Journal of Medical and Biological Engineering (2021) 41:581–591
https://doi.org/10.1007/s40846-021-00656-6

ORIGINAL ARTICLE

Lung Opacity Pneumonia Detection with Improved Residual Networks

Ghazanfar Latif[1,2] · **Faisal Yousif Al Anezi**[3] · **Fadi N. Sibai**[1] · **Jaafar Alghazo**[4]

Received: 13 May 2020 / Accepted: 1 September 2021 / Published online: 8 September 2021

Abstract

Purpose Pneumonia detection is usually done by specialized highly trained specialists through the review of chest radiographs combined with vital signs, laboratory exams, and medical history. Though the presence of pneumonia in images appears in the form of areas of high opacity in the lungs which is difficult to differentiate from opaque areas caused by other conditions. The medical diagnosis of pneumonia is a very costly and time-consuming process. The motivation of this work is to automate the diagnosis process of pneumonia through image processing. In this paper, an automated pneumonia detection system is proposed using improved deep residual networks (ResNet) architectures, which are tested on the dataset of 30,227 DICOM Chest X-rays. Dataset was divided into 80% training and 20% testing, with 20% of the 80% used for training dedicated to validation.

Methods Two residual network models were used (Version 1 & Version 2), and results were also compared with three different CNN models as well as methods found in recent literature.

Results The overall results indicate that the proposed ResNet (Version 2) method achieves higher accuracy than convolution neural networks and other recently proposed methods (Table 5). The proposed ResNet network of a depth of 110 and a batch size of 16 with epochs 80 achieved an average accuracy of 88.67%.

Conclusion An automated method is proposed and implemented in this work for the proper diagnosis of pneumonia using images of the DICOM chest X-rays dataset. The proposed method in this paper outperforms other methods from recent literature.

Keywords Lung opacity · Pneumonia detection · Residual networks · CNN · Chest X-ray

1 Introduction

The human lungs are an integral part of the pulmonary system as well as waste disposal for the body. They are complex organs that are formed by a collection of different cells and tissues; each is specialized in carrying out certain functionalities. Lungs can be affected by a range of diseases that vary from mild to fatal according to symptoms that distort the lungs' behavior. Lung infections can be either acute or chronic and are caused by bacteria, viruses, or fungi. Pneumonia is an acute lung tissue inflammation caused by germs bypassing the nose and entering the lungs. Such germs cause fluids in the air sacs in the lungs with symptoms such as cough, fever, and breathing problems. Pneumonia is the leading killer disease among children under the age of five, causing the death of around 880,000 children in 2016 [1]. Chest radiograph or CT (computed tomography) scan helps healthcare professionals diagnose lung infections and examine lung structure.

Chest X-Rays are primarily used to examine some body organs, such as the lungs and heart. CT scan is one of the commonly performed medical tests as it provides lungs, airways, blood vessels, spinal bones, and chest imagery. It is also used to diagnose and evaluate the treatment of some lung infections, such as pneumonia [2]. X-rays pass through any matter, although some objects absorb it heavily whilst some allow the radiation to pass through them.

✉ Ghazanfar Latif
glatif@pmu.edu.sa

1 Computer Science Department, Prince Mohammad bin Fahd University, Al-Khobar, Saudi Arabia

2 Université du Québec a Chicoutimi, 555 boulevard de l'Université, Chicoutimi, QC G7H2B1, Canada

3 Management Information Department, Prince Mohammad bin Fahd University, Al-Khobar, Saudi Arabia

4 Electrical and Computer Engineering Department, Virginia Military Institute, Lexington, VA, USA

Correspondingly, air appears black, soft tissues appear grey, and bones are white in the X-Ray image.

Chest X-Rays are used for diagnostic purposes to provide doctors with a clear visualization of the case. The symptoms of pneumonia are similar to those of a regular fever, which may lead to inaccurate diagnosis. However, the lungs opacities that are the X-Ray findings of pneumonia have an unclear boundary making it more challenging to detect by the radiologist. Therefore, the proposed algorithm will help in detecting lung opacities that result in the lungs, as there is no clear border where the infection of pneumonia will stop. Moreover, this algorithm will reduce the possibility of the misdiagnosis and allow the radiologist to detect pneumonia faster. Hence, employing machine learning techniques to detect such a disease will enhance the diagnosis process and help radiologists achieve better results. Early diagnosis means early medical intervention, which results in a higher probability of treatment and a higher survival rate. The automation of medical diagnosis using image classification and segmentation is an area of research explored in recent literature, for example, in [3, 4]. Different image processing techniques are proposed to detect and classify Diabetic Retinopathy, Skin Lesions, and Glioma Brain tumors [5–7]. But these are just a few examples of a whole emerging field of automatic medical diagnosis research.

The objective of this work is to provide a solution that enhances the detecting of pneumonia disease using supervised machine learning techniques. The contribution of the work presented is in proposing two Residual Network Deep Learning Architectures (Version 1 & Version 2) for the specific use of automatic pneumonia diagnosis using Chest X-ray images while solving the problems of overfitting and network degradation. Two Deep Residual Networks were proposed and referred to as "Version 1" and "Version 2". The proposed versions were thoroughly tested using the Chest X-ray DICOM images dataset consisting of 30,227 images. The testing was repeated for three conventional CNN models. The results were compared with those obtained from methods proposed in recent literature. Parameter manipulation was also used to arrive at the best architecture for the proposed ResNet Networks, including batches and epochs.

The rest of the paper is organized as follows: Sect. 2 details the literature review, Sect. 3 details the dataset, Sect. 4 describes the Residual Networks and proposed models. Sections 5 and 6 discuss the experimental results and discussions. Finally, Sect. 7 concludes the paper.

2 Literature Review

In [8], the authors developed an algorithm that can detect pneumonia symptoms at a level that exceeds the radiologist's level. The model ChexNet presents a convolutional neural network that contains 121-layers that inputs chest radiographs and produces an output of the localized areas that indicate the presence of pneumonia. The dataset used is Chest X-ray 14, which contains over 112,000 chest X-Ray images of more than 30,000 different patients [9]. Then the results of the model are compared with the radiologist's performance of pneumonia diagnosis. Results show that ChexNet performance has a significant improvement over the radiologist's performance. In [10], the paper introduces the detection of lobar pneumonia statistically while using digitized chest radiograph films. It essentially proposes to have different targeted regions labeled by a vector of wavelet texture measures multiplied by an orthogonal matrix. Moreover, discriminant analysis was used to estimate the existence of any false classification probabilities. Based on results, the paper suggests detecting pneumonia by constructing a discriminant function through the employment of maximum column sum. The method results in a lower misclassification rate with only a 15% error rate.

In [11], the authors propose a method consisting of three main phases. First, apply preprocessing techniques to adjust the display of the image and remove unwanted regions. Second, extract four main features based on specific criteria that involve a shape descriptor, body size ratio, image profile, and histogram pyramid. Then images are classified as frontal/lateral during the training phase using the supervised classifier called support vector machine (SVM). In [12], the proposed research aims to enhance the classification performance of medical images and reduce the required time for training and testing. For images to be classified, the wavelet-based center symmetric local binary pattern (WCS-LBP) features are extracted during training. Next, based on a decision tree, a random forest is used to test the image and classify it into the corresponding category. The dataset used includes CLEF-Med2007 images, and the experimental results showed significant improvement in classification performance. In [13], a semi-supervised learning algorithm is used to classify tuberculosis chest X-rays. The research proposes a method containing labeled and unlabeled data through either self-training, co-training, and tri-training, then classifies according to the majority vote. The suggested algorithm has shown its efficiency with nonparametric statistical tests and proved that the utilization of limited-number labels could be maximized in favor of unlabeled data.

In [14], the proposed method used deep feature fusion to reduce the probability of false-positive results in lung nodule diagnosis. Three modules were proposed to obtain more accurate results. The first module suppresses ribs from chest X-rays to provide a more visible display of the lungs. In the second module, lungs are segmented to locate the region of interest, which is indicative of the abnormalities in the lungs. The third module applies the Laplacian of Gaussian (LoG) to classify suspicious nodules. The modules were tested on

the Japanese Society of Radiological Technology (JSRT) dataset. The results conclude that the preprocessing phase, consisting of the three modules, significantly improves the chest X-ray analysis. With regards to the evaluation of deep convolutional neural networks (DCNN) for the detection of illnesses on chest radiographs for automated classification [15], a total of four datasets that contain 1007 chest radiograph images were used. The applied method partitioned the datasets into training, validation, and test. In cases of disagreements, an experienced radiologist was asked to interpret images and determine any potential cases. As a result, DCNNs were able to classify chest X-rays with an Area Under Curve (AUC) of 0.99, an immense improvement in the accuracy was obtained by consulting a radiologist for false classification cases. Given the importance of the early detection of Cancer, in [16], the authors proposed a method for the classification of lung CT scan images using an artificial neural network. By segmenting the lung from CT images, parameter calculations were done for the segmented part. Feedforward and feedforward backpropagation neural networks are used in the classification process to provide higher accuracy. The first function showed a classification accuracy of 93.3% and a mean square error of 0.998, while the second function showed an accuracy of 93.3% with a mean square error of 0.0942.

Convolutional neural networks (CNN) are widely used in image classification and image recognition due to their high efficiency and accuracy. In [17], the authors proposed classifying frontal chest x-rays using CNNs given an efficiently large data set. GoogleLeNet was used as the CNN and accordingly trained with the help of 3 GPUs for automatic classification of radiographs as either normal or abnormal. The evaluation of the network's overall performance was tested using 2443 radiograph images. The method achieved an integral of 0.964 and specificity of 91%. In [18], the authors used geometric feature descriptors to classify lung abnormalities, such as nodules, and further lower false positives in low-dose computed tomography. Through creating modules for lung nodules then applying the detection, descriptors are applied in the last step; hence, being able to detect possible candidates. The result of the proposed method is an enhancement of 2% for the detection of nodules. In [19], the authors proposed an algorithm that aims to identify abnormalities in chest x-rays. The algorithm utilized supervised machine learning techniques on two types of images classified as normal and abnormal and trained to predict the probability of abnormalities in the radiographs. The algorithm used a data set that consists of 514 abnormal and 431 normal CXRs. Results indicated an increased accuracy in identifying abnormalities in chest X-Rays.

In [20], the authors proposed a method for detecting infectious pulmonary tuberculosis (TB) from chest x-rays through segmentation using the hybrid information-oriented Bayesian classification approach. The method also utilizes other techniques, such as the gradient inverse coefficient of variation (GICOV) that serves as a classification detector and confirmer of a true case of TB cavities. Results indicate high accuracy and a low false-positive rate compared with other non-hybrid approaches. In [21], the authors proposed a method to enhance the active shape model (ASM) with an automatic lung field segmentation approach. Improvement in the performance, with both normal and abnormal chest X-rays, was observed with a 3–6% accuracy enhancement compared to previous ASM techniques. In [22], the authors proposed the use of deep classification techniques of Convolutional neural networks (CNN) to identify abnormalities in chest x-rays. The method used ImageNet for CNN models, pre-trained, and feature extraction performed using a classification approach. The data used for algorithm testing consists of 433 images. Results demonstrate that chest x-rays can be analyzed using the training phase on non-medical images, which introduces a new approach to lung disease detection.

In [23], researchers used the RSNA dataset (21,684 images for training, 2000 images for validation, and 3000 images for testing) employed deep ResNet and Mask-RCNN networks achieving 85.6% accuracy. Donthi et al. [24] broke down the RSNA dataset to 26,000 images for training and 3000 images for testing and employed a CNN to achieve 78.9% accuracy. Sirazitdinov et al. [25] used an ensemble of two networks, RetinaNet and Mask-RCNN, achieving a precision of 0.883 and 0.652 recall. In [26], CNNs were trained with 21,152 images and tested with 4532 images to reach an accuracy of 86.38%.

3 Materials and Methods: Residual Neural Networks

The accuracy of deep neural networks is expected to improve with deeper networks, at least in theory. One problem with deep neural networks with hundreds of layers is that the training error and accuracy drop as the number of network layers increases, reaching a minimum then starting to increase for even larger numbers of layers. This problem is commonly referred to as the neural network degradation problem. This means that deep neural networks cannot grow in depth indefinitely with more non-linear convolution/ReLU layers. ReLU refers to a rectified linear unit, which applies the function $g(x) = \max(0, x)$, that removes negative values from an activation map by setting them to zero. Another problem is that before ResNet, it was difficult for the network to approximate the identity mappings of added non-linear layers.

ResNet was a breakthrough in deep learning and solved the above problems allowing much deeper networks to be

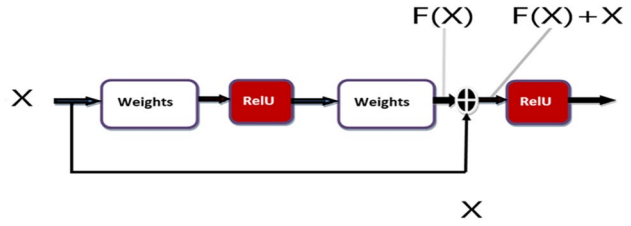

Fig. 1 Simple notation of ResNet block

trained with satisfactory accuracies [27]. ResNet is built with several ResNet blocks, where a ResNet block is displayed in Figs. 1 and 2. A ResNet block is built from 4 layers: A Weights layer ($Zn+1 = Wn+1\ Xn+Yn+1$); the ReLU layer (non-linear, $Xn+1 = H(Zn+1)$), followed by another Weights layer ($Zn+2 = Wn+2\ Xn+1+Yn+2$); the combination of these 3 layers takes an input Xn and outputs F(Xn). Note that these variables represent matrices, and the variable's subscript denotes the layer number. Then, instead of applying ReLU on F(Xn) (=Zn+2), to produce $Xn+2 = H(Zn+2)$ as non-residual networks do, ResNet applies ReLU (4th layer) on F(Xn) (=Zn+2) to produce $Xn+2 = H(Zn+2+Xn)$. In other words, a skip or short cut link bypasses the three layers and passes Xn to an adder, which adds $F(Xn) = H(Zn+2)$ to Xn before it goes through a second ReLU layer to finally generate Xn+2.

The X input to the adder is referred to as the shortcut or skip connection. If X is required to be passed as-is from one layer to another, the skip connection allows a residual learning network to learn that F(X) is 0, which, when added to X, gives X a simple task to perform. Without such skip connection, the network would have to learn that the Weights layer is equivalent to (multiplying X by) the Identity matrix, a more difficult task than teaching the network that F(x) is the 0 matrix. When X does not need to be passed from one layer to another, the network would learn the F(x) value as in normal neural networks using backpropagation. In that case, if the desired output of a layer is D(x), then it would be easier to train F(x) to be the ""residual":" D(X)-X, which after being added to X (via the skip connection) yields the desired D(X). As the skip connection does not involve weights, the gradient value will not change, and the problem of vanishing gradients occurring in the early layers of very deep networks is avoided.

A ResNet architecture puts several ResNet blocks together in a sequence to build deeper neural networks with good accuracy and low training error. Some of these blocks may be pooling blocks as when some Convolution or Weights layers generate an F(X) matrix of different size as the X matrix, in order to add F(X) to X, where X is resized to match the size of the F(X) matrix. This process is achieved by adding F(X) to (W. X), where W is a matrix zero-padded in the rows and columns missing in the original X.

3.1 ResNet Version 1

In this paper, two versions of ResNet, which we refer to as simply "version 1" and "version 2" are proposed. Figure 3 shows the differences between version 1 and version 2 at the block level. Figure 3.a depicts the detailed architectures used in ResNet version 1. ResNet does not suffer from overfitting since it does not introduce additional parameters. This means that ResNet is efficient for deep learning with hundreds of network layers. ResNet Version 1 was designed such that the feature map is divided by two at the beginning of each step using a convolutional layer, and the filter size is doubled such that the convolutional layer, batch normalization layer, and ReLU layer are mapped to $32\times32\times16$, $16\times16\times32$, $8\times8\times64$, respectively. The batch

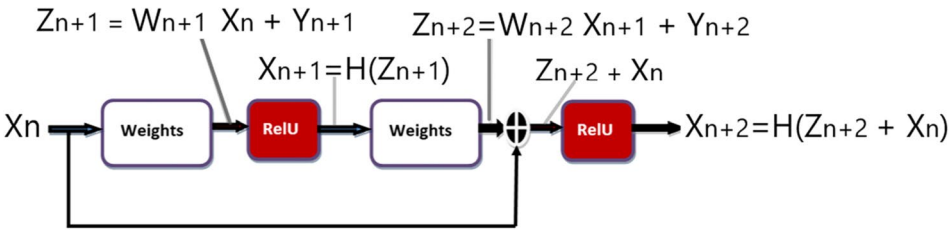

Fig. 2 Detailed notation of ResNet block

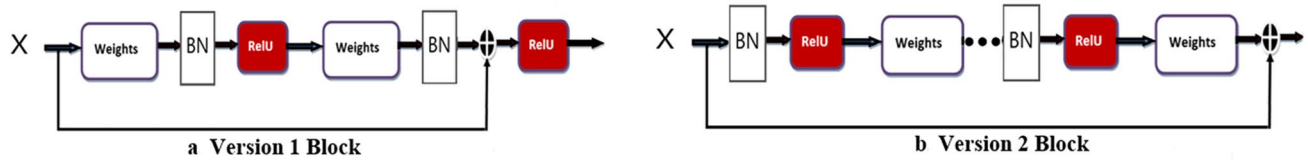

Fig. 3 ResNet Version 1 and Version 2 blocks

normalization block adjusts the input layer to enhance the deep network's performance.

ResNet Architecture Version 1 receives an input image dimension of (48 × 48x3) that is input to the architecture. Each layer in the architecture consists of a convolutional {conv}, Batch Normalization, and Rectified Linear Unit (ReLU).

3.2 ResNet Version 2

Figure 3b depicts the detailed architectures used in ResNet version 2. In version 2, again, the feature maps are halved at the beginning of each stage while doubling the number of filter maps. The stages in version 2 are; convolutional: 32 × 32 × 16, step 1: 32 × 32 × 64, step 2: 16 × 16 × 128, and step 3: 8 × 8 × 256. Epochs sizes were varied to include: 10,20,40, and 80, while batch sizes were chosen to be 16 and 32, respectively.

Note that in both versions 1 and 2, after concatenating multiple blocks together, the same sequence "Weights – Batch Normalization (BN)- ReLU" gets repeated. However, versions 1 and 2 differ as follows:

Initially both versions are identical for the initial sequence of convolution (weights) followed by Batch Normalization followed by an Activation block. Then the two versions differ in the sequences that follow. A major difference between them is that in version 1, the following block sequence is in this order: a convolution block is followed by a Batch Normalization block which is, in turn, followed by an Activation block. In ResNet version 2, the block sequence following the initial block sequence is in this order: a Batch Normalization block is followed by an Activation block, which, in turn, is followed by a convolution block. Thus, version 1 supports post-activation, while version 2 supports pre-activation. ResNet version 1 adds the second ReLU non-linearity after adding x and F(x), while ResNet version 2 has deleted this last ReLU non-linearity (refer to Fig. 3b). This change helps ResNet version 2 pass the output of the addition of the identity mapping and residual mapping with no changes to the next block.

Furthermore, during backpropagation, version 2 of ResNet allows the gradient value at the output layer to be passed back, as is, to the input layer. This step succeeds in eliminating the vanishing gradient issue in deep learning networks with hundreds and even a thousand layers, reduce their training errors, and improve their performance.

The ResNet Version 2 model introduces bottleneck connections whose filter size is calculated, as shown in Fig. 4. In ResNet Version 2, the block size of the shortcut connections is increased to three. The three layers inside a residual function block are convolutional layers of size 1 × 1, 3 × 3, and 1 × 1. The 1 × 1 layers are responsible for decreasing and increasing the input dimensions while 3 × 3 layer becomes the bottleneck with smaller dimensions.

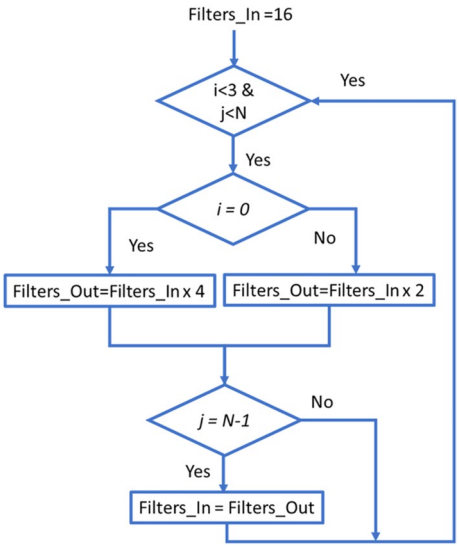

Fig. 4 ResNet Version 2 (conv layer filter size)

4 Experimental Data

The National Institutes of Health Clinical Center provides a publicly accessible large dataset of Chest X-ray DICOM images [28]. An overall of 30,227 CXR 1024 × 1024 size images divided into three classes will be used. The input images are classified into three distinct classes shown in Table 1 which are: A (Normal), B (Pneumonia), and C (Abnormal), along with the number of images of each class within the dataset. Since this study concentrates on diagnosing pneumonia, the dataset will be reduced to two classes: Normal combined with No Lung Opacity (No Pneumonia) and Lung Opacity (Pneumonia).

5 Experimental Results

The experiments were all performed on power computing machines with Graphical Processing Unit (GPU) Nvidia GTX 1080, Central Processing Unit (CPU) of 3.7 Giga-Hertz i7, 32 Gigabyte Random Access Memory (RAM), and using the Windows Server 2016 operating system. The programming of the experiments was done using Python. All phases from preprocessing, classification, and recognition were done using Python 3.7. The experiments are performed using 30,227 CRX images, from which 20% are used for testing, and the rest of the images are further divided into 80% training and a further 20% for validation. As shown in Table 2, ResNet version 1 was implemented

Table 1 Summary of chest X-ray dataset used for experiments

Type	Normal	Lung opacity (pneumonia)	No lung opacity
Target label	0	1	0
Images	8851	11,821	9555
Sample image			

Table 2 Experimental results based on different ResNet version 1 with different models

ResNet blocks	Network depth	Total parameters	Batch size	Epochs	Training Acc (%)	Training loss	Testing Acc (%)	Testing loss
3	3×6+2=20	274,946	16	10	86.30	0.390	83.60	0.462
				20	86.61	0.352	84.20	0.391
				40	87.23	0.323	86.33	0.377
				80	**89.45**	**0.280**	**87.67**	**0.332**
			32	10	81.54	0.484	78.68	0.484
				20	82.41	0.413	80.48	0.450
				40	83.02	0.393	82.30	0.410
				80	83.78	0.379	83.54	0.389
5	5×6+2=32	470,722	16	10	86.02	0.378	85.07	0.396
				20	86.55	0.347	84.80	0.375
				40	87.26	0.315	86.20	0.349
				80	87.69	0.296	85.47	0.343
			32	10	85.98	0.376	82.07	0.484
				20	86.50	0.344	85.60	0.354
				40	86.97	0.319	87.07	0.328
				80	87.74	0.301	86.80	0.322
12	12×6+2=74	1,155,938	16	10	85.71	0.404	84.20	0.424
				20	86.54	0.360	86.73	0.361
				40	86.92	0.320	85.67	0.351
				80	87.75	0.301	84.53	0.383
			32	10	85.51	0.384	85.73	0.365
				20	86.03	0.349	85.67	0.349
				40	86.70	0.332	85.60	0.324
				80	87.22	0.315	86.13	0.345

The bold values indicate the highest accuracy for each model

with a depth of 20, 32, and 74 with 3, 5, and 12 number blocks, respectively. The batch sizes were varied between two values, 16 and 32. For each experiment, Epochs of 10, 20, 40, and 80 were used. There is no overfitting as the accuracies rise and losses drop with increasing epochs. With a ResNet Version 1 network depth of 20, the smaller batch size of 16 resulted in better validation and testing accuracies as well as lower validation and testing losses. As the network depth is increased to 32 with about 470 K parameters and a batch size of 32, the validation and testing accuracies are worse than for a network depth of 20, batch size of 16 with 275 K parameters. The results keep worsening with a higher network depth of 74, batch size of 16, and 1.15 M parameters.

In ResNet V1, each block consists of 6 layers in addition to two layers at the end of the network (Flatten layer and Dense Layer). In ResNet V2, each block consists of 9 layers in addition to two layers at the end of the network (Flatten Layer and Dense Layer). Therefore, in the experimental setup, the combination of networks are chosen according to the same number of blocks but will produce

different depths. The depths of the network are calculated as follows:

V1 Network Depth: $\#\,of\ Blocks \times 6 + 2$

V2 Network Depth: $\#\,of\ Blocks \times 9 + 2$

Therefore, if we assume three blocks are used, the network depth for V1 and V2 will be 20 and 29, respectively. Thus, the best ResNet version 1 network results are achieved for a depth of 20, batch size of 16, and about 275 K parameters. Average testing accuracy of 87.67% was achieved for epochs of 80, and the validation accuracy was also the highest, achieving 89.45%.

ResNet version 2 network depth was implemented with depth 29, 47, 110, and 245, as shown in Table 3. The batch sizes were varied between two values, 16 and 32. Each experiment was done with epochs 10, 20, 40, and 80. There is also clearly no overfitting as the accuracies rise and losses drop with increasing epochs. With a ResNet Version 2 network depth of 29, 47, or 110, and a batch size of 32, acceptable validation and testing accuracies, as well as lower validation and testing losses, were obtained. The best results were achieved with a ResNet version 2 network of the depth of 110, batch size of 16, and about 3.32 M parameters. The average accuracy of 88.67% was achieved with a validation accuracy of 87.23%.

Convolutional neural networks (CNNs) with 9, 10, and 12 layers which included input and output layers, were then explored with batch sizes of 16 or 32, as shown in Table 4. The details of these three models are presented in [5]. For each experiment, Epochs of 10, 20, 40, and 80 were used. Overfitting was detected in the validation results starting at 80 epochs as the validation accuracies drop and validation losses rise as we move from 40 to 80 epochs. This indicates that a CNN with 20 or, in some cases, 40 epochs are sufficient, and the validation should not advance beyond 40 epochs.

As to the testing accuracy and loss, for most CNN models tested, 80 epochs are required to obtain the best accuracy and loss. CNN model 2, with ten layers, and a batch of 16, result in the highest testing accuracy and the second-best testing loss among the CNN models tested. The average accuracy of 81.24% was achieved with a validation accuracy of 80.13%.

Table 5 shows the confusion matrix for the proposed ResNet V2 with 12 blocks (110 network depth) which

Table 3 Experimental results based on different ResNet version 2 with different models

ResNet blocks	Network depth	Total parameters	Batch size	Epochs	Training Acc (%)	Training loss	Testing Acc (%)	Testing loss
3	3×9+2=29	851,042	16	10	86.21	0.374	85.87	0.373
				20	86.09	0.347	84.67	0.380
				40	86.78	0.321	85.80	0.348
				80	87.51	0.307	86.87	0.319
			32	10	79.93	0.474	79.37	0.468
				20	80.54	0.445	80.03	0.447
				40	81.16	0.426	80.20	0.429
				80	83.05	0.385	82.47	0.410
5	5×9+2=47	1,400,866	16	10	85.58	0.385	87.00	0.367
				20	86.21	0.347	86.47	0.348
				40	86.77	0.326	86.27	0.353
				80	87.44	0.309	85.80	0.340
			32	10	85.43	0.406	83.87	0.429
				20	86.03	0.366	84.47	0.408
				40	86.36	0.338	85.53	0.368
				80	86.87	0.324	86.07	0.355
12	12×9+2=110	3,325,250	16	10	85.47	0.473	83.07	0.534
				20	85.56	0.404	84.80	0.418
				40	86.48	0.352	87.27	0.334
				80	**87.23**	**0.327**	**88.67**	**0.319**
			32	10	79.50	0.480	78.40	0.489
				20	80.20	0.449	80.80	0.445
				40	80.95	0.433	82.63	0.425
				80	83.34	0.382	82.83	0.419

The bold values indicate the highest accuracy for each model

Table 4 Experimental results based on three different CNN models

CNN model	Net. depth	Batch size	Epochs	Val. Acc (%)	Val. loss	Test. Acc (%)	Test. loss
1	9 layers	16	10	94.68	0.1356	77.32	1.0540
			20	95.80	0.1102	78.41	1.0649
			40	94.95	0.1254	78.49	1.0503
			80	**81.33**	**0.4004**	**80.88**	**0.4130**
		32	10	98.58	0.0415	76.38	1.6959
			20	99.23	0.0225	78.51	1.6382
			40	98.93	0.0312	77.12	1.6053
			80	**80.00**	**0.4242**	**78.75**	**0.4665**
2	10 layers	16	10	85.57	0.3227	80.08	0.4934
			20	86.15	0.3110	79.86	0.4979
			40	85.64	0.3201	81.20	0.4853
			80	**80.13**	**0.4277**	**81.24**	**0.4139**
		32	10	92.96	0.1716	77.93	1.2705
			20	93.40	0.1647	77.34	0.9442
			40	96.61	0.0893	78.23	0.9053
			80	**80.97**	**0.4184**	**78.67**	**0.4660**
3	12 layers	16	10	88.83	0.2573	78.93	0.5800
			20	87.16	0.2924	79.37	0.5194
			40	89.32	0.2498	76.92	0.6070
			80	**84.71**	**0.3435**	**79.63**	**0.4873**
		32	10	96.54	0.0926	78.19	1.2163
			20	97.00	0.0812	76.18	1.1815
			40	97.56	0.0663	78. 53	1.1118
			80	**79.44**	**0.4368**	**80.28**	**0.4207**

The bold values indicate the highest accuracy for each model

Table 5 Confusion matrix for the proposed ResNet V2 with 12 blocks (110 network depth)

Pneumonia	No pneumonia	
1690	201	Pneumonia
356	2589	No pneumonia

achieved the highest accuracies. The confusion matrix is based on the 20% of the randomly selected test images from the dataset having a total of 4,836 images for two classes. The confusion matrix shows that Pneumonia class has slightly better classification as compared to the No-Pneumonia class.

Among all the models proposed in this paper, ResNet Version 2 achieves the highest Average testing (recognition) accuracy of 88.67% with a network of depth 110, batch size 16, with 1,742,370 parameters, and 80 epochs. Figure 5 shows the training and validation loss curve for the proposed ResNet V2 with 12 blocks (110 network depth), while Fig. 6 shows the training and validation accuracy curve for the proposed ResNet V2 with 12 blocks (110 network depth) which received the highest accuracy for the classification of pneumonia.

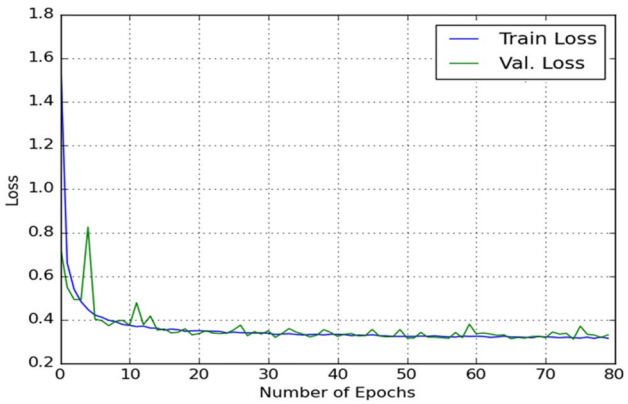

Fig. 5 Training and validation loss curve for the proposed ResNet V2 with 12 blocks (110 network depth)

6 Discussion

LeCun's CNN architecture, named ConvNet, was inspired by Hubel and 'Wiesel's work and resembled the basic structure of 'primate's visual cortex [29]. CNN's went through improvement phases culminating with AlexNet. The improvements focused on optimizing the depth and

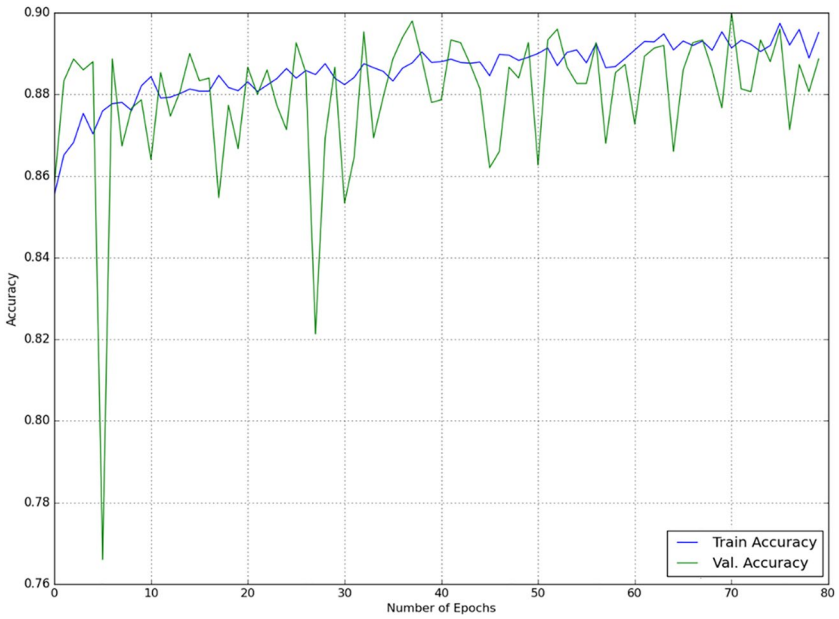

Fig. 6 Training and validation accuracy curve for the proposed ResNet V2 with 12 blocks (110 network depth)

parameters to reduce the computational cost. With such improved and modular CNNs with repeated architectural stages, the need to find filter dimension, stride, paddings in each network layered was eliminated, greatly cutting the computational cost. Recent CNN advances, which added skip connections and cross-layer channel connections, greatly helped improve CNNs convergence rate. Moreover, parallel (not depth-wise) transformations also helped tackle difficult problems. Recent efforts, e.g., GoogleNet, have also targeted reducing CNNs high computational cost and high memory demands.

ResNets allow the training of very deep neural networks with lower computational complexity than other proposed networks [27]. ResNets fell under the categories of multi-path CNNs and deep CNNs and introduced cross-layer connections to pass residual information and keep identity shortcuts enabled. Such a shortcut connection expedites the convergence of deep networks, thereby avoiding the diminishing gradient issue of deep neural networks. The ResNet Model produces better results than CNN models and is considered an improvement to produce reasonable results while cutting the computational cost and memory demands. Based on the results, the ResNet Version 1 and Version 2 both outperforms the 5 CNN models. Also, ResNet Version 2 outperforms ResNet Version 1, which leads to the conclusion that ResNet version 2 outperforms all the models tested in this research. Figure 7 shows the results for ResNet Version 1, ResNet Version 2,

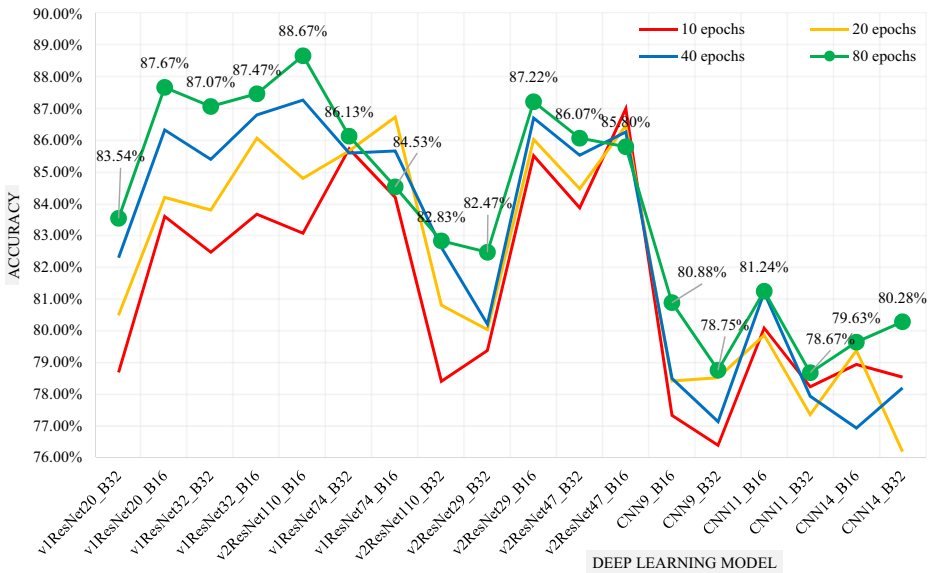

Fig. 7 Comparison between accuracies of different experimented ResNet and CNN models

Table 6 Training and testing computational cost for the proposed models with their accuracies

Proposed model	Train time	Test time	Accuracy (%)
CNN model 1 (9 layers)	26.72 min	1.56 s	80.88
CNN model 2 (10 layers)	28.10 min	1.98 s	81.24
CNN model 3 (12 layers)	36.18 min	2.23 s	80.28
ResNet V1 (3 blocks)	9.09 h	3.16 s	87.67
ResNet V2 (12 blocks)	14.58 h	9.11 s	**88.67**

The bold value indicate the highest accuracy for each model

and the different CNN models. Clearly, the highest accuracies were obtained with 80 epochs. The ResNet Version 2 network with a depth of 110 and a batch size of 16 achieved the highest accuracy (88.67%), which is higher than all other network models tested, taking advantage of the deeper network architecture supported by ResNet. This is followed closely by the ResNet Version 1 with a depth of 20 and batch size of 16 (87.67%), and the ResNet Version 1 with a depth of 110 and batch size of 16. The CNN networks generally performed worse than the ResNet networks.

Table 6 shows the training and testing computational cost for the proposed models with their accuracies. The models in the table are based on the best accuracies achieved, which shows variations in the train time based on the complexity of the model. Although the training time is high for the ResNet V2, the test time is relatively close to the other models. The proposed ResNet V2 with 12 blocks (110 network depth) which received the highest accuracy for the classification of pneumonia, took 14.58 h to train the model and 9.11 s to test the proposed model.

Table 7 shows the comparison between the results of the proposed ResNet Model and other models from recent literature. Our ResNet Version 2 network with a depth of 110 and batch size of 16 results outperforms all previously obtained accuracies published in the literature. All studies included in Table 5 are binary classification problems, same as the one proposed in this study.

7 Conclusion

In this article, improved ResNet Architectures are proposed for the automated diagnosis of pneumonia from chest X-rays. The algorithms proposed in this work are based on using a modified deep ResNet architecture to detect pneumonia in chest radiograph images. The modified networks are tested on the DICOM dataset detailed in the paper. An average recognition rate of 88.67% was achieved, which exceed other methods reported in the recent literation shown in Table 5. Generally, ResNet networks tested performed better than CNN networks tested.

Future work in this area will include developing or modifying more complex algorithms that automatically detect pneumonia not only through image recognition but also through a combination of image recognition, laboratory test, vital signs, and medical history. This will imply the design of complete Artificial Intelligence (AI) based system that can replicate the diagnosis, which is done by medical doctors and practitioners to automate this process without human intervention. The field of deep learning machine learning algorithms and architectures is developing at a swift pace and is envisioned to continue growing so that it is integrated into many fields for automatic recognition. We plan to keep pace or ahead of these developments to develop novel architectures that can detect and diagnose pneumonia and other diseases using various information in various formats to arrive at a diagnosis with accuracy that can exceed that of trained human professionals. Future work will also include the use of the proposed modified ResNet Architecture on other medical diagnosis applications such as MRI brain tumor detection, Skin Cancer Detection, and others.

Table 7 Comparison of proposed method results with recent literature

Source	Method	Dataset	Accuracy (%)
Proposed method	ResNet with 110 layers depth	RSNA Dataset	**88.67**
Al Mubarok et al. [23]	ResNet and mask-RCNN	RSNA Dataset	85.6
Donthi et al. [24]	Convolutional neural networks (CNNs)	RSNA Dataset	78.9
Sirazitdinov et al. [25]	Ensemble of 2 CNNs	RSNA Dataset	Precision 0.883, Recall 0.652
Tang et al. [26]	Convolutional neural networks (CNNs)	RSNA Dataset	86.38
Wesley et al. [30]	Deep learning network	5659 X-rays	72

The bold value indicate the highest accuracy for each model

Acknowledgements The authors would like to thank Dr. Runna Alghazo for proofreading and editing the manuscript to eliminate language issues and enhance readability wherever warranted.

Data availability The dataset being used in this study is publicly published.

Declarations

Conflict of interest The authors have no conflicts of interest.

References

1. Watkins, K., & Sridhar, D. (2018). Pneumonia: A global cause without champions. *The Lancet, 392*(10149), 718–719.
2. Ye, X., Xiao, H., Chen, B., & Zhang, S. (2015). Accuracy of lung ultrasonography versus chest radiography for the diagnosis of adult community-acquired pneumonia: Review of the literature and meta-analysis. *PLoS One*, *10*(6).
3. Nath, S. S., Mishra, G., Kar, J., Chakraborty, S., & Dey, N. (2014, July). A survey of image classification methods and techniques. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)* (pp. 554–557). IEEE.
4. Latif, G., Iskandar, D. A., Alghazo, J., & Jaffar, A. (2018). Improving brain MR image classification for tumor segmentation using phase congruency. *Current Medical Imaging, 14*(6), 914–922.
5. Butt, M. M., Latif, G., Iskandar, D. A., Alghazo, J., & Khan, A. H. (2019). Multi-channel convolutions neural network based diabetic retinopathy detection from fundus images. *Procedia Computer Science, 163*, 283–291.
6. Khan, A. H., Latif, G., Iskandar, D. A., Alghazo, J., & Butt, M. (2018). Segmentation of melanoma skin lesions using anisotropic diffusion and adaptive thresholding. In *Proceedings of the 2018 8th International Conference on Biomedical Engineering and Technology*, pp. 39–45.
7. Latif, G., Iskandar, D. A., Alghazo, J. M., & Mohammad, N. (2018). Enhanced MR image classification using hybrid statistical and wavelets features. *IEEE Access, 7*, 9634–9644.
8. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3462–3471. IEEE.
9. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., & Lungren, M. P. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv:1711.05225
10. Noor, N. M., Rijal, O. M., Yunus, A., & Abu-Bakar, S. A. R. (2010). A discrimination method for the detection of pneumonia using chest radiograph. *Computerized Medical Imaging and Graphics, 34*(2), 160–166.
11. Xue, Z., You, D., Candemir, S., Jaeger, S., Antani, S., Long, L. R., & Thoma, G. R. Chest x-ray image view classification. In *IEEE 28th International Symposium on Computer-Based Medical Systems (CBMS), 2015* 2015 pp. 66–71. IEEE.
12. Ko, B. C., Kim, S. H., & Nam, J. Y. (2011). X-ray image classification using random forests with local wavelet-based CS-local binary patterns. *Journal of Digital Imaging, 24*(6), 1141–1151.
13. Livieris, I., Kanavos, A., Tampakas, V., & Pintelas, P. (2018). An ensemble ssl algorithm for efficient chest x-ray image classification. *Journal of Imaging, 4*(7), 95.
14. Wang, C., Elazab, A., Wu, J., & Hu, Q. (2017). Lung nodule classification using deep feature fusion in chest radiography. *Computerized Medical Imaging and Graphics, 57*, 10–18.
15. Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology, 284*(2), 574–582.
16. Kuruvilla, J., & Gunavathi, K. (2014). Lung cancer classification using neural networks for CT images. *Computer methods and programs in biomedicine, 113*(1), 202–209.
17. Cicero, M., Bilbily, A., Colak, E., Dowdell, T., Gray, B., Perampaladas, K., & Barfett, J. (2017). Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Investigative Radiology, 52*(5), 281–287.
18. Farag, A., Ali, A., Graham, J., Farag, A., Elshazly, S., & Falk, R. (2011). Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest. In *IEEE international symposium on biomedical imaging: From nano to macro*, pp. 169–172. IEEE.
19. Maduskar, P., Muyoyeta, M., Ayles, H., Hogeweg, L., Peters-Bax, L., & van Ginneken, B. (2013). Detection of tuberculosis using digital chest radiography: Automated reading vs. interpretation by clinical officers. *The International Journal of Tuberculosis and Lung Disease, 17*(12), 1613–1620.
20. Antani, S., & Candemir, S. (2015). Automated detection of lung diseases in chest X-rays. US National Library of Medicine.
21. Xu, T., Mandal, M., Long, R., Cheng, I., & Basu, A. (2012). An edge-region force guided active shape approach for automatic lung field detection in chest radiographs. *Computerized Medical Imaging and Graphics, 36*(6), 452–463.
22. Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., & Greenspan, H. Chest pathology detection using deep learning with non-medical training. In *ISBI*, 2015, pp. 294–297.
23. Mubarok, A. F. A., Dominique, J. A. M., & Thias, A. H. (2019). Pneumonia detection with deep convolutional architecture. In *2019 International conference of artificial intelligence and information technology* (*ICAIIT*), pp. 486–489. IEEE
24. Donthi, A, Huang, A., & Tammanagari, A. (2018). Detecting pneumonia with convolutional neural networks. Semanticscholar Org.: Allen Institute for Artificial Intelligence, Seattle, WA, USA
25. Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R., & Ibragimov, B. (2019). Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. *Computers & Electrical Engineering, 78*, 388–399.
26. Tang, Y.-X., You-Bao, T., Yifan, P., Ke, Y., Bagheri, M., Redd, B. A., Brandon, C. J. (2020). Automated abnormality classification of chest radiographs using deep convolutional neural networks. *NPJ Digital Medicine, 3*(1), pp. 1–8.
27. Wu, Z., Shen, C., & Van Den Hengel, A. (2019). Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognition, 90*, 119–133.
28. Shih, G., Wu, C. C., Halabi, S. S., Kohli, M. D., Prevedello, L. M., Cook, T. S., Sharma, A., Amorosa, J. K., Arteaga, V., Galperin-Aizenberg, M., & Gill, R. R. (2019). Augmenting the National Institutes of Health chest radiograph dataset with expert annotations of possible pneumonia. Radiology: Artificial Intelligence, 1(1), e180041.
29. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324.
30. O'Quinn, W., Haddad, R. J., & Moore, D. L. (2019). Pneumonia radiograph diagnosis utilizing deep learning network. In *2019 IEEE 2nd international conference on electronic information and communication technology* (*ICEICT*), pp. 763–767. IEEE