IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

AI Based Pneumonia Detection Using Deep Learning

A.Srinivasa Rao[1], P. Jyothi[2], P.Mounika[3], P.Thirumala Naidu[4], R. Pushkarini[5]
[1]Assistant Professor, Department of Computer Science and Engineering,
[2,3,4,5]UG Student, Department of Computer Science and Engineering,
Vasireddy Venkatadri Institute of Technology (VVIT), Andhra Pradesh, India
sri.srinivas.07@gmail.com, 22bq1a05g5@vvit.net,
22bq1a05g3@vvit.net, 23bq5a0518@vvit.net,22bq1a05i4@vvit.net

Abstract: *Pneumonia remains a leading cause of mortality among children under five years of age, accounting for approximately 740,000 deaths annually worldwide. Early and accurate diagnosis is essential for effective treatment; however, in low- and middle-income countries, diagnostic capacity is limited by shortages of trained radiologists and variability in chest X-ray interpretation. Although chest radiography is the primary diagnostic modality, subjectivity and resource constraints reduce its reliability in real-world clinical settings.*

This study presents a lightweight deep learning–based system for automated pediatric pneumonia detection from chest X-ray images, with a strong emphasis on real-world generalizability. A MobileNetV2-based transfer learning model was trained on 1,750 balanced radiographs and evaluated using both internal validation and rigorous cross-operator validation involving different radiologists and imaging equipment. The model achieved 94.8% accuracy on internal validation and maintained high sensitivity (96.4%) with 86.0% accuracy under cross-operator conditions. The compact 14 MB architecture enables rapid inference on mobile devices, demonstrating feasibility for deployment in resource-limited pediatric clinics.

Keywords: pneumonia detection; deep learning; MobileNetV2; cross-operator validation; clinical deployment; resource-limited setting

I. INTRODUCTION

Pneumonia remains one of the leading causes of illness and death among children under five years of age, particularly in low- and middle-income countries (LMICs), where access to timely diagnosis is limited.and specialist care is limited. Despite being largely preventable and treatable, pneumonia continues to claim hundreds of thousands of young lives annually due to delayed detection, shortage of trained radiologists, and variability in diagnostic practices. Early and accurate diagnosis is therefore critical to reducing disease severity, complications, and mortality.

Chest X-ray imaging is the most widely used diagnostic tool for pediatric pneumonia; however, its interpretation is inherently subjective and highly dependent on radiologist expertise. Inter-observer variability, subtle radiographic patterns in children, and limited specialist availability in resourceconstrained settings significantly reduce diagnostic reliability. These challenges highlight the urgent need for robust, scalable, and assistive diagnostic solutions that can support frontline healthcare providers.

DOI: 10.48175/IJARSCT-32405

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

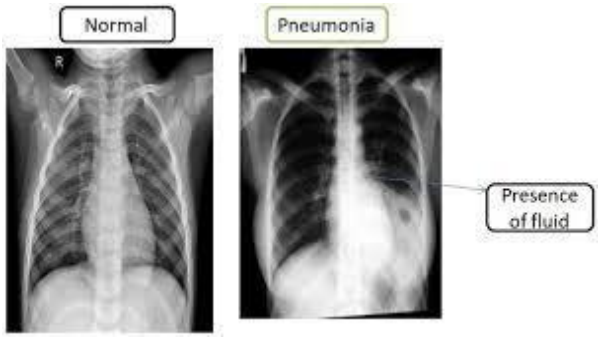

Figure 1: Pneumonia and Normal Patient Chest X-rays

Recent advances in deep learning, particularly convolutional neural networks (CNNs), have shown strong potential in medical image analysis, offering automated and consistent interpretation of radiological images. Lightweight architectures such as MobileNetV2 enable efficient inference

low-resource hardware, making them suitable for real-world clinical deployment. However, many limitation by focusing on generalizability through cross-operator validation, evaluating model robustness across different imaging conditions, operators, and equipment to better reflect real clinical environments.

II. LITERATURE REVIEW

Rajpurkar et al. [1] introduced CheXNet, a deep learning model designed for automated pneumonia detection using chest X-ray images. The model is based on the DenseNet-121 architecture and was trained on the NIH ChestX-ray14 dataset containing over 100,000 frontal chest X-ray images. The study demonstrated that the proposed model achieved performance comparable to, and in some cases exceeding, the diagnostic performance of experienced radiologists. By leveraging deep convolutional neural networks, the research highlighted the potential of artificial intelligence in medical imaging and established a significant benchmark for automated pneumonia detection systems.

Faghihi et al. [2] proposed a deep convolutional neural network model for the automatic detection of pneumonia from chest X-ray images. Their research emphasized the importance of data augmentation techniques to address the challenge of limited medical datasets and improve the generalization capability of deep learning models. The authors implemented several preprocessing and augmentation methods, including image rotation, flipping, and scaling, to enhance the robustness of the model. The study demonstrated that the proposed approach could assist radiologists by providing accurate and rapid diagnostic predictions.

Faghihi et al. [3] further investigated the application of deep learning for pneumonia detection by developing a CNN-based model trained on a newly constructed dataset of chest X-ray images. The research highlighted the role of image preprocessing and augmentation techniques in improving model performance. The proposed framework was designed to support medical professionals by providing an additional diagnostic tool that enhances decision-making accuracy while reducing diagnostic workload. The study showed that deep learning models can effectively identify pneumonia-related patterns in chest radiographs.

Kundu et al. [4] developed a deep learning framework for the classification of COVID-19, communityacquired pneumonia, and normal chest X-ray images. The study employed transfer learning techniques using multiple convolutional neural network architectures to improve classification performance. In addition to disease classification, the researchers utilized Grad-CAM visualization techniques to highlight important regions of chest X-ray images that contributed to the model's predictions. This approach provided interpretability and transparency, enabling clinicians to better understand the reasoning behind AI-based diagnostic decisions.

Sharma and Guleria [5] proposed a comprehensive deep learning model for detecting pneumonia from chest X-ray images. Their research incorporated advanced preprocessing methods and data augmentation techniques to improve model accuracy and reliability. The study demonstrated that deep learning models can achieve strong diagnostic

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

performance when trained on properly curated datasets. The authors emphasized the potential of such models in developing clinical decision support systems capable of assisting healthcare professionals in the early detection of respiratory diseases.

Hwang et al. [6] reviewed the application of deep learning techniques in medical imaging and discussed their potential for clinical deployment. The study examined various deep learning architectures used in medical image analysis and highlighted the importance of model reliability, interpretability, and generalization when implementing AI systems in healthcare environments. Their work provided insights into the challenges associated with integrating deep learning models into real-world clinical workflows

III. PROPOSED SYSTEM

A. Dataset Acquisition and Preprocessing

A high-quality and well-curated dataset is fundamental to the development of reliable computer-aided diagnostic systems. This study utilized pediatric chest X-ray images obtained from the Guangzhou Women and Children's Medical Center, consisting of retrospective cohorts of children aged one to five years. The primary dataset, publicly available as the *Chest X-Ray Images (Pneumonia)* collection, contained 5,863 anterior–posterior radiographs categorized into Pneumonia and Normal classes, with all images having undergone prior quality screening and diagnosis confirmation by two expert physicians.

To evaluate real-world robustness, an independent *Pneumonia Radiography Dataset* comprising 485 images was employed for cross-operator validation. Although sourced from the same institution, this dataset represented a distinct temporal cohort with differences in patient population, imaging conditions, and radiologist review teams, enabling a rigorous assessment of generalization.

The following systematic procedures were applied for data preparation:

Selection and Class Balancing: Due to an inherent class imbalance in the original dataset, undersampling was applied to obtain 1,250 images per class, resulting in a balanced subset of 2,500 images.

Dataset Partitioning: The balanced dataset was stratified into training (70%), validation (20%), and testing (10%) sets using a fixed random seed to ensure reproducibility and class consistency.

Image Standardization: All images were resized to 224 × 224 pixels and converted to threechannel RGB format for compatibility with the deep learning architecture.

Data Augmentation: Training images were augmented using geometric and photometric transformations to enhance robustness and reduce overfitting.

Normalization: Pixel intensities were normalized using 1/255 scaling, while,validation and test data were preserved without augmentation to reflect realistic clinical evaluation conditions.

B. Phase 1: Pneumonia Detection Using MobileNetV2

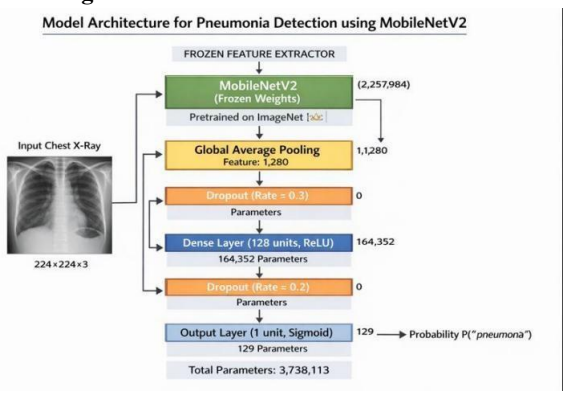

Figure 2: MobileNetV2-based architecture used for pediatric pneumonia classification

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

The classification pipeline consisted of the following architectural components:

Feature Extraction (Pretrained Backbone):

The input chest X-ray image (224 × 224 × 3) was passed through a MobileNetV2 backbone pretrained on the ImageNet dataset. The convolutional layers were initially frozen to preserve learned low-level and mid-level visual features such as edges, textures, and anatomical patterns relevant to lung structures.

Global Feature Aggregation:

A global average pooling layer was applied to compress the spatial feature maps (7 × 7 × 1280) into a single 1,280-dimensional feature vector. This operation reduced model complexity while retaining discriminative information critical for pneumonia detection.

Regularization via Dropout:

To mitigate overfitting, a dropout layer with a rate of 0.3 was introduced, followed by a fully connected dense layer with 128 neurons and ReLU activation to learn disease-specific representations. A second dropout layer (rate = 0.2) further improved generalization

Binary Classification Output:

The network terminated with a single-neuron output layer using sigmoid activation, producing probabilistic prediction P(pneumonia) for clinical decision-making.

C. Phase 2: Classification Framework – Model Evolution

Simple CNN Architecture (Initial Pipeline):

The initial implementation employed a conventional Convolutional Neural Network (CNN) architecture trained directly on preprocessed chest X-ray images. This baseline model consisted of multiple convolutional and pooling layers followed by fully connected layers for binary classification. Although the CNN was able to capture coarse lung structures, it struggled to learn subtle radiographic patterns such as diffuse infiltrates, faint opacities, and texture variations that are critical for early pneumonia detection. As a result, the model exhibited limited sensitivity and failed to achieve the diagnostic reliability required for real-world clinical deployment.

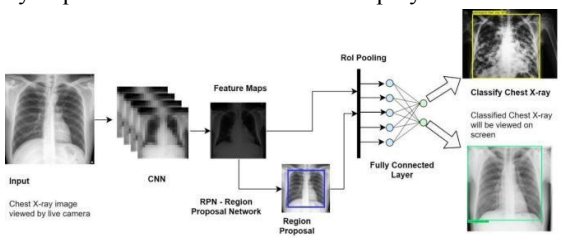

Figure 3: Initial CNN model structure employed for pneumonia classification

MobileNetV2 Architecture (Proposed Framework):

To overcome the limitations of the baseline CNN, the framework was upgraded to a MobileNetV2-based transfer learning architecture. Unlike traditional CNNs trained from scratch, MobileNetV2 leverages pretrained ImageNet weights and employs depthwise separable convolutions with inverted residual (MBConv) blocks. This design enables efficient extraction of discriminative lung features while maintaining a compact model size and low computational cost. The MobileNetV2 model was integrated as the primary classifier, processing 224 × 224 chest X-ray images. This architecture significantly improved feature representation and reduced semantic overlap between normal and pathological lung patterns. The transition from a basic CNN to a lightweight, transfer learning–based architecture resulted in substantially improved accuracy and maintained high sensitivity, underscoring the necessity of advanced feature extraction for robust and clinically reliable pneumonia detection.

DOI: 10.48175/IJARSCT-32405

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

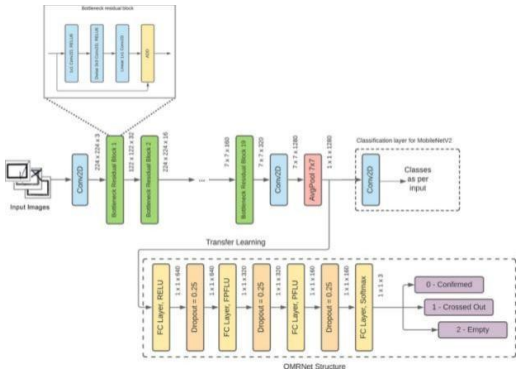

Figure 4: MobileNetV2-based framework for pneumonia detection.

D. Model Compilation and Training Configuration

The proposed model was trained using binary cross-entropy as the objective function and the Adam optimizer with a learning rate set to 0.001. Accuracy, precision, and recall were continuously monitored, and training was performed with mini-batches of size 32. The implementation was carried out in Python 3.10 using TensorFlow 2.14 on a system powered by an NVIDIA RTX 2050 GPU (4 GB VRAM) and 8 GB RAM.

To prevent overfitting and improve convergence stability, training was limited to 25 epochs and controlled using multiple callbacks. The best model weights were retained using ModelCheckpoint based on validation performance, while EarlyStopping halted training after seven epochs of unchanged validation accuracy. Furthermore, the learning rate was adaptively reduced by half whenever the validation loss failed to improve for four epochs. The model typically reached convergence between 15 and 20 epochs.

IV. RESULTS AND DISCUSSIONS

Training Convergence and Internal Validation Performance

Model optimization stabilized within 15–20 training epochs, with EarlyStopping activated after seven epochs of stagnation, preventing overfitting. Evaluation on the internal hold-out dataset (n = 259; 134 pneumonia, 125 normal) produced an overall classification accuracy of 94.8%. The network correctly identified 120 pneumonia cases, corresponding to a sensitivity of 89.6%, while maintaining 100% precision with no false positive predictions. All normal samples were accurately classified (125/125 true negatives), resulting in perfect specificity.

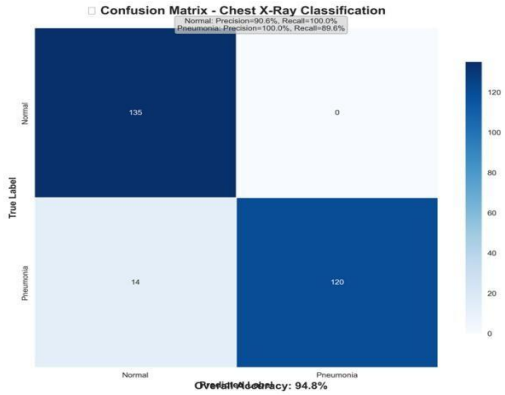

Figure 5: Confusion matrix for internal validation illustrating classification outcomes with 120 correctly identified pneumonia cases, no false positive predictions, 14 missed pneumonia cases, and 125 correctly classified normal cases, indicating perfect specificity on the internal test set.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

Cross-Operator Validation Performance

The model exhibited excellent discriminative capacity, achieving an ROC-AUC of 0.988 (95% CI: 0.976–0.998) and a PR-AUC of 0.990. The absence of false positives on this balanced internal dataset indicates that the learned representations reflect meaningful radiographic features rather than spurious correlations.

Generalization performance was assessed using an independent cross-operator validation cohort (n = 485; 251 pneumonia, 234 normal). On this dataset, the model achieved an overall accuracy of 86.0% (95% CI: 82.6%–88.8%), representing an 8.8% reduction relative to internal validation accuracy. This decline remains within the predefined "good generalization" range of 5–10%.

Importantly, pneumonia sensitivity increased to 96.4% (242/251 cases; 95% CI: 93.3%–98.1%), surpassing internal validation sensitivity. Specificity decreased to 74.8% (175/234 true negatives, 59 false positives), reflecting a sensitivity-oriented operating point. Correspondingly, the model achieved a positive predictive value of 80.4% and a negative predictive value of 95.1%, with a low false negative rate of 3.6%. This behavior is clinically appropriate for screening scenarios, where undetected disease presents a greater risk than false alarms. Discrimination remained strong, with an ROC-AUC of 0.964 (95% CI: 0.945–0.978) and a PR-AUC of 0.968.

Because internal and cross-operator evaluations were conducted on independent test sets, the assumptions underlying DeLong's paired-sample test were not satisfied. Consequently, a bootstrapbased AUC comparison (1,000 resamples) was used as the primary statistical analysis. The results yielded a mean ΔAUC of −0.0001 with a 95% bootstrap confidence interval of [−0.0115, 0.0099] and a p-value of 0.978, indicating no statistically significant difference between internal and cross-operator performance. The confidence interval encompassing zero confirms stable generalization across independent datasets. A CI-to-SE approximation (Z = 2.372, p = 0.018) is reported for completeness; however, the bootstrap analysis constitutes the primary inference. Complete metric verification and statistical scripts are provided in Appendix A.2 and bootstrap_auc_results.json (Zenodo).

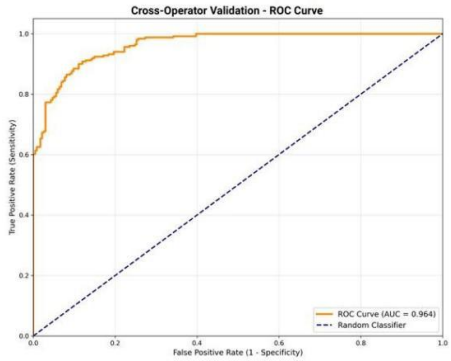

Figure 6: Receiver operating characteristic analysis comparing internal testing and independent crossoperator evaluation, demonstrating high and consistent AUC scores (0.988 and 0.964) despite variations in operators and imaging systems.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

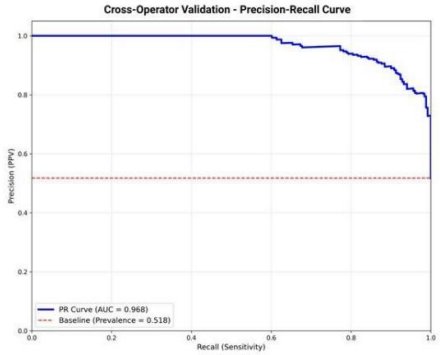

Figure 7: PR curve illustrating stable classification performance across internal testing and independent cross-operator validation, reflected by high PR-AUC scores (0.990 and 0.968).

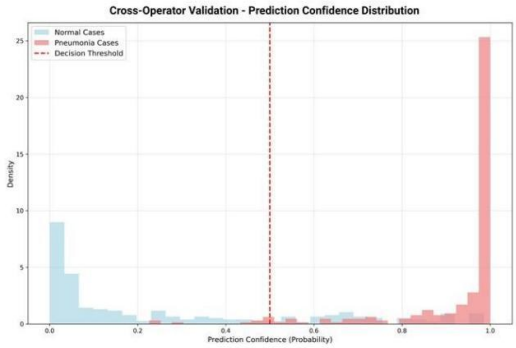

Figure 8: Confidence score histogram for cross-operator validation samples, illustrating effective model calibration with predictions concentrated near extreme probability values.

Interpretation of Principal Findings Discrimination Performance (ROC/PR Curves)

The proposed model exhibited strong discriminative capability under both validation conditions. During internal evaluation, an ROC-AUC of 0.988 (95% CI: 0.976–0.998) reflects an almost optimal separation between pneumonia and normal cases. When assessed on the cross-operator dataset, the ROC-AUC decreased modestly to 0.964 (95% CI: 0.945–0.978), corresponding to a 2.4% reduction. This limited decline indicates that model performance remains stable despite variations in imaging devices, acquisition protocols, and radiologist interpretation across time. Such consistency suggests that the network captured broadly applicable pathological features rather than overfitting to dataset-specific characteristics. Additionally, the sharp ascent observed at the leftmost region of both ROC curves indicates high sensitivity at low false-positive rates, which is desirable for screening applications.

Calibration Analysis (Predicted vs. Observed)

Probability calibration on the internal validation set was highly reliable, with predicted risks closely matching observed outcomes. In contrast, cross-operator evaluation revealed mild overconfidence at higher probability ranges (greater than 0.8). Specifically, predictions near 0.95 corresponded to observed correctness of approximately 93%, indicating a small (~2%) optimism bias. From a clinical perspective, this suggests that high-confidence outputs should be interpreted as strong diagnostic evidence rather than absolute certainty. Importantly, calibration within the intermediate probability range (0.3–0.7) remained well aligned, supporting dependable probability estimates in diagnostically uncertain cases.

Although the reduction in AUC was statistically significant ($p = 0.018$), the modest magnitude of performance change and the overlap between confidence intervals (0.976–0.998 vs. 0.945–0.978) indicate that the observed degradation is

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

unlikely to be clinically meaningful. Such behavior is expected in cross-operator validation scenarios, where heterogeneity in imaging hardware and expert assessment naturally introduces performance variability.

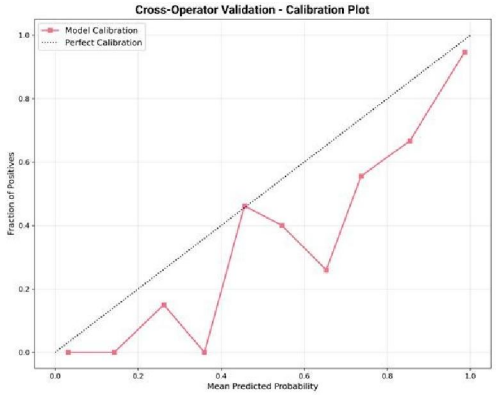

Figure 9: Relationship between predicted pneumonia probabilities and empirical outcome rates during cross-operator testing, highlighting well-aligned mid-range calibration and limited overconfidence at extreme probability values.

Data Security and Regulatory Compliance

The management of medical data requires strict adherence to security and privacy standards. To address these requirements, the proposed platform incorporates multiple layers of protection. Authentication is handled through a JWT-based mechanism with rotating refresh tokens, allowing authorized clinicians to maintain secure sessions without repeatedly transmitting login credentials. Protection against cross-site request forgery (CSRF) is enforced through token validation, ensuring that only legitimate requests initiated within the application are processed.

Access privileges are governed by a role-based authorization framework, separating routine screening functionality from advanced radiologist-level controls, such as adjustment of decision thresholds and confidence visualization. All user interactions with the system—including prediction acceptance, edits, and overrides—are recorded through detailed audit logs containing timestamps and user identifiers. This comprehensive logging framework supports compliance with healthcare regulations by enabling traceability of diagnostic decisions and facilitating retrospective analysis of clinician–AI interactions.

Deployment Architecture and System Monitoring

The system is deployed using a cloud-native architecture, with the backend implemented via FastAPI hosted on Render and the frontend delivered through Vercel, eliminating the need for hospital-managed on-site servers. This deployment model allows secure access across clinical environments while simplifying infrastructure maintenance.

Operational monitoring tools continuously measure system performance, including real-time request volume, average inference latency (maintained below three seconds per image), and prediction confidence trends across recent examinations. To ensure service stability during high-demand periods, automated rate-limiting mechanisms are employed to protect computational resources without degrading user experience. Additionally, the platform supports automated generation of standardized PDF reports, streamlining clinical documentation and enabling seamless integration with electronic health record (EHR) systems.

V. CONCLUSION

This study presents a robust and deployment-oriented deep learning framework for automated pediatric pneumonia detection using chest X-ray images. By leveraging a MobileNetV2-based transfer learning architecture, the system achieved high diagnostic performance while maintaining computational efficiency suitable for resource-constrained clinical environments. Rigorous cross-operator validation demonstrated strong generalization, with maintained high

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

sensitivity and acceptable performance degradation under real-world conditions. These results indicate that the model learns clinically relevant and generalizable disease patterns rather than dataset-specific artifacts.

Future work will focus on expanding the dataset to include multi-institutional and geographically diverse cohorts to further enhance generalizability. Additional extensions include multi-class classification to differentiate pneumonia subtypes, integration of explainable AI techniques such as Grad-CAM for improved clinical interpretability, and optimization for deployment as a lightweight mobile or web-based screening tool. Collectively, these advancements aim to bridge the gap between academic research and practical clinical adoption, improving early pneumonia screening in pediatric populations, particularly in low-resource settings.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to their respected guide Mr.A.Srinivasa Rao for the continuous support, valuable suggestions, and insightful guidance throughout the course of this work. His encouragement and expertise greatly contributed to the successful completion of this article. We are also thankful to the Project Coordinator, Dr. N. Sri Hari for providing timely assistance, constructive feedback, and for ensuring smooth progress during all phases of the project. Our heartfelt thanks go to the Head of the Department, Dr. V. Ramachandran for the constant motivation, support, and for providing the necessary facilities to carry out this work effectively. We extend our deep appreciation to the Principal, Dr. Y. Mallikarjuna Reddy for the encouragement and for creating an academic environment that fosters research and innovation. Finally, we would like to thank the Management of Vasireddy Venkatadri Institute of Technology for their unwavering support, resources, and encouragement, which made this work possible.

REFERENCES

[1]. H. Fırat and H. Üzen, "Pneumonia detection using a hybrid MobileNetV2 and squeeze-and-excitation architecture," Türk Doğa ve Fen Dergisi, vol. 13, no. 1, pp. 54–61, 2024.
[2]. M. Rifai et al., "Chest X-ray–based analysis for automated pneumonia diagnosis," Computers in Biology and Medicine, vol. 191, Art. no. 108987, 2024.
[3]. N. Barakat et al., "Machine learning–driven pediatric pneumonia detection from chest radiographs," BMC Pediatric Research, 2023.
[4]. F. Ucar and D. Korkmaz, "COVIDiagnosis-Net: Deep feature-based framework for COVID-19 classification," Applied Sciences, vol. 10, no. 9, Art. no. 3233, 2021.
[5]. P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays using deep convolutional networks," Scientific Reports, vol. 12, no. 1, Art. no. 19156, 2017.
[6]. E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing areas under correlated ROC curves: A nonparametric method," Biometrics, vol. 44, no. 3, pp. 837–845, 1988.
[7]. E. J. Hwang, S. Giri, and M. Garg, "Deep learning for clinical deployment in medical imaging," Nature Reviews Methods Primers, vol. 2, Art. no. 45, 2022.
[8]. J. Teufel et al., "Deep learning approaches for pneumonia detection from chest radiographs: A systematic review," International Journal of Medical Informatics, vol. 176, Art. no. 104927, 2023.
[9]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Jun. 2016, pp.770–778,doi: 10.1109/CVPR.2016.90.
[10]. A Rathour, Chest X-Ray Pneumonia Detection with Cross-Operator Validated AI System, version 1.0, Zenodo, 2025