

Review

A Deep Learning Review of ResNet Architecture for Lung Disease Identification in CXR Image

Syifa Auliyah Hasanah ¹, Anindya Apriliyanti Pravitasari ^{1,*}, Atje Setiawan Abdullah ², Intan Nurma Yulita ² and Mohammad Hamid Asnawi ¹

¹ Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung 45363, Indonesia; syifa19010@mail.unpad.ac.id (S.A.H.); mohammad19011@mail.unpad.ac.id (M.H.A.)

² Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung 45363, Indonesia; atje.setiawan@unpad.ac.id (A.S.A.); intan.nurma@unpad.ac.id (I.N.Y.)

* Correspondence: anindya.apriliyanti@unpad.ac.id

Abstract: The lungs are two of the most crucial organs in the human body because they are connected to the respiratory and circulatory systems. Lung cancer, COVID-19, pneumonia, and other severe diseases are just a few of the many threats. The patient is subjected to an X-ray examination to evaluate the health of their lungs. A radiologist must interpret the X-ray results. The rapid advancement of technology today can help people in many different ways. One use of deep learning in the health industry is in the detection of diseases, which can decrease the amount of money, time, and energy needed while increasing effectiveness and efficiency. There are other methods that can be used, but in this research, the convolutional neural network (CNN) method is only used with three architectures, namely ResNet-50, ResNet-101, and ResNet-152, to aid radiologists in identifying lung diseases in patients. The 21,885 images that make up the dataset for this study are split into four groups: COVID-19, pneumonia, lung opacity, and normal. The three algorithms have fairly high evaluation scores per the experiment results. F1 scores of 91%, 93%, and 94% are assigned to the ResNet-50, ResNet-101, and ResNet-152 architectures, respectively. Therefore, it is advised to use the ResNet-152 architecture, which has better performance values than the other two designs in this study, to categorize lung diseases experienced by patients.

Keywords: lung diseases; image classification; convolutional neural networks; ResNet-50; ResNet-101; ResNet-152



Citation: Hasanah, S.A.; Pravitasari, A.A.; Abdullah, A.S.; Yulita, I.N.; Asnawi, M.H. A Deep Learning Review of ResNet Architecture for Lung Disease Identification in CXR Image. *Appl. Sci.* **2023**, *13*, 13111. <https://doi.org/10.3390/app132413111>

Academic Editor: Evangelos Spyrou

Received: 17 August 2023

Revised: 9 October 2023

Accepted: 10 October 2023

Published: 8 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Two of the most crucial organs in the human body, the lungs, have a direct impact on both the respiratory and circulatory systems. Injury to the human lungs has a significant impact on the continuity of all bodily functions. Several lung diseases, such as COVID-19, pneumonia, and other diseases, result in a great number of fatalities. The majority of diseases that affect the lungs are brought on by bacteria, viruses, fungi, chemical exposure, lung injury, or even the indirect effects of other diseases. Streptococcus bacteria, Mycoplasma bacteria, adenovirus, rhinovirus, influenza virus, respiratory syncytial virus (RSV), and parainfluenza virus are the pathogens that cause pneumonia [1]. The World Health Organization (WHO) describes pneumonia as a lung-specific acute respiratory disease. Pneumonia is a contagious disease that can affect both children and adults. The most common cause of infection-related death in children worldwide is pneumonia. A total of 4,685,164 individuals under the age of 15 and 9,564,023 individuals over the age of 15 died from pneumonia in 2019. By 2030, there will be about 11 million child deaths worldwide due to pneumonia, according to a study by Johns Hopkins University and Save the Children. The COVID-2019 disease is a respiratory infection brought on by the SARS-CoV-2 virus [2]. The virus spreads through the air from an infected person's lips

or nose when they cough, sneeze, sing, talk, or breathe. Globally, COVID-19 is spreading. Overall, 6,456,988 people died as a result of COVID-19 in December 2019. Lung opacities are an indication of abnormalities in a person's lungs other than cancer [3]. This might be an instance of lung bleeding, infection, or experience of other unidentified issues. Numerous factors with varying degrees of significance can contribute to lung opacities. Viral infections and other lung clouding diseases frequently have short-term effects with little long-term harm. Other conditions, like lung cancer and alveolar hemorrhage, call for more serious treatment.

There are several lung diseases that share the same symptoms but require different treatments, such as COVID-19 and pneumonia [4]. Determining the presence of infection in the early stages of the disease is one of the most challenging aspects of COVID-19 [5]. It is difficult to find similarities between people who have lung-related diseases in terms of their initial personality traits. A crucial first step is the accurate diagnosis of the disease. With the rapid development of technology, several fields, including the medical field, began to use Artificial Intelligence (AI) to support ongoing work processes. Medical imaging is one element that is frequently used in the application of technology in the medical field. Medical images are used to diagnose or detect disease based on image or video data [6]. Image data can be obtained using a variety of techniques, such as magnetic resonance imaging (MRI), X-rays, ultrasonography (USG), endoscopy, computerized tomography (CT) scans, and nuclear medicine [7]. CT scan data are usually used to segment various diseases, such as lung disease. However, segmenting lung nodules on CT images still needs the expertise and experience of a medical professional, specifically a radiologist, and is performed manually by depicting the nodule one slice at a time. This procedure must be repeated several times and is subject to subjectivity, which can lead to misdiagnosis. As a result, accurate automatic lung nodule segmentation has always been at the forefront of computer-assisted diagnosis research [8]. As a result, many studies are competing to achieve good evaluation values for this case, so that the specificity and sensitivity for CT scans are 94% and 37%, respectively. X-ray radiography is frequently the preferred option in clinical settings due to its benefits, including its lower cost and availability in many hospitals, even though CT scans offer higher sensitivity for detection. One of the methods most frequently used to assess a patient's physical condition is an X-ray. A chest X-ray (CXR) examination is performed to diagnose lung disease, the results of which show the condition of the chest cavity, respiratory tract, heart, and lungs.

The presence of unusual findings, or what are typically referred to as abnormal situations, in a radiological examination of a person's lungs performed by a radiologist defines the examination on their condition as being conducted. CXRs are employed to find anomalies in the cavity. Ground-glass opacity (GGO) is one of the anomalous phenomena that is identified [9]. Normal lungs have black CXR findings, but if the CXR findings show a gray area or an area covered with a membrane, the condition is ground-glass opacity. Most people with moderate to severe respiratory conditions show ground-glass opacity, which suggests an infection or inflammation brought on by viruses, fungi, bacteria, or parasites like pneumonia and COVID-19 [10]. Chest X-ray diagnosis and classification of lung diseases is a challenging and intricate task. In order to obtain accurate identification results, extra precision is required. To assist radiology specialists and accelerate the pace of implementation, it is necessary to collaborate with increasingly sophisticated technological advances, such as deep learning models. A convolutional neural network is a deep learning algorithm that uses mathematical operations in network learning, specifically convolution. This is a neural network algorithm that is commonly used for image data.

The Residual Neural Network, also known as ResNet, is a popular CNN architecture. He et al. [11] first proposed this architecture to ease the training of networks that are much deeper than those previously employed, as training becomes more difficult with deeper neural networks. The layers were reformulated as residual learning functions, also referred to as residual blocks. These building blocks, which are present in every layer of the ResNet CNN architecture, form its framework. Utilizing these residual blocks makes

network optimization simpler and allows for significantly better accuracy as depth increases. Furthermore, residual blocks can be used to solve the vanishing gradient problem [12], which occurs when the gradient information that the model is supposed to learn is unable to reach the first layer due to repeated multiplications. As a result, the first layer receives no gradient, rendering the CNN model unable to learn from the calculated errors.

In CNN, various architectures, such as the ResNet architecture, have different variations despite originating from the same source. The only difference between these architectures is the number of layers and parameters, which are all the same. ResNet architectures have layers that correspond to their names; for example, ResNet-50 has 50 layers, ResNet-101 has 101 layers, and ResNet-152 has 152 layers. These layers perform various functions that help the machine learn the data and extract important information for the classification phase. The more layers used in an architecture, the deeper the learning process, and thus the better the performance. However, deeper architectures result in longer training times due to the increased number of parameters. As a result, the purpose of this study is to compare the ResNet-50, ResNet-101, and ResNet-152 architectures using CXR image data to classify pneumonia and COVID-19 diseases.

The contributions of this research are as follows:

1. Our study is useful for conducting a comparative analysis of the three ResNet architectures used, namely ResNet-50, ResNet-101, and ResNet-152, to identify diseases that affect the lungs using X-ray images. By obtaining the highest evaluation score and the lowest error rate for each architecture, it is possible to determine which ResNet architecture has the most optimal modeling, allowing accurate lung disease detection.
2. Our research has the potential to contribute to the academic literature by demonstrating how increasing the number of layers and parameters can improve model accuracy. This study also proves that the models with more layers and parameters resulting in better evaluations and faster performance speeds than previous model architectures.
3. The study conducted can benefit various stakeholders, particularly radiologists. The use of deep learning technology can help increase the efficiency of the work performed; additionally, the quality of the resulting diagnosis can be improved based on the assessment scores. The government's use of technology can help diagnose disease sooner, allowing carrying out of preventive measures and treatment more quickly, reducing costs, time, and energy required.

The rest of this paper is organized as follows: Section 2 contains several research papers on the classification of pneumonia and COVID-19 using the ResNet architecture. Section 3 describes the dataset, data preprocessing, network architecture, and metric evaluation used in this study. Section 4 explains and discusses the outcome of each proposed architecture. Finally, Section 5 brings the paper to a close.

2. Related Works

This section contains research-related material, such as general work on the classification of lung disease data using the ResNet-50, ResNet-101, and ResNet-152 algorithms. These studies form the foundation for our investigation of efficient and effective methods for analyzing lung disorders using classification techniques. Various researchers have contributed to the identification of lung disease using artificial intelligence and X-ray imaging of the lungs as a result of rapid technological advancements.

Medical imaging has become an essential diagnostic tool for a wide range of diseases since the discovery of X-ray imaging in 1895. Ilhem Conrad Rontgen was the first to discover X-rays in 1895 [13]. X-rays are a technique for diagnosing health problems and monitoring a person's health. A two-dimensional image of the inside of the human body is obtained by using long-wave electromagnetic radiation that can penetrate body tissue. Many imaging techniques have been developed since then, including computed tomography, magnetic resonance imaging, ultrasound, and positron emission tomography. The use of medical images for treatment dates back to the discovery of X-rays, and automatic analysis of medical images first appeared in 1960 [14]. Some research is conducted using computers as

a tool, but due to limited access, this work is severely limited. Several researchers from the University of Chicago's Kurt Rossmann Laboratory began developing machine learning systems and medical image analysis techniques in the 1980s [15]. Some disease cases are extremely complex because they necessitate large amounts of data from various image types, such as 2D, which necessitates hundreds of images, and 3D, which necessitates thousands of images. As a result, as the amount of data analyzed grows, so does the workload of radiologists and physicians. Deep learning methods have become an important solution in medical imaging in recent years.

CXR images are an example of medical data in the form of complex pattern images. As a result, a deep learning model is used in this study, with the goal of enabling computers to learn features automatically from data and perform complex tasks without being explicitly programmed. Several studies have shown that network depth is critical because it provides better accuracy results than shallower networks [16,17]. Deeper networks can also introduce new issues like disappearing or exploding gradients, both of which have an adverse effect on the model, making the model more complex and challenging to train [18]. He et al. [11] proposed a deep residual learning framework to address this issue. ResNet is a network that has deeper layers than the VGG network.

The ResNet architecture introduces skip connections, also known as residual connections, which allow the flow of the gradients directly through alternative bypass paths [19]. This addresses the issue of vanishing gradients in deep neural networks, where gradients shrink to zero and impede learning. By using residual blocks, we can train an effective deep neural network. Through residual connections between layers, input can spread more quickly. As a result, we can expand our network, and ResNet offers researchers flexibility in determining the number of layers to use because it has several versions with different numbers of layers, such as ResNet-18 with 18 layers, ResNet-34 with 34 layers, ResNet-50 with 50 layers, and several other types of ResNet. However, it should be kept in mind that as the number of layers increases, so does the number of networks and parameters used.

An experiment was first built in a study conducted by He et al. [11] comparing plain networks with 18 and 34 layers. According to the results of these experiments, 34-layer plain nets have a higher error rate than 18-layer plain nets. However, when residual blocks and skip connections were added to the 34-layer plain net and the 18-layer plain net to create ResNet-18 and ResNet-34, the results changed, indicating that networks with more layers have higher error rates. Then, in the following experiment, the authors changed the ResNet-34 layers by replacing every two-layer block with three-layer bottleneck blocks, yielding ResNet-50. ResNet-101 and ResNet-152 were created as a result of changes to more than three-layer blocks. Despite the fact that the addition reduced the number of FLOPs, ResNet-152 (11.3 billion FLOPs) had a lower complexity than VGG-16 and VGG-19 (15.3 and 19.6 billion FLOPs, respectively). Yessi Jusman's 2023 research comparing ResNet-101 and VGG-19 to identify the severity of prostate cancer [20] found that the ResNet-101 architecture had an average accuracy value of 97.70%, while the VGG-19 architecture had an accuracy value of 96.49%. ResNet-101 outperformed VGG-19 in this case; additionally, ResNet-101 required 34 min and 38 s to train the model, whereas VGG required 56 min and 38 s. ResNet's advantage, with its large number of layers, is the reason why it is widely used in a variety of studies, including medical image data research. For example, Chhabra and Kumar [21] developed a model to detect pneumonia in patients, allowing a more quick and accurate run of the detection process. The ResNet-50 architecture was used in conjunction with transfer learning to obtain images from two distinct classes, normal and pneumonia. The model was then evaluated using the confusion matrix, and the result was that it had a high accuracy value of 94%. ResNet-50 was used by Al-Haija and Adebajo [22] to accurately diagnose breast cancer in histopathological images, demonstrating its potential for improving medical image interpretation. Similarly, Reddy and Juliet [23] used ResNet-50 and transfer learning to achieve proficient malaria cell image classification, demonstrating its utility in disease detection. Furthermore, Shadab et al. [24]

addressed cancer detection using ML and the ResNet-50 architecture, extending the model's potential in the field of histopathology with promising results.

The ResNet-101 architecture was also used to classify medical image data on brain tumor datasets [25]. The study included three types of tumors: meningiomas, gliomas, and pituitary tumors. The preprocessing process includes zero-centering and intensity normalization; additionally, the evaluation results are compared using augmentation in the preprocessing process. The accuracy value produced without augmentation was found to be 89.93%, but when additional augmentation was added to the data, the accuracy rate increased to 93.83%. The use of augmentation during the preprocessing stage was shown to improve model accuracy in Ahuja et al. [26] studies. The use of the ResNet-101 architecture also demonstrated remarkable significance in a variety of domains of medical image analysis, as evidenced by the following studies. Xue et al. [27] took on the difficult task of dealing with noisy labeled medical images, a common issue in real-world datasets, by using ResNet-101 for robust skin lesion classification. Their research demonstrated the architecture's ability to improve classification accuracy, particularly in scenarios with label noise. Abirami et al. [28] investigated COVID-19 classification in the context of the ongoing global health crisis using an innovative application of medical image synthesis using generative adversarial networks (GANs). Their research used ResNet-101 as a key component in generating and classifying synthetic medical images, contributing to the fight against the pandemic and demonstrating the adaptability of the ResNet-101 architecture.

The ResNet-50 and ResNet-101 architectures proved to be versatile and effective tools for medical image analysis, as evidenced by the studies listed below. Athisayamani et al. [29] used ResNet-152, a ResNet variant, to extract features for MRI brain tumor classification. Their work demonstrated the architecture's ability to extract meaningful features from medical images, which were then optimized for dimension reduction to improve brain tumor classification accuracy. In a related field, Roy et al. [30] investigated the use of ResNet-152 for Alzheimer's disease diagnosis. There was research conducted to classify binary classes and multi-classes. It was divided into two classes: Alzheimer's Disease and Cognitive Normal in binary classification, and Alzheimer's Disease, Cognitive Normal, Early Mild Cognitive Impairment, and Late Mild Cognitive Impairment in multiclass classification. ResNet-152 was tested with a variety of medical image datasets and was found to be 99.30% accurate for binary classification and 98.79% accurate for multiclass classification. Data processing was performed with great care, and it played an important role in improving accuracy. The authors contributed to the advancement of medical image-based diagnostic approaches for Alzheimer's disease by leveraging the power of ResNet-152.

Previous research has shown that ResNet-50, -101, and -152 have advantages for accurate medical image analysis. When compared to the previously produced ResNet-34, the resulting ResNet-50, -101, and -152 have a higher level of accuracy [31]. This is supported by experimental results for error rates of various methods, which show that ResNet-50, -101, and -152 have lower error rate values than other methods such as VGG, GoogleLeNet, VGG (v5), PreLU-net, BN-inception, ResNet-34 B, and ResNet-34 C. ResNet-50 has a 5.25% top five error rate, ResNet-101 has a 4.60% top five error rate, and ResNet-152 has a 4.49% top five validation error. This single model outperforms all previous ensemble results. Experiments on the ImageNet dataset were also conducted, and it was discovered that ResNet-152 had a top five error of 3.57% in the test set, earning it first place at ILSVRC 2015.

Abubakar et al. [32] attempted to compare these ResNet-50, -101, and -152 architectures to classify data on human skin burns, which are divided into two categories: burnt skin and injured skin (bruises and pressure sores). Before beginning the classification stage, data were preprocessed by rotating and flipping. Two experiments were carried out during the feature extraction stage, namely by performing additional linear SVM to classify features that did not require additions. Based on these experiments, it was discovered that the ResNet-152 model, whether SVM was added or not, had the highest level of accuracy, namely 99.52% without SVM additions and 99.96% with SVM additions. After ResNet-152, ResNet-101 had the second highest level of accuracy, with 98.81% for those

without additions and 99.87% for those with SVM additions. ResNet-50 had the lowest level of accuracy in this study when compared to the other two ResNets, with 98.33% for the model that was not added and 99.87% for the model that was added. This paper also discussed the discovery that increasing the number of layers in ResNet improves its accuracy. Setiawan [33] investigated the impact of varying image dimensions and exposure fusion framework enhancements on the performance of pneumonia detection using different ResNet architectures, including ResNet-50, ResNet-101, and ResNet-152, in contrast to Abubakar et al. This study discovered that ResNet-50 outperforms the other version of ResNet by systematically analyzing the influence of image dimension and exposure fusion techniques.

Three ResNet architectures, ResNet-50, -101, and -152, are compared in this study. This ResNet series was chosen based on the research summary, which stated that ResNet-50, -101, and -152 are effective for medical image analysis. This study aims to prove that increasing the layer on the ResNet has an effect on the resulting level of accuracy and investigates which of the three architectures is best used in classifying lung disease from CXR.

3. Materials and Methods

In this section, there is an explanation of the analysis process based on the data utilized. The data preparation stage, the research steps according to the technique to be selected, as well as the evaluation of the selected model are performed.

3.1. Dataset

In this study, data in the form of a collection of medical images were used as the input, and lung X-rays, also known as CXR, are commonly used to identify or detect disorders found in the lungs, such as lung disease, pneumonia, lung cancer, and even COVID-19. This study's dataset is renewable and was obtained from M.E.H. Chowdhury et al. [34] and Rahman, T. et al. [35]. These data were obtained from the COVID-19 DATABASE of the Italian Society of Medical and Interventional Radiology (SIRM), the Novel Corona Virus 2019 Dataset developed by Joseph Paul Cohen, Paul Morrison, and Lan Dao on GitHub, and images extracted from 43 different publications. This study's data were obtained from patients, whether they had the disease or not, and included 3616 images of patients infected with COVID-19, 1345 images of pneumonia, 6012 images of lung opacity, and 10,912 images of normal lungs. In Table 1, five samples for each class are presented.

Table 1. Five samples for each class (COVID-19, Lung Opacity, Pneumonia, and Normal).

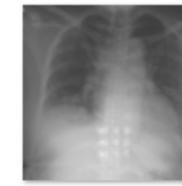
Class	Dataset Samples				
COVID-19					
Lung Opacity					

Table 1. Cont.

Class	Dataset Samples				
Pneumonia					
Normal					

3.2. Data Preprocessing

This study's research steps began with preparing the data used, separating the data into training and test data, followed by the data preparation stage, also known as data pre-processing, modeling using the ResNet-50, ResNet-101, and ResNet-152 architectures, and evaluating the selected model. Some preparation was required before the data entered the modeling process so that the existing data could construct an accurate model and the model could deliver good results. As a result, some transformations or adjustments to the pixel data were required, both to adapt to the model's architecture and to improve the model's accuracy.

Since the data received had varying pixel sizes, all data sizes were converted to 224×224 for equalizing the size before being named according to the class or folder that was formed. To speed up model training and uniform values in the data, 0–255-pixel values in the data were normalized by conversion to a scale of 0–1. After normalizing the data, we used data augmentation to diversify the available data, allowing the model the study of additional data throughout the training process, resulting in better training on the targeted characteristics [2].

The enhancement strengthened the final model's resistance to various types of noise [36]. Shifting, picture reversal, rotation, and image amplification are common augmentation techniques [37]. We changed the image or vertically and horizontally shifted it. Additional versions of the model were formed by adjusting the position of objects in the image, making the model more general. Datasets that were enhanced with an image shift of 10% produced better assessment results than datasets that were not enhanced [38]. As a result, a 10% image shift is used in this investigation.

Flipping is a technique in image augmentation that involves flipping an image along one of its axes, typically a horizontal or vertical axis, to increase the amount of training data and improve a model's performance in pattern recognition or other image processing tasks. Modifications like these can make the trained model more robust and capable of recognizing objects in different orientations. Datasets with horizontally flipped photos can significantly improve dataset diversity, allowing the training of model with more images from the same dataset and achievement of an evaluation score of more than 90% [39]. As a result, horizontal image inversion is used in this investigation. Meanwhile, rotation in image augmentation is the technique of rotating an image at a specific angle around its center point. The goal is to generate variations in perspective from the same image so that models trained with augmented data become more resistant to changes in the orientation of the same object in the image. Using rotated images in the -30° to $+30^\circ$ range [40] results in a higher level of accuracy. As a result, an image rotation of 30° is used in this investigation. Zooming or magnification of an image is also added by zooming in by 20% to improve the

model's assessment value [41]. In Table 2, visualization example of image augmentation are shown.

Table 2. Visualization Example of Image Augmentation Applied to an Image.

Augmentation Type	Visualization
Rotation 30°	
Width Shift 10%	
Height Shift 10%	
Zoom Range 20%	
Horizontal Flip	

3.3. Network Architecture

The modeling process begins after the data is pre-processed. The modeling in this work is performed using the ResNet-50, -101, and -152 architectures. He et al. [11] pioneered this architecture to enable much deeper network training than was previously possible. The greater the depth of the neural network, the more difficult it is to train. As a result, He et al. reformulated the remaining learning function layer, also known as the residual block. This block is present in each layer of the CNN ResNet architecture and serves as the design's foundation. The representation of the residual block can be found in Figure 1. The network is easy to optimize and can gain greater accuracy from that depth by using this residual block [42].

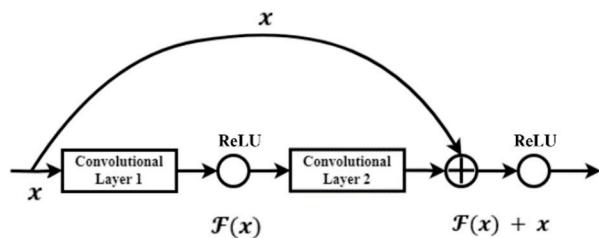


Figure 1. Residual block or skip connection on ResNet.

ResNet, or “residual network”, is one of the Convolutional Neural Network (CNN) topologies. CNN has three basic layers: convolutional layers, pooling layers, and fully connected layers. Furthermore, there are several other procedures that work to improve the accuracy of the CNN algorithm’s classification, such as zero padding, ReLU activation, and batch normalization. ResNet is made up of several layers, including the convolution layer, the pooling layer, the dropout layer, and the fully connected layer.

One advantage of using a residual network is that it can solve the vanishing gradient problem. The presence of residual blocks addresses this issue, as gradients can reach the first layer without undergoing the same number of multiplications as in architectures without residual blocks. These blocks add a shortcut path that acts as an identity function, bypassing the training process for one or more layers indirectly. ResNet has several architectures, and the ResNet 50, 101, and 152-layer architectures are used in this study. Each architecture is depicted in the diagrams below.

When compared to the VGG architecture, the ResNet can create deeper layers while remaining less complex. On the ImageNet test set, Ensemble had an error rate of 3.57% [16]. Convolutional neural networks with the ResNet architecture are a suitable method for effectively classifying CXR images of COVID-19 patients’ lungs due to the advantages of the ResNet architecture, such as maintaining performance even with deeper architectures, lighter computational calculations, and better network training capabilities.

The number of layers is what fundamentally distinguishes ResNet-50, ResNet-101, and ResNet-152. The three networks are named after the number of layers they have. ResNet-50, for example, has 50 layers, ResNet-101 has 101 levels, and ResNet-152 has 152 layers that the data pass through during the modeling process.

Table 3 describes the explanation for each model. The initial convolutional layer with kernel sizes of 77% and 33% is applied to each input image. The ResNet architecture is divided into four stages. The network starts with three residual blocks, each with three layers, in the first stage. The kernel sizes used for convolutional operations in the first-stage blocks’ three layers are 64, 64, and 128, respectively.

According to Table 3, every input that enters the ResNet architecture first proceeds through a convolutional layer. Convolution is a mathematical term for repeatedly applying one function to the output of another function. Convolution uses the output function as a feature map of the input image. This process entails taking a small-number matrix, known as a filter, and applying it to the input image based on the value of the filter. The convolution operation can be calculated using the formula in Equation (1), where f represents the input image and h represents the kernel used. The m and n represent the row and column indexes, respectively.

$$G[m, n] = (f \times h)[m, n] = \sum_j \sum_k h[j, k] f[m - j, n - k]. \quad (1)$$

A ReLU activation function is used in several convolution layers to make learning in the model using training data faster and more effective by keeping positive values and mapping values to zero values. The ReLU activation function’s equation is as follows:

$$f(x) = \max(0, x). \quad (2)$$

Table 3. The ResNet-50, ResNet-101, and ResNet-152 architecture.

	Layer Name	ResNet-50	ResNet-101	ResNet-152
Block 1	Convolutional Layer		$7 \times 7, 64, \text{stride } 2$	
	Maxpooling Layer		$3 \times 3, \text{stride } 2$	
Block 2	Convolutional Layer	$\begin{bmatrix} 1 \times 1 \times 64 \\ 3 \times 3 \times 64 \\ 1 \times 1 \times 256 \\ 1 \times 1 \times 128 \\ 1 \times 1 \times 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 \times 64 \\ 3 \times 3 \times 64 \\ 1 \times 1 \times 256 \\ 1 \times 1 \times 128 \\ 1 \times 1 \times 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 \times 64 \\ 3 \times 3 \times 64 \\ 1 \times 1 \times 256 \\ 1 \times 1 \times 128 \\ 1 \times 1 \times 512 \end{bmatrix} \times 3$
	Convolutional Layer	$\begin{bmatrix} 1 \times 1 \times 256 \\ 3 \times 3 \times 128 \\ 1 \times 1 \times 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1 \times 128 \\ 3 \times 3 \times 128 \\ 1 \times 1 \times 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1 \times 128 \\ 3 \times 3 \times 128 \\ 1 \times 1 \times 512 \end{bmatrix} \times 8$
Block 3	Convolutional Layer	$\begin{bmatrix} 1 \times 1 \times 256 \\ 3 \times 3 \times 256 \\ 1 \times 1 \times 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \times 256 \\ 3 \times 3 \times 256 \\ 1 \times 1 \times 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1 \times 256 \\ 3 \times 3 \times 256 \\ 1 \times 1 \times 1024 \end{bmatrix} \times 36$
	Convolutional Layer	$\begin{bmatrix} 1 \times 1 \times 512 \\ 3 \times 3 \times 512 \\ 1 \times 1 \times 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 \times 512 \\ 3 \times 3 \times 512 \\ 1 \times 1 \times 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 \times 512 \\ 3 \times 3 \times 512 \\ 1 \times 1 \times 2048 \end{bmatrix} \times 3$
Averagepooling Layer		1000 fc, softmax		

The pooling layer is the layer that follows the convolution layer and the ReLU activation function. This layer is made up of a filter with a fixed stride size that moves across the feature map area. To improve computing speed, the pooling layer reduces the number of parameters and spatial size of an image. At each filter change, Max Pooling selects the maximum value in the area to be chosen.

In the subsequent stages, there are differences in the number of layers for each architecture. In the second stage, ResNet-50 and ResNet-101 have four layers, while ResNet-152 has eight layers, with respective kernel sizes of 128, 128, and 512.

ResNet-50 has six layers in the third stage, ResNet-101 has 23 layers, and ResNet-152 has 36 layers, with kernel sizes of 256, 256, and 1024. Each architecture has three layers in the final stage, with kernel sizes of 512, 512, and 2048, respectively. Finally, the network includes an average pooling layer followed by a fully connected layer with 1000 neurons (ImageNet output classes). Average pooling reduces the image by taking the average value of the specified area at each filter change, as the name implies. The fully connected layer connects all activation nerves from the previous layer to the next. The feature map obtained from the convolution and pooling layers is fed into the full connection layer. The equation is output at the full connection layer.

$$h(x) = f\left(b + \sum_i w_i x_i\right). \quad (3)$$

A bottleneck design is used for deeper networks such as ResNet-50, ResNet-101, and ResNet-152. Three layers are stacked together for each residual function. These layers are made up of 11, 33, and 11 convolutions. The 11 convolutional layer is in charge of shrinking and then restoring the dimensions. With smaller input and output dimensions, the 33 layers act as a bottleneck. Skip connections are performed every three layers in the ResNet architecture. Convolutional operations in the residual blocks are performed with a stride size of 2 when transitioning from one stage to the next. As a result, the input size in terms of height and width is cut in half.

3.4. Metric Evaluation

In quantifying the performance of the classification model, and to determine the most suitable technique, the confusion matrix is employed. The confusion table can display and

compare the value to the model's expected value. The precision, recall, and F1-score are typical evaluation matrices used in this study.

Precision is a metric that represents the probability that the data are truly positive, given that they have been classified as positive. Precision is the difference between a correctly predicted positive outcome and an overall positive prediction. Precision can be expressed mathematically as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (4)$$

Recall, also referred to as sensitivity, is a measure used to assess the ability of a test to identify true positives. It is calculated by comparing the number of true positive predictions to the total number of actual positive instances. The F1-score is calculated by summarizing accuracy and sensitivity/recall through the computation of their averages. The formula for recall and F1-score are in Equations (5) and (6), respectively.

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (5)$$

$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})}. \quad (6)$$

In this context, TP (True Positive) and TN (True Negative) represent the sum of data instances whose actual class matches the predicted class. On the other hand, FP (False Positive) and FN (False Negative) denote the sum of data instances whose actual class does not align with the predicted class.

3.5. Experimental Setting

The modeling procedure is carried out at this point in order to obtain weights that can be used to properly categorize CXR. Several hyperparameters, including the learning rate and optimizer used, are chosen for model training in this training. The batch size is 32 and the number of epochs used is 50. The optimizer used in this study is Adam, since it produces good results quickly [43]. The initial learning rate is 1×10^{-6} ; however, Adam calculates the learning rate using an adaptive approach based on various parameters. Adam can be used as an optimization technique as a substitute for the classical stochastic gradient descent procedure to update the weights iteratively based on training data.

A loss function plays a crucial role in assessing the performance of an algorithm in modeling data. It quantifies the disparity between the algorithm's current output and the desired or expected output. Focal loss is a specific type of loss function utilized in multi-class classification tasks, particularly when dealing with imbalanced data. This function is designed to focus on examples that challenge the model's accuracy, preventing it from making unreliable predictions. Consequently, focal loss helps the model improve its performance on challenging examples, making it effective in overcoming these obstacles [44].

Existing data are divided into training data and test data, with a comparison ratio of 80% (16,933 images) for training data, which include 2893 COVID-19 images, 1076 pneumonia images, 4810 lungs opacity images, and 8730 normal lung images. The remaining 20% (4232 images) are for test data, which include 723 COVID-19 images, 269 pneumonia images, 1202 lungs opacity images, and 2183 normal lung images. Additionally, images in the pre-processing stage are classified using the ResNet-50, ResNet-101, and ResNet-152 architectures. The general classification process is depicted in Figure 2 using the ResNet-50, ResNet-101, and ResNet-152 architectures.

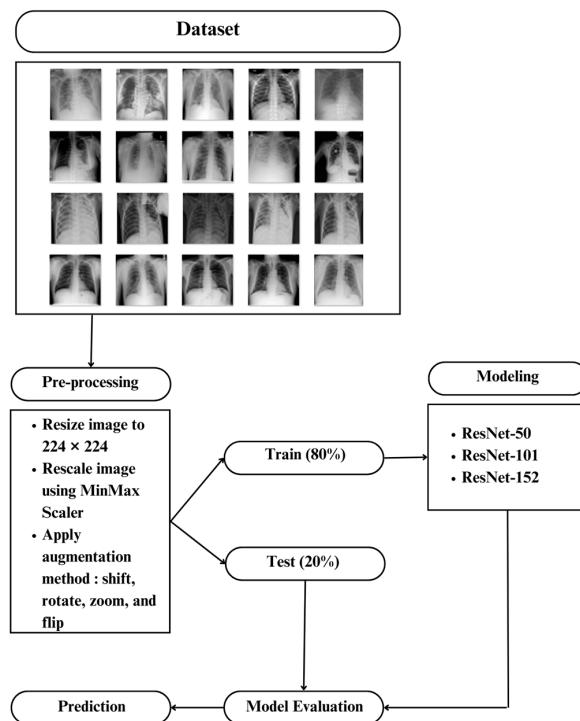


Figure 2. The General Classification process using ResNet-50, ResNet-101, and ResNet-152.

4. Results

The preceding section outlines the data analysis procedure, including the stages of data preparation, research steps aligned with the selected approach, and model evaluation. In this chapter, we offer a comprehensive insight into the process and outcomes of classifying chest X-rays (CXR) using a Convolutional Neural Network (CNN) with the ResNet-50, ResNet-101, and ResNet-152 architectures.

In this section, we delve into the training process of the three ResNet architectures, aiming to achieve precise classification of CXR images. Each incoming image serves as input and undergoes an initial convolutional layer with kernel sizes of 77% and 33%. The ResNet architecture is partitioned into four stages, commencing with the first stage, which comprises three residual blocks, each consisting of three layers. The convolutional operations within the layers of the first-stage blocks employ kernel sizes of 64, 64, and 128, respectively.

The number of layers for each architecture differs in the subsequent stages. In the second stage, ResNet-50 and ResNet-101 have four layers, whereas ResNet-152 boasts eight. In terms of kernel sizes, the second stage employs sizes of 128, 128, and 512 for each architecture, respectively. Moving to the third stage, ResNet-50 employs six layers, ResNet-101 incorporates 23 layers, and ResNet-152 features an impressive 36 layers. Here, the kernel sizes are 256, 256, and 1024 for the respective architectures. In the fourth stage, all three architectures share the same number of layers. The kernel sizes applied in each architecture for this stage are 512, 512, and 2048, respectively. Furthermore, the network incorporates an average pooling layer, followed by a fully connected layer consisting of 1000 neurons that represent the ImageNet classes.

A bottleneck design is used for deeper networks such as ResNet-50, ResNet-101, and ResNet-152. Three layers are stacked together for each residual function. These layers are made up of 11, 33, and 11 convolutions. The dimensions are reduced and then restored by the 1-1 convolution layer. With smaller input and output dimensions, the three-layer system acts as a bottleneck. Skip connections are performed for three layers in the ResNet architecture. The convolution operation in the residual block is performed with a stride size of two, reducing the input size to half of its height and width when transitioning from one stage to the next.

Following the training process, the model is evaluated to determine its performance. The model's performance is determined by making predictions on the test data and comparing the results to the true labels. The resulting confusion matrix is shown below.

According to Table 4, the research findings do not indicate the occurrence of overfitting. The small difference in accuracy values between the training and testing processes demonstrates this. Similarly, there are no significant differences between the training and testing loss values.

Table 4. The comparison of the accuracy and loss of the three models.

Model	Accuracy		Loss	
	Training Set	Testing Set	Training Set	Testing Set
ResNet-50	0.92	0.91	0.172	0.240
ResNet-101	0.94	0.93	0.157	0.214
ResNet-152	0.95	0.94	0.133	0.153

Table 5 compares the classification performance of the three ResNet architecture models on CXR images. To evaluate the performance of each model, we examine the values of each metric used, namely recall, precision, and F1-score. The ResNet-152 architecture achieves the highest performance in each metric by using a learning rate of 10-6. The ResNet-50 architecture, on the other hand, yields the model with the lowest performance in each metric.

Table 5. The comparison of the average recall, precision, and f-1 score of the three models.

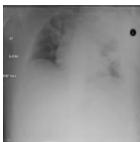
Method	Recall	Precision	F1-Score
ResNet-50	91.75	90.00	90.75
ResNet-101	93.25	93.00	93.25
ResNet-152	94.25	94.25	94.50

Table 6 provides a visual representation of the identification outcomes based on the different ResNet architectures utilized in this study. From the information presented, it is evident that the three ResNet architectures performed reasonably well in recognizing various lung diseases among patients. However, it is important to note that these models are not without their limitations. One notable challenge is distinguishing between the lung opacity and normal classes, as they share similar characteristics, making accurate recognition and classification a challenging task. This highlights the fact that despite significant technological advancements, there is still room for improvement, especially in enhancing model accuracy in scenarios where visual characteristics overlap or are similar.

Table 6. Visual disease identification results.

Dataset							
Actual							
ResNet-50	COVID-19	Lung Opacity	Normal	Pneumonia	Lung Opacity	COVID-19	Lung Opacity
COVID-19	Normal	Lung Opacity	Normal	Pneumonia	Normal	Pneumonia	Normal
ResNet-101	COVID-19	Normal	Lung Opacity	Pneumonia	Lung Opacity	COVID-19	Normal
COVID-19	Normal	Lung Opacity	Lung Opacity	Pneumonia	Lung Opacity	COVID-19	Lung Opacity
ResNet-152	COVID-19	Normal	Lung Opacity	Pneumonia	Lung Opacity	COVID-19	Normal

Table 6. Cont.

Dataset							
							
Actual	COVID-19	Lung Opacity	Normal	Pneumonia	Lung Opacity	COVID-19	Lung Opacity
ResNet-50	COVID-19	COVID-19	Normal	Pneumonia	Lung Opacity	Pneumonia	Normal
ResNet-101	COVID-19	Lung Opacity	Normal	Pneumonia	Lung Opacity	COVID-19	Normal
ResNet-152	COVID-19	Lung Opacity	Normal	Pneumonia	Lung Opacity	COVID-19	Lung Opacity

5. Discussion

In this study, three different ResNet architectures were employed for the classification of four distinct class categories: COVID-19, pneumonia, lung opacity, and normal. The CXR data used in this research were sourced from Chowdhry et al. [34] and Rahman, T. et al. [35] updates. Specifically, the study focused solely on images of normal lungs, pneumonia, lung opacities, and COVID-19 patient lungs, omitting consideration of other lung-related conditions such as fractures, pneumothorax, or heart failure.

The third model was trained using a dataset comprising 16,933 images, which included 2893 COVID-19 images, 1076 pneumonia images, 4810 lung opacification images, and 8730 normal lung images. To ensure model performance and prevent overfitting, model testing was carried out using a test dataset. It is worth noting that this dataset only included four class classifications, but there is hope that additional class categories will be introduced in the future. This expansion would enable the model to identify a wider range of lung diseases comprehensively, thereby enhancing its utility for medical professionals in diagnosing various lung conditions with greater specificity.

In this research, we used three ResNet architectures with different levels of depth. These architectures were ResNet-50, ResNet-101, and ResNet-152. The depth of each architecture corresponded to its respective name. ResNet-50 consisted of one convolution layer 7×7 , followed by maxpooling 3×3 , then entered the residual block section with skip connections, each consisting of three convolution layers of size 1×1 and 3×3 , three for the first part, four for the second part, six for the third part, and three for the fourth part. Therefore, the total layers used in this architecture were 50 layers and produced an f-1 score of 90.75. The second architecture, namely ResNet-101, had the addition of a convolution layer from the ResNet-50 architecture, namely in the third part, the residual block and skip connection, so that three convolution layers measuring 1×1 and 3×3 became 23 pairs. This had an impact on the resulting F1-score value of 93.25. Likewise, for the third architecture, namely ResNet-152, which was a development of the previous ResNet architecture because it increased the number of convolution layers in the second and third residual blocks and skip connections to 8 pairs and 36 pairs, respectively. As with ResNet-101, which experienced an increase in the resulting F1-score value, ResNet-152 produced the highest F1-score value among the other ResNet architectures used, namely 94.50. This experiment illustrates that adding the number of layers from ResNet-50 to ResNet-101 can yield an improvement in model performance of up to 2.5%. A similar result was achieved when extending ResNet-101 by adding layers to ResNet-152, which had an increasing performance of 1.5%.

The three architectures were used to create an X-ray image classification model with high evaluation values. The primary goal of CNN modeling is to achieve maximum accuracy with the least amount of loss. The assessment matrix found that each model had limitations in identifying lung opacity and normal classes due to a very large error value. According to the Centers for Disease Control and Prevention (CDC) [45], this challenge can arise due to the presence of two distinct types of lung opacity: small opacity and large opacity. Small opacity refers to opacities with dimensions less than 10 mm, while

large opacity refers to opacities with dimensions greater than 10 mm. This distinction between the two types of opacities introduces complexity in classification, contributing to the observed difficulties in accurate identification. As a result, the models run the risk of misclassifying lung opacity and normal as having small opacity.

The ResNet-152 architecture demonstrated powerful performance compared to the other architectures across various metrics, including recall, precision, and F1 score, as indicated by the results of the confusion matrix evaluation. This aligns with previous research findings that suggest higher accuracy can be achieved with deeper architectures [32]. The ResNet-152 model's significantly greater number of layers compared to other ResNet architectures allowed it the learning of more complex patterns and consequently achievement of improved performance. According to the analysis results, the time required for one epoch of training was 461 s for ResNet-152, 311 s for ResNet-50, and 549 s for ResNet-101. This discrepancy in training times is attributed to the fact that as the number of layers in an architecture increases, so does the number of parameters used. ResNet-152 boasted 58 million parameters, while ResNet-50 had 23 million, and ResNet-101 had 42 million. Consequently, training times were extended. Therefore, the choice of architecture should be aligned with the specific requirements of the research or use case. While ResNet-152 offered superior performance, it came at the cost of increased computational demands, so researchers should consider their computational resources and objectives when selecting the most appropriate architecture.

Deployment, as a crucial step in facilitating the diagnosis of lung problems in patients by medical personnel, involves the creation of an application-based model for enhanced usability. This process entails the necessity of retraining the model to prevent any potential decline in performance, which could have critical consequences. Moreover, ongoing validation and supervision by radiologists ensures that the model maintains its accuracy over time. However, when considering the broader application of the model to generalize for all lung disease patients across diverse demographics, geographic locations, and healthcare service standards, it is evident that the current dataset remains limited and constrained. Consequently, there is a clear need for the acquisition of additional datasets. These new datasets are essential to train the model effectively, allowing it distinguishing between various types of data variance and ensuring its robust performance in diverse real-world scenarios. Expanding the dataset to encompass a wider range of cases and demographics enhances the model's ability to provide accurate diagnoses and predictions for a more diverse patient population.

6. Conclusions

The models were trained on a dataset consisting of four classes: pneumonia, COVID-19, lung opacity, and normal in a study on the classification of CXR images for pneumonia and COVID-19 using convolutional neural networks with three different architectures, namely ResNet-50, ResNet-101, and ResNet-152. The performance of each architecture was measured by recall, precision, and F1-scores. ResNet-50 achieved the best precision with a value of 92% in the normal and lung opacity classes, the highest recall value of 96% in the pneumonia class, and the F1-score value of 91%. The ResNet-101 architecture produced the highest precision value in the COVID-19 class of 95%, the highest recall value in the pneumonia class of 96%, and a f1-score of 93%. The ResNet-152 architecture produced the highest precision value in the COVID-19 class of 97%, the highest recall value in the pneumonia class of 97%, and a F1-score of 94%. Because ResNet-152 outperformed all other ResNet models in classifying CXR images in pneumonia and the COVID-19 disease, the ResNet-152 is the best model for classifying CXR images in pneumonia and the COVID-19 disease.

To avoid data imbalance, the research data for each class should be improved in the future, particularly for classes with fewer data points than other classes. A larger dataset is expected to be used in future research to identify other lung diseases in a more specific manner.

Future research should result in a more adaptive model that can continuously update during the training process by incorporating new data. Other methods can be combined with the ResNet architecture to produce improved methods and architectures in the future. The study's findings, which are currently in the form of a model, are suggested to be used in an application to help medical professionals effectively diagnose lung diseases under the supervision of radiology experts.

Author Contributions: S.A.H., A.A.P., A.S.A., I.N.Y. and M.H.A. understood and designed the research; S.A.H. analyzed the data and drafted the paper. All authors critically read and revised the draft and approved the final paper. All authors have read and agreed to the published version of the manuscript.

Funding: The authors are grateful to the Research Center for AI and Big Data Universitas Padjadjaran and the Directorate for Research and Community Service (DRPM) Ministry of Research, Technology, and Higher Education Indonesia which supports this research under Higher Education Research Grant no. 094/E5/PG.02.00.PT/2022.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available at [34,35].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McIntosh, K. Community-Acquired Pneumonia in Children. *N. Engl. J. Med.* **2002**, *346*, 429–437. [[CrossRef](#)]
2. Badrahadipura, R.; Septi, S.Q.N.; Fachrel, J.; Yulita, I.N.; Pravitasari, A.A.; Agustian, D. COVID-19 Detection In Chest X-rays Using Inception Resnet-v2. In Proceedings of the 2021 International Conference on Artificial Intelligence and Big Data Analytics, Bandung, Indonesia, 27–29 October 2021; pp. 104–109. [[CrossRef](#)]
3. Fiore, K. Hazy on Ground-Glass Opacities? Here's What They Are. Medpagetoday. Available online: <https://www.medpagetoday.com/pulmonology/generalpulmonary/86751> (accessed on 13 January 2023).
4. Goyal, S.; Singh, R. Detection and classification of lung diseases for pneumonia and COVID-19 using machine and deep learning techniques. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *14*, 3239–3259. [[CrossRef](#)]
5. Tang, N.; Li, D.; Wang, X.; Sun, Z. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *J. Thromb. Haemost.* **2020**, *18*, 844–847. [[CrossRef](#)] [[PubMed](#)]
6. Shi, F.; Wang, J.; Shi, J.; Wu, Z.; Wang, Q.; Tang, Z.; He, K.; Shi, Y.; Shen, D. Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19. *IEEE Rev. Biomed. Eng.* **2020**, *14*, 4–15. [[CrossRef](#)] [[PubMed](#)]
7. Langer, S.G.; Carter, S.J.; Haynor, D.R.; Maravella, K.R.; Mattes, B.D.; Strandness, J.E.D.; Stewart, B.K. Image Acquisition: Ultrasound, Computed Tomography, and Magnetic Resonance Imaging. *World J. Surg.* **2001**, *25*, 1428–1437. [[CrossRef](#)]
8. Zhi, L.; Jiang, W.; Zhang, S.; Zhou, T. Deep neural network pulmonary nodule segmentation methods for CT images: Literature review and experimental comparisons. *Comput. Biol. Med.* **2023**, *164*, 107321. [[CrossRef](#)] [[PubMed](#)]
9. Prakash, S.; B, R. Detection and Classification of Thoracic Diseases in Medical Images Using Artificial Intelligence Techniques: A Systematic Review. *ECS Trans.* **2022**, *107*, 307–340. [[CrossRef](#)]
10. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506. [[CrossRef](#)] [[PubMed](#)]
11. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *arXiv* **2015**, arXiv:1512.03385.
12. Hu, Y.; Huber, A.; Anumula, J.; Liu, S.-C. Overcoming the vanishing gradient problem in plain recurrent networks. *arXiv* **2019**, arXiv:1801.06105.
13. Sarton, G. Wilhelm Conrad Röntgen and the Early History of the Roentgen Rays. Otto Glasser, Margret Boveri. *Isis* **1934**, *22*, 256–259. [[CrossRef](#)]
14. Winsberg, F.; Elkin, M.; Macy, J.; Bordaz, V.; Weymouth, W. Detection of Radiographic Abnormalities in Mammograms by Means of Optical Scanning and Computer Analysis. *Radiology* **1967**, *89*, 211–215. [[CrossRef](#)]
15. Chen, C.-M.; Chou, Y.-H.; Tagawa, N.; Do, Y. Computer-Aided Detection and Diagnosis in Medical Imaging. *Comput. Math. Methods Med.* **2013**, *2013*, 790608. [[CrossRef](#)]
16. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2015**, arXiv:1409.1556.
17. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A.; Liu, W.; et al. Going deeper with convolutions. *arXiv* **2014**, arXiv:1409.4842.
18. Glorot, X.; Bengio, B. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; Volume 9.

19. Drozdzal, M.; Vorontsov, E.; Chartrand, G.; Kadoury, S.; Pal, C. The Importance of Skip Connections in Biomedical Image Segmentation. In *Deep Learning and Data Labeling for Medical Applications*; Carneiro, G., Mateus, D., Peter, L., Bradley, A., Tavares, J.M.R.S., Belagiannis, V., Papa, J.P., Nascimento, J.C., Loog, M., Lu, Z., et al., Eds.; DLMIA LABELS 2016 2016; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; pp. 179–187. [[CrossRef](#)]
20. Jusman, Y. Comparison of Prostate Cell Image Classification Using CNN: ResNet-101 and VGG-19. In Proceedings of the IEEE 13th International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 25–26 August 2023. [[CrossRef](#)]
21. Chhabra, M.; Kumar, R. An Efficient ResNet-50 based Intelligent Deep Learning Model to Predict Pneumonia from Medical Images. In Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 7–9 April 2022; pp. 1714–1721. [[CrossRef](#)]
22. Abu Al-Haija, Q.; Adebanjo, A. Breast Cancer Diagnosis in Histopathological Images Using ResNet-50 Convolutional Neural Network. In Proceedings of the IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Vancouver, BC, Canada, 9–12 September 2020; pp. 1–7. [[CrossRef](#)]
23. Reddy, A.S.B.; Juliet, D.S. Transfer Learning with ResNet-50 for Malaria Cell-Image Classification. In Proceedings of the International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 4–6 April 2019; pp. 945–949. [[CrossRef](#)]
24. Alam Shadab, S.; Ansari, M.; Singh, N.; Verma, A.; Tripathi, P.; Mehrotra, R. Detection of cancer from histopathology medical image data using ML with CNN ResNet-50 architecture. *Comput. Intell. Healthc. Appl.* **2022**, *237–254*. [[CrossRef](#)]
25. Ghosal, P.; Nandanwar, L.; Kanchan, S.; Bhadra, A.; Chakraborty, J.; Nandi, D. Brain Tumor Classification Using ResNet-101 Based Squeeze and Excitation Deep Neural Network. In Proceedings of the Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India, 25–28 February 2019; pp. 1–6. [[CrossRef](#)]
26. Lashgari, E.; Liang, D.; Maoz, U. Data augmentation for deep-learning-based electroencephalography. *J. Neurosci. Methods* **2020**, *346*, 108885. [[CrossRef](#)]
27. Xue, C.; Dou, Q.; Shi, X.; Chen, H.; Heng, P.-A. Robust Learning at Noisy Labeled Medical Images: Applied to Skin Lesion Classification. In Proceedings of the IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 1280–1283. [[CrossRef](#)]
28. Abirami, R.N.; Vincent, P.M.D.R.; Rajinikanth, V.; Kadry, S. COVID-19 Classification Using Medical Image Synthesis by Generative Adversarial Networks. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **2022**, *30*, 385–401. [[CrossRef](#)]
29. Athisayamani, S.; Antonyswamy, R.S.; Sarveshwaran, V.; Almeshari, M.; Alzamil, Y.; Ravi, V. Feature Extraction Using a Residual Deep Convolutional Neural Network (ResNet-152) and Optimized Feature Dimension Reduction for MRI Brain Tumor Classification. *Diagnostics* **2023**, *13*, 668. [[CrossRef](#)] [[PubMed](#)]
30. Roy, P.; Chisty, M.O.; Fattah, H.A. Alzheimer’s Disease Diagnosis from MRI images using ResNet-152 Neural Network Architecture. In Proceedings of the 5th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 17–19 December 2021; pp. 1–6. [[CrossRef](#)]
31. Ma, L.; Shuai, R.; Ran, X.; Liu, W.; Ye, C. Combining DC-GAN with ResNet for blood cell image classification. *Med. Biol. Eng. Comput.* **2020**, *58*, 1251–1264. [[CrossRef](#)]
32. Abubakar, A.; Ajuji, M.; Yahya, I.U. Comparison of Deep Transfer Learning Techniques in Human Skin Burns Discrimination. *Appl. Syst. Innov.* **2020**, *3*, 20. [[CrossRef](#)]
33. Setiawan, A.W. The Effect of Image Dimension and Exposure Fusion Framework Enhancement in Pneumonia Detection Using Residual Neural Network. In Proceedings of the International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 17–18 September 2022; pp. 41–45. [[CrossRef](#)]
34. Chowdhury, M.E.H.; Rahman, T.; Khandakar, A.; Mazhar, R.; Kadir, M.A.; Bin Mahbub, Z.; Islam, K.R.; Khan, M.S.; Iqbal, A.; Al Emadi, N.; et al. Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access* **2020**, *8*, 132665–132676. [[CrossRef](#)]
35. Rahman, T.; Khandakar, A.; Qiblawey, Y.; Tahir, A.; Kiranyaz, S.; Kashem, S.B.A.; Islam, M.T.; Al Maadeed, S.; Zughayer, S.M.; Khan, M.S.; et al. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Comput. Biol. Med.* **2021**, *132*, 104319. [[CrossRef](#)]
36. Elgendi, M.; Nasir, M.U.; Tang, Q.; Smith, D.; Grenier, J.-P.; Batte, C.; Spieler, B.; Leslie, W.D.; Menon, C.; Fletcher, R.R.; et al. The Effectiveness of Image Augmentation in Deep Learning Networks for Detecting COVID-19: A Geometric Transformation Perspective. *Front. Med.* **2021**, *8*, 629134. [[CrossRef](#)]
37. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
38. Ahuja, S.; Panigrahi, B.K.; Dey, N.; Rajinikanth, V.; Gandhi, T.K. Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices. *Appl. Intell.* **2020**, *51*, 571–585. [[CrossRef](#)] [[PubMed](#)]
39. Panwar, H.; Gupta, P.; Siddiqui, M.K.; Morales-Menendez, R.; Singh, V. Application of deep learning for fast detection of COVID-19 in X-rays using nCOVnet. *Chaos Solitons Fractals* **2020**, *138*, 109944. [[CrossRef](#)]
40. Taylor, L.; Nitschke, G. Improving Deep Learning with Generic Data Augmentation. In Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 18–21 November 2018; pp. 1542–1547. [[CrossRef](#)]
41. Monshi, M.M.A.; Poon, J.; Chung, V.; Monshi, F.M. CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR. *Comput. Biol. Med.* **2021**, *133*, 104375. [[CrossRef](#)] [[PubMed](#)]

42. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [[CrossRef](#)]
43. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
44. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. *arXiv* **2018**, arXiv:1708.02002.
45. Centers for Disease Control and Prevention. Study Syllabus for Classification of Radiographs of Pneumoconioses. Available online: <https://www.cdc.gov/niosh/learning/b-reader/radiograph/subset4/1.html> (accessed on 30 April 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.