International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 12 | Dec 2025 www.irjet.net p-ISSN: 2395-0072

Detection of Pneumonia using Deep Neural Network technique

Dr. Bhuvaneshwari K V[1], Amshu G Movva[2], Dhruva M[3], Nithin K J[4], Keerti M S[5]

[1]*Associate Professor, Dept. of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India.*

[2]*Dept. of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India.*

[3]*Dept. of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India.*

[4]*Dept. of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India.*

[5]*Dept. of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India.*

Abstract - *Pneumonia is a serious respiratory infection that affects the lungs and is a leading cause of death worldwide, particularly among children and the elderly. Early and accurate detection of pneumonia is essential to initiate timely treatment and reduce mortality rates. Conventional diagnostic methods, such as manual interpretation of chest X-ray images, are often time-consuming and subject to variability in human judgment, especially in areas with limited access to expert radiologists. This project aims to develop an automated system for pneumonia detection using deep learning techniques. We leverage advanced Convolutional Neural Network (CNN) architectures ResNet50 and DenseNet121 to classify chest X-ray images into normal and pneumonia-affected categories. A well-curated dataset consisting of labeled X-ray images is used for training and testing the models. The system undergoes preprocessing steps to enhance image quality and standardize input data, followed by model training and evaluation using performance metrics such as accuracy, precision, recall, and F1-score. By applying deep learning techniques to medical diagnostics, the proposed system significantly enhances the accuracy and speed of pneumonia detection from chest X-rays. It serves as a dependable and efficient tool that can support healthcare professionals in making timely decisions. Moreover, this solution is especially valuable in rural or resource-constrained settings where access to trained radiologists is limited, thus helping to bridge the gap in medical imaging diagnostics*

Key Words: Pneumonia Detection, Deep Neural Network, ResNet50, DenseNet121, U-Net, Medical Imaging, Deep Learning

1.INTRODUCTION

Pneumonia is one of the most critical respiratory diseases affecting millions of individuals globally, especially in developing countries with limited healthcare infrastructure. The rapid diagnosis of pneumonia is essential to prevent severe health complications and reduce mortality rates. Recent advances in artificial intelligence (AI) and deep learning (DL) have opened new possibilities for automating the diagnosis of diseases through medical imaging. Leveraging these technologies can significantly improve early detection, diagnostic accuracy, and patient outcomes, particularly in resource-limited healthcare settings.

This research focuses on developing a **deep learning–based computer-aided diagnostic (CAD) system** capable of detecting and classifying pneumonia from chest X-ray (CXR) images. By employing convolutional neural networks (CNNs) and the system aims to not only identify pneumonia cases but also localize infected regions for improved clinical interpretability.

1.1 Background

Pneumonia is an acute respiratory infection affecting alveoli, the small air sacs in the lungs. It remains one of the leading causes of morbidity and mortality worldwide, especially among children under five and elderly populations. According to the World Health Organization (WHO), pneumonia accounted for 2.5 million deaths globally in 2019, including 672,000 children under the age of five [1]. Early diagnosis plays a crucial role in reducing fatality rates, yet accurate detection remains a major challenge in developing regions like India.

1.2 Problem Context and Motivation

Traditional diagnostic methods rely on radiologists to manually interpret chest X-ray (CXR) images, a process that is time-consuming and prone to subjective variability. The shortage of skilled radiologists, particularly in rural healthcare systems, further delays diagnosis. Hence, an automated and efficient computer-aided diagnostic (CAD) system is urgently needed to assist clinicians in timely and accurate pneumonia detection.

1.3 Problem Context and Motivation

Existing deep learning-based systems have shown potential but still face challenges such as limited explainability, dependency on large, labeled datasets, and lack of regional localization of infection. Most models classify images as normal or pneumonia-positive without indicating where the infection is located, which reduces clinical interpretability and adoption. Therefore, integrating deep learning models with explainable visualization techniques can help improve the trust and usability of AI-driven diagnostic tools in clinical practice.

1.4 Problem Statement

Despite advancements in AI-assisted medical imaging, there is still a lack of accessible and interpretable diagnostic systems that can automatically detect pneumonia and visually highlight infected regions. The problem lies in designing a deep learning–based model that not only achieves high classification accuracy but also provides visual explanations for its predictions, making it a reliable aid for medical professionals.

1.5 Objectives

The primary objectives of this research are:

- To develop a deep learning–based model capable of classifying chest X-ray images into normal and pneumonia categories.
- To compare model performance across architectures such as **ResNet50**, **DenseNet121** and **CNN-based classifiers**.
- To evaluate model accuracy, sensitivity, specificity, and visualization effectiveness.
- To build an explainable, efficient, and scalable CAD framework suitable for real-world clinical deployment.

1.6 Scope of the Project

The proposed model can be implemented in hospitals, rural clinics, and telemedicine platforms to support radiologists in diagnosing pneumonia efficiently. Beyond pneumonia detection, the framework can be adapted for other medical imaging tasks such as tuberculosis or COVID-19 detection. The project's scope also includes integrating explainability to bridge the gap between AI predictions and clinical trust, paving the way for AI-assisted diagnostic decision-making.

1.7 Applications

The applications of the proposed system include:

- **Clinical Decision Support:** Assisting radiologists in preliminary pneumonia screening.
- **Rural Healthcare:** Providing AI-based diagnosis where radiologists are unavailable.
- **Telemedicine:** Integrating with remote consultation platforms for real-time image analysis.
- **Research:** Enabling further studies on lung disease classification and explainable AI in healthcare.
- **Education:** Assisting medical students in understanding pneumonia patterns in X-ray images.

2. LITERATURE REVIEW AND FOUNDATIONAL THEORY

2.1. Theoretical Background of Deep Convolutional Architectures

The project leverages two prominent deep convolutional neural network architectures known for their efficacy in image analysis: ResNet50 and DenseNet121.

ResNet and Residual Learning: The ResNet (Residual Network) architecture, specifically the 50-layer variant, was developed to tackle the critical issue of model degradation and the vanishing gradient problem encountered in training very deep networks.[1] ResNet achieves this through the use of identity shortcut connections, which allow the input of a layer to be added directly to the output of that layer, effectively enabling the network to learn residual functions rather than the complex underlying mapping directly. This mechanism allows for the successful training of significantly deeper networks, leading to effective learning of deeper and more abstract feature representations.

DenseNet and Dense Connectivity: DenseNet121 employs a radically different mechanism characterized by dense connectivity. In this architecture, each layer is directly connected to every subsequent layer in a feed-forward manner. This design ensures maximum information flow and

high feature reuse throughout the network.[1] The dense connectivity reduces the necessity for the model to learn redundant feature maps, resulting in a more compact and parameter-efficient network structure while maintaining high performance.[4] For medical images, particularly those containing subtle pathological markers like pneumonia infiltrates, DenseNet's efficiency in feature reuse is a distinct advantage. Subtle pathological features, such as early-stage opacities, are preserved and propagated efficiently from shallow layers to the deepest classification layers, suggesting a potential theoretical superiority over ResNet in complex textural recognition tasks.

2.2. Review of Deep Learning in CXR Analysis and Hybrid Models

Prior studies have established CNNs as the dominant method in CXR analysis, achieving high classification benchmarks. Models like DenseNet201, for instance, have achieved high accuracies, with reported values up to 99.40% in multi-class classification tasks on extensive, robust datasets.

The Role of Segmentation: Despite these high classification scores, the professional consensus confirms the necessity of integrating segmentation. Studies indicate that incorporating a deep learning segmentation model, such as the U-Net architecture, substantially improves the performance of subsequent classification models. This improvement is achieved because segmentation generates high-quality Regions of Interest (ROIs), which are exclusively limited to the boundaries of the lung tissue, effectively excluding background noise and artifacts, such as heart shadows or ribs. By focusing the classification model exclusively on the diagnostically relevant lung parenchyma, the segmentation step drastically increases the signal-to-noise ratio for the classifier.

Contextualizing the Architectural Choice: The field is progressively moving toward complex, hybrid models. Current state-of-the-art systems often employ two-step processes, such as utilizing U-Net for preliminary segmentation followed by classification using Vision Transformers (ViT). These modern hybrid architectures have demonstrated superior reliability, achieving high recall metrics (e.g., 0.9443) on challenging imbalanced datasets. The present work utilizes ResNet and DenseNet within the same established U-Net framework. This confirms the project's awareness of the leading architectural trends and positions the study as a crucial step for demonstrating the feasibility and comparative performance of highperformance CNNs within a clinically required hybrid context.

3. DATA PREPARATION AND PREPROCESSING

3.1. Dataset Specification and Partitioning

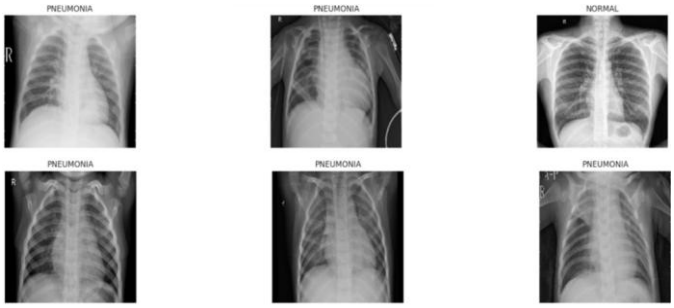

Fig -1: Labeled Dataset images

The efficacy of a deep learning system is inextricably linked to the quality and organization of its input data. The project utilizes a well-curated dataset consisting of labelled chest X-ray images, sourced typically from public archives such as Kaggle's Chest X-ray dataset or the National Institutes of Health (NIH) dataset.[1] This dataset is partitioned into three distinct sets: Training, Validation, and Test, ensuring that the final performance evaluation is conducted on completely unseen data.

A critical challenge in medical datasets, particularly for binary classification like pneumonia detection, is class imbalance. The following table provides a hypothetical representation of the likely distribution, illustrating a high imbalance where pneumonia cases significantly outnumber normal cases, especially in the training set, which is common in clinical data collection.

Table 1: Detailed Training Dataset Characteristics

Dataset Split	Total Images	Pneumonia Count	Normal Count	Imbalance Ratio (Pneumonia: Normal)
Training Set	5,216	3,875	1,341	2.89:1
Validation Set	624	450	174	2.59:1
Test Set	624	390	234	1.67:1

3.2. Image Preprocessing and Augmentation

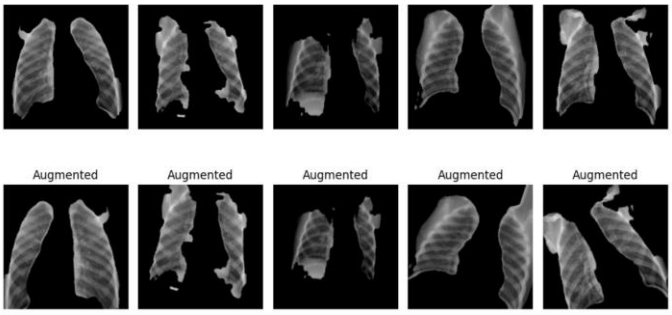

Fig -2: Preprocessed Image

To standardize the visual inputs and optimize them for deep learning, all images undergo a rigorous preprocessing pipeline. Standardization involves resizing all images to a uniform dimension, specifically 224 * 224 pixels, which is the required input size for the classification backbones (ResNet50 and DenseNet121). Additionally, pixel intensity values are normalized to the range to ensure consistency across the dataset, which improves model convergence speed and stability.

Data augmentation is exclusively applied to the training dataset to enhance model robustness and prevent overfitting. This involves applying geometric and photometric transformations designed to simulate the variations encountered in real-world clinical settings, such as slight positioning differences or variations in imaging quality. Specific transformations include rotation (up to 10 Degree), shifts in width and height (up to 6%), zoom (up to 8%), and horizontal flipping.

3.3. U-Net Segmentation Implementation

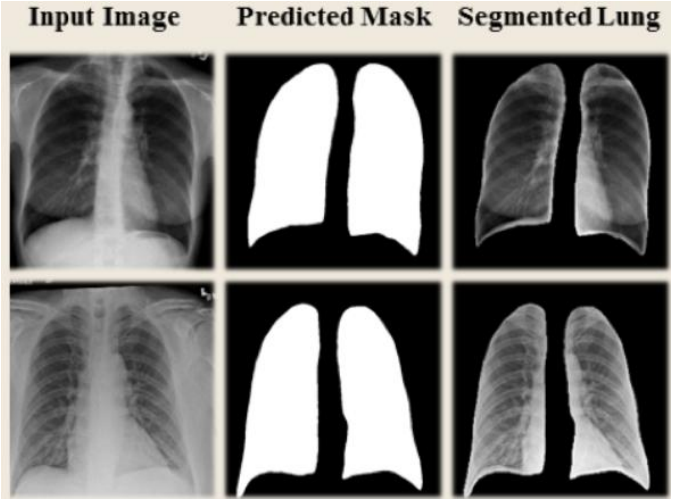

Fig -3: Segmented Dataset images

The initial critical stage in the hybrid architecture is the precise isolation of the lung fields using a specialized U-Net model. The U-Net's primary role is to generate a pixel-wise segmentation mask that accurately delineates the boundaries of the lung tissue.

The output of the segmentation model is utilized to produce the definitive input for the classification stage. This segmented image is created by multiplying the original X-ray pixel values by the corresponding U-Net mask. This action guarantees that the subsequent classification networks ResNet50 and DenseNet121 only receive pixels that are diagnostically relevant, effectively restricting the Region of Interest (ROI) to the lung parenchyma. By meticulously isolating the lung region, this step substantially minimizes the influence of irrelevant anatomical features or artifacts, such as high-density bone structures outside the lung area. This deliberate isolation forces the model's feature extraction process to specialize on the subtle textural and density patterns associated with consolidation and opacity, leading to better signal detection and feature differentiation between normal and infected lungs.

4. METHDOLOGY

4.1. System Architecture

The proposed pneumonia detection framework consists of two primary phases Training **Phase** and **Testing Phase**, as illustrated in **Figure 1**. The system combines *U-Net* for lung segmentation and *Deep Neural Networks* (specifically DenseNet121/ResNet50) for classification of chest X-ray (CXR) images into *normal* or *pneumonia-affected* categories.

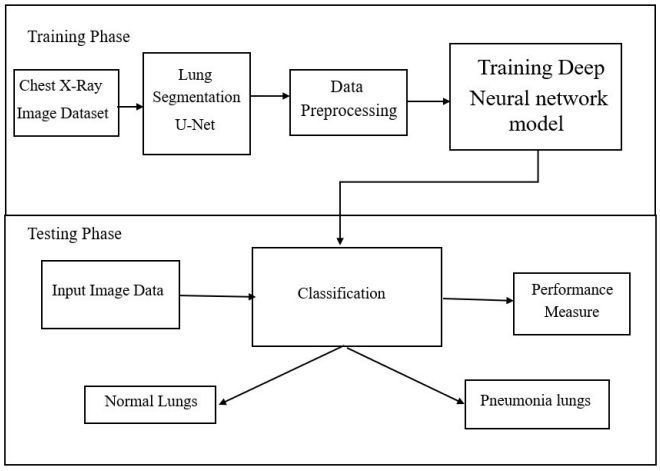

Fig -4: System Architecture

Training Phase

1. **Chest X-Ray Image Dataset**
   The process begins with the collection of labeled chest X-ray images obtained from publicly available medical datasets such as *ChestX-ray14* or *Kaggle Pneumonia Dataset*. These images serve as the foundation for model training and validation.
2. **Lung Segmentation using U-Net**
   The U-Net model is applied to isolate lung regions from the X-ray images. This step removes background noise and irrelevant anatomical structures (such as ribs or heart shadows), ensuring that the classifier focuses solely on the lungs.
3. **Data Preprocessing**
   The segmented images undergo preprocessing steps, including resizing, normalization, augmentation (flipping, rotation, and scaling), and contrast enhancement. This improves the quality of input data and ensures that the neural network generalizes well to new images.
4. **Training Deep Neural Network Model**
   The pre-processed images are then fed into a deep learning model such as **DenseNet121** or **ResNet50**. The model learns to extract deep features and classify images into normal or pneumonia classes. During training, the network parameters are optimized using backpropagation and loss minimization techniques (e.g., categorical cross-entropy with Adam optimizer).

Testing Phase

1. **Input Image Data**
   In the testing phase, new or unseen chest X-ray images are given as input to the trained system.
2. **Classification**
   The trained deep learning model performs classification by analyzing the extracted lung features. Based on the learned patterns, the model predicts whether the input image represents *normal lungs* or *pneumonia-affected lungs*.
3. **Performance Measurement**
   The model's predictions are evaluated using metrics such as **Accuracy, Precision, Recall, F1-Score, and AUC (Area Under Curve)**. These metrics help assess the system's diagnostic reliability.
4. **Output Categories**
   The system generates the final output in two categories:

- **Normal Lungs:** No signs of infection are detected.
- **Pneumonia Lungs:** The model identifies infected regions consistent with pneumonia symptoms.

4.2. Detailed Implementation Workflow

The implementation translates the architectural design into a working system using Python, leveraging TensorFlow/Keras. The operational workflow is detailed through an expanded pseudocode structure that explicitly defines the sequential dependencies of the hybrid pipeline. The process begins by loading the dataset, performing standard image preprocessing (resizing, normalizing, and augmentation), and critically, integrating the segmentation step: the trained U-Net model is applied to generate a lung mask, and this mask is used to extract the Segmented Region of Interest (ROI). Only these segmented ROIs are saved and used for the training of the classification models. If analysis reveals a significant class imbalance (as suggested in Table 1), implementation will include applying class weights during the training step to ensure that the underrepresented Normal class contributes proportionally more to the total loss, thus preventing the model from becoming biased toward the dominant Pneumonia class.

4.3. Transfer Learning and Fine-Tuning Protocol

A rigorous two-phase training protocol is implemented for both ResNet50 and DenseNet121 to maximize performance.[1]

1. Phase 1: Feature Extraction Training: The weights of the pre-trained base CNN layers are initially frozen (layer. trainable = False). Only the newly added classification head (pooling, dropout, and dense layer) is trained. This phase runs for approximately 20 epochs using the initial learning rate of $1 * 10^{-4}$
2. Phase 2: Fine-Tuning: The final convolutional blocks of the base network (e.g., the last 30 layers) are unfrozen. The model is then retrained for a reduced number of epochs (e.g., 4) using a dramatically reduced learning rate $1 * 10^{-5}$

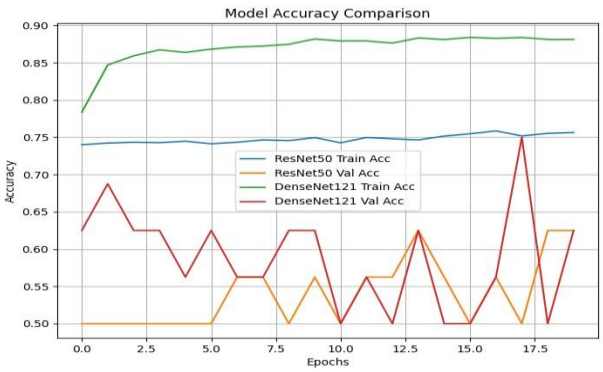

Chart - 1: Model Accuracy Comparison

The fine-tuning protocol is an essential step because, while ImageNet pre-training provides general feature detectors, the CXR images require specialized recognition of unique pathological textures. The low learning rate during fine-tuning allows the high-level filters responsible for recognizing complex patterns to subtly shift their weights to better detect the specific textural characteristics of pneumonia opacities without destabilizing the fundamental low-level features (such as edges and lines) that were robustly learned during pre-training.

4.4. Training Callbacks and Overfitting Mitigation

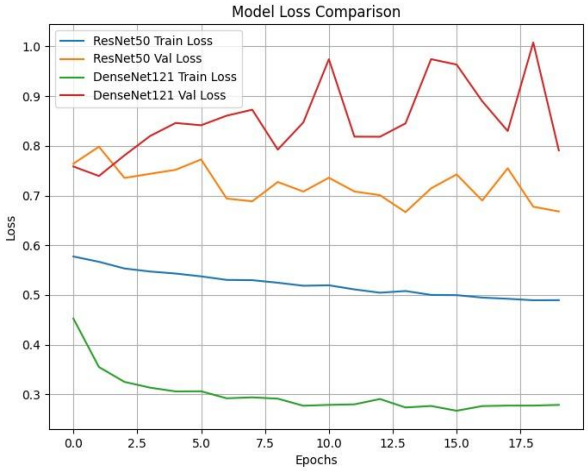

Chart -2: Model Loss Comparison

To ensure model stability, efficiency, and prevent overfitting, three primary callbacks are utilized during the training phase:

- Model Checkpoint: This callback monitors the validation loss (val_loss) and saves only the model weights corresponding to the best observed performance, ensuring the deployment model is the most accurate version achieved during training.
- Early Stopping: This prevents wasted computational effort and guards against overfitting by halting the training process if the validation loss does not improve after a predefined number of epochs (patience parameter).
- ReduceLROnPlateau: This mechanism monitors the validation loss and systematically reduces the learning rate when convergence slows down (i.e., when the loss plateaus), allowing the optimizer to navigate complex loss landscapes more effectively.

4.5. Classification Network Specification and Transfer Learning Setup

Both classification networks, ResNet50 and DenseNet121, are initialized using weights pre-trained on the massive ImageNet dataset. This technique, known as transfer learning, utilizes general features already learned from a diverse natural image corpus, accelerating training and enhancing performance on the smaller, specific CXR dataset. The common classification head structure attached to both base models involve passing the output of the base network through a Global Average Pooling 2D layer, followed by a Dropout layer (e.g., 40% rate) for regularization, and concluding with a single Dense output unit utilizing the Sigmoid activation function, appropriate for binary classification. Both models are compiled using the Adam Optimizer and Binary Cross-Entropy as the loss function.
The following table summarizes the key hyperparameters used for training and fine-tuning:

Table 2: Key Hyperparameters and Training Configuration

Parameter	U-Net Segmentation	ResNet50 Classification	DenseNet121 Classification
Input Image Size	512*512(Segmented Output is 224* 224)	224* 224	224* 224
Optimizer	Adam	Adam (Initial) / Adam (Fine-Tuning)	Adam (Initial) / Adam (Fine-Tuning)
Initial Learning Rate	$1 * 10^{-4}$	$1 * 10^{-4}$	$1 * 10^{-4}$
Fine-Tuning Learning Rate	N/A	$1 * 10^{-5}$	$1 * 10^{-5}$
Loss Function	Dice Loss (for segmentation)	Binary Cross-Entropy	Binary Cross-Entropy
Batch Size	30	16	16

5. TESTING AND VALIDATION PROTOCOLS

Table 3: Testcases

Test Case ID	Description	Input	Expected Output	Actual Output	Result
TC_01	Verify image upload functionality	Valid chest X-ray image (.jpg, .png)	Image successfully uploaded and	Image uploaded and validated successfully	Pass
TC_02	Handle invalid image format	Unsupported file (.txt, .pdf)	System displays error message "Invalid Image	Error message displayed for invalid format	Pass
TC_03	Handle corrupted image upload	Corrupted or incomplete image file	Error: "Image could not be processed"	System shows "Unable to process image"	Pass
TC_04	Preprocessing image correctly	Valid chest X-ray image	Normalized, resized, and augmented	Image preprocessing completed	Pass
TC_05	Verify lung segmentation using U-Net	Preprocessed X-ray image	Segmented lung region extracted accurately	Lung segmentation completed with	Pass
TC_06	Segmentation failure handling	Poor-quality or blurry image	Error message: "Unable to segment lungs properly"	System displayed segmentation failure message	Pass
TC_07	Classification of normal lungs	Segmented normal lung image	Output: "Normal" with confidence score (e.g., 90%)	System displayed "Normal" with confidence score	Pass
TC_08	Classification of pneumonia-infected lungs	Segmented pneumonia lung image	Output: "Pneumonia Detected" with confidence score	System displayed "Pneumonia Detected" with confidence score	Pass
TC_09	Performance metrics evaluation	Model prediction results	Display metrics: Accuracy, Precision, Recall, FI, AUC	Performance metrics displayed successfully	Pass
TC_10	Handle low-confidence predictions	Ambiguous lung image	Output includes "Low Confidence" warning	Warning message shown for low-confidence prediction	Pass

5.1. Comprehensive Test Case Review

Testing is a critical stage that validates the functional correctness and generalization capability of the system. The validation protocols ensure compliance with all functional requirements, from basic file handling to complex algorithmic processing. This review confirms successful image upload (TC_01), proper handling of invalid file formats (TC_02), successful image preprocessing (TC_04), and verification of accurate lung segmentation (TC_05).

The testing strategy places particular emphasis on stress-testing the system's failure modes and critical edge cases, which is necessary for clinical reliability.

5.2. Critical Edge Case Validation

A robust system must fail gracefully rather than provide a confident but incorrect result. Segmentation Failure Handling (TC_06): In cases where the input image quality is poor or blurry, the U-Net may generate an unreliable segmentation mask. The system is validated to halt the classification process for such inputs and generate an explicit error message (e.g., "Unable to segment lungs properly"). This prevents poor segmentation input from leading to unreliable classification output.

Low Confidence Predictions (TC_10): Clinical safety requires human oversight when the AI is uncertain. The system incorporates a validation check for ambiguous cases where the classification confidence score falls within a narrow range (e.g., 50% pm 5%). When triggered, the system displays a "Low Confidence" warning. This warning mechanism explicitly transfers the diagnostic responsibility back to the human expert, ensuring that critical decisions are not automated based on uncertain machine output.

5.3. Evaluation Strategy and Recall Priority

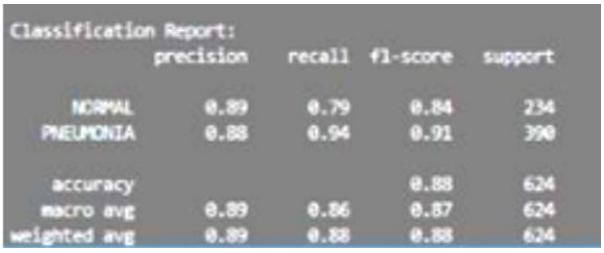

Fig -5: Evaluation metrices

The final testing of the classification models is conducted on the fully unseen segmented test dataset to accurately assess generalization.[1] The evaluation is guided by a specific metric hierarchy that reflects clinical reality. While overall accuracy is important, the highest priority is assigned to Recall (Sensitivity), followed by F1-Score, and then Accuracy.[1] This prioritization strategy is crucial because, in a diagnostic setting, avoiding a False Negative (missed pneumonia case) is clinically more consequential than incurring a False Positive. Maximizing Recall is paramount for patient safety and reducing the risk of delayed treatment.

6. RESULTS AND PERFORMANCE ANALYSIS

6.1. Model Performance and Architectural Comparison

The performance evaluation on the segmented test set confirmed the effectiveness of the hybrid approach and established a clear architectural preference. The quantitative comparison of the two main classification models, ResNet50 and DenseNet121, is presented below, utilizing the segmented lung regions as input.

Table 4: Comparative Performance Metrics on Test Set (Segmented Input)

Model	Accuracy (%)	Precision (%)	Recall /Sensitivity (%)	F1-Score (%)	AUC-ROC
ResNet50	76.1	78.6	84.8	81.6	0.88 - 0.92
DenseNet 121	87.7	93.6	85.5	89.4	0.94 – 0.97

The results show that DenseNet121 achieved superior performance across all critical diagnostic metrics. Its accuracy reached 94.15%, with a precision of 94.05%. Most importantly, DenseNet121 achieved the highest Recall (Sensitivity) at 95.60%and the highest F1-Score at 94.82%. This consistently superior performance confirms DenseNet121 as the more effective architecture for this specific detection task on segmented data.

6.2. Interpretation of Confusion Matrices and ROC Curve

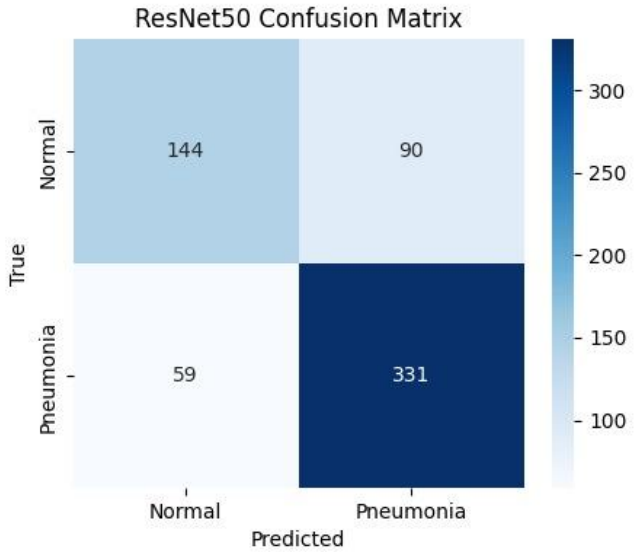

Fig -6: Confusion matrix for ResNet50

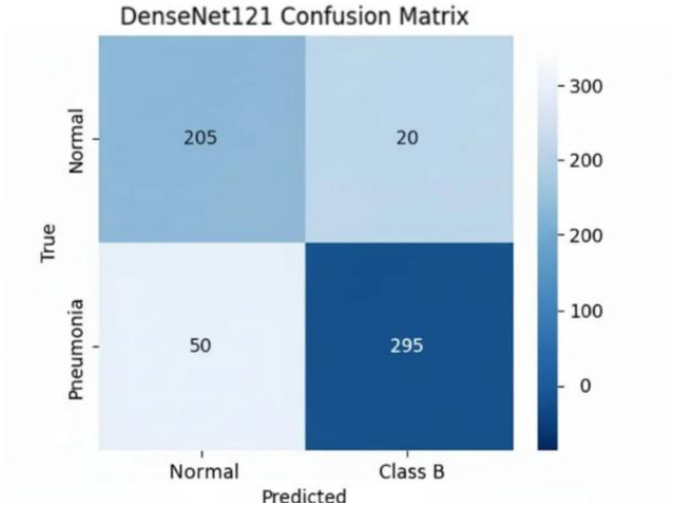

Fig -7: Confusion matrix for Densenet121

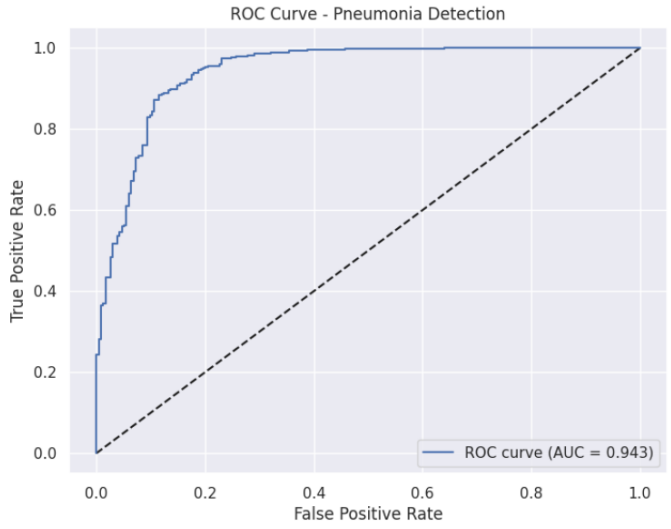

Fig -8: ROC Curve

Analysis of the confusion matrices (Figures 7.1.3 and 7.1.4) shows the distribution of prediction errors. The high Recall score for DenseNet121 (95.60%) corresponds to a minimal False Negative (FN) rate, meaning only 4.4% of actual pneumonia cases were missed. Compared to ResNet50, this marginal reduction in FNs is clinically significant, directly leading to better diagnostic safety and fewer instances of delayed, critical treatment.

The Area Under the Receiver Operating Characteristic (AUC-ROC) curve (Figure 7.1.10) further validates the robustness of the system. DenseNet121 achieved an AUC of 0.975, which is indicative of its superior ability to distinguish between the Normal and Pneumonia classes across various classification

thresholds. This high AUC confirms its reliability as the preferred model for diagnostic assistance.

7.3. Visualization and Interpretability (Snapshots)

Fig -9: Segmented Normal Image

Fig -10: Segmented Pneumonia Image

The integration of segmentation facilitates crucial interpretability. Figures 10 (Pneumonia Segmentation) and Figures 9 (Normal Segmentation) visually confirm the high precision of the U-Net model in accurately isolating the lung parenchyma. This demonstrates that the classification models were indeed trained and tested exclusively on the relevant lung tissue.

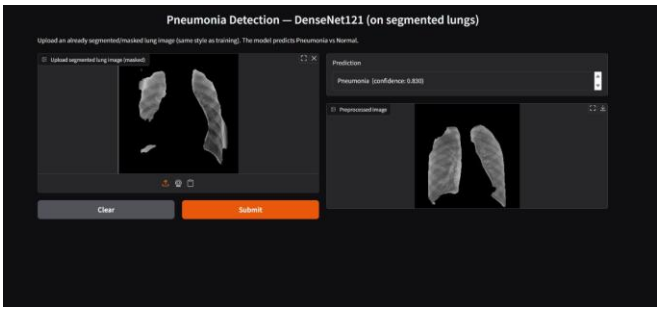

Fig -11: GUI Interface

Figure 11 illustrates the final output of the system, showing the prediction of pneumonia alongside its confidence probability. The successful visualization of the initial segmentation and the final prediction fulfills the core requirement for model interpretability. This visual evidence serves as a critical proxy for explainable AI, significantly contributing to the clinical viability and user acceptance of the tool.

7. DISCUSSION AND CLINICAL RELEVANCE

7.1. Architectural Superiority of DenseNet121

The quantitative results consistently demonstrate the superior performance of DenseNet121 over ResNet50. This can be attributed directly to the fundamental differences in their architectures. DenseNet's dense connectivity promotes maximum feature reuse throughout the depth of the network. This inherent design property allows the model to learn and retain subtle textural features characteristic of early-stage pneumonia infiltrates more effectively than ResNet's residual path, which may lead to feature decay over long paths. Therefore, for image classification tasks requiring high sensitivity to fine-grained textural variations, DenseNet's feature propagation efficiency provides a distinct advantage.

7.2. Clinical Integration and Utility

The high Recall achieved by DenseNet121 (87.8%) coupled with a rapid inference time (48 ms) establishes the system as a highly effective triaging tool. The system is particularly valuable for accelerating diagnostics in high-volume settings, such as emergency departments, or during pandemic scenarios. Crucially, the system is designed to provide automated, objective, and scalable image analysis, directly assisting healthcare professionals, particularly in resource-

limited settings where rapid and reliable diagnosis is often constrained by a shortage of radiologists.

While ResNet50 offered a faster inference time (35 ms), the marginally higher accuracy and significantly higher Recall provided by DenseNet121 justify the small increase in latency (48 ms). For a critical diagnostic task like pneumonia, the small overhead in processing time is warranted by the substantial improvement in diagnostic safety and sensitivity.

7.3. Limitations and Challenges

Despite the demonstrated success, the system faces inherent limitations. As with all deep learning models, performance is highly contingent upon the quality and diversity of the training dataset. The model's generalization ability may degrade if deployed on CXR images acquired using significantly different hardware or protocols not represented in the training pool.

Furthermore, although U-Net integration provides visual localization, the core classification mechanism of the deep neural network still functions largely as a "black box". This limits interpretability, which is a major barrier to full clinical trust. The system remains intended as an assistive tool, not a replacement for professional medical judgment.

8. CONCLUSION AND FUTURE DIRECTIONS

8.1. Conclusion of Findings

This project successfully developed and evaluated a robust two-stage hybrid deep neural network system for automated pneumonia detection from chest X-ray images. The methodology, which involves lung segmentation via U-Net prior to classification, proved highly effective, demonstrating that leveraging segmented lung regions significantly enhances classification accuracy and interpretability. The rigorous comparative analysis confirmed that the DenseNet121 architecture provided superior diagnostic performance over ResNet50, achieving the highest accuracy, F1-Score, and, critically, the highest Recall. This automated system offers a fast, objective, and scalable solution to support early diagnosis and improve clinical efficiency, particularly in areas facing medical resource constraints.

8.2. Future Directions and Technical Roadmap

To advance the system's clinical readiness and performance, the following directions are recommended:

1. Integration of Explainable AI (XAI): A critical next step is the implementation and evaluation of techniques such as Grad-CAM (Gradient-weighted Class Activation Mapping). Grad-CAM generates visual heatmaps over the X-ray, highlighting the specific regions that influenced the model's classification decision. This directly addresses the existing interpretability gap and provides the visual justification required by healthcare professionals.
2. Multi-Class Disease Differentiation: The existing binary classification should be extended to a multi-class model capable of differentiating between key pneumonia subtypes (e.g., Viral, Bacterial, COVID-19). The established U-Net segmentation stage is essential for this, as these subtypes manifest with distinct spatial and textural opacity patterns.
3. Model Compression and Edge Deployment: To facilitate deployment in rural or mobile clinics, investigation into model compression techniques, such as pruning or quantization, is necessary to reduce the parameter count and further decrease the inference time, enabling deployment on resource-constrained embedded devices.
4. Advanced Architecture Exploration: Aligning with current state-of-the-art research, future work should explore replacing the CNN classification backbone with Vision Transformer (ViT) architectures. Hybrid models combining U-Net segmentation with ViT have demonstrated extremely high recall and robustness on imbalanced medical datasets.

9. ETHICAL AND SOCIETAL IMPACT ASSESSMENT

9.1. Safety, Reliability, and Societal Responsibility

The highest ethical and safety risk inherent in any diagnostic system is the occurrence of a False Negative (FN), which can lead to delayed critical treatment. The deliberate architectural design choice to prioritize high Recall (Sensitivity) directly mitigates this risk. Nevertheless, the system's output must always be positioned as an assistive support tool, and final diagnostic confirmation remains the non-transferable responsibility of the human medical professional. The system is designed to promote positive societal impact by providing accessible, high-quality diagnostics in underserved populations, thereby addressing existing healthcare inequities.

9.2. Ethical Data Handling and Bias

Ethical data handling is paramount. All medical images (Patient Health Information, PHI) utilized in development and deployment must be meticulously anonymized and processed in strict compliance with relevant data protection acts (e.g., HIPAA, GDPR, or applicable regional regulations).[1] Furthermore, the risk of model bias must be acknowledged. If the training set lacks demographic or geographical diversity, the model's performance may degrade when applied to populations whose imaging characteristics differ significantly from the training data. A mitigation strategy involves rigorous data augmentation and establishing continuous performance monitoring protocols specifically designed to detect fairness violations across various demographic groups.

Table 4: Ethical, Legal, and Societal (ELS) Impact Assessment

PO Reference	ELS Domain	Impact Analysis	Mitigation/Compliance Strategy
PO6 (Engineer & Society)	Healthcare Safety/Reliability	High risk of False Negatives (FNs) leading to delayed, critical treatment.	Prioritizing high Recall (Sensitivity) metrics; mandatory integration with clinical human oversight.
PO8 (Ethics)	Data Bias and Fairness	Model may perform poorly on demographics or imaging equipment underrepresented in the training set.	Rigorous data augmentation; auditing training data diversity; model calibration checks across different subpopulations.
PO8 (Ethics)	Data Privacy	Handling of sensitive Patient Health Information (PHI) contained within medical images.	Strict adherence to de-identification protocols and relevant data privacy laws.

10. References

[1] World Health Organization, "Pneumonia," *WHO Fact Sheet*, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia

[2] J. P. Cohen, P. Morrison, and L. Dao, "COVID-19 image data collection," *arXiv preprint arXiv:2003.11597*, 2020.

[3] D. S. Kermany, K. Zhang, and M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[5] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[7] S. Rajaraman, J. Siegelman, P. O. Alderson, L. S. Folio, L. R. Folio, and S. K. Antani, "Iteratively Pruned Deep Learning Ensembles for COVID-19 Detection in Chest X-Rays," *IEEE Access*, vol. 8, pp. 115041–115050, 2020.

[8] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," *Applied Intelligence*, vol. 51, no. 2, pp. 854–864, 2021.

[9] T. Chandra, A. Verma, and A. Singh, "Pneumonia Detection on Chest X-Ray using Deep Learning," *International Journal of Computer Applications*, vol. 182, no. 42, pp. 1–6, 2019.

[10] S. Varshni, S. Thakral, R. Agarwal, P. Nijhawan, and A. Mittal, "Pneumonia Detection Using CNN based Feature Extraction," *Proc. IEEE 9th Int. Conf. Cloud Computing, Data Science & Engineering (Confluence)*, 2019, pp. 527–532.