PNEUMONIA DETECTION WITH DEEP LEARNING

Lazaros Panitsidis, Konstantinos Kravaritis

International Hellenic University

ABSTRACT

Pneumonia remains a substantial health danger, especially among young children, due to the difficulty and variability in identifying the infection with standard imaging tools such as chest X-rays. The purpose of this study is to determine how well Convolutional Neural Networks (CNNs), a type of Deep Learning (DL) model, can detect pneumonia in chest X-ray images. A number of CNN architectures, including ResNet-18 and ResNet-50, were assessed, emphasizing the advantages of ensemble learning for enhancing diagnostic precision. With an accuracy of 86.472%, the ensemble models outperformed the individual models, demonstrating how these sophisticated DL approaches can help medical professionals diagnose pneumonia.

Index Terms — Pneumonia detection, Deep Learning, Convolutional Neural Networks, Chest X-ray, ResNet, Ensemble Learning.

1. INTRODUCTION

Pneumonia is a disease that affects the respiratory system. Its causes can be attributed to many infectious agents like fungi, bacteria and viruses. According to the WHO **[1]** pneumonia is the single most prevalent infectious cause of death for children aged between 1 and 5 years old, noting a death rate of 14% in 2019.

Chest-X-ray (CXRs) and CT-scan images are among the most commonly and cost-effective used methods to detect and diagnose pneumonia. Identifying however and diagnosing pneumonia from CXRs can be a difficult and time-consuming process, which simultaneously can lead to conflict amongst radiologists when trying to interpret the results of CXRs. This is where machine learning models like Deep Neural Networks come in to assist. Recent DL advancements have helped achieve and perhaps even significantly exceed human performance in many activities, including image identification and classification. (DNN) models have traditionally been built and tested by human professionals in a continuous trial-and-error technique that takes time, resources, and expertise **[2]**. These models have shown promising results and may improve the efficacy and accuracy especially in the case of pneumonia diagnosis. On the subject of pneumonia CXRs, accurate diagnoses are quite difficult to achieve with the sheer volume of images available in datasets which are used to train ML models. Deep Learning with models like Convolutional Neural Networks are the perfect candidates to improve the accuracy of predictions and the trustworthiness of the results in the eyes of medical practitioners which down the line can rely on them for aid in their work.

2. THE DATASET

The dataset utilized in the Detect Pneumonia (Spring 2024) is divided into a training portion of 4672 images and a testing portion of 1168 images. The images originate from chest X-rays and all have different sizes. The training images are split into 3 classes: class 0 which indicates a healthy individual, class 1 which indicates an individual suffering from bacterial pneumonia and class 3 which indicates an individual with viral pneumonia. From the training images 1227 images belong to class 0, 2238 images belong to class 1 and 1207 images belong to class 2.

3. DESCRIPTION OF THE MODELS

ResNet is an abbreviation for Residual Network. It represents a family of extremely deep CNN architectures that won the ILSVRC-2015 challenges for image recognition, object detection and localization **[3]**.

It is a form of neural network that serves as the foundation for a variety of computer image identification and classification tasks. Compared to conventional neural networks, the ResNet approach can train networks with up to 150 layers. Convolution neural networks are difficult to train due to disappearing gradients. Vanishing gradients is an issue in which gradient backward propagation occurs. Additionally, repetitive multiplication can make the value of the gradient factor extremely small. **[4]**

The ResNet model is one of the better options for addressing this issue. ResNet models are classified into numerous categories based on the number of layers. As the network model progresses through the layers, its performance diminishes. It is not straightforward to enhance network depth by simply stacking layers one after the other. **[5]**

While specifically developed for image classification problem solving, it uses the concept of matrix addition by which the data from initial layers are passed to deeper layers. This procedure does not support any other additional parameters as the previous layer outcome is added to the onward layer.

According to Alshazly et al **[6]** the cornerstone for constructing deep residual networks is the residual module. The left path of the residual module in Figure 3a is composed of two convolutional layers, which apply 3×3 kernels and preserve the spatial dimensions. Batch normalization and ReLU activation are also applied. The right path is the skip connection where the input is added to the output of the left path. This variant is used in the ResNet18 model. Another variant of the residual module named the bottleneck residual module is depicted in Figure 3b, in which the input signal also passes through two branches. However, the left path performs a series of convolutions using 1×1 and 3×3 kernel sizes, along with batch normalization and ReLU activation. The right path is the skip connection, which connects the module's input to an addition operation with the output of the left path. This variant is utilized in ResNet50 and ResNet101 models.

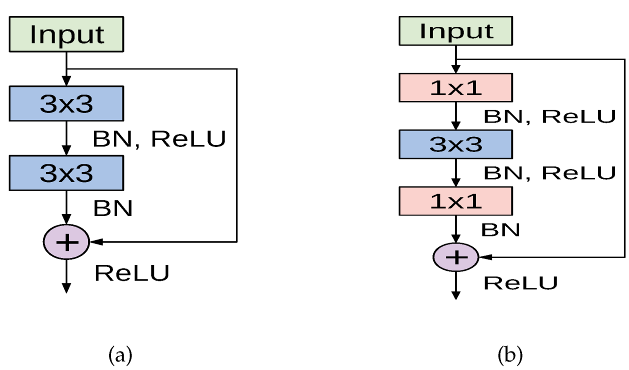

Figure 1. The basic residual module used in ResNet18 (**a**), and the bottleneck residual module utilized in ResNet50 and ResNet101 (**b**), both as introduced in **[7]**.

There are several types of convolution neural networks based on residual networks. Layer-based models include ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152, ResNet-164, and ResNet-202.

3.1. ResNet-18

ResNet-18 is a convolutional neural network (CNN) architecture widely used in image classification tasks and serves as a backbone for more complex models in various computer vision applications, such as object detection and segmentation. It is a part of the Residual Networks (ResNets) family, which was introduced by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun in their 2015 paper "Deep Residual Learning for Image Recognition." **[8]**

ResNet-18 consists of 18 layers, which include:

- One initial convolutional layer.
- Four groups of residual blocks, each containing two convolutional layers.
- A fully connected layer at the end for classification.

Layer Structure:

- The initial layer is a 7x7 convolutional layer followed by a max pooling layer.
- Each residual block consists of two 3x3 convolutional layers with a ReLU activation function and a batch normalization layer.
- The network ends with a global average pooling layer and a fully connected layer for the final classification.

3.2. ResNet-34

ResNet-34 is another variant of the ResNet architecture, following the same principles as ResNet-18 but with a deeper network structure. It was introduced in the same paper, "Deep Residual Learning for Image Recognition," by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. **[9]**

Here are the key features of ResNet-34:

- Depth: ResNet-34 has a total of 34 layers, hence the name. These layers are organized into groups of residual blocks, similar to ResNet-18 but with more blocks.
- Layer Structure: Like other ResNet variants, ResNet-34 consists of convolutional layers, residual blocks, and a final classification layer. The residual blocks contain multiple convolutional layers with shortcut connections to help with the flow of gradients during training.
- Increased Capacity: With its deeper architecture, ResNet-34 has a larger capacity to capture more complex features compared to shallower networks like ResNet-18.
- Performance: ResNet-34 typically achieves better performance on image recognition tasks compared to ResNet-18, especially on more challenging datasets and tasks that require finer-grained distinctions between classes.

- Applications: ResNet-34, like other ResNet variants, is widely used in various computer vision applications, including image classification, object detection, and segmentation.

3.3. ResNet-50

Developed by researchers at Microsoft Research Asia, ResNet-50 is renowned for its depth and efficiency in image classification tasks.

A breakdown of the components within the residual block follows:

ReLU Activation **[10]**: The ReLU (Rectified Linear Unit) activation function is applied after each convolutional layer and the batch normalization layers. ReLU allows only positive values to pass through, introducing non-linearity into the network, which is essential for the network to learn complex patterns in the data.

Bottleneck Convolution Layers **[11]**: the block consists of three convolutional layers with batch normalization and ReLU activation after each.:

The first convolutional layer likely uses a filter size of 1x1 and reduces the number of channels in the input data. This dimensionality reduction helps to compress the data and improve computational efficiency without sacrificing too much information.

The second convolutional layer might use a filter size of 3x3 to extract spatial features from the data.
The third convolutional layer again uses a filter size of 1x1 to restore the original number of channels before the output is added to the shortcut connection.

There are almost twenty three million parameters available for training in the ResNet-50 model. ResNet-34 and ResNet-50 differ in the design of their building blocks. These were changed to a holdup intention due to scheduling constraints, based on the time required for layer training. ResNet-50 uses a three-layer stack, whereas ResNet-34 uses a two-layer stack. As a result, ResNet-50 has higher accuracy than ResNet-34.**[12]**

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 2. Architectures of Resnet for ImageNet dataset. The residual building blocks are shown in brackets with the numbers of stacked blocks **[13]**

4. EXPERIMENTS AND RESULTS

4.1. Experiments

We started by testing different architectures and ended up with ResNet-18, ResNet-34, ResNet-50 which, due to being smaller and less complex, gave better results given the volume of data we had at our disposal.

We emphasized on image augmentation due to the fact that these models are data driven. Hence the better the data and with more diversity, the better the results would be, and the models would generalize more effectively to unknown data. This image generation was performed every time before data was entered into the model, and it was achieved with random combinations of the parameters we had set, such as angle, slope, shadows, shear factor e.t.c...., with the goal of making the data introduced each time unique. In addition to this, we tried using the pytorch libraries to alter the transformation of the images before introducing them to the neural network. Examples are the size of the image, rotation, random cutting or zooming and others, in order to avoid overfitting and improve generalization.

Then we modified the learning rate and evaluated two schedulers: ReduceLROnPlateau and CyclicLR. The first is based on the premise that we want a high learning rate at first, but then we would like it to decrease so that it learns slower, more carefully and therefore might be better. Whereas the second is based on the idea that it may be constantly stuck in a local minimum and we can help it get unstuck by finding a lower local minimum or the gradient's global minimum if possible.

Finally, our ultimate goal was to create ensemble models, with various combinations of the models that produced a better result. This is because, based on theory, the average of the variance of the ensemble is less than the average of the sum of the separate models.

4.2. Results

Several important insights were obtained from our investigation of hyperparameters and architectures. The ensemble models outperformed all individual models, with V6EB3 (86.472%), V7EB3 (86.130%), and V5EB2 (86.301%) consistently achieving the best accuracy. This demonstrates how well ensemble learning works to lower variance and possibly take advantage of the various model strengths that are applicable to our goal.

Performance differed amongst individual models. Using ReduceLROnPlateau (RLRP), ResNet-18 (V3R18) outperformed ResNet-50 (V3R50) in terms of accuracy, achieving 84.589% with a higher batch size (32). This shows that while ResNet-50 with RLRP produced decent results (84.246%) at a modest batch size (60) and a longer training duration (53 epochs), smaller models such as ResNet-18 may perform better in certain scenarios based on how training time and accuracy are balanced. Noteworthy is the possibility of improving ResNet-18's (V3R18) performance with additional hyperparameter adjustment, particularly with regard to learning rate.

In the following table, the models with the top 10 accuracies are shown.

Model name	Samples per class (*1000)	Batch size	Learning rate	Epochs trained	Accuracy (%)
V6EB3	V3R18, V3R50, V7R50				86.472
V7EB3	V3R18, V3R50, V1R50				86.130
V5EB2	V3R18, V3R50				86.301
V2EB4	V3R18, V3R50, V7R50, V15R50				85.787
V1R50	30	9	0.0001	7	85.445
V3R18	30	32	0.001 CLR	14	84.589
V3R50	50	60	0.0001 RLRP	53	84.246
V7R50	30	120	0.0001	10	83.732
V15R50	30	32	0.0001 CLR	24	83.732
V4R50	50	9	0.0001 RLRP	9	83.39

5. CONCLUSIONS

In this paper, we investigated how well different deep learning models could identify pneumonia from chest X-ray images. Comparing various methodologies, learning rate schedulers, and the effects of ensemble approaches were all part of our analysis. The ensemble models performed better than the individual models; specifically, V6EB3, V5EB2, and V7EB3 achieved the highest accuracies of 86.472%, 86.301% and 86.130%, respectively. This demonstrates how ensemble learning may be used to maximize the advantages of different models in order to improve performance as a whole.

The results further emphasized how crucial model selection and hyperparameter adjustment are. For example, ResNet-18 fared better than ResNet-50 when used with the ReduceLROnPlateau scheduler in some scenarios, indicating that smaller models may be more efficient with appropriate tuning. In order to improve performance, the CyclicLR scheduler demonstrated potential in assisting models in escaping local minima.

Our results imply that deep learning models can greatly increase the accuracy of pneumonia identification from chest X-rays, especially when coupled with ensembles. In order to improve the models' ability to generalize, future work will concentrate on improving these models even more, investigating new hyperparameter changes, and growing the dataset to include larger and more varied samples.

All things considered, this study provides insightful information about how to optimize deep learning models for medical image analysis, paving the way for the development of automated diagnostic systems in the medical field that are more trustworthy and accurate.

6. REFERENCES

[1] Pneumonia in children (2019). https://www.who.int/news-room/fact-sheets/detail/pneumonia

[2] Stephen O, Sain M, Maduh UJ, Jeong DU. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. *Journal of Healthcare Engineering*. Published March 27, 2019. https://www.hindawi.com/journals/jhe/2019/4180949/

[3] Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*. 2015;115(3):211-252. doi:https://doi.org/10.1007/s11263-015-0816-y

[4] Elkorany AS, Elsharkawy ZF. COVIDetection-Net: A Tailored COVID-19 Detection from Chest Radiography Images Using Deep Learning. *Optik*. Published online February 2021:166405. doi:https://doi.org/10.1016/j.ijleo.2021.166405

[5],[6] Alshazly H, Linse C, Barth E, Martinetz T. Explainable COVID-19 Detection Using Chest CT Scans and Deep Learning. *Sensors*. 2021;21(2):455. doi:https://doi.org/10.3390/s21020455

[7], [8], [9] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. openaccess.thecvf.com. Published 2016. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

[10], [11] What is ResNet-50? Roboflow Blog. Published March 13, 2024. https://blog.roboflow.com/what-is-resnet-50/

[12] Qian H, Zhou X, Zheng M. Abnormal Behavior Detection and Recognition Method Based on Improved ResNet Model. *Computers, Materials & Continua*. 2020;65(3):2153-2167. doi:https://doi.org/10.32604/cmc.2020.011843

[13] 1.Le K. A quick overview of ResNet models. Medium. Published March 15, 2022. Accessed June 10, 2024. https://lekhuyen.medium.com/a-quick-overview-of-resnet-models-f8ed277ae81e