check for **updates**

.

Deep Learning Models for Classification of Pediatric Chest X-ray Images using VGG-16 and ResNet-50

Rosa Alba Pugliesi
MD, Klinikum Ludwigsburg (Germany)
https://orcid.org/0000-0001-5108-2104

Abstract

Medical imaging plays a crucial role in diagnosing various diseases, including pneumonia, in pediatric patients. In this research, we investigate the performance of two convolutional neural networks (CNNs), VGG-16 and ResNet-50, for the task of pneumonia detection using chest X-ray images of pediatric patients aged one to five years old. The dataset consists of anterior-posterior chest X-ray images, categorized into two classes: Pneumonia and Normal. For preprocessing, we standardized the dataset by setting each sample mean to zero and dividing the inputs by the standard deviation of the dataset. Additionally, we applied ZCA whitening to further enhance the data. The two CNN models, VGG-16 and ResNet-50, were trained and evaluated on the dataset. VGG-16, with 16 layers, can classify images into 1000 object categories, while ResNet-50 is a deeper CNN with 50 layers. The experimental results demonstrate that the ResNet-50 model outperformed the VGG-16 model in terms of accuracy and loss during testing. The VGG-16 model achieved a testing accuracy of 74.9% with a testing loss of 48.8%, whereas the ResNet-50 model achieved a significantly higher testing accuracy of 88.9% with a lower testing loss of 28.9%. This study highlights the efficacy of deep learning models in pediatric pneumonia detection and underscores the superior performance of ResNet-50 over VGG-16. These findings have significant implications for developing more accurate and efficient diagnostic tools to aid medical professionals in diagnosing pneumonia in pediatric patients.

Keywords: Medical imaging, Pneumonia detection, Convolutional neural networks (CNNs), Pediatric patients, ResNet-50, VGG-16

Introduction

Pneumonia, a prevalent and serious inflammatory lung condition, targets the delicate alveoli, small air sacs responsible for gas exchange within the lungs [1], [2]. Characterized by an array of distressing symptoms, such as productive or dry cough, chest pain, fever, and breathing difficulties, pneumonia's impact on individuals can vary significantly. Primarily triggered by viral or bacterial infections, it can also be caused by other microorganisms, certain medications, or underlying health conditions like autoimmune diseases. A range of risk factors can predispose individuals to pneumonia, including cystic fibrosis, chronic obstructive pulmonary disease (COPD), asthma, diabetes, heart failure, a history of smoking, compromised cough reflexes following a stroke, and a weakened immune system [3], [4]. When infectious agents, such as bacteria or viruses, infiltrate the alveoli, the body's immune response is triggered, leading to an influx of immune cells and the release of inflammatory mediators. As a result, the alveoli become filled with pus and cellular debris, hindering efficient gas exchange and causing respiratory distress. In severe cases, the infection can spread to neighboring lung tissues, causing consolidation, abscess formation, or even pleural effusion. The severity and course of pneumonia depend on the

type of pathogen, the individual's overall health status, and the timeliness and appropriateness of medical intervention [5].

Pneumonia can present differently across age groups, with distinct symptoms observed in children under 5 years old, infants, and older individuals. Children in this age range may experience rapid breathing or wheezing, which can be concerning for caregivers and may be accompanied by other signs of respiratory distress. In infants, the symptoms can be more subtle, with some appearing asymptomatic. However, they may exhibit symptoms like vomiting, lethargy, or difficulties with feeding, which can be indicative of an underlying respiratory infection [6], [7]. Conversely, older individuals may display milder symptoms, making pneumonia potentially harder to detect in this population.

In the field of medical diagnostics, image analysis has emerged as a critical tool for the accurate detection of pneumonia, especially in pediatric patients [8]. The ability to promptly and precisely identify pneumonia is of utmost importance as it allows for timely treatment and management of the disease, ultimately leading to improved patient outcomes. Radiologists, healthcare professionals, and specialized computer algorithms all play pivotal roles in this process. By carefully reviewing chest X-ray images, these experts can pinpoint characteristic patterns that may indicate pneumonia, such as focal opacities or diffuse haziness in the lungs [9], [10]. Considering the complexities of pediatric cases, healthcare professionals further enhance their diagnostic accuracy by factoring in the clinical history and symptoms of the young patients [11].

Over the past few years, the domains of deep learning and artificial intelligence have garnered immense attention and popularity [12]. Deep learning, in particular, has emerged as a prominent subset of machine learning, drawing inspiration from the workings of the human brain. It achieves this by employing intricate networks of interconnected layers known as neural networks. Each layer progressively processes different sets of information, ultimately producing the desired output. This unique approach has positioned deep learning as a powerful force within the machine learning industry, especially when harnessed with vast amounts of data, often referred to as Big Data [13], [14].

The healthcare industry, with its abundance of data and critical need for efficient and accurate results, represents an ideal environment for the application of advanced machine learning techniques like deep learning [15]. The sheer volume of data that can be collected in healthcare, ranging from patient records and medical imaging to genetic data and clinical trial results, presents an unprecedented opportunity for deep learning algorithms to extract meaningful insights and patterns [16].

With the growing adoption of electronic health records and digital health technologies, the healthcare sector has become more data-driven than ever before. Deep learning techniques can harness this wealth of data to improve decision-making processes, reduce medical errors, and enhance patient outcomes. By automating and augmenting various tasks, such as medical image analysis, drug discovery, and patient risk stratification, deep learning can support healthcare professionals in making faster, more informed decisions. This, in turn, can lead to optimized treatment strategies and more efficient healthcare delivery.

In recent years, the medical community has witnessed a remarkable transformation in pneumonia detection thanks to advancements in artificial intelligence and machine learning technologies. This breakthrough has given rise to powerful computer algorithms that can

be trained on vast datasets of annotated chest X-ray images, enabling them to recognize patterns associated with pneumonia with remarkable precision. Deep learning techniques have proven particularly effective in the analysis of chest X-ray images in pediatric patients. By utilizing neural networks with multiple layers, these algorithms can automatically extract intricate features from the X-ray images, allowing for the detection of subtle abnormalities that might otherwise go unnoticed by the human eye [17].

The integration of artificial intelligence into pneumonia detection workflows has shown great promise in revolutionizing pediatric healthcare. With the ability to process large volumes of data rapidly, computer algorithms can efficiently analyze chest X-ray images and provide valuable insights to healthcare professionals. This augmented analysis not only expedites the diagnosis process but also helps in reducing the workload of radiologists and medical practitioners, enabling them to focus on other critical aspects of patient care. Furthermore, the use of AI-powered image analysis systems promotes consistency and standardization in diagnosing pneumonia cases, minimizing the potential for human error and subjectivity in interpretation [18].

Traditional manual assessment of chest X-ray images by human radiologists can be time-consuming and prone to errors. In contrast, computer algorithms can swiftly process massive datasets of images, significantly reducing the time required for analysis and decision-making. Moreover, these algorithms can detect subtle patterns and variations that might be challenging for human eyes to discern, enhancing the overall accuracy and sensitivity of pneumonia detection .

The objective of this research is to investigate the performance of two convolutional neural networks (CNNs), VGG-16 and ResNet-50, for the task of pneumonia detection using chest X-ray images of pediatric patients aged one to five years old. The study aims to evaluate the effectiveness of these deep learning models in accurately classifying chest X-ray images into two classes: Pneumonia and Normal.

Materials and Methods

Convolutional Neural Network (ConvNet/CNN)
Convolutional neural networks (CNNs) have driven significant advancements within the realm of deep learning [19]. These networks have found widespread application in image recognition tasks, and their growing popularity can be attributed to several key factors. One compelling reason behind the popularity of CNN models is their ability to delve deeper into image analysis compared to traditional neural networks (NNs). This is made possible by their architectural design, which facilitates the extraction of highly intricate and nuanced features from images [20]. The typical structure of a CNN consists of three essential layers: the input layer, the hidden layer, and the output layer [21].

When working with CNNs, raw images are initially fed into the input layer, from where they are seamlessly passed on to the hidden layer. It is within this hidden layer that the crucial task of feature extraction takes place. This layer is composed of three integral components: the convolution layer, the pooling layer, and the fully connected layer.

The cornerstone of a CNN model lies in its feature extraction capabilities, and this is where the convolution layer comes into play [22]. By convoluting the raw images with a sliding window containing a fixed-size filter, this layer automatically discerns and extracts

39

relevant features from the input data. This ability to automatically learn and identify crucial patterns is what empowers CNNs to excel in image recognition tasks. Next, the pooling layer steps in to address the issue of dimensionality. As the feature maps obtained from the convolution layer can be quite extensive, the pooling layer performs a vital function in reducing their dimensionality while preserving the most significant features. This process of dimensionality reduction not only helps manage computational complexity but also enhances the network's ability to recognize patterns across diverse images. The fully connected layer serves as the concluding segment of the CNN model [23]. As it implies, this layer fully connects the previously processed feature maps, effectively flattening them to perform the classification task. By doing so, the CNN can accurately categorize the input images based on the extracted features, making it a powerful tool for tasks like image classification, object detection, and more [24].

VGG16

VGG16 stands for the Visual Geometry Group 16, a renowned CNN architecture highly regarded in the field of computer vision. This model is widely acclaimed for its exceptional performance in image-related tasks [25]. The creators of VGG16 meticulously analyzed existing networks and made significant advancements by incorporating a compact (3×3) convolution filter architecture, surpassing previous state-of-the-art configurations.

The "16" in VGG16 denotes the remarkable depth of the network, comprising 16 layers with trainable weights. With an extensive structure, VGG16 boasts an impressive scale, featuring an astounding 138 million parameters, making it one of the largest and most powerful neural networks available today in the realm of computer vision. VGG-16 follows a distinct architecture for image processing. It takes a fixed-size RGB image of 244x244 pixels as input. Before passing through VGG-16, each pixel's RGB value is subtracted by its mean value as part of the preprocessing step.

Once preprocessing is complete, the images go through a series of convolutional layers, each equipped with small receptive-field filters of size 3x3. In some configurations, the filter size is set to 1x1, implying a linear transformation of the input channels followed by a non-linear activation [26].

The convolution operation's default stride is set to 1, enabling precise feature extraction. Spatial pooling is then performed through five max-pooling layers that come after multiple convolutional layers [27]. This pooling step aids in reducing spatial dimensions, making the subsequent computations more efficient.

ResNet-50

ResNet, short for Residual Network, represents a distinctive class of convolutional neural networks (CNNs) known for its groundbreaking approach to handling deep network architectures . ResNet introduces a revolutionary idea of residual learning, which enables the successful training of extremely deep neural networks without suffering from the vanishing gradient problem [28]. This problem had previously hindered the efficacy of deep networks due to the challenges of propagating gradients through multiple layers during backpropagation. The ingenious solution lies in the inclusion of shortcut connections, also known as skip connections or identity mappings. By bypassing certain layers and connecting them directly to deeper layers, ResNet creates a residual block that preserves the original information while allowing the network to learn residual features .

This unique design effectively transforms a conventional network into a residual network, providing the foundation for various ResNet variants with differing numbers of layers.

One of the most widely used variants is ResNet-50, denoting a specific architecture with a remarkable depth of 50 layers. Comprising 48 convolutional layers, one MaxPool layer, and one average pool layer, ResNet-50 demonstrates the power of stacking multiple residual blocks to create a robust and expressive CNN. Through its deep structure, ResNet-50 excels in learning intricate patterns and features from data, making it highly suitable for challenging computer vision tasks, such as image classification, object detection, and segmentation [29], [30]. The ResNet-50 architecture has become a benchmark model for assessing the performance of various deep learning models in the field of image analysis, showcasing its versatility and reliability.

Initially, the ResNet series began with the ResNet-34 architecture, which consisted of 34 weighted layers. This earlier version laid the foundation for the subsequent advancements in residual learning. By pioneering the concept of shortcut connections, ResNet-34 set the stage for tackling the limitations of traditional deep networks. The shortcut connections not only introduced greater depth to CNNs but also enabled smoother information flow during training, mitigating the risk of vanishing gradients and promoting more efficient optimization.

Data

Chest X-ray images (anterior-posterior) collected from Mendeley data were chosen from a collection of past cases involving pediatric patients aged one to five years at Guangzhou Women and Children's Medical Center, Guangzhou [31]. These X-ray scans were conducted during the course of the patients' regular clinical care. To ensure the accuracy and reliability of the analysis conducted on the chest X-ray images, a rigorous quality control process was implemented. This involved eliminating all scans of low quality or those that were unreadable [32]. For further evaluation and grading, two expert physicians assessed the diagnoses of the selected images before they were deemed suitable for training the AI system. Moreover, to account for any potential grading errors, an additional examination was conducted on the evaluation set by a third expert [33].

Preprocessing

The training set for the dataset comprises a total of 5216 images, while the testing set consists of 624 images, and the validation set contains 16 images. Focusing on the presence of pneumonia and normal cases in the dataset, we observe that there are a total of 4273 images representing pneumonia cases, and 1583 images depicting normal cases.

In the image processing pipeline, we applied three fundamental techniques: image dilation, image erosion, and the Sobel filter [34], [35]. Image dilation is employed to expand the boundaries of objects within an image, enhancing their prominence and filling gaps between them. Conversely, image erosion works to shrink the object boundaries, eradicating small irregularities and noise [36], [37]. These two complementary operations help fine-tune and manipulate the image's structural elements effectively. Furthermore, we utilized the Sobel filter, a powerful edge detection tool that highlights edges within the image by computing gradients in both the horizontal and vertical directions. Employing both vertical and horizontal Sobel filters, we effectively highlight and extract the edges present within the image. Upon edge extraction, the subsequent step involves determining

41

the edge that lies farthest from the center of the image, which is approximated by identifying the vertical center of the image. This crucial information aids in locating the region of interest, where the most prominent edge lies.

With the edge farthest from the center identified, we proceed to perform a vertical cropping operation. By cropping along this particular edge, we can isolate and extract the relevant section from the original image. This cropping process strategically focuses on the most significant edge, enabling us to retain the essential features of the image while discarding irrelevant elements.

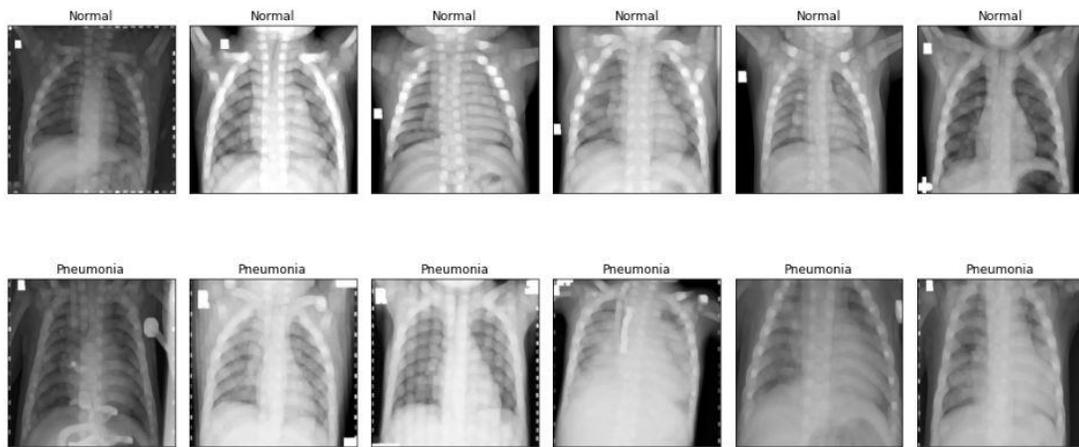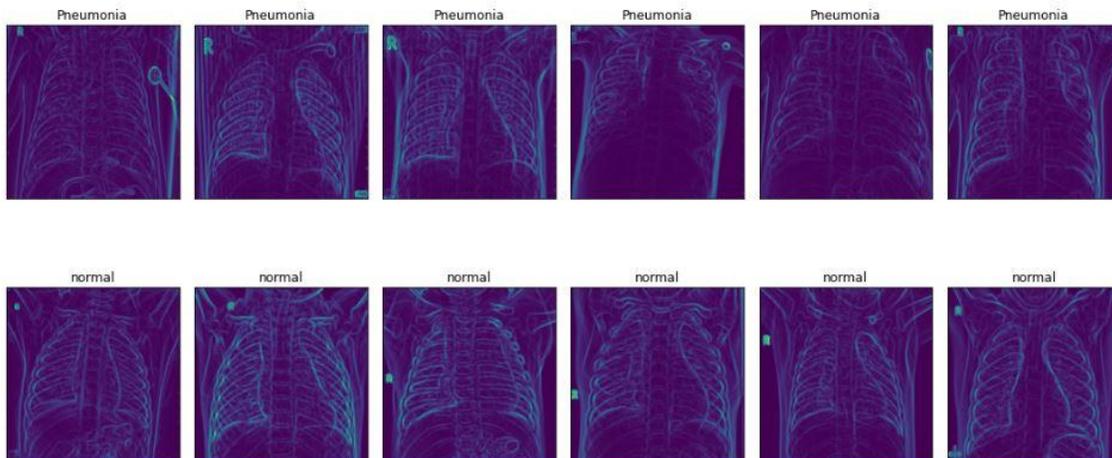Figure 1. input Image Dialation

Figure 2. Sober Filter on input images

Results

The VGG16 model used for the task has a total of 14,714,688 parameters. All of these parameters are trainable, meaning they can be adjusted during the training process to improve the model's performance. There are no non-trainable parameters in this specific configuration. After evaluating the model on the testing dataset, it achieved an accuracy of 74.9%, which indicates the proportion of correctly classified samples. Additionally, the testing loss, a measure of the model's prediction error, is recorded as 48.8.

Figure 3. VGG16 (Model accuracy and loss)
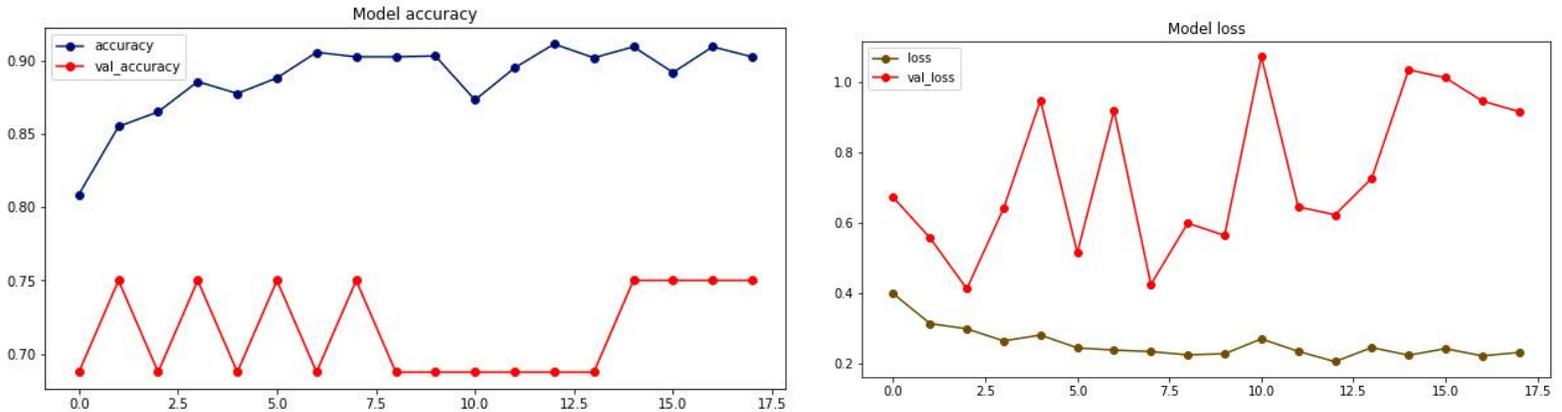
The ResNet50 model was employed for the task, and upon evaluation on the testing dataset, it achieved an impressive testing accuracy of 88.9%. This accuracy metric reflects the proportion of correctly classified samples by the model. Additionally, the testing loss, a measure of the model's prediction error, was recorded as 28.9, which indicates how well the model's predictions align with the actual ground truth values. These results demonstrate the strong performance and robustness of the ResNet50 model in handling the given task, showcasing its ability to accurately classify and generalize to new data.

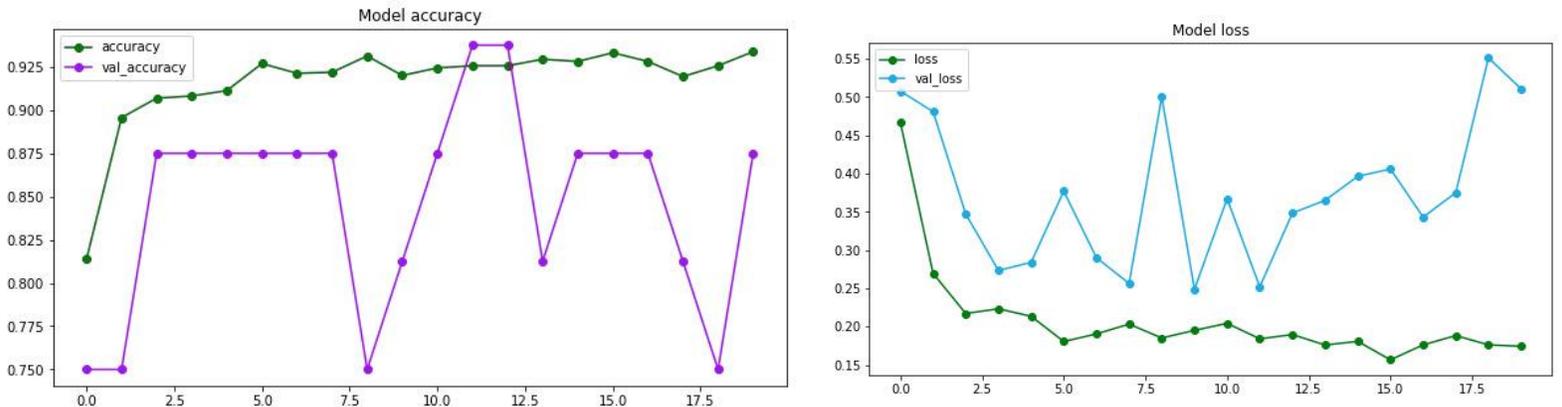Figure 4. VGG16 (Model accuracy and loss)

43

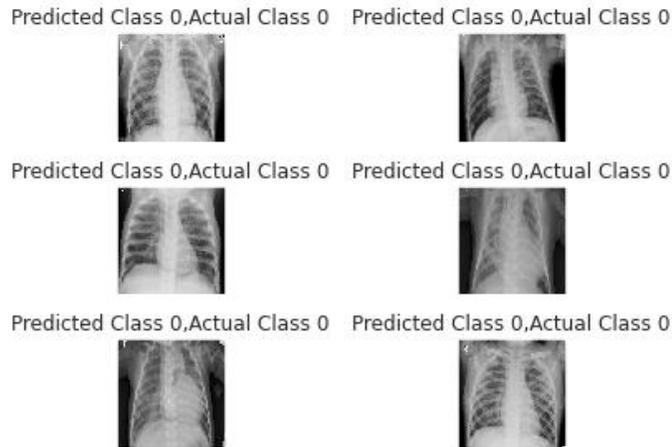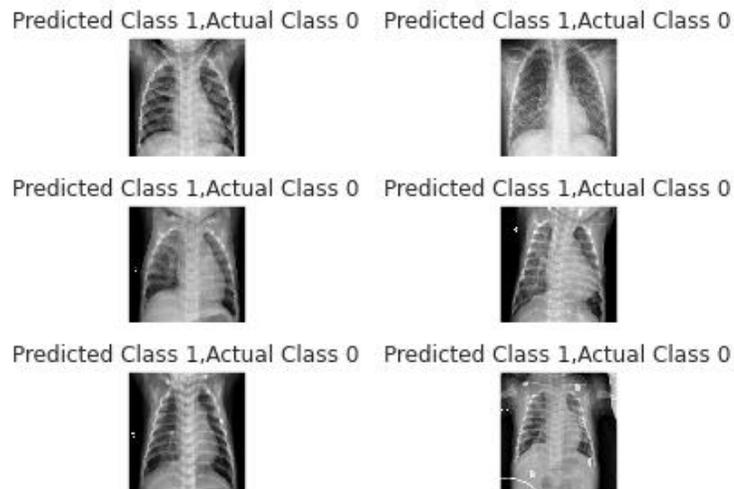Figure 5. Prediction performance demonstration Correctly predicted class

Figure 6. Prediction performance demonstration Correctly predicted class

Conclusion

This research emphasizes the crucial role of medical imaging and deep learning in pediatric pneumonia detection. Using convolutional neural networks, specifically VGG-16 and ResNet-50, our study investigated the potential of these models in accurately classifying chest X-ray images into pneumonia and normal categories. The results clearly demonstrate that ResNet-50 outperformed VGG-16 in terms of accuracy and loss during testing. With a testing accuracy of 88.9% and a testing loss of 28.9%, ResNet-50 exhibited superior performance compared to VGG-16's testing accuracy of 74.9% and testing loss of 48.8%.

The study applies standardization and other preprocessing techniques to enhance the data, the impact of these preprocessing techniques on the CNN models' performance is not extensively explored. Alternative preprocessing methods could potentially yield different results, and a sensitivity analysis of different preprocessing techniques would strengthen the study's conclusions. The study solely evaluates the performance of the CNN models on the dataset used for training and testing. The absence of an external validation dataset from a different source or institution limits the assessment of the models' generalizability to unseen data.

The study does not address the potential consequences of false positives in the context of clinical decision-making. False positives may lead to unnecessary follow-up tests or treatments, which could impact healthcare resources and patient well-being. A comprehensive analysis of false positives and their implications would add value to the study's applicability in clinical practice. While the study compares VGG-16 and ResNet-50, it does not compare the performance of these CNN models with other pneumonia detection methods, such as traditional machine learning algorithms or radiologist interpretations. A comparison with other approaches would provide a broader perspective on the superiority of deep learning models in this specific application.

Our study contributes to the growing body of evidence supporting the integration of deep learning models in pediatric pneumonia detection. The promising results obtained with the ResNet-50 model underscore its potentials in medical imaging landscape and improve patient care. Accurate and efficient pneumonia diagnosis in pediatric patients is crucial for timely treatment and better health outcomes. As technology advances, the adoption of deep learning architectures in clinical practice could lead to more effective disease detection and better resource allocation in healthcare settings.

References

[1] O. Ruuskanen, E. Lahti, L. C. Jennings, and D. R. Murdoch, "Viral pneumonia," *Lancet*, 2011.
[2] G. S. Sawicki, F. L. Lu, and C. Valim, "Necrotising pneumonia is an increasingly detected complication of pneumonia in children," *European*, 2008.
[3] D. M. Musher and A. R. Thorner, "Community-acquired pneumonia," *N. Engl. J. Med.*, 2014.
[4] I. Rudan, C. Boschi-Pinto, and Z. Biloglav, "Epidemiology and etiology of childhood pneumonia," *Bull. World Health Organ.*, 2008.
[5] J. Rello and E. Diaz, "Pneumonia in the intensive care unit," *Crit. Care Med.*, 2003.
[6] E. Prina, O. T. Ranzani, and A. Torres, "Community-acquired pneumonia," *Lancet*, 2015.
[7] G. H. McCracken Jr, "Etiology and treatment of pneumonia," *Pediatr. Infect. Dis. J.*, 2000.
[8] W. R. Hendee and E. R. Ritenour, "Medical imaging physics," 2003.
[9] H. Kasban and M. A. M. El-Bendary, "A comparative study of medical imaging techniques," *International Journal of*, 2015.
[10] E. A. Krupinski, "The importance of perception research in medical imaging," *Radiat. Med.*, 2000.
[11] G. N. Hounsfield, "Computed medical imaging," *Science*, 1980.

[12] R. R. Dixit, "Factors Influencing Healthtech Literacy: An Empirical Analysis of Socioeconomic, Demographic, Technological, and Health-Related Variables," *Applied Research in Artificial Intelligence and Cloud*, 2018.

[13] M. I. Pramanik, R. Y. K. Lau, H. Demirkan, and M. A. K. Azad, "Smart health: Big data enabled health paradigm within smart cities," *Expert Syst. Appl.*, vol. 87, pp. 370–383, Nov. 2017.

[14] K.-H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," *Nat Biomed Eng*, vol. 2, no. 10, pp. 719–731, Oct. 2018.

[15] H. M. Krumholz, "Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system," *Health Aff.* , vol. 33, no. 7, pp. 1163–1170, Jul. 2014.

[16] R. R. Dixit, R. P. Schumaker, and M. A. Veronin, "A Decision Tree Analysis of Opioid and Prescription Drug Interactions Leading to Death Using the FAERS Database," 2018, pp. 67–67.

[17] W. R. Hendee, G. J. Becker, J. P. Borgstede, and J. Bosma, "Addressing overutilization in medical imaging," *Radiology*, 2010.

[18] J. G. Lee, S. Jun, Y. W. Cho, and H. Lee, "Deep learning in medical imaging: general overview," *Korean journal of*, 2017.

[19] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[20] H. H. Aghdam and E. J. Heravi, "Guide to convolutional neural networks," *New York, NY: Springer*, 2017.

[21] J. Wu, "Introduction to convolutional neural networks," *National Key Lab for Novel Software Technology*, 2017.

[22] A. Lavin and S. Gray, "Fast algorithms for convolutional neural networks," *Proceedings of the IEEE conference on*, 2016.

[23] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018.

[24] J. Bouvrie, "Notes on convolutional neural networks," 2006.

[25] E. Rezende, G. Ruppert, T. Carvalho, A. Theophilo, F. Ramos, and P. de Geus, "Malicious software classification using VGG16 deep neural network's bottleneck features," in *Advances in Intelligent Systems and Computing*, Cham: Springer International Publishing, 2018, pp. 51–59.

[26] H. Qassim, D. Feinzimer, and A. Verma, "Residual Squeeze VGG16," *arXiv [cs.CV]*, 05-May-2017.

[27] X. Xie, X. Han, Q. Liao, and G. Shi, "Visualization and Pruning of SSD with the base network VGG16," in *Proceedings of the 2017 International Conference on Deep Learning Technologies*, Chengdu, China, 2017, pp. 90–94.

[28] T. Akiba, S. Suzuki, and K. Fukuda, "Extremely Large Minibatch SGD: Training ResNet-50 on ImageNet in 15 Minutes," *arXiv [cs.DC]*, 12-Nov-2017.

[29] M. Sankupellay and D. Konovalov, "Bird call recognition using deep convolutional neural network, ResNet-50," *Proc. Acoustics*, 2018.

[30] H. Mikami, H. Suganuma, P. U-chupala, Y. Tanaka, and Y. Kageyama, "Massively Distributed SGD: ImageNet/ResNet-50 Training in a Flash," *arXiv [cs.LG]*, 13-Nov-2018.

[31] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (oct) and chest x-ray images for classification," *Mendeley data*, 2018.

[32] D. S. Kermany *et al.*, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122-1131.e9, Feb. 2018.

[33] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (OCT) and chest X-ray images for classification (2018)," *Mendeley Data, v2 https://doi. org/10.17632*, 1763.

[34] J. L. Mitchell, "Image processing with 1.4 pixel shaders in Direct3D," *Direct3D ShaderX: Vertex and Pixel Shader Tips and*, 2002.

[35] D. Bernecker, "Image Processing," in *Medical Imaging Systems: An Introductory Guide*, A. Maier, S. Steidl, V. Christlein, and J. Hornegger, Eds. Cham (CH): Springer, 2018.

[36] K. Thenmozhi and U. S. Reddy, "Image processing techniques for insect shape detection in field crops," in *2017 International Conference on Inventive Computing and Informatics (ICICI)*, 2017, pp. 699–704.

[37] Z. Vahabi, M. Vafadoost, and S. Gharibzadeh, "The new approach to automatic detection of Optic Disc from non-dilated retinal images," in *2010 17th Iranian Conference of Biomedical Engineering (ICBME)*, 2010, pp. 1–6.