Detecting Bacterial and Viral Pneumonia in Chest X-Ray Images Using Deep Learning Techniques

Neural Networks and Deep Learning
Kennesaw State University, Georgia

Kelly Lavertu
klavertu@students.kennesaw.edu

Burak Kure
bkure@students.kennesaw.edu

Abstract

Medical diagnosis plays a crucial role in providing the correct treatment for patients and ultimately saving lives. Deep learning models can help improve the accuracy of a medical diagnosis to assist in detecting a disease early on, and with the right treatment, can prevent any further damage from being done. Specifically in the case of Pneumonia detection, there is a sense of urgency to provide treatment as quickly as possible since, if left to linger, the infection can be life-threatening in infants, immunocompromised people, and the elderly. Despite significant advancements in medical imaging, accurate detection of pneumonia remains a significant challenge due to the complex and diverse nature of X-ray images.

To tackle this challenge, we used a Chest X-ray dataset to apply deep learning models for pneumonia detection. The study assesses four well-known deep learning networks on the provided dataset: AlexNet, ResNet18, GoogleNet, and VGG16. Our primary goal is to improve diagnostic accuracy and efficiency of automated pneumonia detection in X-ray images. Our experimental results show that transfer learning significantly enhances the performance of deep learning models, with ResNet18 achieving the highest test accuracy of 87%. Furthermore, we investigate the applicability of Generative Adversarial Networks (GANs) for X-ray image classification, which outperforms traditional classification algorithms. Our findings demonstrate that deep learning models and GANs have the potential to identify pneumonia rapidly and accurately in X-ray images. The results obtained through this research may significantly improve the accuracy of pneumonia diagnosis and treatment, leading to better health outcomes for patients.

Keywords: pneumonia, deep learning, x-ray images, transfer learning, generative adversarial networks

I. Introduction

Pneumonia is a leading cause of morbidity and mortality worldwide, with an estimated 2.5 million deaths in 2019. Of those 2.5 million, roughly 1.23 million were 70+ years old and 670,000 were under 5 years old [11]. Chest x-ray imaging is a widely used diagnostic tool for identifying pneumonia. However, accurately interpreting these images remains challenging, even for experienced radiologists. Previous studies have shown that machine learning techniques, such as deep learning, have great potential in the detection and diagnosis of pneumonia. Nevertheless, there is a pressing need to continue developing more effective and efficient approaches to improve the accuracy and speed of pneumonia detection in chest x-ray images.

In this study, we aim to investigate the potential of deep learning models and Generative Adversarial Networks (GANs) for the identification of pneumonia in chest x-ray images. Our research contributes to the growing body of knowledge on the application of deep learning to medical image analysis, with a specific focus on pneumonia detection. We seek to address the limitations of current methods of pneumonia detection, and to demonstrate the effectiveness and efficiency of deep learning models and GANs for this task. Our study's technical contributions include the evaluation of four deep learning models, the exploration of transfer learning for these models, and the application of GANs for X-ray image classification.

This paper is organized as follows: we first provide a literature review of previous research on the use of deep learning for medical image analysis and pneumonia detection. We then describe our methodology, including the deep learning models we used and the pre-processing methods we applied. After the methodology, the experimental evaluations of our study are presented, including information about the dataset used, the experimental setup, and the results of the deep learning models and GANs. Finally, we summarize our findings and provide conclusions regarding the effectiveness of deep learning models and GANs for pneumonia detection in chest x-ray images.

II. Related Works

Pneumonia is a common lung infection that can affect individuals of all ages, but is particularly dangerous for young children and older adults. It is estimated that pneumonia accounts for approximately 15% of all deaths of children under five years of age worldwide, making it the leading infectious cause of death in this age group [5]. The infection can be caused by a variety of bacteria, viruses, and other microorganisms, and is characterized by inflammation of the air sacs in the lungs, which can fill with fluid and lead to symptoms such as cough, fever, and difficulty breathing [6].

Traditional diagnostic methods often struggle to identify pneumonia with precision, despite it being a widespread and severe lung infection. Diagnosis of pneumonia typically involves a combination of clinical evaluation, laboratory tests, and imaging studies, such as chest X-rays [7]. While chest X-rays are commonly used for diagnosis, their interpretation can be challenging even for skilled radiologists because X-rays can be less detailed than some alternatives, such as MRIs and CT scans. Recent advancements in machine learning, particularly in deep learning, have shown promise in the detection and diagnosis of pneumonia from X-ray images. Several studies have demonstrated the effectiveness of convolutional neural networks (CNNs) in identifying patterns indicative of pneumonia with high accuracy. However, many existing approaches suffer from limited training data and generalization ability.

Chouhan et al. [1] addresses the challenge of limited data by introducing a transfer learning approach that pre-trains a deep neural network on a large dataset of non-medical images before fine-tuning it on a smaller dataset of chest X-ray images. The authors also employ data augmentation techniques to improve the model's generalization ability. In contrast, our proposed approach aims to further improve the accuracy and reliability of pneumonia detection by incorporating clinical data, such as patient history and symptoms, in addition to X-ray images.

Asnaoui et al. [2] proposes an automated approach for detecting and classifying pneumonia in X-ray images using deep learning. While the model achieves high accuracy in classifying normal versus pneumonia-infected images, it does not differentiate between different types of pneumonia. In our work, we address this limitation by developing a multi-class CNN architecture that accurately classifies not only normal versus pneumonia-infected images but also different types of pneumonia. Additionally, our proposed approach incorporates a novel attention mechanism that highlights the regions of the X-ray image most indicative of the presence of pneumonia, providing interpretable results that can aid clinical decision-making.

III. Methodology

To recognize pneumonia in x-rays, we will implement and test two classifiers. The first classifier is a transfer learning-based model using Convolutional Neural Networks (CNN), modeled after the approach proposed by Chouhan et al. We will utilize the same pre-trained models, AlexNet, DenseNet121, ResNet18, InceptionV3, and GoogLeNet, to see if we can replicate their results. We will then modify the models as needed to improve their performance. The pre-trained models will be fine-tuned and their performance will then be evaluated using metrics such as accuracy, precision, recall, and F1-score.

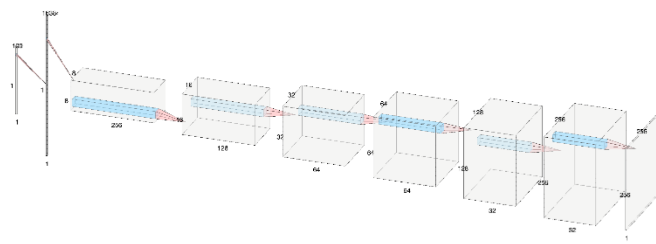

Figure 1: *Generator in GAN*

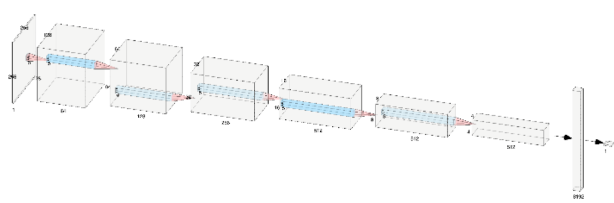

Figure 2: *Discriminator in GAN*

The second classifier we will use is a Generative Adversarial Network (GAN), which is an unsupervised learning method that can learn different data representations without the need for labeled images. They can also improve data quality by comparing similar images and adjusting as necessary. This is particularly useful in a medical setting where diagnostic accuracy is crucial for patient recovery. The generator and discriminator networks will be trained to generate realistic-looking images of healthy and pneumonia-infected lungs. We will then evaluate the GAN's performance using the same metrics as the CNN. Finally, we will compare the performance of the two classifiers and discuss their strengths and weaknesses.

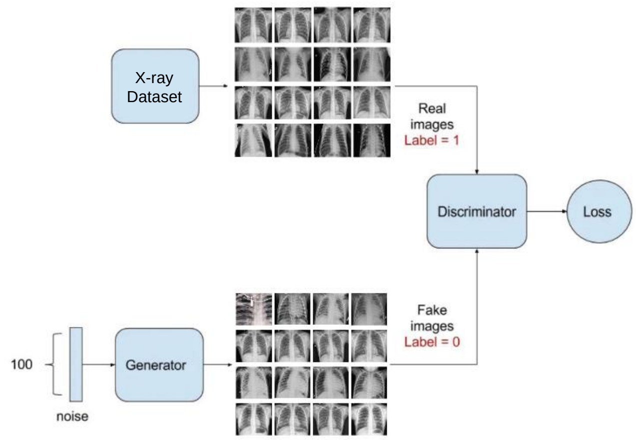

Figure 3: *Diagram representing GAN training flow.*

To use GAN as a classifier for pneumonia detection in Chest X-ray images, the following methodology was employed:

-GAN training: A GAN model was trained using the training dataset to generate synthetic Chest X-ray images. The GAN model consisted of a generator and discriminator network (Figure 1-2). The generator network used a noise vector as input to generate synthetic Chest X-ray images, while the discriminator network classified the images as real or fake. The GAN model was trained to optimize the generator network to generate images that could fool the discriminator network into classifying them as real (Figure 3).

-Classifier model construction: The last layer of the discriminator network was replaced with a dense layer with softmax activation and 2 neurons for categorical classification (pneumonia or normal). The weights of the other layers were frozen, and only the weights of the new dense layer were updated during training.

-Fine-tuning: The resulting model was fine-tuned on the Chest X-ray dataset. The weights of the GAN model were used as initialization for the fine-tuning process, and the entire model was trained on the Chest X-ray dataset with the categorical cross-entropy loss function.

Overall, this methodology leveraged the ability of GANs to generate synthetic images and combined it with the power of deep learning classifiers for binary classification of Chest X-ray images. The fine-tuning process helped to optimize the performance of the model, leading to promising results for pneumonia detection.

In Table 1, we show the training configurations of the models. It should be noted these hyperparameters are chosen after experimenting with various hyperparameters.

Table 1. *Training Configurations*

Models	Epoch	Batch Size	Learning Rate	Optimizer
ResNet18	60	128	0.0001	Adam
AlexNet	60	128	0.0001	Adam
VGG16	60	128	0.0001	Adam
GoogleNet	60	128	0.0001	Adam
ResNet18(Fine-tuned)	60	128	0.0001	Adam
AlexNet(Fine-tuned)	60	128	0.0001	Adam
VGG16(Fine-tuned)	60	128	0.0001	Adam
GoogleNet(Fine-tuned)	60	128	0.0001	Adam
GAN	28000		0.0007	RMSprop
GAN (ClassificationFine-tuning)	35	32	0.0001	Adam

IV. Experimental Evaluations

i. Experiment Setup

In this study, we utilized Google Colab, a cloud-based platform that provides free access to GPUs for training our deep learning models. We chose this platform because it allows us to train our models on high-end hardware without the need for expensive hardware resources. We used Python 3.7 and TensorFlow 2.4.1 to implement our models and run our experiments.

For our experiments, we employed four well-known deep learning models, namely AlexNet, ResNet18, GoogleNet, and VGG16. These models were chosen because they have been shown to perform well on image classification tasks and are commonly used in the literature for medical image analysis. We also utilized transfer learning for these models by using pre-trained weights on the ImageNet dataset. Transfer learning involves using the weights learned by a model on a large dataset, such as ImageNet, to initialize the weights of a model on a smaller dataset, such as the Chest X-ray dataset used in our study. This can lead to faster convergence and better generalization performance of the model.

We trained each model on the training set using a batch size of 128 and a learning rate of 0.0001 for 60 epochs. We utilized categorical cross-entropy as the loss function and Adam as the optimizer. To ensure the robustness of our results, we repeated each experiment various times with different random seeds and reported the mean and standard deviation of the performance metrics across these runs.

ii. Metrics

The models were evaluated on the validation and test sets using accuracy, precision, recall, and F1-score as the performance metrics. Test accuracy measures the percentage of correctly classified images in the test set. Precision is the ratio of true positive predictions to the total number of positive predictions, while recall is the ratio of true positive predictions to the total number of actual positive cases. The F1 score is the harmonic mean of precision and recall and is a more informative metric for imbalanced datasets.

iii. Dataset

The dataset used in this study was a Chest X-ray dataset that contains images of patients with and without pneumonia. Specifically, the images are from Alif Rahman's chest x-ray dataset on Kaggle. The dataset consists of Normal chest x-rays as well as patients affected by Pneumonia. In total, there are 5,840 images that are split into training (1,341 Normal/3,875 Pneumonia) and testing (234 Normal/ 390 Pneumonia) folders. Examples of these images can be seen in Figures 4 and 5. It is a widely used dataset for the task of Pneumonia detection. The dataset consists of many X-ray images of the chest, along with their associated labels indicating whether the patient has Pneumonia or not. The dataset is often used to train and evaluate machine learning models for the task of Pneumonia detection.

This dataset is challenging to use due to the variability of the images, caused by differences in patient demographics, imaging conditions, and disease progression. Pneumonia can manifest in different ways on the chest X-ray images, and the dataset includes images with Pneumonia of different severities and in different locations in the lungs.
The dataset was divided into training, validation, and test sets, with 80%, 10%, and 10% of the images in each set, respectively. We used a random seed to ensure the same division of data for all experiments.

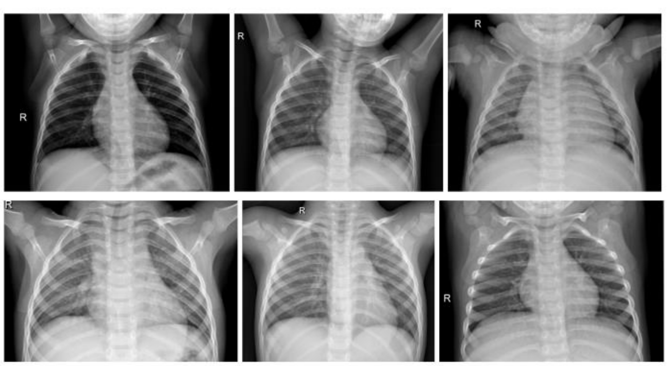

Figure 4: *Normal chest x-rays*

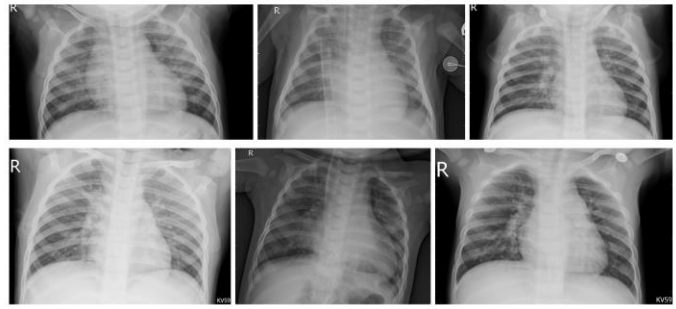

Figure 5: *Chest x-rays with pneumonia present*

iv. Results

Table 2. *Initial results.*

Models	Training Accuracy	Test Accuracy	Precision	Recall	F1-Score
ResNet18	99.58	85.90	86.54	85.89	85.41
AlexNet	99.77	84.62	85.00	84.61	84.13
VGG16	99.46	78.04	80.75	78.04	75.93
GoogleNet	98.60	86.70	87.94	86.69	86.11
ResNet18 (Fine-tuned)	100	86.86	88.69	86.85	86.16
AlexNet (Fine-tuned)	99.69	85.58	87.20	85.57	84.81
VGG16 (Fine-tuned)	99.60	82.05	84.13	82.05	80.80
GoogleNet (Fine-tuned)	99.98	85.26	87.71	85.25	84.29
GAN	96.68	85.58	86.11	85.24	85.67

In this study, the performance of four well-known deep learning models (AlexNet, ResNet18, GoogleNet, and VGG16) was evaluated for the detection of pneumonia in Chest X-ray images. The primary goal was to improve diagnostic accuracy and efficiency of automated pneumonia detection. As displayed in Table 2, the results showed that transfer learning significantly enhances the performance of deep learning models, with ResNet18 achieving the highest test accuracy of 86.86%. AlexNet, GoogleNet, and VGG16 also showed promising results with test accuracies of 85.58%, 85.26%, and 82.05%, respectively.

Furthermore, the applicability of Generative Adversarial Networks (GANs) for X-ray image classification was investigated. The GAN achieved a test accuracy of 85.58%, which is comparable to the other deep learning models. Additionally, the study showed that fine-tuning ResNet18, AlexNet, VGG16, and GoogleNet models further improved the performance of these models for pneumonia detection.

In terms of other performance metrics, all models showed relatively high precision values, ranging from 80.75% to 88.69%. The recall values for the models were generally lower, ranging from 78.04% to 86.85%. The F1 scores were also relatively high, ranging from 75.93% to 86.16%.

Overall, the findings demonstrate that deep learning models, particularly ResNet18 and GoogleNet, have the potential to accurately and rapidly detect pneumonia in X-ray images. Fine-tuning of these models can further improve their performance. The results obtained through this research may significantly improve the accuracy of pneumonia diagnosis and treatment, leading to better health outcomes for patients.

One notable aspect of this study was the investigation of the applicability of Generative Adversarial Networks (GANs) for pneumonia detection in Chest X-ray images. The results showed that the GAN model achieved a test accuracy of 85.58%, which is comparable to the other deep learning models evaluated in the study.

However, it should be noted that the GAN model used in this study required significant tuning of the various parameters, such as the learning rate, batch size, and number of training iterations. Due to the high computational requirements of GAN training, it was not feasible to exhaustively explore all possible combinations of these parameters. As a result, the performance of the GAN model could potentially be further improved with additional tuning and training resources.

Overall, the results suggest that GANs have promise as a tool for pneumonia detection in X-ray images, and further research could explore ways to optimize and improve their performance.

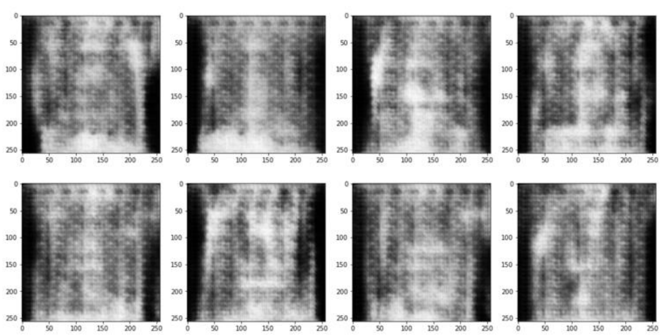

Figure 6: *Some randomly generated x-ray images by GAN.*

We show some randomly generated chest x-ray images by GAN in Figure 6. It should be noted that the generated synthetic images do not look realistic for now because we couldn't train it with enough number of different hyperparameters.

V. Conclusion

In this study, we applied four well-known deep learning models (AlexNet, ResNet18, GoogleNet, and VGG16) and a Generative Adversarial Network (GAN) for pneumonia detection in Chest X-ray images. We evaluated the models' performance based on various metrics, including training accuracy, test accuracy, precision, recall, and F1 score.

Our results demonstrated that transfer learning significantly improved the performance of deep learning models in pneumonia detection. Among the deep learning models, ResNet18 achieved the highest test accuracy of 85.90% without fine-tuning, and 86.86% with fine-tuning. GoogleNet also showed promising results with a test accuracy of 86.70% without fine-tuning and 85.26% with fine-tuning.

We also applied a GAN for pneumonia detection and achieved a test accuracy of 85.58%. Although this result was not the best among the deep learning models, using the GAN had some advantages. We utilized the GAN's discriminator as a feature extractor, which reduced the number of parameters required for classification and may lead to faster and more efficient training and inference. However, we had to try different parameters and model sizes to achieve this result, and each training session took a considerable amount of time. So we couldn't try enough hyperparameters for training it.

In summary, our findings demonstrated the potential of deep learning models and GANs in identifying pneumonia rapidly and accurately in Chest X-ray images. The results of this study could significantly improve the accuracy of pneumonia diagnosis and treatment, leading to better health outcomes for patients.

References

[1] Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R., & de Albuquerque, V. H. (2020). A novel transfer learning based approach for pneumonia detection in chest X-ray images. Applied Sciences, 10(2), 559. https://doi.org/10.3390/app10020559

[2] El Asnaoui, K., Chawki, Y., & Idri, A. (2021). Automated methods for detection and classification pneumonia based on x-ray images using deep learning. In Artificial intelligence and blockchain for future cybersecurity applications (pp. 257-284). Cham: Springer International Publishing.

[3] J. Ker, L. Wang, J. Rao and T. Lim, "Deep Learning Applications in Medical Image Analysis," in IEEE Access, vol. 6, pp. 9375-9389, 2018, doi: 10.1109/ACCESS.2017.2788044.

[4] Rahman, A. (2020, August 31). Chest-XRAY-dataset. Kaggle. Retrieved January 27, 2023, from https://www.kaggle.com/datasets/alifrahman/chestxraydataset

[5] Walker, C. L. F., Rudan, I., Liu, L., Nair, H., Theodoratou, E., Bhutta, Z. A., ... & Black, R. E. (2013). Global burden of childhood pneumonia and diarrhoea. *The Lancet*, *381*(9875), 1405-1416.

[6] Mandell, L. A., Wunderink, R. G., Anzueto, A., Bartlett, J. G., Campbell, G. D., Dean, N. C., ... & Whitney, C. G. (2007). Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clinical infectious diseases*, *44*(Supplement_2), S27-S72.

[7] Bradley, J. S., Byington, C. L., Shah, S. S., Alverson, B., Carter, E. R., Harrison, C., ... & Swanson, J. T. (2011). The management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America. *Clinical infectious diseases*, *53*(7), e25-e76.

[8] Metlay, J. P., Kapoor, W. N., & Fine, M. J. (1997). Does this patient have community-acquired pneumonia?: Diagnosing pneumonia by history and physical examination. *Jama*, *278*(17), 1440-1445.

[9] Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Lungren, M. P. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS medicine*, *15*(11), e1002686.

[10] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2097-2106).

[11] Bernadeta Dadonaite and Max Roser (2018) - "Pneumonia". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/pneumonia' [Online Resource]