pneumonia/Covid-19 classification and localization: a systematic review with qualitative and quantitative analysis. *Chronic Dis Transl Med*. 2022;8:154-171. doi:10.1002/cdt3.17