Proceedings of the 3rd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/138/2025.21363

Performance Analysis of Residual Networks for Pneumonia Diagnosis on Chest X-Rays

Tanjing Zhang[1,a,*]

[1]*Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA*
a. tanjingzhang54@outlook.com
**corresponding author*

Abstract: ResNets have demonstrated exceptional capabilities in image classification tasks, making them a promising tool for clinical applications. This study evaluates the performance of deep learning Residual Networks (ResNets) in classifying chest X-ray images for pneumonia diagnosis. Five ResNet variants: ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152 are compared on their performance training on the same dataset of chest X-ray images, which included positive and negative diagnoses. The models were trained on a subset of images and tested on a separate set to assess their accuracy and consistency. Our results indicate that ResNet50 achieves the highest accuracy in predictions, while ResNet34 demonstrates the most consistent performance across trials. These findings provide valuable insights into selecting the most suitable ResNet variant for binary classification tasks involving chest X-rays, particularly when considering factors such as computational resource availability and accuracy requirements.

Keywords: Residual Network, Medical Image Classification, Pneumonia, X-ray images, Deep Learning

1. Introduction

Pneumonia is a type of infection that affects one or both lungs that can be caused by bacteria, viruses, or fungi. It often presents as a serious condition in which the air sacs fill with pus and other fluids, impairing normal breathing [1]. Diagnosis typically relies on a combination of symptoms, laboratory tests, and medical imaging. Early, accurate, and immediate diagnosis is crucial for effective management and improving patient outcomes. However, providing immediate and precise pneumonia diagnostic can be challenging, often leading to delays in initiating appropriate treatment.

Medical imaging, specifically chest X-rays, is a standard non-invasive method used by healthcare professionals to diagnose and classify pneumonia. However, manual interpretation has limitations due to variability in expertise and workload among medical professionals, which can affect diagnostic accuracy and efficiency [2].

To address these issues, machine learning algorithms, particularly Convolutional Neural Networks (CNNs), offer promising solutions. CNNs excel in image analysis tasks by learning complex patterns from labeled training data and applying that knowledge to make reliable predictions on new images [3]. Among CNN architectures, Residual Networks (ResNets) stand out due to their innovative use

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/138/2025.21363

of identity mappings, which enable them to maintain high accuracy even in deeper architectures. This makes ResNets particularly effective for medical imaging tasks.

In this project, we explore the application of different ResNet variants to classify pneumonia using an existing dataset from PneumoniaMNIST [4], which contains 5,856 chest X-ray images. These images are categorized into positive/negative cases. By comparing the performance of various ResNet models, we aim to assess their effectiveness in aiding the diagnosis and classification of pneumonia, thereby supporting clinical decision-making, and potentially enhancing patient care.

This paper is divided into the following parts:

1. Literature Review: A discussion of current applications and the benefits of using convolutional neural networks in medical imaging, as well as the challenges associated with their use in clinical diagnoses.
2. Methodology: A detailed explanation of the experimental approach, including how the performance of different ResNet variants is compared.
3. Results: A presentation of the data collected during the experiment.
4. Discussion: An interpretation of the results, focusing on the strengths and limitations of ResNet models in solving pneumonia classification tasks.
5. Conclusion and Future Directions: A summary of the findings and recommendations for further research in this domain.

2. Literature Review

Medical images are fundamental to diagnosing pneumonia and many other conditions. According to Nambu[5] et al., the primary role of imaging examinations is to confirm the diagnosis of pneumonia and also provides information for treatment plans.

Residual Networks (ResNets) have shown excellent performance in replicating manual diagnostics. In a review by Xu et al. [6], ResNet applications were highlighted as prominent in disease detection tasks, including lung tumors, skin diseases, breast diseases, and brain diseases. These tasks typically fall under the category of classification, which determines whether a disease is present (positive) or absent (negative) and can further classify subtypes of diseases.

The performance of ResNet in medical image classification can be also studied in details through its variants. Sarwinda [7] et al. have shown that in colon gland image classification. ResNet50 appears to have an overall better performance than ResNet18 in terms of accuracy, sensitivity, and specificity when dealing datasets with different train-to-test ratios. Athisayamani [8] has conducted studies which shows that ResNet-152 and softmax classifier are used to perform the feature classification process for different kinds of brain tumors with precision rating of 96.81%.

However, a major challenge in medical imaging is the variability in image resolution across different equipment, as there is no universal imaging standard. This inconsistency impacts not only clinical diagnostics but also the performance of neural networks used in automated analysis. Mainly different choices of ResNet differs by how many convolutions layer appearing as a label behind the name of neural network. Typically, neural network with smaller number of layers would have fewer parameters and faster training time, and vice versa. Tang [9] et al. has discovered that Light-weight neural networks optimized for specific resolutions often suffer significant performance drops when trained or tested on images of different resolutions, as higher or lower resolutions can distort crucial features, leading to flawed outputs. Low-resolution images can hinder the detection of small, critical lesions, ultimately reducing a network's effectiveness in precise diagnostics. Additionally, it's more difficult to make an accurate diagnostic since chest X-ray examinations for pneumonia detection are prone to subjective variability, shown by Neuman [10] et al.

DOI: 10.54254/2755-2721/138/2025.21363

3. Methodology

Different variations of ResNet were evaluated for performance using the same dataset of pneumonia X-ray images from PneumoniaMNIST [4], a subset of MedMNIST. MedMNIST is a collection of standardized biomedical images designed for machine learning training. The dataset includes pre-processed images of uniform dimensions (224x224 pixels) that are properly labeled as positive or negative based on diagnostic results. The dataset is divided into a training set of 4,708 X-ray images and a testing set of 624 X-ray images.

These images were used to train a specific type of neural network: ResNet. ResNet was developed to address challenges in traditional CNN-based architectures, such as the vanishing or exploding gradient problem. These issues become more pronounced as convolutional layers are added for solving complex problems, often leading to increased error rates during both training and testing. ResNet introduces a key innovation: identity mapping, where the original mapping is recasted onto the underlying mapping from the convolutional network. [11]

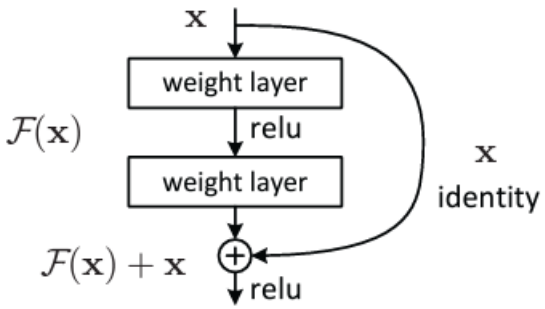

Figure 1: Residual learning: a building block [11]

The addition of identity mappings ensures that possible malfunctioning convolutional layers do not significantly degrade the neural network's performance. In this study, five ResNet variants, ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152, provided by PyTorch library, were evaluated on the PneumoniaMNIST testing set. The suffix in each variant's name corresponds to the number of convolutional layers in the network, with fewer layers offering a more lightweight architecture but generally lower accuracy. Different variants are evaluated based on preditction accuracy, the percentage of true positive and true negative predictions over all cases [12]. Each variant goes through five trials where in each trial the training pictures are divided into 32 batches and go through the ResNet for 15 epochs. Then then maximum testing accuracy is selected out of those 15 epochs and that would be the maximum testing accuracy of that trial. Then the maximum testing accuracies are compared again each other and standard deviations of the maximums are taken for the trials for the quantification of performance consistency.

4. Result

The experiment evaluated the performance of various ResNet-based neural network architectures on classifying pneumonia using the PneumoniaMNIST dataset. PneumoniaMNIST is a subset of the MedMNIST medical image collection that consists of standardized and pre-processed biomedical images tailored for machine learning applications. Each image is a grayscale X-ray scan resized to dimensions with binary class labels of positive or negative diagnostic. A sample image of each category is listed below:

DOI: 10.54254/2755-2721/138/2025.21363

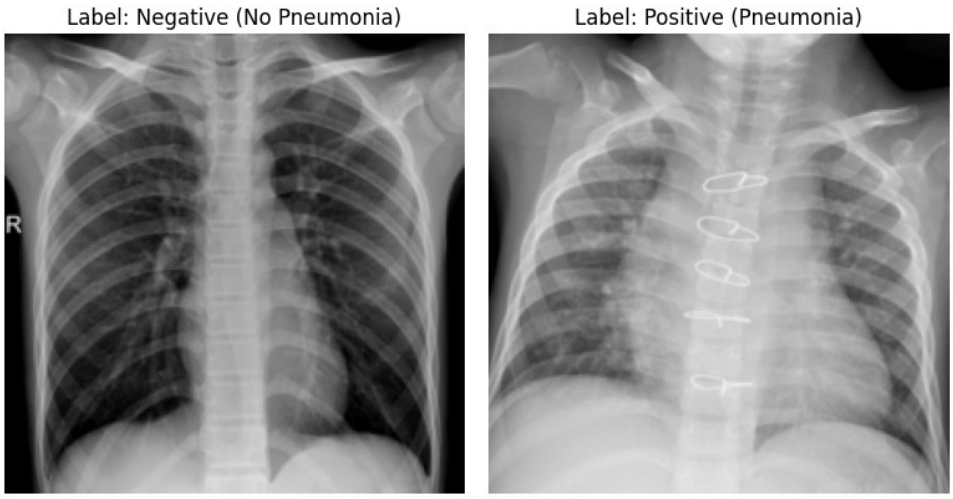

Figure 2: Sample Images from positive and negative pneumonia diagnostic from PneumoniaMNIST

The number of samples picture in this collection is 5856 pictures with 4708 for testing, 524 for validation, and 624 for testing. The MedMNIST data were first preprocessed and encapsulated into data loader form. In this process, the images are cropped into uniform sizes of 224 x 224 since there are small size deviations among the images, and the grayscale channel is converted to RGB channels to fit the neural network input. The training dataset includes 4,708 X-ray images, randomly shuffled and divided into mini-batches of size 32. The testing dataset consists of 624 X-ray images, which are exclusively used for evaluating model performance.

The ResNet models used in this study were sourced from PyTorch's torchvision.models module. These architectures are widely known for their robustness in deep learning applications, particularly in image classification tasks. The following ResNet variants were evaluated: ResNet18, ResNet34, ResNet50, ResNet101, ResNet152, where the number adjacent to network name represents its number of convolutional layers of that particular neural network variant. A larger number means that the neural network is more complex that usually results in higher accuracy but requires higher computational resources.

Each ResNet model was trained using a narrowed training scope, where a randomly selected batch (148 images) from the training dataset was used in each epoch. This approach reduces computational demands while focusing on smaller subsets of data (Repeat 5 times).

The models were trained for 15 epochs. Cross-entropy loss was used to evaluate classification error throughout the epochs. Stochastic Gradient Descent (SGD) with a learning rate of 0.01 and a momentum of 0.9. Performances of each variant are quantified with accuracy and area under the curve (AUC), where accuracy shows how the number of correct predictions over the total number of trials and AUC shows the ability to distinguish true positive and false positives.

The ResNet models were evaluated based on their performance on the PneumoniaMNIST testing pool of 624 images. Key observations include:

Table 1: Test accuracy from ResNet variants

	Maximum Test Accuracy	Average Maximum Test Accuracy	Standard Deviation
ResNet18	92.79%	91.22%	0.0135
ResNet34	92.63%	91.61%	0.0082
ResNet50	93.11%	90.93%	0.0210

DOI: 10.54254/2755-2721/138/2025.21363

Table 1: (continued).

ResNet101	92.95%	91.73%	0.0088
ResNet152	91.51%	90.13%	0.0141

5. Discussion

Residual blocks are a fundamental feature of ResNet, enabling the network to "skip" layers to address the vanishing gradient problem. This design allows ResNet to achieve outstanding performance in image classification tasks [13]. The aim of this study is to evaluate how different ResNet variants perform on the specific task of pneumonia X-ray image classification.

The ResNet variants used in this experiment differ by the number of convolutional layers in each neural network. The number appended to each ResNet variant (e.g., ResNet18, ResNet50) indicates the total number of convolutional layers. A larger number generally corresponds to a more complex network that can achieve higher accuracy but requires greater computational resources.

PneumoniaMNIST, the dataset used in this study, is a subset of MedMNIST—a standardized collection of 2D and 3D biomedical images sourced from a wide range of clinical sites. The images have been visually reviewed to ensure the absence of personal identifiers.

Our results indicate that ResNet50 achieved the highest level of accuracy in predicting pneumonia from lung region X-ray images, reaching 93.11% accuracy. Interestingly, more complex variants, such as ResNet101 and ResNet152, showed lower performance. Their lower performance may be attributed to overfitting or insufficient data to fully utilize the added complexity of these models. In contrast, ResNet34 demonstrated consistent performance, achieving a maximum accuracy consistency of 0.0082 across five trials.

The image selected for this dataset is collected in different clinical sites across the world with no inclination or biased information. They are grayscale X-ray images cropped to the size of 224x224. All the network variants are directed applied from Pytorch.

6. Conclusion

This research consists of a cross comparison of the performance in medical image classification across different Residual Network variants. The performance of ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152 on specific task of pneumonia classification using X-ray images from the PneumoniaMNIST dataset is quantified in maximum testing accuracy, average maximum testing accuracy, and the standard deviations. This dataset consists of visually reviewed, anonymized biomedical images labeled as positive (pneumonia) or negative (no pneumonia). Each image was pre-processed to a standardized format to ensure compatibility with the ResNet models.

It is shown that ResNet50 achieved the highest testing accuracy in classification, up to 93.11%, meaning that this model is the most suitable and efficient for this task. More complex architectures, ResNet101 and ResNet152 in this case, show a lower performance, possibly due to overfitting or insufficient data. Meanwhile, ResNet34 shows the most consistent performance by showing the lowest variations across the maximum testing accuracy in five trials. ResNet34 would serve as a reliable alternative when the computation resources are not abundant.

These results demonstrated that ResNet50 is the most suitable for performing grayscale image binary classification task such as the one performed in this study. However, the performance decline in deeper ResNet variants suggests that model complexity should align with the dataset size and variability to prevent overfitting. A simpler model is also considered a practical option since it shows consistent performance.

DOI: 10.54254/2755-2721/138/2025.21363

In conclusion, this study highlights the importance of selecting an appropriate ResNet variant for specific tasks and datasets. Future work may involve experimenting with different training strategies or evaluations and additional medical image datasets to further validate and generalize these findings.

References

[1] *Johns Hopkins Medicine. (n.d.). Pneumonia. Johns Hopkins Medicine. https://www.hopkinsmedicine.org/health/conditions-and-diseases/pneumonia*

[2] *Ahmad, H. K., Milne, M. R., Buchlak, Q. D., Ektas, N., Sanderson, G., Chamtie, H., Karunasena, S., Chiang, J., Holt, X., Tang, C. H. M., Seah, J. C. Y., Bottrell, G., Esmaili, N., Brotchie, P., & Jones, C. (2023). Machine learning augmented interpretation of chest x-rays: A systematic review. Diagnostics, 13(4), 743. https://doi.org/10.3390/diagnostics13040743*

[3] *Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. Insights into Imaging, 9(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9*

[4] *Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., & Ni, B. (2023). MedMNIST v2—A large-scale lightweight benchmark for 2D and 3D biomedical image classification. Scientific Data, 10(1), 41. https://doi.org/10.1038/s41597-022-01721-8*

[5] *Nambu, A. (2014). Imaging of community-acquired pneumonia: Roles of imaging examinations, imaging diagnosis of specific pathogens and discrimination from noninfectious diseases. World Journal of Radiology, 6(10), 779. https://doi.org/10.4329/wjr.v6.i10.779*

[6] *Xu, W., Fu, Y.-L., & Zhu, D. (2023). ResNet and its application to medical image processing: Research progress and challenges. Computer Methods and Programs in Biomedicine, 240, 107660. https://doi.org/10.1016/j.cmpb.2023.107660*

[7] *Sarwinda, D., Paradisa, R. H., Bustamam, A., & Anggia, P. (2021). Deep learning in image classification using residual network (Resnet) variants for detection of colorectal cancer. Procedia Computer Science, 179, 423–431. https://doi.org/10.1016/j.procs.2021.01.025*

[8] *Athisayamani, S., Antonyswamy, R. S., Sarveshwaran, V., Almeshari, M., Alzamil, Y., & Ravi, V. (2023). Feature extraction using a residual deep convolutional neural network (ResNet-152) and optimized feature dimension reduction for mri brain tumor classification. Diagnostics, 13(4), 668. https://doi.org/10.3390/diagnostics13040668*

[9] *Tang, S., Jing, C., Jiang, Y., Yang, K., Huang, Z., Wu, H., Cui, C., Shi, S., Ye, X., Tian, H., Song, D., Xu, J., & Dong, F. (2023). The effect of image resolution on convolutional neural networks in breast ultrasound. Heliyon, 9(8), e19253. https://doi.org/10.1016/j.heliyon.2023.e19253*

[10] *Neuman, M. I., Lee, E. Y., Bixby, S., Diperna, S., Hellinger, J., Markowitz, R., Servaes, S., Monuteaux, M. C., & Shah, S. S. (2012). Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children. Journal of Hospital Medicine, 7(4), 294–298. https://doi.org/10.1002/jhm.955*

[11] *He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. arXiv. https://doi.org/10.48550/ARXIV.1512.03385*

[12] *Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010*

[13] *Mastering resnet: Deep learning breakthrough in image recognition. (n.d.). Retrieved January 21, 2025, from https://www.ikomia.ai/blog/mastering-resnet-deep-learning-image-recognition*