Asian Journal of Applied Science and Technology (AJAST)
Volume 8, Issue 4, Pages 21-30, Oct-Dec 2024

Impact of Dataset Size on the Performance of Deep Learning Models: A Case Study on Lung Ultrasound and X-ray Image Classification

S.E. Morsy[1*], P.E. Salah[2], M.A. Bibars[3], N.M. Abd-Elsalam[4], A.H. Kandil[5], A. Elbialy[6] & A.M. Youssef[7]

[1,2,3]PhD Candidate, [4]PhD., [5,6,7]Professor, [1-7]Department of Systems and Biomedical Engineering, Cairo University, Egypt.
Corresponding Author (S.E. Morsy) Email: shereen.morsy.m@eng-st.cu.edu.eg[*]

DOI: https://doi.org/10.38177/ajast.2024.8403

Article Received: 08 August 2024 *Article Accepted: 22 October 2024* *Article Published: 27 October 2024*

ABSTRACT

Pneumonia is a severe respiratory condition that can affect individuals of all ages. Its diagnosis traditionally relies on chest X-rays, interpreted by experienced physicians, a process that can be both time-intensive and demanding. Ultrasound imaging has emerged as an alternative diagnostic tool. Meanwhile, recent advances in deep learning have introduced automated frameworks capable of analyzing medical images to classify cases as either normal or pneumonia, offering a faster, reliable alternative comparable to expert evaluations. This study presents a comprehensive comparative analysis of multiple deep learning models for classifying both ultrasound and chest X-ray images. The results demonstrate that a Convolutional Neural Network (CNN) achieved the highest accuracy for ultrasound-based diagnosis with accuracy of 99.9%, whereas the VGG-16 model outperformed others in X-ray image classification. To enhance performance, six different X-ray datasets were combined and augmented. VGG-16 achieved the highest accuracy (95.7%).

Keywords: Pneumonia; Ultrasound; X-ray; Artificial intelligence (AI); Deep learning; Convolutional neural network (CNN); Pre-trained models; Transfer learning; Lung; Medical image classification.

1. Introduction

Pneumonia is a lower respiratory tract infection characterized by symptoms such as fever, chills, cough with sputum production, chest pain, and shortness of breath. It occurs when the air sacs (alveoli) in the lungs become inflamed, often due to infection, resulting in impaired oxygen exchange [1]. Pneumonia can affect one or more sections of the lungs, leading to conditions such as double or multilobar pneumonia. Although anyone can develop pneumonia, it mostly affects young children and the elderly [2]. The disease manifests in different forms, including bacterial, viral, and mycoplasma pneumonia, with bacterial pneumonia being the most severe and requiring prompt medical attention. Symptoms of bacterial pneumonia may develop gradually or appear suddenly [3].

The World Health Organization (WHO) announced that Pneumonia is the most common infectious cause of death in children. As of 2021, pneumonia remains a major global health issue, particularly for children under five, causing approximately 500,000 deaths annually in this age group, with 740,180 children under five dying from it in 2019 [4]. Despite significant progress in healthcare and vaccination efforts, pneumonia continues to disproportionately affect low-income regions, particularly South Asia and Sub-Saharan Africa, where access to preventive measures and healthcare remains limited.

Pneumonia diagnosis relied on physical examinations, patient history, and symptom evaluation. However, these clinical assessments alone are insufficient, necessitating imaging techniques for confirmation. Chest X-ray remains the gold standard for pneumonia diagnosis, as recommended by international guidelines [5- 7] alongside laboratory tests, sputum cultures, and clinical history [8] Radiologists diagnose pneumonia by identifying characteristic signs on X-rays, such as lung infiltrates, dense opacity, poorly defined centrilobular nodules, and airspace consolidation [9]. However, manual interpretation of X-rays can be time-consuming and labor-intensive. Rajpurkar et al. [10]

demonstrated that deep learning algorithms could label X-ray images within 1.5 minutes with accuracy comparable to radiologists. Despite its effectiveness, chest radiography presents challenges, including high costs, radiation exposure, and limited feasibility for critically ill patients. Lung ultrasound (LUS) has emerged as a valuable bedside tool for assessing pulmonary conditions, offering higher sensitivity than X-rays in detecting pleural effusions [11]. LUS can also identify pulmonary changes associated with pneumonia, particularly when these changes involve the outer pleural surface.

This study aims to perform a comprehensive comparative analysis of deep learning models for classifying pneumonia from medical images. First, we developed a model to diagnose pneumonia using ultrasound images, employing pre-trained models for classification and validating the results with three-fold cross-validation. Additionally, for X-ray image classification, we implemented data augmentation techniques and utilized six datasets. The results highlight the effectiveness of both approaches in accurately distinguishing between normal and pneumonia cases.

1.1. Study Objectives

The objectives of this study can be formulated as follows: (1) Compare the performance of different deep learning models on Ultrasound and X-ray image modalities, (2) Evaluate the generalization capability of deep learning models, focusing on their ability to generalize across training, validation, and testing datasets, (3) Investigate the extent of overfitting challenges in deep learning models, particularly when applied to smaller datasets, and examine how this impacts real-world applicability, (4) Identify models that demonstrate strong performance on smaller datasets (limited data), and (5) Recommend suitable architectures for deployment in medical image classification that balance accuracy and generalization.

2. Related Work

Born et al. [12] employed a CNN network to classify ultrasound images into normal, COVID-19, and pneumonia categories, achieving an accuracy of 89% through 5-fold cross-validation. For X-ray classification, Dey et al. [13] explored multiple pretrained models, including VGG-16/19, AlexNet, and ResNet50. They achieved 95.7% accuracy using a hybrid method combining VGG-19 features with Random Forest but only 86.07% with individual deep learning models. Their work, which relied on the Kermany et al. dataset [14] of 7,150 images, showed fluctuations across 12 epochs, with VGG-19 taking 12.48 ± 4 minutes per run. Ibrahim et al. [15] improved upon Dey's research by adding four datasets from GitHub and Kaggle to the Kermany dataset, accumulating 11,568 images. They used AlexNet but reported lower accuracies—94.93% for viral pneumonia vs. normal and 91.43% for bacterial pneumonia vs. normal—without specifying training time. One dataset link used in their study was also faulty. Narayanan et al. [16] compared multiple pretrained models, such as InceptionV3, DenseNet201, MobileNet versions, and achieved the highest accuracy (96.43%) on the Kermany dataset with MobileNetV3Large after 100 epochs. Although they highlighted the efficiency of MobileNetV3, the training time was not detailed. Villanueva et al. [17] used VGG-16/19, ResNet50, and InceptionV3 on 5,840 X-ray images, finding that InceptionV3 yielded 72.9% accuracy and an F1-score of 82%, though they suggested performance could improve with more data. In 2024, Mardianto et al. [18] compared CNNs and SVMs for pneumonia detection. CNNs achieved 91% overall

accuracy, outperforming SVMs' 79%, with CNNs also excelling in distinguishing bacterial from viral pneumonia at 92% accuracy, compared to 88% for SVMs.

3. Materials and Methods

3.1. Ultrasound Dataset

The used ultrasound dataset [12] consists of 63 lung video recordings, including 35 videos of pneumonia cases and 28 videos of normal lungs, captured using a convex probe. From these videos, individual frames were extracted, resulting in 4,104 pneumonia images and 3,668 normal lung images. The data was split into 5,113 images for training, 1,520 for validation, and 1,139 for testing. As a preprocessing step, images were resized to 224 × 224 pixels, normalized to a 0–1 range, and standardized. For avoiding overfitting problem, data augmentation strategies were utilized that boost model performance and generalizability. These augmentations comprised a range of transformations such as rotations, translations, zooming, shearing, and flipping, all of which increased the diversity of the training data and reduced the likelihood of the model "memorizing" the dataset.

3.2. X-Ray Datasets

This study combines six distinct datasets for pneumonia classification using chest X-ray images. The first dataset, by Cohen et al. [19], initially contained 679 X-rays, which increased to 930 images, with non-pneumonia categories (e.g., COVID-19) excluded for this study. The second dataset, RSNA pneumonia detection dataset [20], contributed 29,684 DICOM images, but noisy samples—those labelled incorrectly as normal—were removed to avoid confusion during training. The third dataset, Figure1 Chest X-ray [21], is small, with only 55 pneumonia images, but was included to improve the model's generalizability. Similarly, the fourth dataset, Actualmed Covid-19 Chest X-ray [22], provides 238 scans, with only pneumonia cases retained. The fifth dataset, Covid-19 Radiography [23, 24] supplies 1,345 pneumonia and 10,192 normal cases, with irrelevant categories omitted. Many of the normal cases in the study come from this source. Finally, the Kermany et al. dataset [25] offers 4,273 pneumonia and 1,584 normal cases, with high-quality images. After filtering unnecessary or noisy cases, the combined dataset includes 20,626 normal cases and 12,301 pneumonia cases. The data was split into training (75%) and testing (25%) sets. Data augmentation techniques, such as horizontal flipping and shifting, were applied to the training set to enhance model performance. This resulted in an expanded training dataset with 31,218 normal and 17,444 pneumonia images, used to train multiple models with varying architectures.

3.3. Deep Learning Architectures

Deep learning architectures, particularly Convolutional Neural Networks (CNNs), have become fundamental tools for image classification tasks. These networks are designed to automatically learn hierarchical features from data, making them well-suited for tasks like classification of medical images, where precise patterns are critical. many pre-trained models were investigated to apply transfer learning to the classification task. In this study, a CNN network was built from scratch, and several pre-trained models were employed by leveraging the transfer learning technique. We implemented the models using the TensorFlow framework [26] and the Keras library [27] The implementation and model training code were developed in Python and run on our local machine.

3.3.1. CNN Model

A conventional CNN consists of multiple layers, including convolutional layers (for feature extraction), pooling layers (for down-sampling), and fully connected layers (for classification). These architectures learn increasingly complex features at each layer—starting from edges and shapes in the lower layers to more abstract patterns in deeper layers. Here, as illustrated in Figure 1, the model was developed with four convolutional layers, each followed by batch normalization and max pooling to reduce dimensionality and control overfitting. The activation function used in all hidden layers is the Rectified Linear Unit (ReLU). After the convolutional layers, a flattening layer converts the multi-dimensional data into a one-dimensional vector. This is followed by two fully connected (FC) layers with 1,024 and 512 neurons, respectively, both utilizing ReLU activation. The final output layer uses the SoftMax function to perform classification across multiple categories (pneumonia and normal).

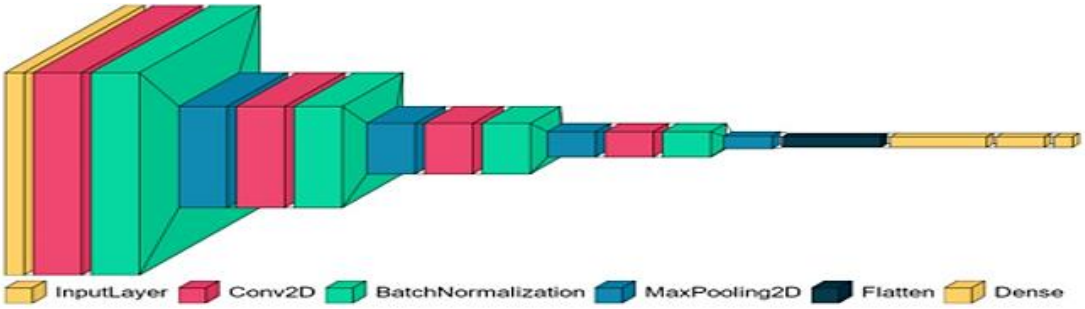

Figure 1. The Architecture of the Convolution Neural Network (CNN) Model

(a) CNN with Transfer Learning

Transfer learning has proven highly successful in achieving state-of-the-art performance in medical image classification tasks, as it mitigates the need for large, labeled datasets and reduces the risk of overfitting. Various techniques can be applied for transfer learning, such as utilizing a pre-trained CNN as a fixed feature extractor or fine-tuning its parameters for a specific image classification task. Since these CNN models are trained on large-scale datasets like ImageNet, they can effectively serve as feature extractors, as depicted in Figure 2. In this study, the pre-trained models employed include VGG16, ResNet50, InceptionResNetV2, EfficientNetB3, InceptionV3, and MobileNetV2.

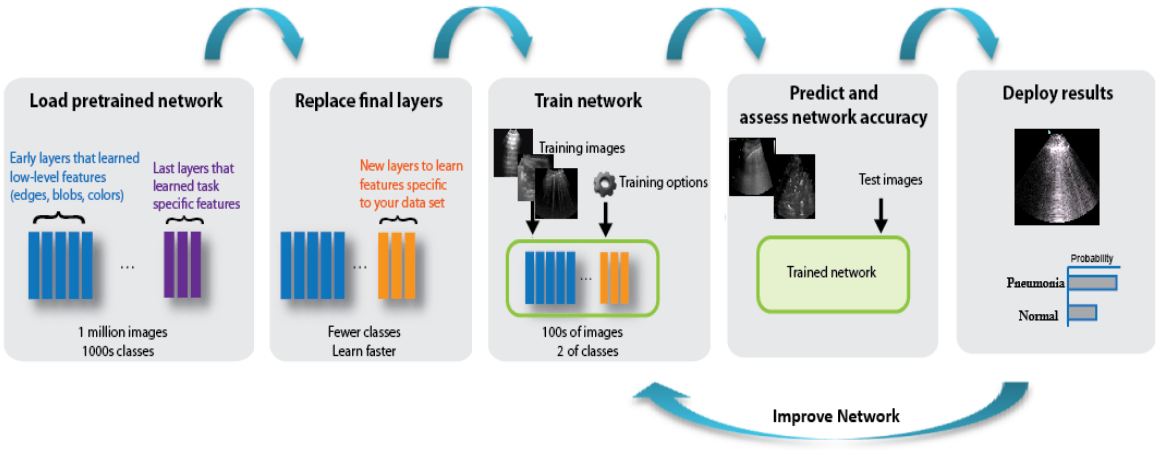

Figure 2. Transfer Learning steps from pre-trained models

The model architecture is presented in Figure 3, for lung images classification task into Pneumonia and Normal cases. The lung images are input into the pre-trained model, and the resulting feature maps are subsequently classified using a CNN classifier, which consists of average pooling followed by three fully connected layers. A dropout rate of 0.2 was tuned to prevent overfitting and ended with an output layer that employs SoftMax activation for classification. In other words, the network base model was frozen, and the fully connected layers were trained on the data with a very small learning rate (stochastic gradient descent, 0.0001). Thus, the fully connected layer's random weights were replaced by data specific weights. This methodology was followed in the all the models.

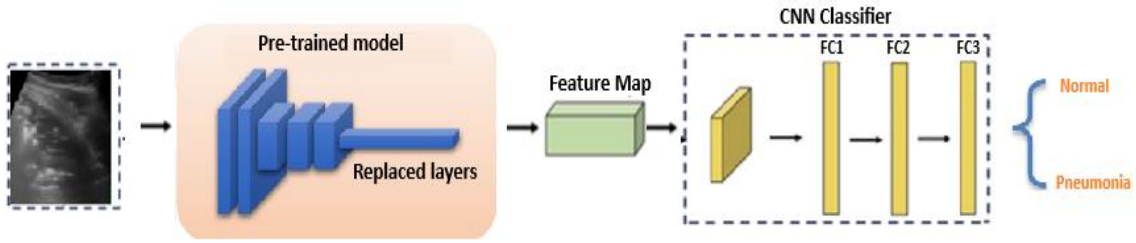

Figure 3. The Architecture of the CNN Transfer Learning Model

4. Results and Discussion

The evaluation of the model's performance is based on the accuracy of classification using the testing dataset, which is calculated using the formula in (1):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \ldots(1)$$

Accuracy serves as a fundamental metric for assessing the effectiveness of a classification model, providing insights into its ability to correctly identify both positive and negative cases. In this equation, TP (True Positives) and TN (True Negatives) represent the correctly classified instances, while FP (False Positives) and FN (False Negatives) denote the incorrectly classified instances. To assess the classification performance, the US dataset was divided into training, validation, and testing sets. While the X-ray dataset was split into training and testing. Although evaluating the network on the training dataset is possible, it does not provide a reliable indication of the network's performance as a predictive model since it has already encountered this data. Instead, performance should be assessed using a separate dataset, referred to as the testing dataset, which has not been seen during training. This approach offers a more accurate estimation of the network's ability to make predictions on unseen data in the future.

Figure 4 highlights the performance of several deep learning models for the lung US classification task, after being trained for 50 epochs, revealing diverse outcomes in terms of accuracy across training and validation phases, while Figure 5 presents accuracy, F1-score, and precision of testing phase. The CNN architecture demonstrates exceptional consistency, achieving 99.9% on both training and validation accuracy. This suggests that the model generalizes well, without signs of overfitting. VGG16, MobileNetV2, and ResNet50, however, demonstrate signs of significant overfitting. Although VGG16 achieves 99.8% on the training set, its validation accuracy drops to 99.7%, and its evaluation metrics (accuracy: 52.41%, F1: 52.41%, precision: 52.41%) suggest poor performance in real-world settings. Also, the drastically low scores of MobileNetV2 model (Accuracy: 49.6%, F1-Score: 49.61%,

Precision: 49.63%) reflect severe issues with its generalization capability, even though the training and validation accuracies were near perfect (99.9%). This discrepancy points to serious performance degradation on the test set, which could be caused by overfitting to specific patterns in the training and validation data. ResNet50 fares even worse, with a steep drop from 97% training accuracy to only 55% validation accuracy, accompanied by very low accuracy (47.15%) and precision (22.2%). This indicates that the model may not have effectively captured generalizable patterns.

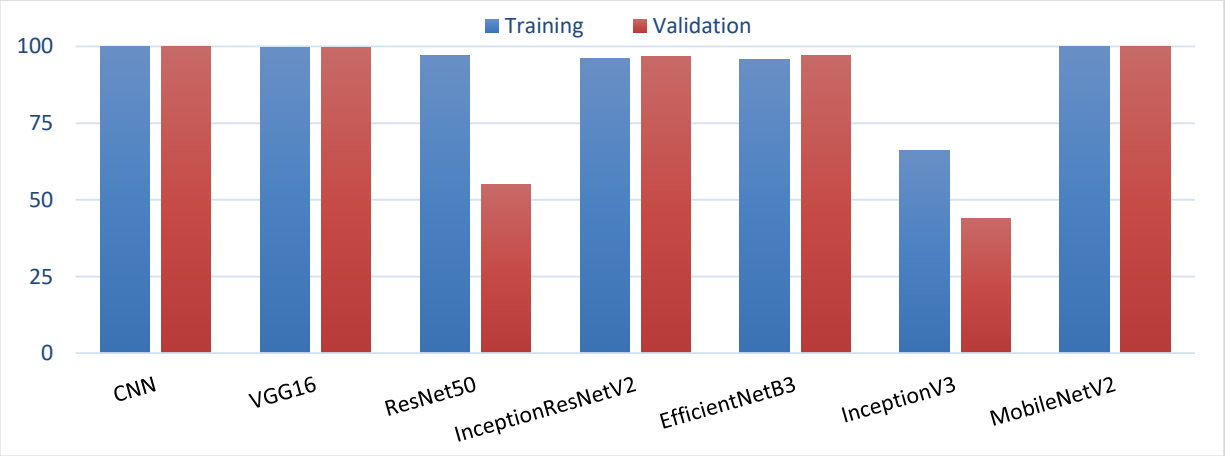

Figure 4. The Performance of several Deep learning models for the Lung US Classification task during the Training phase

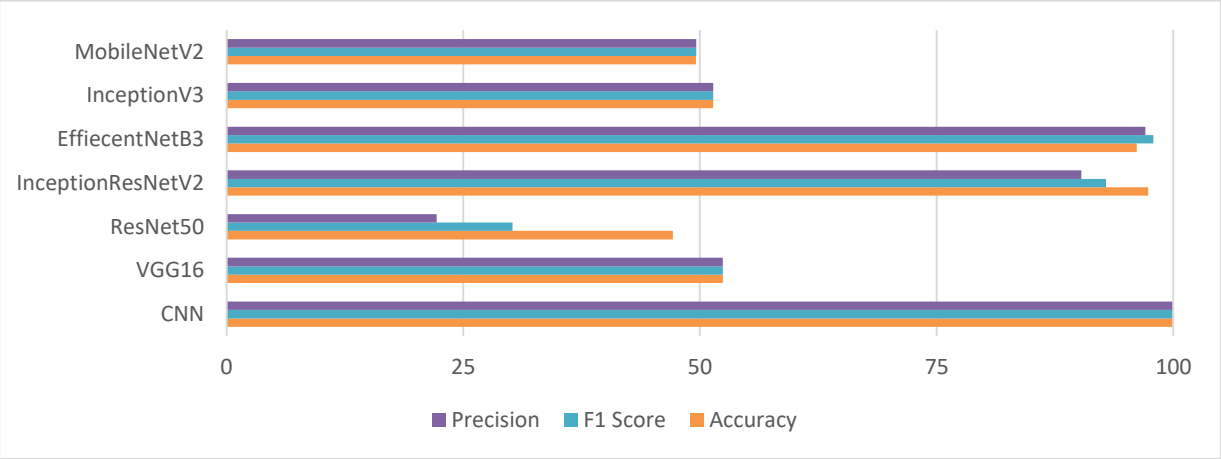

Figure 5. Performance Metrices of several Deep learning models for the Lung US Classification task on the Testing phase

On the other hand, InceptionResNetV2 and EfficientNetB3 perform well, with validation metrics either matching or slightly exceeding their training scores. Specifically, EfficientNetB3 achieves 95.9% in training and 97% in validation, reflecting a well-trained model with minimal overfitting. InceptionResNetV2 follows a similar trend, with a training score of 96.2% and validation accuracy of 96.7%. The test Accuracy of 97.3%, F1-score of 92.9% and precision of 90.3% further underscore its reliable performance.

InceptionV3 struggles with both training and validation, obtaining 66% and 44% accuracy, respectively, which results in poor generalization (accuracy: 51.4%, F1-score: 51.4%). This suggests that the model might require

tuning or additional data preprocessing to improve learning. These results demonstrate varying degrees of performance across these models. CNN, EfficientNetB3, and InceptionResNetV2 emerge as the best performers, showing consistent results across different metrics. In contrast, VGG16, MobileNetV2, InceptionV3, and ResNet50 reveal challenges with overfitting and weak generalization. This analysis underscores the importance of balancing training and validation performance to avoid overfitting and ensure real-world applicability.

On the other hand, the X-ray results are shown in Table 1, that shows a comparative result for training and validation accuracies of different deep learning models. It is important to mention that all these models are trained for only 10 epochs.

Table 1. Training and Testing Accuracies of X-Ray Classification task

Architecture	**Training Accuracy**	**Testing Accuracy**
VGG16	99.5%	95.7%
InceptionV3	93.5%	92.9%
InceptionResNetV2	99.5%	93.3%
EfficientNetB3	95.9%	94.8%
ResNet50	96.4%	94.6%

The results for X-ray image classification, where models were trained with significantly larger datasets, paint a different picture. VGG16 stands out with 95.7% test accuracy, showing much-improved generalization compared to its performance on ultrasound data. InceptionV3, which struggled earlier, now achieves a respectable 92.9% on the test set, while InceptionResNetV2 also shows strong performance with a 93.3% test accuracy. EfficientNetB3 and ResNet50 follow closely, with test accuracies of 94.8% and 94.6%, respectively. These models demonstrate better overall learning and generalization, thanks to the availability of more comprehensive datasets for X-ray classification. Comparing both modalities' results, it can be concluded that VGG16, ResNet50, and InceptionV3 show significant improvements on the X-ray dataset, confirming that they require larger datasets to generalize well. On the other hand, EfficientNetB3 and InceptionResNetV2 are more robust architectures, capable of performing well even on smaller datasets.

In summary, the disparity between the results highlights the impact of dataset size and quality on model performance. Models that struggled with overfitting or underfitting on smaller ultrasound datasets were able to perform much better with larger and more diverse X-ray datasets. This comparison emphasizes the importance of both dataset size and task-specific tuning when deploying deep learning models for medical image classification. Overall, CNN, InceptionResNetV2, and EfficientNetB3 stand out as reliable models for the classification task, showing strong generalization and balanced performance. On the other hand, VGG16, MobileNetV2, InceptionV3, and ResNet50 require additional regularization to overcome overfitting, while. Selecting the best model for deployment will likely require focusing on architectures like CNN or EfficientNetB3, which demonstrate the most consistent performance across datasets.

5. Conclusion

The task of classifying lung images from ultrasound (US) and X-rays into pneumonia or normal cases demonstrates the challenges and nuances of deep learning in medical image analysis. Our study reveals that the performance of models is heavily influenced by the nature of the data, model architecture, and dataset size. The findings underscore that no single model or approach is universally optimal across different imaging modalities. While CNN architectures showed strong potential for ultrasound data, pretrained models like VGG16 and EfficientNetB3 excelled in X-ray classification due to richer datasets. Therefore, achieving high performance in pneumonia classification requires not only careful selection of models but also access to well-curated datasets tailored to the imaging modality. This highlights the need for continued refinement of architectures and collection of diverse data to ensure accurate and generalizable medical diagnoses. So, it is recommended for future applications, to use EfficientNetB3 or InceptionResNetV2 when working with smaller datasets, as these models have shown consistent generalization. Meanwhile, VGG16, ResNet50, and InceptionV3 can be utilized effectively for larger datasets but may require careful tuning or data augmentation if applied to smaller datasets to avoid overfitting. For future work, it is strongly recommended to expand dataset collection efforts to enhance model training and improve robustness. Additionally, advanced regularization techniques should be developed to mitigate overfitting, alongside an exploration of refined fine-tuning strategies for better results. Furthermore, investigating ensemble learning approaches could significantly enhance accuracy and reliability in medical image diagnosis. It is also vital to develop lightweight model variants, particularly for models like EfficientNetB3 and InceptionResNetV2, to facilitate deployment in low-resource settings where computational power and memory are limited but accurate diagnostics are crucial.

Declarations

Source of Funding

This study has been financially supported by the Information Technology Industry Development Agency (ITIDA).

Competing Interests Statement

The authors declare no competing financial, professional, or personal interests.

Consent for publication

The authors declare that they consented to the publication of this study.

Acknowledgment

All authors gratefully acknowledge the Information Technology Industry Development Agency (ITIDA) for providing necessary financial support.

References

[1] American Thoracic Society (2016). What is Pneumonia. American Journal of Respiratory and Critical Care Medicine, 193: 1–2. https://www.scribd.com/document/499899650/what-is-pneumonia.

[2] Mani, C.S. (2017). Acute Pneumonia and Its Complications. Principles and Practice of Pediatric Infectious Diseases, Pages 234–237. https://doi.org/10.1016/b978-0-323-40181-4.00034-7.

[3] American Lung Association (2024). Pneumonia Symptoms and Diagnosis. Accessed: Sep. 25, 2024. https://www.lung.org/lung-health-diseases/lung-disease-lookup/pneumonia/symptoms-and-diagnosis.

[4] World Health Organization. (2024). Pneumonia. Accessed: Sep. 25, 2024. https://www.who.int/news-room/fact-sheets/detail/pneumonia.

[5] Levy, M.L., Le Jeune, I., Woodhead, M.A., Macfarlaned, J.T., & Lim, W.S. (2010). Primary care summary of the British Thoracic Society Guidelines for the management of community-acquired pneumonia in adults: 2009 update. Primary Care Respiratory Journal, 19(1): 21–27. https://doi.org/10.4104/pcrj.2010.00014.

[6] Mandell, L.A., et al. (2007). Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clinical Infectious Diseases, 44(S2): S27–72. https://doi.org/10.1086/511159.

[7] Hoare, Z., & Lim, W.S. (2006). Pneumonia: Update on diagnosis and management. BMJ, 332(7549): 1077–1079. https://doi.org/10.1136/bmj.332.7549.1077.

[8] Htun, T.P., Sun, Y., Chua, H.L., & Pang, J. (2019). Clinical features for diagnosis of pneumonia among adults in primary care setting: A systematic and meta-review. Scientific Reports, 9(1): 7600. https://doi.org/10.1038/s41598-019-44145-y.

[9] Koo, H.J., Lim, S., Choe, J., Choi, S.H., Sung, H., & Do, K.H. (2018). Radiographic and CT Features of Viral Pneumonia. RadioGraphics, 38(3): 719–739. https://doi.org/10.1148/rg.2018170048.

[10] Rajpurkar, P., et al. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Medicine, 15(11): e1002686. https://doi.org/10.1371/journal.pmed.1002686.

[11] Long, L., Zhao, H.T., Zhang, Z.Y., Wang, G.Y., & Zhao, H.L. (2017). Lung ultrasound for the diagnosis of pneumonia in adults: A meta-analysis. Medicine, 96(3): 1–6. https://doi.org/10.1097/md.0000000000005713.

[12] Born, J., Brändle, G., Cossio, M., Disdier, M., Goulet, J., & Roulin, J. (2021). POCOVID-NET: Automatic Detection of COVID-19 from a New Lung Ultrasound Imaging Dataset (POCUS).

[13] Dey, N., Zhang, Y.D., Rajinikanth, V., Pugalenthi, R., & Raja, N.S.M. (2021). Customized VGG19 Architecture for Pneumonia Detection in Chest X-Rays. Pattern Recognition Letters, 143: 67–74. https://doi.org/10.1016/j.patrec.2020.12.010.

[14] Kermany, D., Zhang, K., & Goldbaum, M. (2018). Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification. https://doi.org/10.17632/rscbjbr9sj.2.

[15] Ibrahim, A.U., Ozsoz, M., Serte, S., Al-Turjman, F., & Yakoi, P.S. (2024). Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19. Cognitive Computation, 16(4): 1589–1601. https://doi.org/10.1007/s12559-020-09787-5.

[16] Narayanan, B., Kuriakose, V.A., & Sreekumar, K. (2022). A Comparative Analysis of Pneumonia Detection Using Various Models of Transfer Learning. In P. Karuppusamy, I. Perikos, & F.P. García Márquez (Eds.), Ubiquitous Intelligent Systems, Pages 143–155, Springer Singapore.

[17] Iparraguirre-Villanueva, O., Guevara-Ponce, V., Roque Paredes, O., Sierra-Liñan, F., Zapata-Paulini, J., & Cabanillas-Carbonell, M. (2022). Convolutional neural networks with transfer learning for pneumonia detection. International Journal of Advanced Computer Science and Applications, 13(9). https://doi.org/10.14569/ijacsa.2022.0130963.

[18] Mardianto, M., Yoani, A., Soewignjo, S., Putra, I., & Dewi, D.A. (2024). Classification of Pneumonia from Chest X-ray images using Support Vector Machine and Convolutional Neural Network. International Journal of Advanced Computer Science and Applications, 15(6). https://doi.org/10.14569/ijacsa.2024.01506104.

[19] Cohen, J.P., Morrison, P., Dao, L., Roth, K., Duong, T.Q., & Ghassemi, M. (2020). COVID-19 Image Data Collection: Prospective Predictions are the Future. ArXiv Preprint arXiv:2006.11988.

[20] 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. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Pages 2097–2106.

[21] Wang, L., Wong, A., Lin, Z.Q., McInnis, P., Chung, A., & Gunraj, H. (2020). Figure 1 COVID-19 Chest X-ray Dataset Initiative. https://github.com/agchung/figure1-covid-chestxray-dataset.

[22] Wang, L., Wong, A., Lin, Z.Q., McInnis, P., Chung, A., & Gunraj, H. (2020). Actualmed COVID-19 Chest X-ray Dataset Initiative. https://github.com/agchung/actualmed-covid-chestxray-dataset.

[23] Chowdhury, M.E.H., et al. (2020). Can AI Help in Screening Viral and COVID-19 Pneumonia? IEEE Access, 8: 132665–132676. https://doi.org/10.1109/access.2020.3010287.

[24] Rahman, T., et al. (2021). Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection Using Chest X-ray Images. Computers in Biology and Medicine, 132: 104319. https://doi.org/10.1016/j.compbiomed.2021.104319.

[25] Kermany, D., Zhang, K., & Goldbaum, M. (2018). Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images.

[26] TensorFlow: An end-to-end platform for machine learning. Accessed: Sep. 25, 2024. https://www.tensorflow.org/.

[27] Keras: The Python deep learning API. Accessed: Sep. 25, 2024. https://keras.io/.