



INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv



Using Various Convolutional Neural Network to Detect Pneumonia from Chest X-Ray Images: A Systematic Literature Review

Darnell Kikoo ^{a,*}, Bryan Tamin ^a, Stephen Hardjadilaga ^a, Anderies ^a, Irene Anindaputri Iswanto ^a

^a Computer Science Department, School of Computer Science, Bina Nusantara University Jakarta, Indonesia 11480, Indonesia

Corresponding author: *darnell.kikoo@binus.ac.id

Abstract—Pneumonia is one of the world's top causes of mortality, especially for children. Chest X-rays serve an important part in diagnosing pneumonia due to the cost-effectiveness and quick advancement of the technology. Detecting Pneumonia through Chest X-rays (CXR) is a challenging and time-consuming process requiring trained professionals. This issue has been solved by the development of automation technology which is machine learning. Moreover, Deep Learning (DL), a machine learning specification that uses an algorithm that resembles the human brain, can predict more accurately and is now dependable enough to predict pneumonia. As time passes, another Deep Learning improvement has been made to produce a new method called Transfer Learning, that is done by extracting specific layers from some pre-trained network to be used on other datasets, which reduces the training time and improves the model performance. Although numerous algorithms are already available for pneumonia identification, a comprehensive literature evaluation and clinical recommendations are still small in numbers. This research will assist practitioners in choosing some of the best procedures from the recent research, reviewing the available datasets, and comprehending the outcomes gained in this domain. The reviewed papers show that the best score for predicting pneumonia using DL from CXR was 99.4% accuracy. The exceptional techniques and results from the reviewed papers served as great references for future research.

Keywords—Pneumonia; lung diseases; X-rays; CXR; radiography; deep learning; convolutional neural network; CNN; classification.

Manuscript received 5 Jul. 2022; revised 30 Oct. 2022; accepted 29 Dec. 2022. Date of publication 30 Jun. 2023. International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

Pneumonia is the major cause of death globally, contributing to around 16% of all deaths of children under five years worldwide [1], the world's leading cause of death among young children [2]. In Indonesia itself, it ranks second as the cause of the toddler's death. In 2019 only, it was reported that there were more than 400,000 pneumonia cases in this country [3]. Pneumonia is an illness that causes the air sacs in one or both lungs to become inflamed. The air sacs may get filled with fluid or pus, resulting in a cough with phlegm or pus, fever, chills, and trouble breathing. Pneumonia can be caused by several species, including bacteria, viruses, and fungi [4].

Since several species can cause pneumonia, the first word indicates the species that can cause pneumonia. For example, bacterial pneumonia is caused by bacteria, a virus causes viral pneumonia, and fungal pneumonia is pneumonia usually caused by breathing certain fungal spores [5].

The early diagnosis of pneumonia may start with a simple lung check with a stethoscope; if the lung is suspected of pneumonia, additional tests might be required. In clinics,

many imaging modalities such as magnetic resonance imaging (MRI), computed tomography, and CXR are used to diagnose pneumonia (CT). CXR is the most often used method to identify pneumonia globally due to its low cost and ease of access [6].

Chest radiographs give valuable information about a patient's status; yet even highly experienced radiologists have difficulty recognizing pneumonia with CXRs since these pictures contain comparable opacities for other lung diseases such as lung cancer and excess fluid. Therefore, traditional pneumonia identification by CXR pictures is time consuming and imprecise, potentially delaying the diagnosis and treatment process. As a result of the critical need for early complexity of interpreting these images, computer-aided detection (CAD) systems represent an exciting approach for automated detection that can assist physicians in overcoming the aforementioned issues and improving detection accuracy in a clinical setting [7].

CAD systems can complement medical staff decision-making by combining parts of computer vision and machine learning (ML) with images such as radiographical images to

detect and bring out patterns [8]. A typical CAD system consecutively analyzes input data (i.e., CXRs), extracting and classifying the characteristics. The first stage is to preprocess the CXR data. The second stage extracts feature from the input photos using Gaussian filters, morphological operations, and edge detection. The collected characteristics are then classified in the third stage using an appropriate classifier, such as a support vector machine (SVM), random forest (RF) method, or neural network. In the years that followed, Machine Learning (ML) and Deep Learning (DL) accelerated their growth. Compared to ML, DL requires more data; however, it is more humanlike and able to predict by itself without much human intervention [9].

One of the DL models, Artificial Neural Networks (ANN), make a comeback for their name in the 2010s. Even though it was first created in 1943, ANN that Google creates can still accomplish amazing results by beating the Go board game world champion [9]. This event is recorded as one of the most notable performances in artificial intelligence and expedited AI progress in research, especially ANN and DL, which are now used for various things, such as processing sensor data, recognizing images, speech, and many more. Unsurprisingly, AI, particularly deep learning, and its use in medicine are advancing quickly in recent years, owing to CNNs' capacity to identify pictures correctly. Nowadays, artificial intelligence is employed in a variety of medical specialties, and one example is radiology. AI and machine learning are generally utilized in domains that need a huge collection of medical data, typically in the form of digital pictures, that can be used to train models later. The application of artificial intelligence in support systems for medical image processing has been recommended to increase reporting accuracy, consistency, and time efficiency [9]. Therefore, in this systematic literature review, we aimed to evaluate various methods and datasets connected to Convolutional Neural Network using Chest X-ray images.

II. MATERIALS AND METHOD

A. Review Method

Digital libraries can be used to do preliminary searches for primary studies, but this is insufficient for a comprehensive assessment. Other sources of evidence must be looked into, such as journals and conference papers [10]. We can use certain keywords that correlate with the title or content. The search string is ("pneumonia" OR "lung diseases") AND ("X-rays" OR "CXR" OR "radiography") AND ("deep learning" OR "convolutional neural network" OR "CNN" OR

"classification"). Based on the primary studies collected, we will create the research questions that drive the entire systematic review methodology [10]. To ease the process of data extraction and data analysis, we created the following research questions:

- **RQ1:** Which are the most used Chest X-Ray Image dataset for pneumonia detection?
- **RQ2:** What data preprocessing methods are commonly used for pneumonia detection?
- **RQ3:** What combination of techniques improves the performance of a Deep Convolutional Neural Network?

To filter the primary studies that have been collected before, we created an inclusion and exclusion criteria table, as shown in Table 1, that is relevant to our research objective.

TABLE I
INCLUSION AND EXCLUSION CRITERIA

Inclusion	Exclusion
Studies that use Chest X-Ray images dataset	Studies that do not use CNN for Pneumonia Detection
Studies that are published in a credible conference or journal	Studies that are not written in English
Studies that are published between 2015 - 2022 inclusively	Studies in which the suggested approach has not been validated
Studies that are written in English	
Studies that use CNN for Pneumonia Detection	

B. Review Analysis

After collecting the papers, we assess the eligible papers based on our research questions. Then, we compare the methodologies utilized to evaluate the advantages and disadvantages. The following is our analysis of the publications we read.

1) RQ1: Which is the Most used Chest X-Ray Image Dataset for Pneumonia Detection?

Based on Table 1, it can be concluded that from the paper we have reviewed before, Kermanshah's Chest X-Ray Images dataset [26] is the most used dataset, with 15 papers using the dataset. Then the second most used dataset is the Kaggle RSNA Pneumonia Detection dataset, with five papers that use the dataset. The third most used dataset is Chest X-Ray14 (NIH) dataset, with three papers that use the dataset, and the fourth most used dataset is the JSRT-complied dataset, with two papers that use the dataset. Below are the details for the four most-used datasets, as shown in Table 2.

TABLE II
DATASET COMPARISON TABLE

Reference	Dataset	Count	Total Images	Num of Classes	Classes
[11]–[25]	Kermanshah's Chest X-Ray Images for Classification dataset / Chest X-Ray pneumonia image Kaggle dataset [26]	15	5,856	3	viral pneumonia, bacterial pneumonia, normal lung
[27]–[31]	Kaggle RSNA Pneumonia Detection Challenge dataset	5	25,684	9	pneumonia and non-pneumonia
[32]–[34]	Chest X-Ray 14 (NIH) dataset	3	112,120	15	atelectasis, consolidation, infiltration, pneumothorax, edema, emphysema, fibrosis, effusion, pneumonia, pleural thickening,

					cardiomegaly, nodule, mass, hernia, and No Findings
[35], [36]	JSRT-complied dataset	2	247	2	pneumonia and non-pneumonia
[17]	Coheny's dataset	1	646	5	viral pneumonia, bacterial pneumonia, fungal pneumonia, lipoid aspiration pneumonia, and unknown
[37]	Private dataset	1	5,216	2	pneumonia infected (bacterial or/and viral pneumonia) and non-pneumonia
[32]	Indiana University (Open-I) dataset [38]	1	7,470	5	pulmonary edema, cardiac hypertrophy, pleural effusion, and pulmonary opacity, normal
[32]	MSH (Mount Sinai Hospital) dataset	1	48,915	9	pneumonia, emphysema, effusion, consolidation, nodule, atelectasis, edema, cardiomegaly, and hernia
[36]	MC dataset	1	138	2	normal and tuberculosis (TB)
[39]	GitHub and Kaggle combined dataset	1	4,563	3	covid-19, non-covid pneumonia, and Normal/Other Diseases

- Kermany's Chest X-Ray Images dataset from Kaggle.

Kermany, Zhang, and Goldbaum [26] studied "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification," which contains around 1.2 GB of data. There are three files containing the test data, train data, and validation data, which have two subfolders for each image category. There are 5,863 X-Ray images (JPEG) classified as normal, bacterial pneumonia, and viral pneumonia.

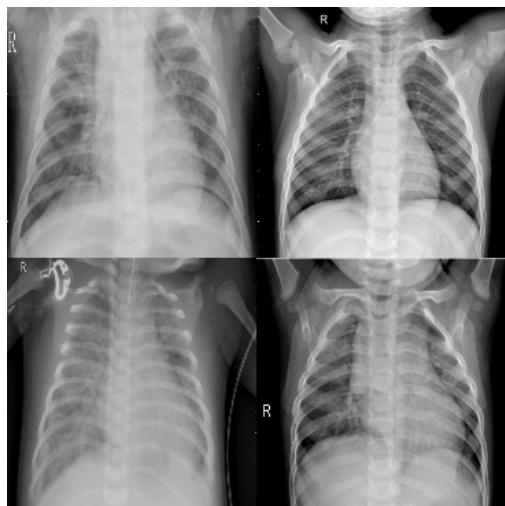


Fig. 1 Sample Images from Kermany's Dataset

- ChestX-ray14 Dataset.

Wang et al [6] made this dataset freely available on Kaggle. It comprises 112,120 X-ray images of the frontal chest taken from 30,805 people. Each radiography image in the collection is associated with one or more of fourteen different thoracic diseases. Natural Language Processing (NLP) generated these labels by text-mining disease categorization from the related radiological reports. They are thought to be more than 90% accurate. Apart from being referred to as ChestX-ray14 [31], [33], this dataset is sometimes referred to as the National Institute of Health (NIH) dataset.[30], [34], since this dataset is available from the National Institutes of Health. Due to the dataset's vastness, it frequently contains subsets. For instance, have a look at the RSNA pneumonia detection dataset [29]–[31].

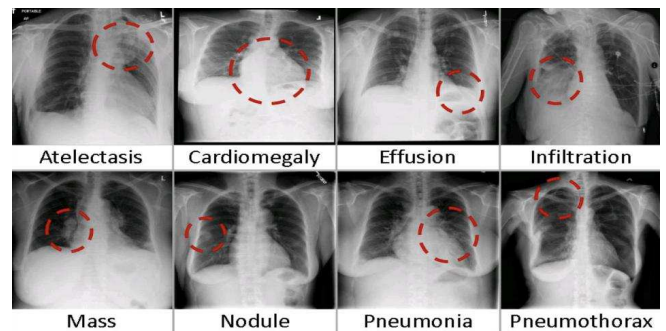


Fig. 2 Sample Images from ChestX-ray14 Dataset

- RSNA Pneumonia Detection Dataset

The training set consists of about 25,684 chest x-ray images, and the test set has 1,000 images, each with a resolution of 1024 x 1024. The photographs are grouped according to the conclusion of their diagnosis. It's worth mentioning that the dataset includes a substantially greater proportion of pneumonia-negative images, and this dataset is the subset of ChestX-ray8 dataset [6].

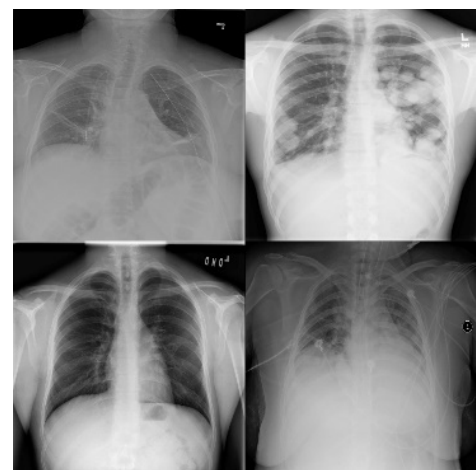


Fig. 3 Sample Images from RSNA Dataset

- JSRT-complied Dataset

The Japanese Society of Radiological Technology (JSRT) created the dataset. It contains 247 Chest X-Ray images, 154 of which include pulmonary nodules (100 malignant and 54 benign), and 93 do not. Each X-ray picture has a resolution of 2048 x 2048 pixels, while the grayscale image has a color depth of 12 bits.

- Coheny Dataset.

The dataset was created by Cohen et al. [40], consisting of 1500 images with five different classes. The classes consist of viral pneumonia, bacterial pneumonia, fungal pneumonia, lipoid aspiration pneumonia, and unknown. The total number of images in the dataset is 646. There are 506 Viral Pneumonia images, 46 Bacterial Pneumonia images, 26 Fungal Pneumonia images, and 9 Lipoid Aspiration Pneumonia images; the rest is unknown.

- Indiana University (Open-I) Dataset.

Gathered by Demner-Fushman et al. [38], Indiana University (IU) Dataset was collected from numerous hospitals associated with the IU School of Medicine. This dataset consists of 7,470 chest radiographs with 512 x 512 pixels and 3,955 corresponding reports. The photos show diseases such as pulmonary edema, cardiac hypertrophy, pleural effusion, and pulmonary opacity which are labeled with the perspective (frontal or lateral). However, this dataset only contains 40 Pneumonia sample images.

- MSH (Mount Sinai Hospital) Dataset.

From 2009 to 2016, Mount Sinai Hospital approved and collected a total of 42,936 frontal and 6,519 lateral chest radiographs combined into 48915 total images, consisting of 18,993 females and 29,922 males. The data did not contain labels, but it had patient identifiers. The total proportion of diseases includes Pneumonia 34.2%, Emphysema 3.1%, Effusion 46.1%, Consolidation 59.7%, Nodule 1.3%, Atelectasis 39.4%, Edema 16.9%, Cardiomegaly 33.7%, Hernia 0.5%. In addition, 34.2% of the data are pneumonia samples, with 18,993 being females. Nonetheless, this dataset is not available to the public, Zahi Fayad, can be reached at zahi.fayad@mssm.edu if researchers are interested in gaining access to Mount Sinai data via the Imaging Research Warehouse Initiative [41].

- MC Dataset.

The MC Dataset was compiled in collaboration with the Department of Health and Human Services in Montgomery County, Maryland, USA. There are 138 PNG frontal X-ray images in total, including 58 TB manifestations and 80 normal cases. It also includes information such as the patient's gender, age, and any lung anomalies. It contained 63 data from males, 74 from females, and one unknown gender. The resolution is either 4892 x 4020 or 4020 x 4892 pixels. However, this dataset does not contain any Pneumonia samples; hence, it is not relevant for Pneumonia detection [42].

2) *RQ2: What data preprocessing methods are commonly used for pneumonia detection?*

In deep learning, some preprocessing methods must be inserted into the model. All these data preprocessing combinations are believed to improve the model performance and generalize the CNN to the real dirty data (See Table 3). A result mentioned that preprocessing is able to improve recognition accuracy. This means the model will be able to focus on the features of each image, increasing the weight on the specific part which is infected by pneumonia. Below is some explanation about the common preprocessing different researchers use.

TABLE III
PREPROCESSING TECHNIQUE COMPARISON TABLE

Reference	Preprocessing Technique	Count
[11], [12], [20], [21], [28], [32], [34], [36], [39]	Data Augmentation	9
[14], [18], [19], [23]–[25], [29], [37]	Data Augmentation + Image Preprocessing	8
[13], [15], [16], [21], [22], [33], [35]	Image Preprocessing	7
[11], [17], [27]	Without Data Preprocessing	3
[31]	Image Preprocessing + Data Sorting	1

- Data Augmentation

The term "data augmentation" refers to a collection of strategies for artificially increasing the volume of data by producing additional data points from current data. This may be accomplished by making minor adjustments to existing data or by employing deep learning models to produce new data points. It is projected that combining data augmentation technology and deep learning algorithms would improve detection accuracy. This preprocessing algorithm performs vertical and horizontal flips, random degree rotation, random brightness, gamma adjustment, random Gaussian noise, and blur on the positive sample image. With adding supplementary regularization and generalization, this specific improvement method may be used to strengthen the training process of our network.

Data augmentation techniques are widely utilized to improve the performance of deep learning systems. The purpose of data augmentation is to increase the performance of the convolutional neural network model that is used to classify chest x-rays [28]. Images are usually resized before they get augmented. Other than generalizing the model, it can also help prevent the model from overfitting training data since the real-world CXR images are dirty and still limited. Here are some of the most common data augmentation techniques used in general computer vision classification, which can be classified into position and color augmentation. Some common data augmentation techniques that are used are scaling, geometric transformations, flipping, shearing, and rotation.

- Image Preprocessing

There are several forms of image preprocessing, but in general, the goal of image preprocessing is to improve the clarity of the image or to rescale it [13], [14], [33], [37], [15], [16], [18], [19], [21], [23]–[25] so that the model can accept it and projected to increase training efficacy and have a good influence on the prediction outcomes. When working with Chest X-Ray radiograph, A low-quality radiograph gives less information on the patient's condition and might be the greatest obstacle to an accurate interpretation [31]. Therefore, Filters [35] or image enhancement [29], [31] can be used to improve the photographs, and this enhancement can enrich the network with the first-stage data set, where the distribution of the first-stage training process serves as an extra regularization and generalization technique [29]. Additionally, there are approaches such as cropping photographs to eliminate superfluous data [9]. For some cases

in pneumonia detection, lung parts in the CXR have been segmented away [36] so that it throws away all the noise in the image. Finally, the model can focus on extracting the features directly from the chest x-ray images. Filters such as the Sobel filter and the Laplacians filter can be used to process the image. Those two filters are known as edge detection filters that remove the dimension of the information of the pictures by using contour extraction and then emphasize the feature quantity [35].

- Data Sorting

The data sorting algorithm classifies photos into low- and high-quality categories. The subjective quality of a photograph governs the resources and processing required for accurate analysis and so acts as a major initial impediment to an expedited workflow. One way to reduce the time spent on corrective actions is to create sorting automation while

utilizing a simple convolutional neural network and some images already presorted [31].

3) *RQ3: What combination of techniques improves the performance of various Convolutional Neural Network?*

From Table 4, it is apparent that researchers use Transfer Learning in their method. Many researchers have used transfer learning to get certain layers from some pre-trained neural networks to be used on other datasets, which are proven to reduce the time needed and improve the neural network's performance. This idea lies in the fact that we can utilize the information from one problem to be used for other problems that are still related [17]. Through transfer learning, creating a more accurate model on small datasets seem possible by using the features from neural networks that have been trained.

TABLE IV
COMPARISON TABLE FOR TECHNIQUE COMBINATIONS AND ITS PERFORMANCE

Dataset	Reference	Methods	Performance
Kermamy's Dataset	[11]	Basic CNN Architecture with Data augmentation	Accuracy: 93.24%
	[12]	Inception ResNetV2, Xception, DenseNet201, and VGG19 with Fine Tuning and Data Augmentation	Accuracy of Inception ResNetV2: 94.20%, Accuracy of Xception: 92.80%, Accuracy of DenseNet201: 92.70%, and Accuracy of VGG19: 93.80%
	[13]	Feature Extraction of InceptionV3 CNN Architecture + Machine Learning Classifier (SVM)	83.49% accuracy, 93.1% AUC
	[14]	Data Augmentation + 5 Ensembled Feature Extraction (Transfer Learning)	96.39% accuracy, 99.34% AUC, 93.28% precision, 99.62% recall
	[15]	ResNet50 + Feature Vector + Dual Added Layer	97.82% accuracy, 98.84% AUC
	[16]	Feature Extraction of AlexNet + SVM Classifier	99.4% accuracy, 98.3% sensitivity, 100% specificity
	[18]	Basic CNN and Transfer Learning model (VGG16, VGG19, and Inception V3) with Data augmentation + resizing	Accuracy of Basic CNN: 97%, Accuracy of VGG16: 98%, Accuracy of VGG19: 97%, and Accuracy of Inception V3: 98%
	[19]	CNN with Data augmentation + resizing	Accuracy: 88.90%
	[20]	CNN with Data augmentation and without data preprocessing	Accuracy With augmentation: 83.38%, Accuracy without augmentation: 80.25%
	[21]	Image pre-processing + 3-layer CNN	88.68% accuracy
	[22]	Image pre-processing + VGG16 + modified multilayer perceptron	97.3% AUC, 97.9% sensitivity, 96.6% specificity, 97.4% accuracy, 97.9% F1 score
	[23]	Augmentation + Customized CNN	93.75% accuracy, 92% precision, 98% sensitivity, 85% specificity
	[24]	Augmentation + with Dropout Layer	90.68% accuracy
	[25]	Offline Augmentation + REsNet50 Architecture adjusted with additional layers	97.65% Accuracy
	Kermamy's Dataset and Cohen's Dataset	[17]	Feature Extraction of Trained Model with ResNet50 Network
RSNA Dataset	[27]	Residual Network and Mask Regional CNN without preprocessing	Accuracy of Residual Network: 85.60% and Accuracy of Mask-RCNN: 78.06%
	[28]	CNN with data augmentation	Accuracy: 92.47%
	[29]	Data augmentation + image enhancement + ensemble of RetinaNet and Mask R-CNN	81.3% Recall, 22.83% mAP
	[30]	Data augmentation + weighted voting + ensemble of Mask R-CNN and RetinaNet	21.75% mAP
	[31]	ResNet-18 (Data sorting) + CycleGAN(image enhancement) + ResNet-18 (Predictor)	92.1% AUC
	[32]	DCNN ResNet-50 without preprocessing	Internal AUC: 93.1% and External AUC 0.815%
ChestX-Ray 14 Dataset, IU Dataset, and MSH Dataset			

ChestX-Ray 14 Dataset	[32]	DCNN (DenseNet) with Data augmentation	76% AUC
	[33]	Image pre-processing + DenseNet (feature extractor) + SVM (classifier)	80.02% AUC
	[34]	Data augmentation + VGG16	68% accuracy, 69% specificity
JSRT Dataset	[35]	Image pre-processing + 3-middle-layer CNN	100% AUC
MC Dataset and JSRT Dataset	[36]	Data Augmentation + Image Segmentation + AlexNet	80.48% accuracy with 77.55% sensitivity
Private Dataset	[37]	Transfer Learning model (VGG16) with Data augmentation + resizing	AUC: 96%, Accuracy 50%
GitHub and Kaggle Combined Dataset	[39]	Data augmentation + 4-layer CNN + AlexNet	98.33% accuracy, 95% precision, 100% recall

Based on the information, with the limited amount of CXR images available, transfer learning is a preferable method to develop an accurate pneumonia detection classifier. Overall, most of the techniques use data augmentation, and image preprocessing as the preprocessing method and are combined with CNN. It is evident that the combination of techniques that utilize the preprocessing method overall got a higher score than those without the preprocessing method.

III. RESULTS AND DISCUSSION

This section shows a comparison and examines the current methodologies and datasets used to identify pneumonia from CXRs. The examination focuses mostly on guaranteeing durability and utility.

The reference for the authors, methods that are used in the papers, dataset, and performance, followed by the advantages and challenges, are presented in Table 5. Numerous architectures are used, and they are based on the architectures such as DenseNet201, ResNet-18, and VGG16. Furthermore, the preprocessing methods are also varied, for example, data augmentation, image preprocessing, and data sorting. In addition, there are also some papers that do not use any preprocessing method.

In terms of accuracy, a considerable number of papers achieve more than 90% in accuracy, while some use different metrics, and some score badly. Nonetheless, variations of datasets are used; therefore, the performance cannot be ranked easily.

TABLE V
PAPER'S TECHNIQUE AND RESULT OVERVIEW

Reference	Methods	Challenges
[11]	Basic CNN Architecture with Data augmentation	Since the author uses a different approach from transfer-based learning or a more complex model, they encountered challenges in finding a suitable classification architecture.
[32]	DCNN (DenseNet) with Data augmentation	Only frontal radiograph images were present
[32]	DCNN ResNet-50 without preprocessing	The CNNs model does not perform well with external data compared with internal data
[12]	Inception ResNetV2, Xception, DenseNet201, and VGG19 with Fine Tuning and Data Augmentation	The author did some experiments with a different method, such as without data augmentation, with data augmentation, and fine-tuning. Each of these approaches has its own advantage and disadvantage. When using fine-tuning, it can achieve higher accuracy compared to without doing any data augmentation. However, fine-tuning has a slower training speed compared to the other approaches.
[18]	Basic CNN and Transfer Learning model (VGG16, VGG19, and Inception V3) with Data augmentation + resizing	The performance of the model is very high with the dataset used by the author. Although the performance is very good, the model can be justified by testing using different datasets to see the model's performance.
[19]	CNN with Data augmentation + resizing	The performance almost reached 90%; however, the model is likely to overfit due to the small size of the training dataset, which can affect the model performance.
[20]	CNN with Data augmentation and without data preprocessing	The performance of both models, with or without data augmentation, outputs a similar result.
[37]	Transfer Learning model (VGG16) with Data augmentation + resizing	Although the AUC scores around 96%, the accuracy could be improved since it only reaches 50%.
[27]	Residual Network and Mask Regional CNN without preprocessing	Due to the low-quality image, the mask-RCNN did not perform well since the pneumonia features are hard to obtain. Because of the unbalanced data between pneumonia and non-pneumonia can affect each of the model's performance.
[28]	CNN with data augmentation	The model performance can be improved.
[29]	Data augmentation + image enhancement + ensemble of RetinaNet and Mask R-CNN	The detection performance is still inadequate because the data is small and the pneumonia position is tiny, making the detection difficult.
[30]	Data augmentation + weighted voting + ensemble of Mask R-CNN and RetinaNet	The model performance can be improved.

[31]	ResNet-18 (Data sorting) + CycleGAN(image enhancement) + ResNet-18 (Predictor)	The model performance can be improved.
[35]	Image pre-processing + 3-middle-layer CNN	Lack of data, a small number of data used. The performance was not high, and the range of uses was limited.
[39]	Data augmentation + 4-layer CNN + AlexNet	Lack of data, a small number of data used (100 on each class). Currently not reliable and has not met the standard of the medical industry.
[33]	Image pre-processing + DenseNet (feature extractor) + SVM (classifier)	Only used frontal chest X-rays. Very high computational power because a lot of convolutional layers are used.
[34]	Data augmentation + VGG16	The performance needs improvement.
[21]	Image pre-processing + 3-layer CNN	Creating the model is time and cost-consuming, while additional parameters such as blood supply in oxygen and historical data of the patient must be considered to detect the pneumonia
[22]	Image pre-processing + VGG16 + modified multilayer perceptron	Limited dataset for generalization purposes.
[23]	Augmentation + Customized CNN	Creating self customized Deep Convolutional Neural Network costs too much time and might cause overfitting.
[24]	Augmentation + with Dropout Layer	Self-proposed CNN architectures are prone to overfitting, and the data is still insufficient.
[25]	Offline Augmentation + REsNet50 Architecture adjusted with additional layers	The proposed architecture is prone to overfitting if it is trained on higher epochs and has to add additional preprocessing methods, such as regularized and dropout functions
[13]	Feature Extraction of InceptionV3 CNN Architecture + Machine Learning Classifier (SVM)	Most papers are only able to focus on the feature extraction rather than the performance of each classifier. In addition, machine learning algorithms still cannot beat Neural networks in terms of sensitivity.
[14]	Data Augmentation + 5 Ensembled Feature Extraction (Transfer Learning)	The computation is expensive to train five great transfer learning algorithms, and this method offers zero to no explanation for how the classification decision is made.
[15]	ResNet50 + Feature Vector + Dual Added Layer	A large variation of results shows concerns about the CNN's ability to generalize with the world pneumonia lungs data. Further investigation into the generalizability is needed.
[16]	Feature Extraction of AlexNet + SVM Classifier	It only represents two transfer learning algorithms and cannot cover general transfer learning.
[17]	Feature Extraction of Trained Model with ResNet50 Network	Hard to generalize two different datasets as one since the first one covers only normal pneumonia, and the second one covers the newest COVID-19 pneumonia
[36]	Data Augmentation + Image Segmentation + AlexNet	Data between the normal and pneumonia lungs are not balanced, hence have to generate or do additional data augmentation to create a generalized model

Table 5 above represents the various methods used by different papers from different institutions. Some might use the same dataset and produce different model performances. In addition, the metrics are also distinct varying in accuracy, AUC, recall, precision, sensitivity, specificity, and also F1 score. The factor that impacts the model performances mostly are datasets. The most used datasets are Kermany's Dataset, which contains around 5856 datasets for both training and validation, while the next came from ChestX-ray14 Dataset with 112,120 X-Ray Images.

In addition, the most frequently used method for data preprocessing is data augmentations. Several studies stated that it could help improve the Deep Convolutional Neural Network (DCNN) model to be more generalized and robust to outliers and noisy data. Besides that, the algorithm or architecture of the neural networks also directly affects the final performance. The ResNet, DenseNet, and VGG architectural neural network are commonly used as a method for transfer learning to improve the model's performance and reduce the long waiting time in the training process.

It is also apparent from Figure 3. that there are still many challenges that researchers face. The first problem can be related to the insufficient amounts of datasets. Small amounts of images that are used to train a DCNN, might cause several

problems. Some researchers had only received frontal chest x-ray images for the training data[32], [33], causing the model to not be able to generalize to the real-world images[15], [22]. Some others also had their model overfit to the training data[14], [18]–[20], [25]. With small-sized images and unbalanced classes [27], [36], they were not able to create a good model, even when data augmentations had already been done. In some cases, parameter tuning did not improve model performances effectively [20]. Lastly, some datasets contain low-quality images[27]. This challenge can't even be solved just by data augmentation, and hence output a bad-performing neural network, with the model getting too underfitted. While many researchers had their model overfitted, some of them also had a really bad model performance which can still be improved in some cases [28], [30], [31], [37].

The next general one is the time cost. Most researchers who used self-created or transfer learning model [11], [16] used most of their time either creating the neural network architecture or in the training process[12], [14], [21], [24], [33]. While these processes are time-consuming, there is no other way since they want to create a generalized and good-performing neural network. Also, high computational powers are needed in order to process thousands of images with a total of ten-thousands per image.

IV. CONCLUSION

Overall, we succeeded in reviewing up to 27 technical papers related to pneumonia classification using Deep Convolutional Neural Network (DCNN). We focused on finding the use of datasets of different researchers and found the most frequently used dataset is the one provided by Kermany et al. [26]. Although it is used frequently, some of the researchers still mentioned the problem of not having sufficient amounts of datasets. Further research can be done to achieve additional data, either by gathering, generating, or integrating different sources of data. Next, we also studied the use of data augmentations and preprocessing methods, which turned out to be successful in increasing the DCNN model performances. From this statement, we suggest that all researchers should consider data augmentations as their preprocessing methods to save time in finding the best preprocessing options. Lastly, we covered that the best method to create the DCNN architecture is the transfer learning, which accounts for up to 99.4% accuracy. This can be done through feature extractions of different pre-trained models. With deep learning algorithms improving from time to time, there is no doubt that there will be new models with better architectures and performances. Therefore, our recommendation for other researchers is to make some improvements by keep finding new combinations of transfer learning models from pre-trained DCNN to get an even better result.

REFERENCES

- [1] UNICEF, "Levels and Trends in Child Mortality Report 2017," 2017. [Online]. Available: https://www.unicef.org/media/48871/file/Child_Mortality_Report_2017.pdf
- [2] ATS, Top 20 pneumonia facts—2019 - american thoracic society, <https://www.thoracic.org/patients/patient-resources/resources/top-pneumonia-facts.pdf> (accessed Mar. 21, 2022).
- [3] VOA, "Pneumonia Pembunuh Balita Nomor 2 di Indonesia." <https://www.voaindonesia.com/a/pneumonia-pembunuh-no-2-balita-di-indonesia/5665607.html> (accessed Mar. 21, 2022).
- [4] "Pneumonia," Mayo Clinic, <https://www.mayoclinic.org/diseases-conditions/pneumonia/symptoms-causes/syc-20354204> (accessed Mar. 21, 2022).
- [5] K. Miller, "The 6 Different Types of Pneumonia, Explained by Doctors," 2021. <https://www.health.com/condition/pneumonia/types-of-pneumonia> (accessed Jun. 06, 2022).
- [6] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 3462–3471, 2017, doi: 10.1109/CVPR.2017.369.
- [7] Y. Li, Z. Zhang, C. Dai, Q. Dong, and S. Badrigilan, "Accuracy of deep learning for automated detection of pneumonia using chest X-Ray images: A systematic review and meta-analysis," *Comput. Biol. Med.*, vol. 123, no. June, p. 103898, 2020, doi: 10.1016/j.combiomed.2020.103898.
- [8] W. Khan, N. Zaki, and L. Ali, "Intelligent Pneumonia Identification from Chest X-Rays: A Systematic Literature Review," *IEEE Access*, vol. 9, pp. 51747–51771, 2021, doi: 10.1109/ACCESS.2021.3069937.
- [9] P. ran Liu, L. Lu, J. yao Zhang, T. tong Huo, S. xiang Liu, and Z. wei Ye, "Application of Artificial Intelligence in Medicine: An Overview," *Curr. Med. Sci.*, vol. 41, no. 6, pp. 1105–1115, 2021, doi: 10.1007/s11596-021-2474-3.
- [10] B. A. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering. EBSE Technical Report EBSE-2007-01. School of Computer Science and Mathematics, Keele University," no. January, pp. 1–57, 2007.
- [11] I. Lahsaini, M. El Habib Daho, and M. A. Chikh, "Convolutional Neural Network for Chest X-ray Pneumonia Detection," *ACM Int. Conf. Proceeding Ser.*, pp. 16–20, 2020, doi: 10.1145/3432867.3432873.
- [12] Z. Jiang, "Chest X-ray Pneumonia Detection Based on Convolutional Neural Networks," *Proc. - 2020 Int. Conf. Big Data, Artif. Intell. Internet Things Eng. ICBAIE 2020*, pp. 341–344, 2020, doi: 10.1109/ICBAIE49996.2020.00077.
- [13] S. L. K. Yee and W. J. K. Raymond, "Pneumonia Diagnosis Using Chest X-ray Images and Machine Learning," *ACM Int. Conf. Proceeding Ser.*, pp. 101–105, 2020, doi: 10.1145/3397391.3397412.
- [14] V. Chouhan et al., "A novel transfer learning based approach for pneumonia detection in chest X-ray images," *Appl. Sci.*, vol. 10, no. 2, 2020, doi: 10.3390/app10020559.
- [15] G. Ali, A. Shahin, M. Elhadidi, and M. Elattar, "Convolutional Neural Network with Attention Modules for Pneumonia Detection," *2020 Int. Conf. Innov. Intell. Informatics, Comput. Technol. 3ICT 2020*, pp. 0–5, 2020, doi: 10.1109/3ICT51146.2020.9311985.
- [16] S. Jamil, M. S. Abbas, Fawad, M. F. Zia, and M. U. Rahman, "A Deep Convolutional Neural Network Based Framework for Pneumonia Detection," *2021 Int. Conf. Digit. Futur. Transform. Technol. ICODT2 2021*, pp. 2–6, 2021, doi: 10.1109/ICoDT252288.2021.9441539.
- [17] I. Katsamenis, E. Protopapadakis, A. Vouloimos, A. Doulamis, and N. Doulamis, "Transfer Learning for COVID-19 Pneumonia Detection and Classification in Chest X-ray Images," *ACM Int. Conf. Proceeding Ser.*, pp. 170–174, 2020, doi: 10.1145/3437120.3437300.
- [18] G. Labhane, R. Pansare, S. Maheshwari, R. Tiwari, and A. Shukla, "Detection of Pediatric Pneumonia from Chest X-Ray Images using CNN and Transfer Learning," *Proc. 3rd Int. Conf. Emerg. Technol. Comput. Eng. Mach. Learn. Internet Things, ICETCE 2020*, no. February, pp. 85–92, 2020, doi: 10.1109/ICETCE48199.2020.9091755.
- [19] L. Račić, T. Popović, S. Čakić, and S. Šandi, "Pneumonia Detection Using Deep Learning Based on Convolutional Neural Network," *2021 25th Int. Conf. Inf. Technol. IT 2021*, no. February, pp. 17–20, 2021, doi: 10.1109/IT51528.2021.9390137.
- [20] S. A. Khoiriyah, A. Basofi, and A. Fariza, "Convolutional Neural Network for Automatic Pneumonia Detection in Chest Radiography," *IES 2020 - Int. Electron. Symp. Role Auton. Intell. Syst. Hum. Life Conf.*, pp. 476–480, 2020, doi: 10.1109/IES50839.2020.9231540.
- [21] A. Gawali, P. Bide, V. Kate, C. Kothastane, and E. Hirani, "Deep Learning Approach to detect Pneumonia," *Proc. 4th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2020*, pp. 1277–1284, 2020, doi: 10.1109/ICECA49313.2020.9297393.
- [22] J. R. Ferreira, D. Armando Cardona Cardenas, R. A. Moreno, M. De Fatima De Sa Rebelo, J. E. Krieger, and M. Antonio Gutierrez, "Multi-View Ensemble Convolutional Neural Network to Improve Classification of Pneumonia in Low Contrast Chest X-Ray Images," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2020-July, pp. 1238–1241, 2020, doi: 10.1109/EMBC44109.2020.9176517.
- [23] R. Siddiqi, "Automated pneumonia diagnosis using a customized sequential convolutional neural network," *ACM Int. Conf. Proceeding Ser.*, pp. 64–70, 2019, doi: 10.1145/3342999.3343001.
- [24] H. Sharma, J. S. Jain, P. Bansal, and S. Gupta, "Feature extraction and classification of chest X-ray images using CNN to detect pneumonia," *Proc. Conflu. 2020 - 10th Int. Conf. Cloud Comput. Data Sci. Eng.*, pp. 227–231, 2020, doi: 10.1109/Confluence47617.2020.9057809.
- [25] T. A. Youssef, B. Aissam, D. Khalid, B. Imane, and J. El Miloud, "Classification of chest pneumonia from x-ray images using new architecture based on ResNet," *2020 IEEE 2nd Int. Conf. Electron. Control. Optim. Comput. Sci. ICECOCS 2020*, 2020, doi: 10.1109/ICECOCS50124.2020.9314567.
- [26] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (oct) and chest x-ray images for classification," *Mendeley data*, vol. 2, no. 2, 2018.
- [27] A. F. Al Mubarak, J. A. M. Dominique, and A. H. Thias, "Pneumonia detection with deep convolutional architecture," *Proceeding - 2019 Int. Conf. Artif. Intell. Inf. Technol. ICAIIT 2019*, pp. 486–489, 2019, doi: 10.1109/ICAIIIT.2019.8834476.
- [28] S. Shah, H. Mehta, and P. Sonawane, "Pneumonia detection using convolutional neural networks," *Proc. 3rd Int. Conf. Smart Syst. Inven. Technol. ICSSIT 2020*, no. Icssit, pp. 933–939, 2020, doi: 10.1109/ICSSIT48917.2020.9214289.
- [29] L. Mao, T. Yumeng, and C. Lina, "Pneumonia Detection in chest X-rays: A deep learning approach based on ensemble RetinaNet and Mask R-CNN," *Proc. - 2020 8th Int. Conf. Adv. Cloud Big Data, CBD 2020*, pp. 213–218, 2020, doi: 10.1109/CBD51900.2020.00046.

- [30] H. Ko, H. Ha, H. Cho, K. Seo, and J. Lee, "Pneumonia Detection with Weighted Voting Ensemble of CNN Models," *2019 2nd Int. Conf. Artif. Intell. Big Data, ICAIBD 2019*, pp. 306–310, 2019, doi: 10.1109/ICAIBD.2019.8837042.
- [31] Z. Wang, J. Hall, and R. J. Haddad, "Improving pneumonia diagnosis accuracy via systematic convolutional neural network-based image enhancement," *Conf. Proc. - IEEE SOUTHEASTCON*, vol. 2021-March, pp. 0–5, 2021, doi: 10.1109/SoutheastCon45413.2021.9401810.
- [32] A. Tilve, S. Nayak, S. Vernekar, D. Turi, P. R. Shetgaonkar, and S. Aswale, "Pneumonia Detection Using Deep Learning Approaches," *Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020*, pp. 1–8, 2020, doi: 10.1109/ic-ETITE47903.2020.152.
- [33] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, "Pneumonia Detection Using CNN based Feature Extraction," *Proc. 2019 3rd IEEE Int. Conf. Electr. Comput. Commun. Technol. ICECCT 2019*, pp. 1–7, 2019, doi: 10.1109/ICECCT.2019.8869364.
- [34] M. Aledhari, S. Joji, M. Hefeida, and F. Saeed, "Optimized CNN-based Diagnosis System to Detect the Pneumonia from Chest Radiographs," *Proc. - 2019 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2019*, pp. 2405–2412, 2019, doi: 10.1109/BIBM47256.2019.8983114.
- [35] D. Hirahara, E. Yuda, T. Takahara, and Y. Kobayashi, "Fundamental study on preliminary image processing at time development of CNN using chest radiography," *2019 IEEE 1st Glob. Conf. Life Sci. Technol. LifeTech 2019*, pp. 102–104, 2019, doi: 10.1109/LifeTech.2019.8884064.
- [36] X. Gu, L. Pan, H. Liang, and R. Yang, "Classification of bacterial and viral childhood pneumonia using deep learning in chest radiography," *ACM Int. Conf. Proceeding Ser.*, pp. 88–93, 2018, doi: 10.1145/3195588.3195597.
- [37] P. Naveen and B. Diwan, "Pre-trained VGG-16 with CNN architecture to classify X-Rays images into normal or pneumonia," *2021 Int. Conf. Emerg. Smart Comput. Informatics, ESCI 2021*, pp. 102–105, 2021, doi: 10.1109/ESCI50559.2021.9396997.
- [38] D. Demner-Fushman *et al.*, "Preparing a collection of radiology examinations for distribution and retrieval," *J. Am. Med. Informatics Assoc.*, vol. 23, no. 2, pp. 304–310, 2016, doi: 10.1093/jamia/ocv080.
- [39] A. Panwar, A. Dagar, V. Pal, and V. Kumar, "COVID 19, pneumonia and other disease classification using chest X-ray images," *2021 2nd Int. Conf. Emerg. Technol. INCET 2021*, pp. 23–26, 2021, doi: 10.1109/INCET51464.2021.9456192.
- [40] Joseph Paul Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "COVID-19 Image Data Collection: Prospective Predictions Are the Future," 2020, [Online]. Available: <https://github.com/ieee8023/covid-chestxray-dataset>
- [41] J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study," *PLoS Med.*, vol. 15, no. 11, pp. 1–17, 2018, doi: 10.1371/journal.pmed.1002683.
- [42] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wang, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases.," *Quant. Imaging Med. Surg.*, vol. 4, no. 6, pp. 475–7, 2014, doi: 10.3978/j.issn.2223-4292.2014.11.20.