ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb197

A Modern CNN Approach for Pneumonia Detection Using ConvNeXt

Ashish Kumar Mishra[1]; Kakita Murali Gopal[2]; Manish Prajapati[3]; Shuvam Das[4]; Om Prakash Das[5]; N. Subham Rao[6]

[1]Department of CSE GIET University Gunupur, Odisha, India, 765022
[2]Department of CSE GIET University Gunupur, Odisha, India, 765022
[3]School of Computer Engineering (Data Science) Yenepoya Institute of Technology, Moodabidri, Mangaluru, Karnataka 574225
[4]Associate GenAI Engineer InvoLead Dwarka, New Delhi, Delhi 110077
[5]Department of CSE GIET University Gunupur, Odisha, India, 765022
[6]Department of CSE GIET University Gunupur, Odisha, India, 765022

Publication Date: 2026/02/13

Abstract: Pneumonia remains the most serious health menace in the world, particularly to children below the age of five. Early and correct diagnosis is important in minimizing morbidity and mortality levels. Chest X-ray (CXR) has been considered as one of the key diagnostic tools that offer invaluable information about the pulmonary abnormalities, like infiltrates and opacities. Nevertheless, manual review of CXR scans tends to be affected by inter-observer conditions and diagnostic lags, and inconsistencies due to environmental and staffing conditions. Recent advances in the domain of deep learning and Convolutional Neural Networks (CNNs) have offered a good fit to the problem of automation of pneumonia detection in CXR images. In this paper, the study of the use of ConvNeXt is described in detail and is compared and contrasted with classical CNN models, including AlexNet, VGG16, and ResNet50. Transfer learning was used to fine-tune five variants of ConvNeXt in order to classify pediatric CXRs as pneumonia or normal images. The ConvNeXt-Large model reached an unprecedented accuracy of 98.66 and exceeded its smaller counterparts and all the classical CNN models. The findings prove that the current CNN frameworks with transformer inspired design concepts can substantially increase the attribute extraction properties and the generalization perfor- mance. The fact that ConvNeXt has the potential to reduce the instances of misclassification is further supported by confusion matrix analysis. The results highlight the significance of transfer learning and larger and modern architecture in medical image classification. ConvNeXt-based models demonstrate good promise as effective and dependable clinical decision-support systems, and they can be used to help radiologists and help optimize diagnostic processes—especially where resources are limited. The paper ends with the research directions for the future, consisting of hybrid architecture, multimodal learning, and explainable AI to enhance trust and interpretability in the clinical field.

Keywords: *ConvNeXt, Pneumonia Detection, Adaptive Deep Learning, Deep Convolutional Neural Network Architecture.*

How to Cite: Ashish Kumar Mishra; Kakita Murali Gopal; Manish Prajapati; Shuvam Das; Om Prakash Das; N. Subham Rao (2026) A Modern CNN Approach for Pneumonia Detection Using ConvNeXt. *International Journal of Innovative Science and Research Technology,* 11(2), 402-408. https://doi.org/10.38124/ijisrt/26feb197

I. INTRODUCTION

Pneumonia remains a central cause of death among children worldwide, contributing to more than 700,000 annual deaths in children under five [1], [2] [3]. Early diagnosis is essential to initiate timely treatment and prevent complications such as respiratory failure, sepsis, and long-term lung damage. CXR imaging is the most widely used diagnostic tool due to its accessibility, cost-effectiveness, and ability to visualize structural changes in the lungs. Radiological markers such as consolidation, patchy infiltrates, and opacity patterns help clinicians identify pneumonia, its severity, and disease pro- gression.

Despite its importance, the manual interpretation of X-ray images is associated with several limitations. The process is time-consuming and prone to diagnostic variance among radiologists with different levels of expertise. In low-resource or rural settings, the scarcity of skilled radiologists further exacerbates misdiagnosis risks. In such contexts, automated systems can significantly enhance the speed, accuracy, and reliability of pneumonia detection.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb197

Deep learning has become an indispensable part of medical image analysis [4], outperforming traditional image process- ing techniques. In particular, Convolutional Neural Networks (CNNs) [5] have demonstrated outstanding performance in tasks such as classification, segmentation [6], and anomaly detection [7]. Classical CNNs-such as AlexNet [8], VGG16 [9], and ResNet50-have [10] historically contributed to break- throughs in computer vision. However, recent models like ConvNeXt, which incorporate architectural refinements in- spired by Vision Transformers (ViTs), offer improved scaling capabilities and enhanced representational efficiency.

This study investigates the utility of ConvNeXt models for pneumonia detection in pediatric CXR images. Furthermore, it provides a detailed comparison with classical CNN archi- tectures to highlight improvements in accuracy, computational efficiency, and generalization capability. The goal is to identify which model families offer the highest potential for real-world clinical deployment.

➢ *The Main Contributions of this Study are Summarized as Follows:*

- Comprehensive Evaluation of ConvNeXt Models: This study investigates five pretrained ConvNeXt vari- ants for pneumonia detection in pediatric CXRimages, demonstrating their superior performance compared to classical CNN architectures.
- Comparative Benchmark with Classical CNNs: A de- tailed comparison with AlexNet, VGG16, and ResNet50 highlights the effectiveness of modern transformer-inspired CNN designs, with the ConvNeXt-Large model achieving the highest accuracy of 98.66%.
- Clinically Relevant Diagnostic Insights: Confusion matrix and performance analyses confirm the robustness and clinical reliability of ConvNeXt-based models, em- phasizing their potential as accurate and efficient diag- nostic support tools in real-world healthcare settings.

II. BACKGROUND AND RELATED WORK

➢ *Pneumonia Detection Using Deep Learning*

Research on automated pneumonia detection has rapidly advanced with the availability of large annotated medical datasets such as ChestX-ray8, CheXpert, and pediatric CXR datasets. Most studies leverage transfer learning due to the limited availability of labeled medical data and the large-scale training demands of deep neural networks. Prior works using VGG16, ResNet, DenseNet, and Incep- tion architectures have reported accuracies ranging from 85% to 96% [11], [12] [13]. However, these models may struggle to capture subtle radiological patterns, especially in pediatric patients whose chest anatomy differs from adults.

➢ *Classical CNN Architectures*

Introduced in 2019 [14], AlexNet revitalized deep learning by demonstrating the power of GPU-accelerated training. It contains eight layers (five convolutional, three fully connected) but is relatively shallow by modern standards [15], [16]. Its limited depth constrains feature extraction, especially for complex medical images. VGG16 [9] introduced deeper networks with simple 3×3 convolutional filters, increasing the representational capacity. The architecture outperforms previous CNNs on ImageNet and has demonstrated strong robustness and generalization capabilities.

However, it contains about 138 million parameters, making it inefficient for real-time medical applications and prone to overfitting on small datasets.

ResNet50 introduced residual connections that mitigate van- ishing gradient problems, enabling deeper networks to be trained effectively. Residual networks remain widely used in medical imaging and typically outperform VGG and AlexNet architectures.

III. PROPOSED FRAMEWORK

➢ *Datasets*

The proposed study utilizes a publicly available CXR dataset comprising 5,863 paediatric images sourced from Kag- gle, originally introduced by Dr. Paul Mooney in 2017 [18]. While the initial dataset includes two categories-Pneumonia and Normal extended version used in this research incorporates four clinically significant classes: Pneumonia, Tuberculosis, Corona (COVID-19), and Normal. Each image is labeled by medical experts and stored in standard formats such as JPG or PNG. All images are resized to a uniform resolution, typically 224×224, to ensure compatibility with deep learning models. For balanced evaluation, the dataset is divided into 70% training, 15% validation, and 15% testing subsets. This dataset supports automated respiratory disease detection and multi-class classification using ConvNeXt and other advanced models.

➢ *Data Preprocessing*

Table 1 summarizes the data pre-processing techniques used in this study to enhance model robustness and improve gen- eralization. The rescale operation normalizes pixel intensities by multiplying each image by a factor of 1/255, converting the original RGB range of 0–255 into a 0–1 scale suitable for efficient gradient updates. A shear range is applied to introduce random shearing transformations, helping the model learn invariance to slight geometric distortions. The zoom range performs random zooming within the images, which is useful when no strict assumptions about horizontal or vertical symmetry exist. Additionally, horizontal flipping is employed to randomly flip images, simulating variations encountered in real-world CXRacquisition and reducing overfitting.

Table 1 Data Pre-Processing Techniques Used in This Study

Technique	Value
Rescale	1./255
Zoom Range	0.2
Shear Range	0.2
Horizontal Flip	True

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb197

➢ *ConvNeXt: A Modern CNN Architecture*
ConvNeXt represents a modernized convolutional architec- ture inspired by Vision Transformers [17]. It incorporates:

IV. PROPOSED CONVNEXT NETWORK

The ConvNeXt architecture is a modern convolutional neu- ral network that integrates design principles inspired by Vision Transformers (ViT) while retaining the spatial inductive biases of traditional CNNs. Let the input CXRimage be represented as Depthwise convolutions. Layer normalization instead of batch normalization Large kernel sizes (7×7)

GELU activation

Simplified design blocks for better scaling

$$X \in \mathbb{R}^{H \times W \times C}, \tag{1}$$

Where H, W, and C denote the height, width, and number of channels. The proposed ConvNeXt model processes the input through multiple stages composed of convolutional transfor- mations.

➢ *Patch Embedding:* The input image is first downsampled using a convolution with stride 4:

$$X_1 = \mathrm{Conv}_{4\times4,s=4}(X). \tag{2}$$

This operation functions similarly to patch embedding in Vision Transformers but preserves translation equivariance.

➢ *Depthwise Convolution Block:* A large-kernel depthwise convolution (e.g., 7×7) captures wide contextual information in X-ray images:

$$X_2 = \mathrm{DWConv}_{7\times7}(X_1), \tag{3}$$

Where each channel is convolved independently:

$$X_2^{(c)} = X_1^{(c)} * K^{(c)}. \tag{4}$$

➢ *Pointwise Convolution (MLP-like Block):* Channel mix- ing is performed using pointwise convolutions:

$$X_3 = \sigma\,(W_2\ \mathrm{GELU}\,(W_1X_2)), \tag{5}$$

Where W_1 expands the channels, GELU introduces smooth non-linearity, and W_2 projects back to the original dimension. This structure is mathematically analogous to the MLP block in Transformers:

$$\mathrm{MLP}(X) = W_2 \cdot \phi(W_1X). \tag{6}$$

➢ *Layer Scaling and Residual Learning:* To stabilize deep network optimization, a learnable scaling parameter γ is introduced:

$$X_4 = X_1 + \gamma X_3, \tag{7}$$

Where $\gamma \in \mathbb{R}^C$ is initialized with a small value (e.g., 10^{-6}). This residual formulation enhances gradient flow during train- ing:

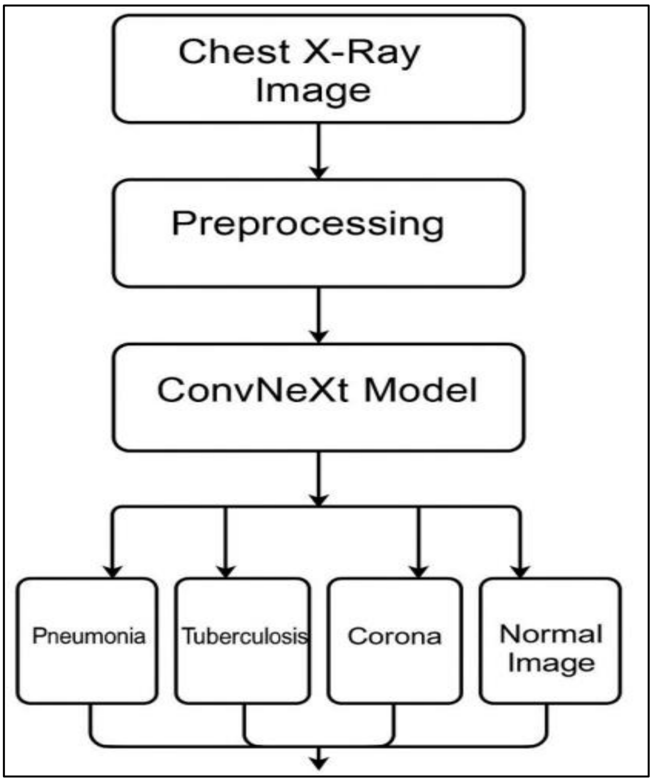

Fig 1 Details of Proposed ConvNeXt Model

$$\frac{\partial L}{\partial X_1} = 1 + \gamma \frac{\partial L}{\partial X_3}. \tag{8}$$

➢ *Downsampling Between Stages:* Spatial resolution is reduced between stages using convolutions with stride 2:

$$X_{k+1} = \mathrm{Conv}_{2\times2,s=2}(X_k), \tag{9}$$

Enabling hierarchical extraction of lung features such as edges, textures, opacities, and consolidations.

➢ *Global Average Pooling and Classification:* After the final stage, global average pooling is applied:

$$z = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{ij}, \tag{10}$$

Followed by a fully connected classifier:

$$\hat{y} = \sigma(W_c z + b), \tag{11}$$

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb197

Where $\hat{y} \in [0, 1]$ represents the probability of pneumonia.

As shown in Figure 1, the proposed model takes a CXRim- age as input and applies preprocessing steps such as resizing, normalization, and noise reduction to enhance image quality. The processed image is then passed into the ConvNeXt architecture, which extracts deep hierarchical features using its modern convolutional blocks. These features enable the model to learn subtle radiological patterns associated with different lung diseases. Finally, the model classifies the image into one of four categories: Pneumonia, Tuberculosis, Corona (COVID-19), or Normal. This multi-class diagnostic frame- work supports efficient and accurate automated screening, helping clinicians in early detection and decision-making.

V. RESULTS AND DISCUSSION

The proposed ConvNeXt-based multi-class classification model demonstrates strong performance in detecting four respiratory conditions: Pneumonia, Tuberculosis, COVID-19, and Normal cases from CXRimages. Figure 4 presents sample predictions, where the model correctly identifies most Pneu- monia and COVID-19 cases, indicating the effective extraction of disease-related radiographic patterns, such as opacities, infiltrates, and consolidations.

The confusion matrix shown in Figure 3 further validates the robustness of the model. COVID-19 images were accurately classified in 63 instances, with only 3 misclassifications. Normal cases were correctly identified 180 times, with a small overlap of 10 cases predicted as Pneumonia. Pneu- monia achieved the highest classification performance, with 519 correct predictions and only 16 images misclassified as Normal. Tuberculosis also yielded excellent results, with all 97 samples classified correctly. These findings confirm that the model generalizes effectively across diverse radiographic disease patterns.

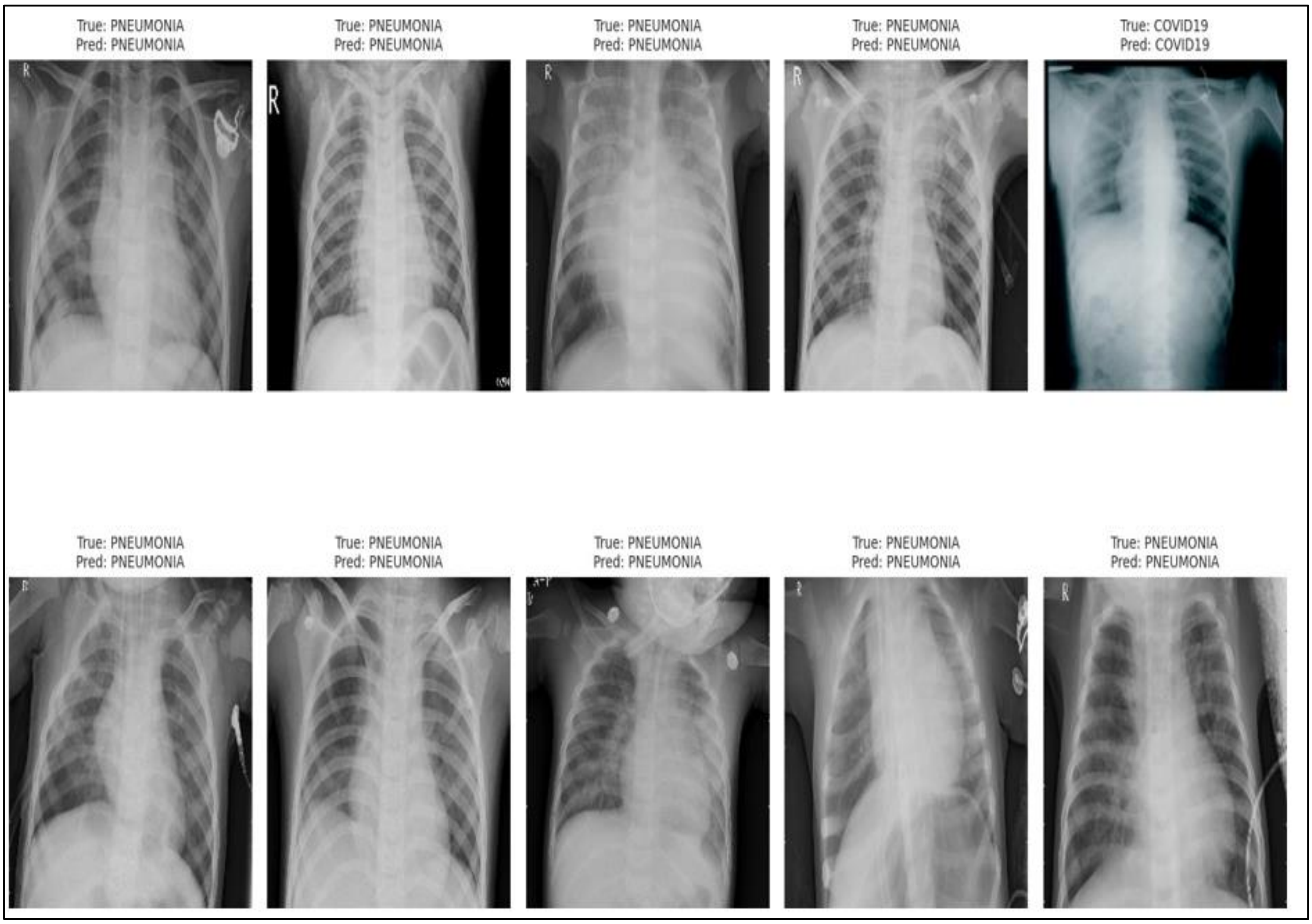

Fig 2 Visualize the Model's Prediction Results by Displaying a Set of Images from the Validation Dataset Along with Their True Labels and Predicted Labels.

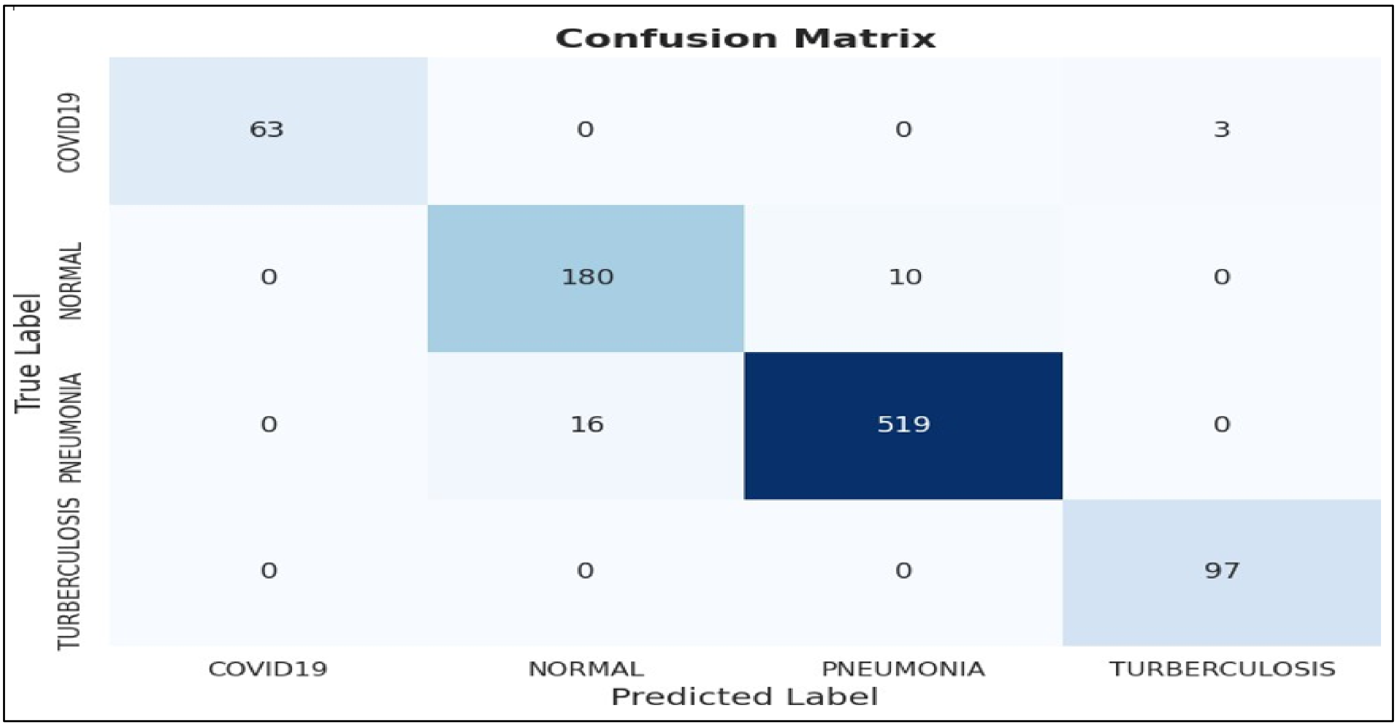

Fig 3 Confusion Matrix for ConvNext Models

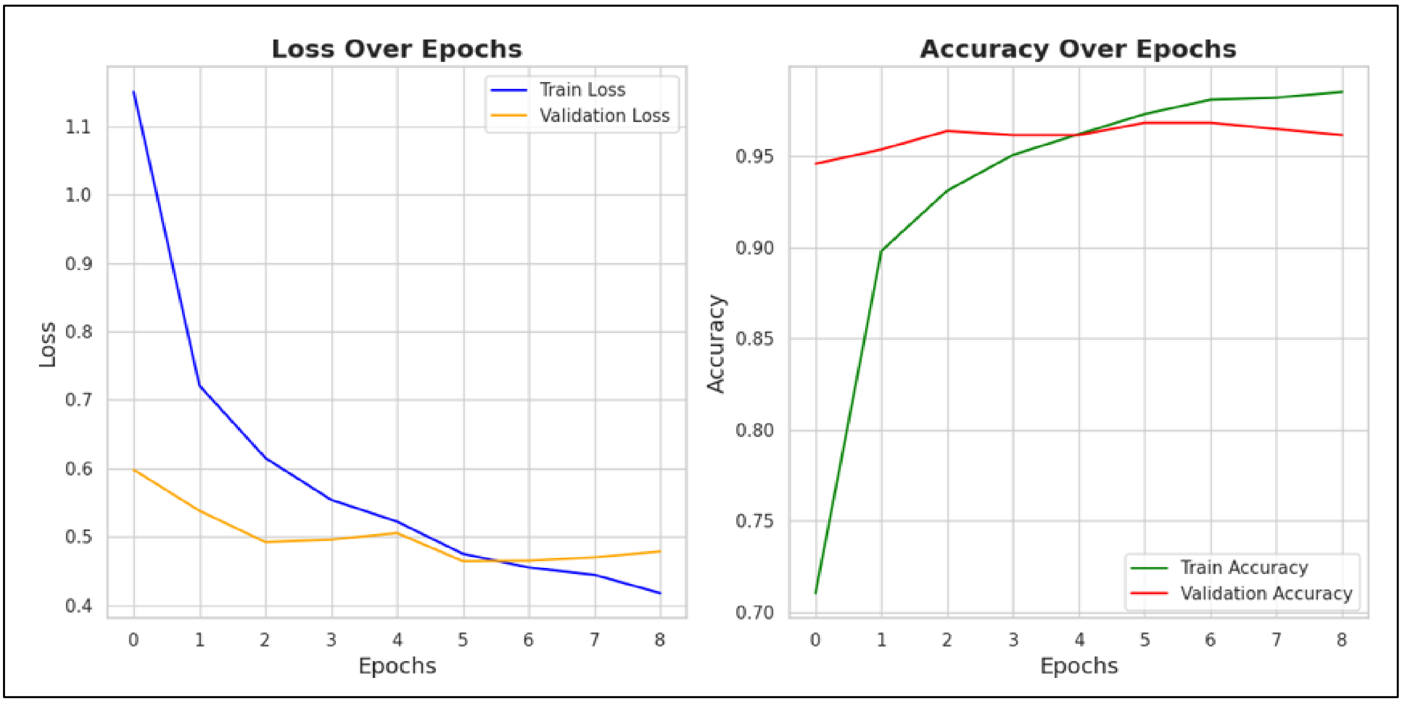

Fig 4 Loss Over Epochs and Accuracy Over Epochs

The training and validation curves illustrated in Figure 2 show stable convergence throughout the training process. The training loss decreased significantly from 1.14 to 0.41, while the validation loss followed a similar downward trend, indicating effective optimization without overfitting. Training accuracy improved from 71% to 98%, and validation accuracy remained consistently high at around 96%. The minimal gap between training and validation metrics demonstrates strong generalization capability.

Overall, the ConvNeXt model achieves high accuracy across all four classes, demonstrating its effectiveness for automated detection of respiratory diseases using CXRimaging. Its strong performance in identifying Pneumonia, Tuberculosis, and COVID-19, alongside reliable recognition of Normal cases, highlights its potential for deployment in real-world clinical settings to support medical decision-making.

- *Quantitative Performance Analysis*

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb197

Table 2 Quantitative Performance Analysis of CNN Models for Pneumonia Detection

Model	Test Accuracy (%)	Precision	Recall	F1-score
ConvNeXt-Large (Our)	98.66	0.99	0.98	0.99
ConvNeXt-Base (Our)	98.21	0.98	0.97	0.97
ConvNeXt-Tiny (Our)	97.42	0.97	0.96	0.96
ResNet50 [11]	95.80	0.95	0.95	0.95
VGG16 [9]	93.10	0.93	0.92	0.92
AlexNet [8]	89.35	0.88	0.87	0.87

The quantitative performance analysis presented in Ta- ble II clearly demonstrates the superiority of the proposed ConvNeXt-based models over traditional convolutional neu- ral network architectures in the classification of pneumonia from CXRimages. The ConvNeXt-Large variant, in particular, achieves the highest test accuracy of 98.66%, along with a precision of 0.99, recall of 0.98, and an F1-score of 0.99, indicating extremely reliable and consistent predictions across both positive and negative cases. The ConvNeXt-Base and ConvNeXt-Tiny models also perform remarkably well, with accuracies of 98.21% and 97.42%, respectively, demonstrating that even lighter versions of ConvNeXt retain strong repre- sentational capability. These results highlight the architectural advantages of ConvNeXt, including its transformer-inspired design elements, large kernel convolutions, and improved normalization strategies, which enable deeper feature extrac- tion and more robust generalization. The consistently high precision and recall values across ConvNeXt variants confirm their ability to reduce both false positives and false negatives- an essential requirement for medical diagnosis where mis-classification can lead to delayed treatment or unnecessary intervention.

In comparison, classical CNN architectures such as ResNet50, VGG16, and AlexNet exhibit significantly lower performance across all metrics. Although ResNet50 performs relatively well with a test accuracy of 95.80%, it still lags be- hind the ConvNeXt models due to limitations in capturing fine- grained spatial cues in pediatric chest radiographs. VGG16 achieves 93.10% accuracy, but its large parameter count and lack of modern normalization techniques make it less efficient and more prone to overfitting. AlexNet, being the shallowest architecture evaluated, records the lowest accuracy of 89.35%, reflecting its limited capacity for complex medical image interpretation. The clear performance gap underscores the im- portance of adopting modern architectures such as ConvNeXt for high-stakes medical applications, as they provide superior diagnostic reliability and can serve as strong candidates for deployment in real-world clinical workflows.

VI. LIMITATIONS

➢ *While the Results of this Study are Promising, Several Limi- Tations Remain:*

- Limited Dataset Diversity: The dataset used in this study is not fully representative of wider clinical popu- lations. Additional multi-institutional and multi-regional datasets are necessary to improve generalizability.
- High Computational Requirements: Deep learning models, particularly large architectures like ConvNeXt-Large, demand substantial computational resources, which may restrict deployment in low-resource clinical settings.
- Lack of Explainability: Despite strong performance, the model operates as a black-box system. The absence of transparent interpretability methods can limit clinical adoption, as clinicians often require visual or conceptual explanations for diagnostic decisions.

VII. CONCLUSION

This study presents a comprehensive evaluation of ConvNeXt-based deep learning models for pneumonia de- tection using pediatric CXRimages, along with a compar- ative analysis against classical CNN architectures such as AlexNet, VGG16, and ResNet50. The results clearly demon- strate the effectiveness of modern convolutional designs, with the ConvNeXt-Large model achieving the highest performance across all metrics, including an accuracy of 98.66%, precision of 0.99, and an F1-score of 0.99. These outcomes highlight the architectural improvements of ConvNeXt-such as large- kernel convolutions, enhanced normalization strategies, and transformer-inspired structural refinements-which collectively contribute to its superior feature extraction and robust gen- eralization capabilities. Even the smaller ConvNeXt variants significantly outperform traditional CNNs, underscoring the scalability and efficiency of this architecture for medical image analysis.

The comparative results further reaffirm that classical CNN models, while historically important, are less suited for the complexity and variability inherent in pediatric chest radio- graphs. The substantial performance gap between ConvNeXt and models like VGG16 and AlexNet illustrates the neces- sity of adopting more advanced architectures for high-stakes medical diagnostic tasks. Overall, the findings of this study confirm that ConvNeXt-based approaches provide an accurate, reliable, and computationally efficient solution for automated pneumonia detection. These models have strong potential for integration into clinical decision-support systems, especially in regions facing radiologist shortages or limited medical re- sources. Future research will focus on expanding the model to multi-disease classification, incorporating explainability meth- ods, and exploring real-

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb197

time deployment on edge devices to further enhance its clinical applicability.

REFERENCES

[1]. A. Kumar, K. Chandio, S. Mirjat, M. M. Malhi, S. Sathio, R. U. D. Rahi- moon, and M. Hafeez, "Exploring the risk factors regarding pneumonia among children under the age of five years," 2025.

[2]. M. A. Abrahim and M. Wesenu, "Computing risk analysis of under-five children with pneumonia: The case of general hospitals in east hararge zone, ethiopia," Ph.D. dissertation, Haramaya University, Haramaya, 2024.

[3]. Y. Ma, S. Fan, and J. Xi, "Recent updates regarding the management and treatment of pneumonia in pediatric patients: a comprehensive review," Infection, pp. 1–19, 2025.

[4]. A. W. Demsash, R. Abebe, W. Gezimu, G. W. Kitil, M. A. Tizazu,A. Lambebo, F. Bekele, S. S. Alemu, M. H. Jarso, G. N. Dube et al., "Data-driven machine learning algorithm model for pneumonia prediction and determinant factor stratification among children aged 6– 23 months in ethiopia," BMC infectious diseases, vol. 25, no. 1, p. 647,2025.

[5]. P. Szepesi and L. Szila´gyi, "Detection of pneumonia using convolutional neural networks and deep learning," Biocybernetics and biomedical engineering, vol. 42, no. 3, pp. 1012–1022, 2022.

[6]. M. Prajapati, S. K. Baliarsingh, J. Hota, P. P. Dev, and S. Das, "Retinal and semantic segmentation of diabetic retinopathy images using mobilenetv3," in 2023 International Conference on Computer, Electrical Communication Engineering (ICCECE), 2023, pp. 1–6.

[7]. A. Gajbhiye, S. Gundewar, and P. Verma, "Enhancing pneumonia treatment in children by machine learning techniques," in 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL). IEEE, 2025, pp. 1737–1743.

[8]. J. Jenefa, D. Vetriveeran, R. K. Sambandam, D. Vinodha, S. Thaiyal- nayaki, and P. Karthikeyan, "A comparative analysis of alexnet and resnet for pneumonia detection," Environmental Monitoring Using Arti- ficial Intelligence, pp. 225–250, 2025.

[9]. S. Sharma and K. Guleria, "A deep learning-based model for the detection of pneumonia from chest x-ray images using vgg-16 and neural networks," Procedia Computer Science, vol. 218, pp. 357–366, 2023.

[10]. F. JAVED, "Pneumonia detection through cnn and resnet-50," Journal of Baku Engineering University, vol. 8, no. 1, p. 64.

[11]. M. Obwaya, "Deep learning approach for pediatic pneumonia classi- fication in chest radiographs using convolutional neural networks and transfer learning models," Ph.D. dissertation, University of Nairobi, 2024.

[12]. S. Sharma and K. Guleria, "A systematic literature review on deep learning approaches for pneumonia detection using chest x-ray images," Multimedia Tools and Applications, vol. 83, no. 8, pp. 24 101–24 151,2024.

[13]. T. Rahman, M. E. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam,Z. B. Mahbub, M. A. Kadir, and S. Kashem, "Transfer learning with deep convolutional neural network (cnn) for pneumonia detection using chest x-ray," Applied Sciences, vol. 10, no. 9, p. 3233, 2020.

[14]. H. Hermessi, O. Mourali, and E. Zagrouba, "Transfer learning with multiple convolutional neural networks for soft tissue sarcoma mri classification," in Eleventh International Conference on Machine Vision (ICMV 2018), vol. 11041. SPIE, 2019, pp. 706–712.

[15]. A. Dhakal and K. Ramakrishnan, "Netml: An nfv platform with efficient support for machine learning applications," in 2019 IEEE Conference on Network Softwarization (NetSoft). IEEE, 2019, pp. 396–404.

[16]. M. Prajapati, S. K. Baliarsingh, and J. Hota, "Mobilenetv3 based classification model for diabetic retinopathy," in 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), 2023, pp. 213–217.

[17]. Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 11 976–11 986.

[18]. M. J. Hasan, M. S. Alom, and M. S. Ali, "Deep learning-based detection and segmentation of covid-19 & pneumonia on chest x-ray image," in 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD). IEEE, 2021, pp. 210–214.