European Journal of Computer Science and Information Technology, 12 (7), 1-11, 2024
Print ISSN: 2054-0957 (Print),
Online ISSN: 2054-0965 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

A Comparative Study on the Detection of Pneumonia in Chest X-Ray Images Utilizing Deep Learning Models

Muhammad Maruf Billah [1], Abdullah Al Rakib[2], Asif Shakil Ahamed[3], Sakib Chowdhury[4], Satu Mitro[5]
[1,2,3,4&5]School of Artificial Intelligence, Nanjing University of Information Science & Technology, Nanjing, China.
Corresponding author email: maruf.swpu@gmail.com

doi: https://doi.org/10.37745/ejcsit.2013/vol12n7111 Published October 27 2024

Citation: Billah M.M., Al Rakib A., Ahamed A.S., Chowdhury S., Mitro S. (2024) A Comparative Study on the Detection of Pneumonia in Chest X-Ray Images Utilizing Deep Learning Models, *European Journal of Computer Science and Information Technology*, 12 (7), 1-11

Abstract: *Pneumonia continues to pose a notable public health issue on a global scale, highlighting the crucial need for precise and prompt identification to lessen its effects on patient prognosis. Chest X-ray imaging is a common diagnostic tool for identifying pneumonia due to its non-invasiveness and wide availability. Recently, convolutional neural networks (CNNs), a type of deep learning method, have shown promise in automating the detection of pneumonia from X-ray images. In this paper, we present a comprehensive comparative study of three popular deep learning models—VGG16, Inception V3, and ResNet 50V2—for pneumonia detection in X-ray datasets. The dataset used consists of 5,863 chest X-ray images collected from Kaggle, which are classified into two main categories: pneumonia and normal. The ResNet 50V2 model did very well in the experiments; it was able to correctly identify pneumonia 95.97% of the time, which was better than both VGG16 (93.58%) and Inception V3 (93.38%). We also conduct an analysis of performance metrics such as validation loss, validation accuracy, recall, and precision, and calculate the area under the receiver operating characteristic (ROC) curve (AUC) for each model. We also talk about how to compare ROC curves and precision-recall curves, which lets you see how well the models do at telling the difference between things and how well they do across a number of evaluation metrics. Our study contributes to the development of efficient and high-performance deep learning models for improving the diagnosis and treatment processes for pneumonia patients in the medical field.*

Keywords: Pneumonia, Chest X-ray, Comparative Analysis, VGG16, Convolutional Neural Networks, Inception V3, Deep Learning, ResNet 50V2

INTRODUCTION

Pneumonia stands as a prominent contributor to the rates of illness and death on a global scale; the impact is particularly significant on vulnerable populations such as children, the elderly, and individuals with pre-existing health conditions [1]. Timely and precise identification of pneumonia is critical in initiating prompt therapeutic measures and preventing serious complications. While clinical assessment and laboratory tests play a vital role in diagnosis, chest X-ray imaging remains the cornerstone for confirming the presence of pneumonia [2].

As humanity grapples with a multitude of illnesses, pneumonia stands out as "one of the prevalent acute respiratory and pulmonary ailments" and ranks among the leading causes of mortality worldwide. The lack of current medical technology in early pneumonia detection has resulted in the loss of a minimum of 1.4 million young lives, with approximately 1 million adults requiring hospitalization, leading to around 50,000 deaths annually in the United States alone. However, the diagnosis of pneumonia depends not only on the efficiency of medical technologies employed but also on the availability of skilled radiologists [3].

The emergence of deep learning methodologies has sparked a surge in enthusiasm for utilizing artificial intelligence (AI) to streamline the analysis of medical imagery, such as chest X-rays, to identify pneumonia. [4]. Convolutional neural networks (CNNs), a category of advanced learning models in the field of artificial intelligence, have shown impressive efficacy in tasks involving image classification, demonstrating a high level of appropriateness in the identification of pneumonia from radiographic images of the chest [5]. Nevertheless, the choice of a CNN structure may have a notable impact on both the model's performance and capacity for generalization.

This paper presents a comparative analysis of three commonly used CNN architectures—VGG16, Inception V3, and ResNet 50V2—used for the detection of pneumonia in chest X-ray images. We conduct the assessment and comparison of these models' efficacy on an extensive dataset from Kaggle, which includes both pneumonia and normal chest X-ray images. The aim is to pinpoint the model exhibiting the highest level of accuracy and efficiency in pneumonia detection, thus streamlining diagnostic processes in clinical environments.

RELATED WORK

Researchers have dedicated several academic inquiries to examining the use of deep learning frameworks for identifying pneumonia from radiographic images of the thorax. Rajpurkar et al. [6] introduced and trained the CheXNet CNN model on an extensive dataset of chest X-ray images

annotated with a variety of thoracic ailments, including pneumonia. The outcomes of their study exhibited performance comparable to that of radiologists in pneumonia detection, underscoring the promising prospects of deep learning in the realm of medical imaging diagnostics.

Similarly, Wang et al. [7] developed the ChestX-ray8 dataset, which comprises a substantial collection of chest X-ray images. They introduced novel weakly supervised learning techniques for pneumonia classification and localization. This research contribution provided significant insights into the challenges and potentials of leveraging deep learning to diagnose pneumonia.Shah A et al. [8] carried out a study examining various deep learning models designed for the purpose of detecting pneumonia in chest X-ray (CXR) images. The assessment encompasses an examination of the effectiveness, limitations, and efficiency of the existing models. This study provides significant perspectives on the application of deep learning in categorizing and identifying pneumonia within the framework of COVID-19. It is critical to recognize the limited scope of this research, which focuses on a select few DL models for pneumonia detection.
Padash S et al. [9] illustrated various publicly accessible datasets on pneumonia and presented an overview of DL models suitable for these datasets. However, this manuscript has solely assessed a limited quantity of research on pneumonia detection algorithms, lacking a comparative evaluation of their efficacy or an exploration of existing gaps in the literature.Lakhani and Sundaram [10] introduced a CNN-centered method for the automated categorization of pulmonary tuberculosis from chest X-rays, illustrating encouraging outcomes in disease identification, while prior research focused on specific deep learning structures customized for identifying pneumonia.

Despite the valuable insights offered by current literature on the use of deep learning in pneumonia detection, there is a need for comparative research to assess the efficacy of various CNN structures on standardized datasets, a gap that we aim to fill in this study.

METHODOLOGY

Dataset:

We employed a dataset sourced from Kaggle [11], comprising 5,863 chest X-ray images that were categorized into two distinct classes: pneumonia and normal. The dataset underwent segmentation into training, validation, and test sets, ensuring equitable representation of both classes within each subset. We sourced the anterior-posterior chest X-ray images for this study from historical cohorts of pediatric patients aged one to five years at Guangzhou Women and Children's Medical Center in Guangzhou. We integrated all chest X-ray imaging protocols into the standard clinical management we offered to the patients.

Model Architectures:

We utilized three widely recognized convolutional neural network (CNN) structures for the purpose of detecting pneumonia: VGG16, Inception V3, and ResNet 50V2. These architectures had been pre-trained using the ImageNet dataset before being fine-tuned on our specific pneumonia dataset to tailor them for the binary classification task.

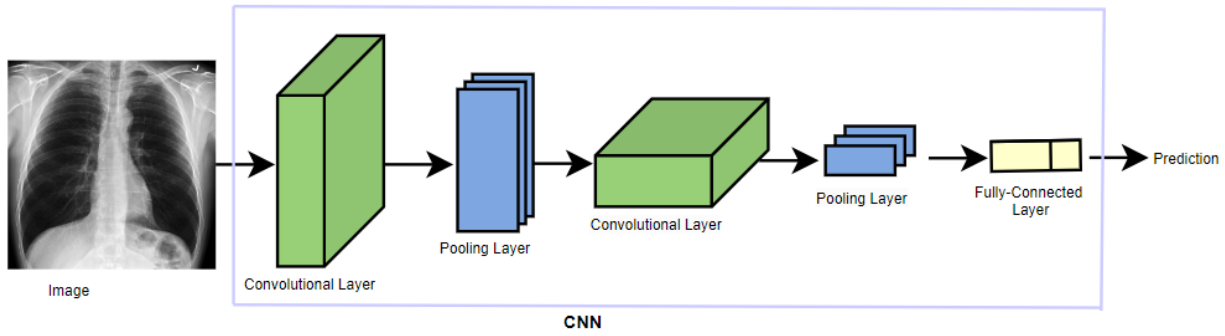

Figure 1: CNN Architecture

Training and Evaluation:

The optimization of the models was carried out through the utilization of the Adam optimizer, alongside a learning rate scheduler and an early stopping mechanism based on validation loss. We evaluated the models' efficacy using traditional performance metrics such as accuracy, precision, recall, and the area under the ROC curve (AUC). We delineate the evaluation criteria for the prediction of the proposed models as follows:

The formulation of Accuracy calculation is defined as (TP + TN) / (TP + TN + FP + FN).

Recall is computed as TP / (TP + FN).
Precision is determined as TP / (TP + FP).

The metric of accuracy serves as an indicator of the algorithm's ability to make correct predictions. However, in skewed data distribution scenarios, a high accuracy rate may not guarantee effective discrimination between distinct categories. In the healthcare domain, we can universally apply image classification techniques across all classes in the dataset. The concept of "precision" denotes the proportion of accurately predicted positive labels; the depiction of the accuracy rate in predictions compared to the overall number of predictions generated by the model is known as "precision." Conversely, "recall" refers to the proportion of actual positives correctly recognized by the model. The calculation of the Area Under the ROC Curve (AUC) offers a thorough overview of the performance of a binary classification model, considering both its sensitivity (true positive rate) and specificity (1-ffalse positive rate). A higher AUC value signifies enhanced

differentiation between positive and negative instances, rendering it a crucial metric for assessing model efficacy in healthcare and various other domains.

RESULTS

The experimental results demonstrated varying performance among the three CNN architectures for pneumonia detection. The ResNet 50V2 model exhibited the highest recognition accuracy of 95.97%, followed by VGG16 (93.58%) and Inception V3 (93.38%). Furthermore, the confusion matrix analysis demonstrated high recall (97.69%) and precision (86.79%) for the ResNet 50V2 model, underscoring their robustness in accurately identifying true positive cases and minimizing the occurrence of false positives.

The ROC and Precision-Recall curves showed that the ResNet 50V2 model was better at telling the difference between things than VGG16 and Inception V3. It also had higher AUC values. These findings highlight the effectiveness of the ResNet 50V2 architecture in pneumonia detection from chest x-ray images.

The VGG16 model's training and validation loss decreased significantly during the initial epochs and stabilized towards the end, indicating effective learning and minimal overfitting. The second graph demonstrates that the training and validation accuracy increases quickly and plateaus at high values, with stable performance and good generalization throughout the training process shown in Figure 2. Overall, the graphs indicate that the VGG16 model is well-trained, with high accuracy and stable performance.

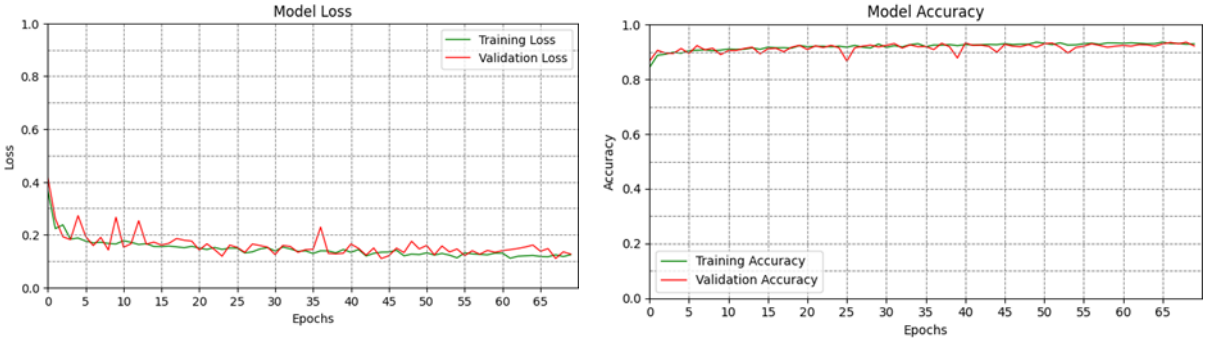

Figure 2: The graph of the VGG16 model shows the accuracy and the loss.

The first graph shows that the ResNet 50V2 model's training and validation losses drop by a lot and then stay low. This means that it is learning well with little overfitting. The second graph demonstrates that both training and validation accuracy increase quickly and plateau at high values, reflecting robust and consistent performance that is shown in figure 3.

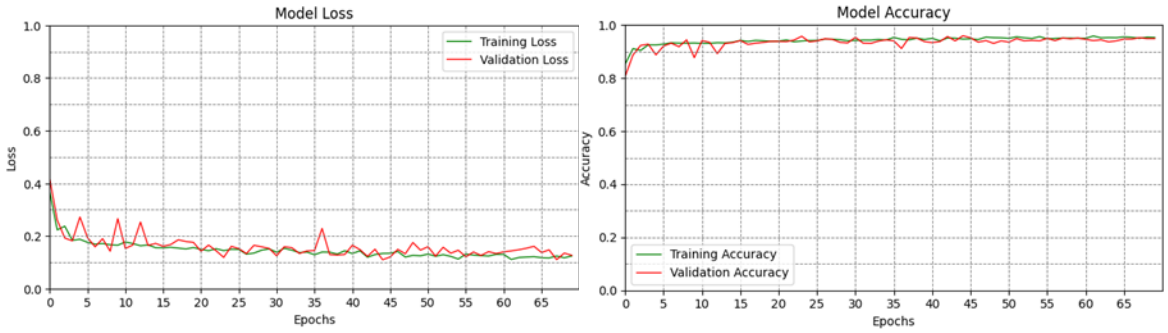

Figure 3: The graph of the ResNet 50V2 model shows the accuracy and the loss.

The Inception V3 model's training and validation loss goes down and stays low with only small changes, as shown in the first graph of Figure 4. This means that the model is learning well. The second graph demonstrates that both training and validation accuracy remain high and stable throughout the epochs, reflecting consistent performance and excellent generalization.

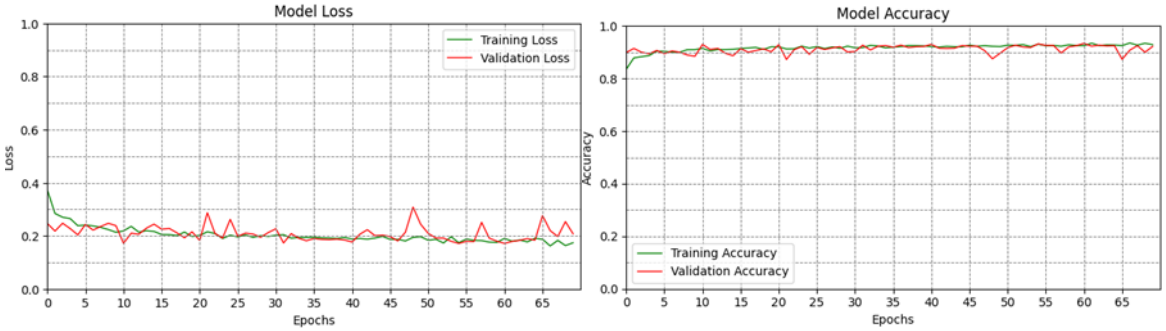

Figure 4: The graph of the Inception V3 model shows the accuracy and the loss.

The confusion matrix for the ResNet 50V2 model shows a high true positive rate (61.06%) and a low false negative rate (1.44%), with recall at 97.69% and precision at 86.79%. The true negative rate is 28.21%, and the false positive rate is 9.29%, as shown in figure 5. The AUC of 97.37% indicates the model's excellent discriminative ability in distinguishing between classes.

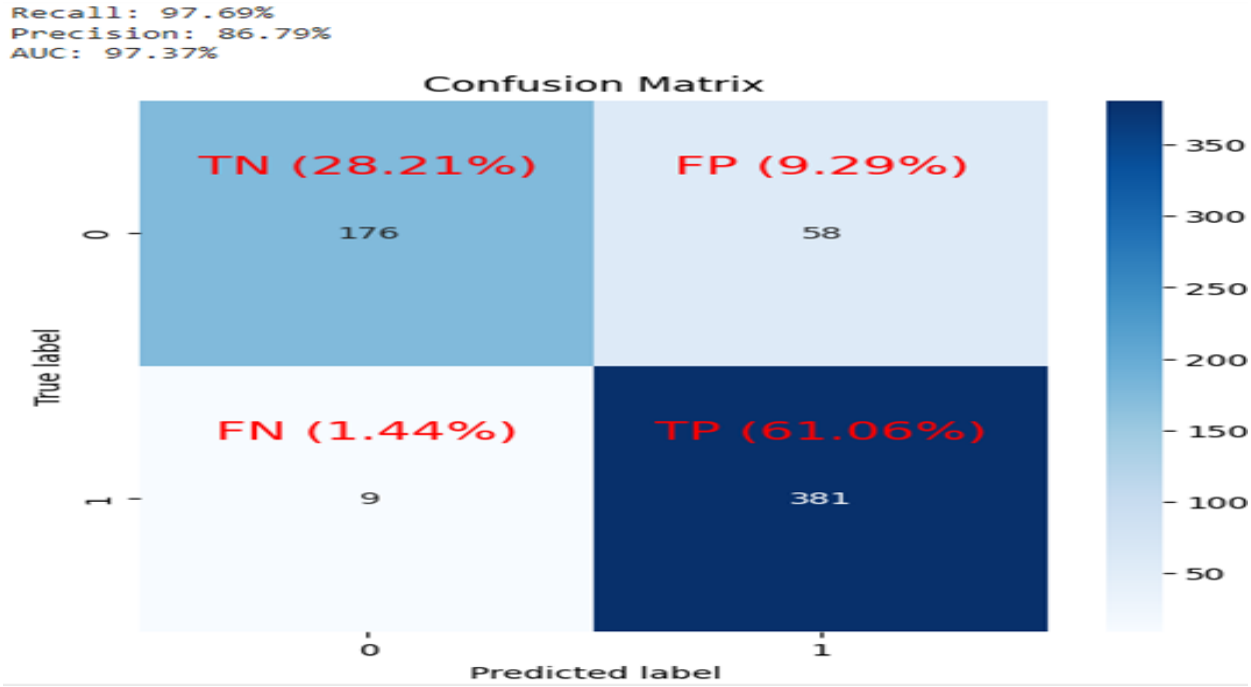

Figure 5: Confusion matrix of ResNet 50V2 model confusion matrix

The images depict the results of chest X-ray tests using the ResNet 50V2 model. The top row displays correctly classified normal cases with confidence levels ranging from 69.60% to 94.94%, as shown in figure 6. The bottom row shows correctly classified pneumonia cases with high confidence levels ranging from 97.65% to 99.77%.

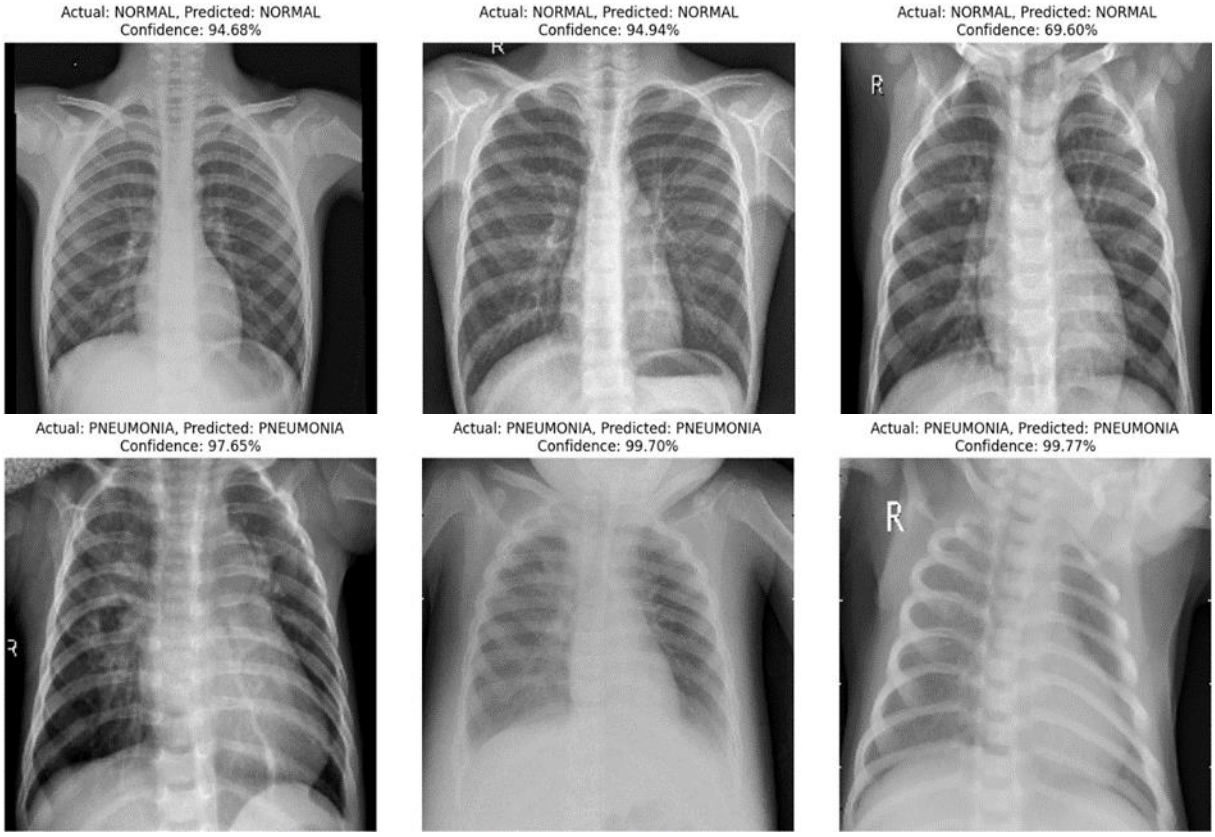

Figure 6: Images after testing ResNet50V2

We are comparing the loss and accuracy of 3 models in Figure 7. The first graph shows that ResNet 50V2 has the lowest and most stable validation loss compared to VGG16 and Inception V3, indicating superior performance. The second graph demonstrates that ResNet 50V2 achieves the highest validation accuracy, with VGG16 slightly lower and Inception V3 showing more fluctuations. Overall, ResNet 50V2 outperforms the other models in both validation loss and accuracy.

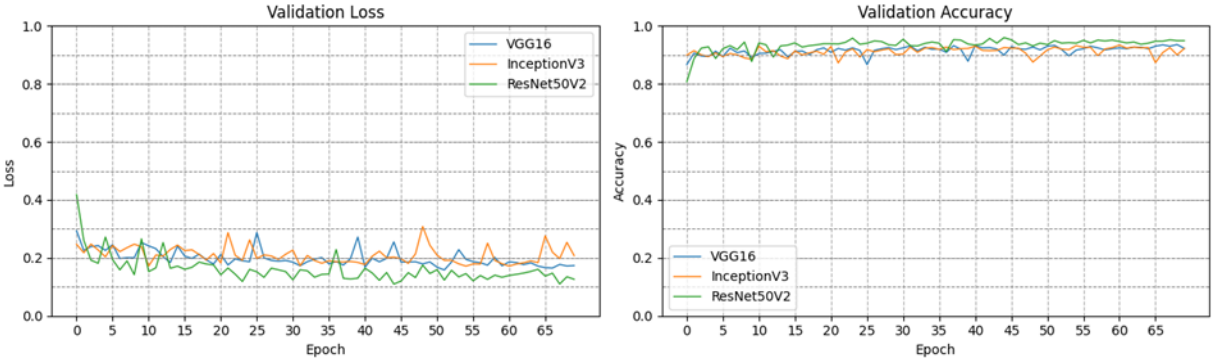

Figure 7: Chart comparing Loss and Accuracy of 3 models

The ROC curve shows that ResNet 50V2 has the highest AUC (0.97), indicating a superior true positive rate compared to VGG16 and Inception V3, both with an AUC of 0.93. The precision-

recall curve also highlights ResNet 50V2's superior performance, with the highest AUC (0.98) compared to VGG16 (0.95) and Inception V3 (0.96) that we show in figure 8. Overall, ResNet 50V2 demonstrates the best discriminative power and precision among the three models.

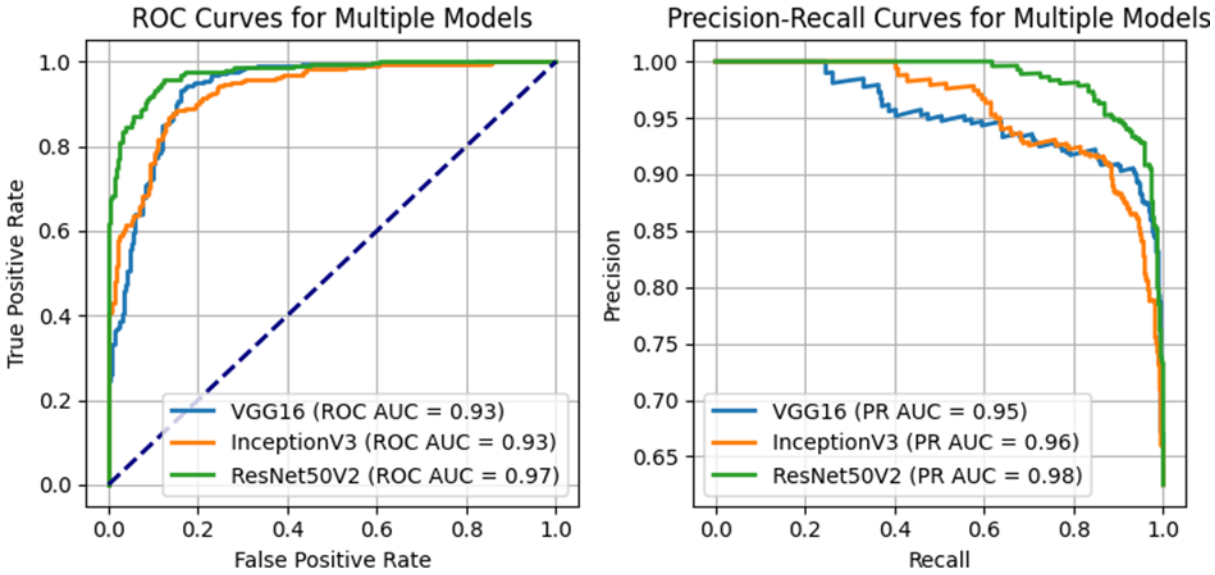

Figure 8: Compare ROC and Precision-Recall graphs on 3 models

DISCUSSION

Our study underscores the importance of selecting appropriate CNN architectures for pneumonia detection tasks. Although all three models demonstrated praiseworthy results, it was the ResNet 50V2 model that stood out as the highest achiever in terms of accuracy, recall, precision, and AUC. The deeper architecture and skip connections in ResNet 50V2 likely contributed to its superior performance by enabling better feature representation and extraction.

Comparing ROC curves and precision-recall curves also taught us a lot about the trade-offs between sensitivity and specificity, which helped doctors choose the best model for their diagnostic needs. These curves highlighted ResNet 50V2's ability to maintain high sensitivity and specificity, making it a reliable tool for medical diagnoses where accuracy is critical.

The confusion matrix for ResNet 50V2 further validates its robustness, showing a high true positive rate and a low false negative rate, which is essential for reducing missed pneumonia cases. This aspect is critical in clinical settings, where the cost of false negatives can be very high.

In the future, researchers could look into ensemble learning methods and model interpretability techniques to make pneumonia detection systems even more reliable and clear. Utilizing these techniques could potentially enhance the precision of diagnoses and offer a more lucid understanding of the decision-making mechanisms employed by the models, thereby increasing trust and adoption in clinical settings. Furthermore, expanding this study to include broader and

more diverse datasets could help to authenticate the universality of these results across various populations and imaging circumstances.

Another possible direction for future research could be to incorporate multimodal data, such as the amalgamation of chest X-ray images with patient history and clinical symptoms, to improve the accuracy of diagnostic procedures. The application of transfer learning from other medical imaging domains could also be investigated to leverage pre-trained models for enhanced feature extraction. Ultimately, the implementation of these models in real-world clinical settings necessitates continuous monitoring and evaluation to uphold their effectiveness and reliability over time. This approach will assist in maintaining high patient care standards and adapting to any changes in the clinical workflow or patient population.

CONCLUSION

Our comparison study shows that deep learning models, especially the ResNet 50V2 architecture, are good at finding pneumonia on chest X-rays. The findings underscore the potential of AI-driven diagnostic systems to augment clinical decision-making and improve patient outcomes in the management of pneumonia. As AI continues to evolve, leveraging advanced computational techniques for medical image analysis holds promise for enhancing healthcare delivery and population health management. Higher accuracy, recall, precision, and AUC values for ResNet 50V2 show that it works better than other networks, which proves that it can be used in clinical settings. These results suggest that integrating AI models into diagnostic workflows can significantly enhance the early detection and treatment of pneumonia, ultimately reducing morbidity and mortality rates associated with this condition.

FUTURE WORK

Future investigations may focus on various essential areas to expand upon the findings of this investigation. Firstly, delving into ensemble learning methodologies could potentially enhance the efficacy of models by combining the advantages of numerous architectures. Secondly, investigating model interpretability techniques will be crucial to ensuring that AI-driven diagnostics are transparent and trustworthy, facilitating greater acceptance among healthcare professionals. Additionally, broadening the scope of this study to encompass larger and more varied datasets will contribute to confirming the applicability of the models across diverse populations and imaging scenarios. Furthermore, incorporating multimodal data, such as fusing chest X-ray images with patient medical history and clinical manifestations, could improve diagnostic precision and provide a more comprehensive assessment of patient well-being. Additionally, by utilizing transfer learning from other medical imaging fields, we can enhance the performance of pre-trained models by improving their feature extraction skills, thereby requiring less training data. Finally, to ensure the long-term efficacy and dependability of these models, ongoing surveillance and assessment must accompany their implementation in actual clinical settings. This approach will help maintain high standards of patient care and adapt to any changes

in clinical workflows or patient populations, ensuring that AI-driven diagnostic systems remain a valuable tool in healthcare.

References

[1] P. Rajpurkar *et al.*, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," Nov. 2017, [Online]. Available: http://arxiv.org/abs/1711.05225

[2] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017. doi: 10.1109/CVPR.2017.369.

[3] Y. Li, Z. Zhang, C. Dai, Q. Dong, and S. Badrigilan, "Accuracy of deep learning for automated detection of pneumonia using chest X-Ray images: A systematic review and meta-analysis," *Computers in Biology and Medicine*, vol. 123. 2020. doi: 10.1016/j.compbiomed.2020.103898.

[4] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, 2017, doi: 10.1148/radiol.2017162326.

[5] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA - Journal of the American Medical Association*, vol. 316, no. 22, 2016, doi: 10.1001/jama.2016.17216.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015. doi: 10.1007/978-3-319-24574-4_28.

[7] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016. doi: 10.1109/CVPR.2016.308.

[8] A. Shah and M. Shah, "Advancement of deep learning in pneumonia/Covid-19 classification and localization: A systematic review with qualitative and quantitative analysis," *Chronic Diseases and Translational Medicine*, vol. 8, no. 3. 2022. doi: 10.1002/cdt3.17.

[9] S. Padash, M. R. Mohebbian, S. J. Adams, R. D. E. Henderson, and P. Babyn, "Pediatric chest radiograph interpretation: how far has artificial intelligence come? A systematic literature review," *Pediatric Radiology*, vol. 52, no. 8. 2022. doi: 10.1007/s00247-022-05368-w.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[11] ChestX-Ray images (Kaggle). https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia