Efficient Machine Learning Approaches for Pneumonia Detection on Chest Radiographs

Ibtasam Ur Rehman*, Warda Maqbool*, Maryam Shahzadi*, Malaika Khan*,
Abdulraqeb Alhammadi†, Jibran K. Yousafzai‡,
*Ho Chi Minh City University of Technology, Ho Chi
*The Millennium Universal College (TMUC), Islamabad, Pakistan
†Faculty of Artificial Intelligence, University of Technology Malaysia
‡College of Engineering and Applied Sciences, American University of Kuwait

***Abstract*—This research presents a machine learning-based framework for automated pneumonia detection from chest X-ray images. Three classifiers—SVM, KNN, and Random Forest—were developed and evaluated on 5,856 chest X-rays using PCA for dimensionality reduction. The SVM model demonstrated superior performance with 94.97% accuracy, 0.9846 AUC-ROC, and 87.09% MCC, excelling across clinical metrics including 97.54% sensitivity and 88.01% specificity. The approach addresses the dataset's class imbalance and includes extensive validation through cross-validation and bootstrapping analyses. The implementation provides interpretable confidence scores and visual explanations, offering a reliable and deployable solution for pneumonia screening, particularly in resource-constrained healthcare settings [1].**

***Index Terms*—Pneumonia Detection, Chest X-Ray Analysis, Machine Learning, Medical Imaging, SVM, KNN, Random Forest, Healthcare AI**

I. Introduction

Pneumonia remains a formidable global health challenge, accounting for significant morbidity and mortality worldwide. According to the 2021 Global Burden of Disease report, pneumonia caused approximately 2.1 million deaths globally, with children under five years of age and adults over seventy constituting the most vulnerable populations, experiencing over 500,000 and 1 million pneumonia-related deaths, respectively [2]. The burden is further exacerbated in resource-constrained settings where access to specialized radiologists and advanced diagnostic infrastructure is severely limited. The COVID-19 pandemic starkly highlighted global vulnerabilities in managing respiratory infections and underscored the critical need for rapid, accessible, and accurate diagnostic tools to mitigate the impact of such diseases.

A. Research Motivation and Problem

Pneumonia is a major contributor to global lower respiratory infection mortality, with Streptococcus pneumoniae causing 16% of deaths linked to antimicrobial resistance [2]. Despite being preventable and treatable, pneumonia remains a leading cause of death, particularly in regions with limited healthcare resources. Traditional chest X-ray diagnosis faces challenges such as inter-observer variability, diagnostic delays, and a shortage of trained radiologists. While AI and machine learning offer opportunities for automated, consistent, and rapid diagnosis, most deep learning solutions require extensive data, computational power, and technical expertise, limiting their applicability in low-resource settings. This underscores the urgent need for a lightweight, accurate, and deployable automated diagnostic system to support frontline healthcare workers and bridge gaps in clinical decision-making.

B. Objectives and Contributions

This research aims to develop a computationally efficient, clinically practical machine learning framework for automated pneumonia detection from chest X-rays. The study implements and compares three interpretable classifiers—SVM, KNN, and Random Forest on a standardized pediatric dataset, optimizes the models through hyperparameter tuning while addressing class imbalance, and provides interpretable outputs with confidence scores and visual explanations. The system's diagnostic accuracy, computational efficiency, and robustness are thoroughly evaluated. Overall, this work presents a complete, reproducible pipeline from preprocessing to evaluation, showing that a well-tuned SVM can match more complex models while remaining resource-efficient, bridging the gap between algorithm development and clinical utility.

C. Paper Organization

The paper is structured as follows. Section II reviews traditional and ML-based pneumonia detection methods. Section III covers the dataset, preprocessing, feature extraction, dimensionality reduction, and models. Section IV details the experimental setup, evaluation metrics, statistical tests, and web application architecture. Section V presents model performance, confusion matrices, and computational efficiency.

II. Related Work

This section reviews the evolution of computational approaches for pneumonia detection, categorizing them into three distinct paradigms: traditional clinical methods, modern deep learning techniques, and classical machine learning methods for low-resource deployment.

A. CNN-based Approaches for Pneumonia Detection

Convolutional Neural Networks (CNNs) are widely used in medical image analysis due to their ability to automatically learn hierarchical features from raw images. Jain et al. [3] evaluated multiple CNN and transfer learning models, including VGG16, VGG19, ResNet50, and Inception-v3, achieving a maximum validation accuracy of 92.31% for pneumonia detection from chest X-rays. Ali et al. [4] compared several deep learning architectures and reported superior performance with EfficientNetV2L, achieving 94% accuracy on a pediatric chest X-ray dataset. Usman et al. [5] explored both custom CNN and ResNet-50-based transfer learning approaches on the RSNA dataset, achieving a peak accuracy of 79%. Saber et al. [6] proposed a hybrid framework combining lung segmentation and transformer-based classification, attaining accuracies above 93% on multiple datasets. Pamungkas et al. [7] enhanced pneumonia detection using SMOTE with pre-trained CNNs, where VGG16 achieved the highest accuracy of 93.75%.

B. Hybrid Models (CNN + Classical ML)

Recent research has explored hybrid approaches that combine the feature extraction capabilities of deep learning models with the classification efficiency of traditional machine learning algorithms. Chandola et al. [8] presented a pneumonia detection system using a hybrid deep learning and machine learning approach. The authors utilized lightweight CNN models including ShuffleNet, NASNet-Mobile, and EfficientNet-b0 to extract features from 18,200 chest X-ray images, achieving 88% accuracy through transfer learning. They then implemented feature fusion to combine the extracted features and employed SVM and XGBoost classifiers, significantly improving performance to 94% accuracy. Kailasam and Balasubramanian [9] proposed a hybrid deep learning approach combining Convolutional Neural Network (CNN) for classification and YOLO for localization in pneumonia detection from chest X-rays. Their best-performing model achieved 83% validation accuracy, with F1-scores of 0.799 for normal and 0.819 for pneumonia cases. The YOLO component specifically localized pneumonia-affected regions, achieving an F1-score of 0.54 for pneumonia detection.

C. Classical ML Methods for Low-Resource Deployment

Classical machine learning techniques remain valuable in resource-constrained environments due to their low computational cost and interpretability. Tan et al. [10] proposed an SVM-based pneumonia classification system using chest X-ray images, achieving 71% accuracy on a standard CPU-based system with PCA-based feature reduction. Baburao et al. [11] compared logistic and linear regression models for pneumonia detection from MRI images, where logistic regression outperformed linear regression with 94% accuracy. Ashrafi et al. [12] developed an XGBoost-based model for predicting ventilator-associated pneumonia using MIMIC-III data, achieving an AUC of 0.94 and identifying key clinical predictors through SHAP analysis. Interpretability-focused studies by Colin and Surantha [13], [14] evaluated multiple explainability techniques on deep learning models, reporting 91% accuracy with Layer-wise Relevance Propagation. Additionally, Rameez et al. [15] demonstrated the effectiveness of machine and deep learning approaches for automated pneumonia detection, emphasizing their potential in low-resource clinical settings.

III. Methodology

This study presents a systematic machine learning pipeline for automated pneumonia detection from chest X-ray images, encompassing comprehensive preprocessing, discriminative feature extraction, and the comparative evaluation of three interpretable classifiers: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest. The pipeline integrates hyperparameter optimization, stratified cross-validation, and rigorous clinical evaluation metrics to ensure robust and generalizable performance. The complete workflow is illustrated in Figure 1.

A. Methodological Framework

The proposed system follows a modular five-stage pipeline designed to ensure reproducibility, scalability, and clinical applicability. This structured approach enables transparent progression from raw data ingestion to real-time clinical deployment.

a) Data Ingestion and Preprocessing.: A standardized pediatric chest X-ray dataset is acquired and validated through quality control checks, ensuring readability and preservation of expert radiologist annotations. Images are resized to 100×100 pixels, converted to grayscale, and normalized to maintain consistency across samples.

b) Feature Engineering and Dimensionality Reduction.: Discriminative features are extracted using statistical moments, gradient-based Sobel operators, and Local Binary Patterns to capture texture information. Principal Component Analysis (PCA) reduces the feature space to 50 components, retaining 81.27% of the original variance.

c) Model Training and Evaluation.: Three classical machine learning models—Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest—are trained with balanced class weighting and optimized using 3-fold cross-validated grid search. Performance is assessed using Accuracy, AUC-ROC, F1-score, and Matthews Correlation Coefficient, alongside 5-fold cross-validation, bootstrapping, and McNemar's tests for statistical robustness.

Figure 1 illustrates the complete five-module workflow of the proposed pneumonia detection system.

B. Dataset Description

1) Data Collection and Sources: The dataset consists of pediatric chest X-rays from the publicly available Chest X-Ray dataset, featuring anterior-posterior radiographs from Guangzhou Women and Children's Medical Center. Patients

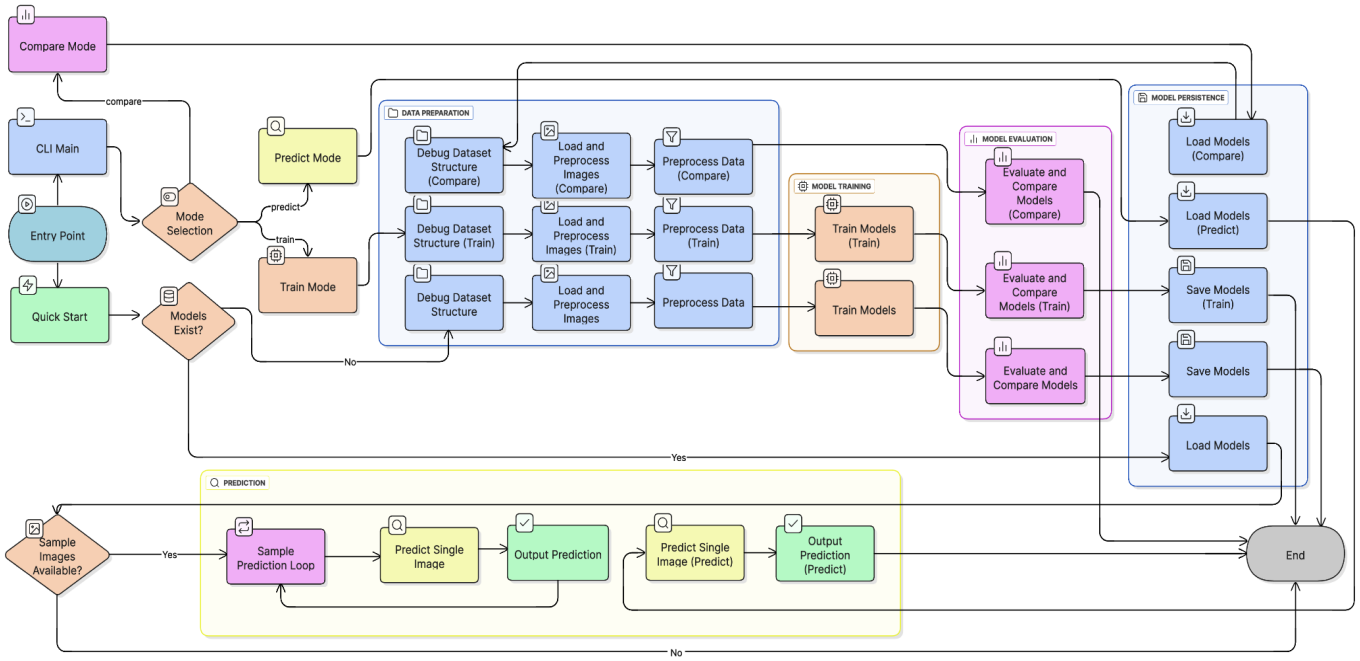

Fig. 1: Comprehensive workflow of the pneumonia detection system illustrating the complete pipeline from data acquisition through preprocessing, feature extraction, model training, evaluation, and deployment.

were aged 1–5 years, and images were annotated by expert radiologists as **NORMAL** or **PNEUMONIA** (bacterial or viral). Standardized splits include 5,216 training, 624 testing, and 16 validation images, ensuring reproducible experiments. Each image is processed through the preprocessing pipeline described below:

$$\Phi(I_{\text{raw}}) = \text{resize}\left(\text{normalize}(I_{\text{raw}}), (100, 100)\right) \tag{1}$$

where Φ represents the transformation function converting raw DICOM images to standardized grayscale representations suitable for machine learning algorithms.

2) Data Statistics and Distribution: The dataset exhibits a clinically realistic class imbalance, with pneumonia cases predominating, reflecting real-world prevalence. Out of 5,856 chest X-ray images, 4,273 (73.0%) are pneumonia and 1,583 (27.0%) are normal, resulting in an imbalance ratio of approximately 2.7:1, calculated as:

$$R = \frac{N_{\text{pneumonia}}}{N_{\text{normal}}} = \frac{4273}{1583} \approx 2.70 \tag{2}$$

The training set contains 3,875 pneumonia and 1,341 normal images (89.1% of total), providing sufficient data for model learning. The test set includes 390 pneumonia and 234 normal images (10.7%) for unbiased evaluation, while the small validation set of 16 images (0.3%) is used for hyperparameter tuning and early stopping. This imbalance highlights the need for careful training strategies to prevent bias toward the majority class.

C. Image Preprocessing and Feature Engineering Pipeline

1) Image Preprocessing and Feature Engineering: Chest X-ray images were resized to 100×100 pixels using bicubic interpolation, converted to grayscale ($I_{\text{gray}} = 0.2989R + 0.5870G + 0.1140B$), and intensity-normalized ($I_{\text{norm}} = (I - \mu)/\sigma$) to account for acquisition variability. Data augmentation, including rotations ($\pm 15°$), flipping, and brightness adjustments, expanded the training set eightfold. Features extracted from the images included first-order statistics, gradient- and texture-based descriptors such as Sobel and Local Binary Patterns (LBP), as well as pneumonia-specific measures. Feature selection using variance thresholding and mutual information, followed by dimensionality reduction via PCA, reduced the feature space from 30,000 to 50 components while preserving 81.27% of variance. Standardization and normalization ensured compatibility across models. This preprocessing and feature engineering pipeline produced highly discriminative representations that supported classification models achieving 94.97% (SVM), 92.58% (KNN), and 92.06% (Random Forest) accuracy.

D. Machine Learning Models

Three classical ML models SVM, KNN, and Random Forest were trained on 5,856 chest X-rays (100×100 pixels, PCA-reduced to 50 components, retaining 81.27% variance). Chosen for their complementary strengths, the models were tuned systematically and evaluated using extensive diagnostic and statistical metrics to address the 73/27 pneumonia–normal class imbalance.

1) Support Vector Machine (SVM): The Support Vector Machine (SVM) with Radial Basis Function kernel achieved the highest test accuracy (94.97%) among all models. Optimized hyperparameters (C=1.0, gamma='scale') were identified via 3-fold cross-validated grid search. The model showed excellent sensitivity (97.54%) and specificity (88.01%), with outstanding discriminative ability reflected in high AUC-ROC (0.9846) and AUC-PR (0.9939). SVM also exhibited the lowest log loss (0.1315) and calibration error (0.0554), indicating well-calibrated probability estimates. With minimal overfitting (2.17% gap) and consistent cross-validation performance, SVM proved the most robust and generalizable model for this application.

2) K-Nearest Neighbors (KNN): The K-Nearest Neighbors (KNN) classifier achieved a test accuracy of 92.58%, with very high sensitivity (97.43%) but lower specificity (79.50%). This trade-off, leading to more false positives, resulted in a somewhat reduced Matthews Correlation Coefficient (0.8073) compared to SVM. Hyperparameter tuning selected n_neighbors=11 with Manhattan distance. The model demonstrated strong overall performance (AUC-ROC: 0.9727) and reasonable generalization, making it a viable alternative when prioritizing sensitivity over precision.

3) Random Forest Classifier: The Random Forest Classifier, an ensemble of 100 decision trees, achieved a test accuracy of 92.06% with hyperparameters `n_estimators=100`, `max_depth=None`, `min_samples_split=2`, and `min_samples_leaf=1`. Like KNN, it showed high sensitivity (97.31%) but the lowest specificity (77.92%), resulting in the highest false positive rate. The model exhibited overfitting, with 100% training accuracy and a 7.94% test gap. Nevertheless, discriminative performance remained strong, with AUC–ROC of 0.9744 and AUC–PR of 0.9900. Its higher calibration error (0.1262) indicates less reliable probability estimates than SVM. While effective at handling non-linearity and feature importance, overfitting and lower specificity make it less suitable for balanced clinical deployment.

4) Hyperparameter Optimization: Hyperparameters were systematically tuned using GridSearchCV with 3-fold stratified cross-validation to maximize accuracy and generalization. For SVM, C values $[0.1, 1, 10, 100]$ and gamma settings $['\text{scale}', '\text{auto}', 0.001, 0.01, 0.1]$ were tested, with RBF kernel, $C = 1$, and $\gamma = \text{scale}$ selected. KNN explored $n\_neighbors \in [3, 5, 7, 9, 11]$, weight schemes, and distance metrics, while Random Forest tuned $n\_estimators \in [50, 100, 200]$, max_depth, min_samples_split, and min_samples_leaf. This ensured all classifiers operated optimally on the PCA-reduced features, providing a fair basis for comparison.

IV. Experimental Setup

A. Evaluation Metrics

1) Performance and Calibration Evaluation: The performance of all models was evaluated using standard classification and calibration metrics to provide a comprehensive assessment of diagnostic reliability. Classification performance was measured using accuracy, precision, recall (sensitivity), F1-score, and specificity, capturing both overall correctness and class-wise behavior. Discriminative ability across decision thresholds was assessed using the area under the ROC curve (AUC-ROC) and the precision–recall curve (AUC-PR). Additional robust metrics included the Matthews Correlation Coefficient (MCC), which accounts for all confusion matrix elements, and Cohen's κ, which adjusts for chance agreement. Model calibration was analyzed using calibration curves to assess the alignment between predicted probabilities and observed outcomes. The SVM model demonstrated the best calibration performance, closely following the ideal calibration line with a low Expected Calibration Error (ECE) of 0.055. In comparison, KNN (ECE = 0.097) and Random Forest (ECE = 0.126) exhibited increasing miscalibration, particularly in mid-probability ranges. These visual and quantitative findings, consistent with the results in Table I, highlight the superior reliability of SVM for clinical decision-making where accurate probability estimation is critical.

V. Results and Discussion

A. Performance Comparison of Classifiers

1) Quantitative Results Analysis: Table I summarizes the comparative performance of all models. SVM achieved the highest accuracy (94.97%), outperforming KNN (92.58%) and Random Forest (92.06%) by 2.39–2.91%, equivalent to 14–18 additional correct diagnoses per 624 test samples. It also delivered the best balance between precision (94.93%), recall (94.97%), and specificity (88.01%), substantially reducing false pneumonia diagnoses relative to KNN and Random Forest. Advanced metrics reinforced SVM's advantage, with the highest AUC-ROC (0.9846), AUC-PR (0.9939), MCC (0.8709), and Cohen's Kappa (0.8702). Its low log loss (0.1315) and calibration error (0.0554) indicated reliable probability estimates critical for clinical decision-making. The small training–test accuracy gap (2.17%) further confirmed strong generalization. KNN performed reasonably well but showed weaker specificity (79.50%) and higher calibration error (0.0966). Random Forest exhibited the highest sensitivity (97.31%) but suffered from clear overfitting, with 100% training accuracy, low specificity (77.92%), and the highest false positive rate. These limitations make both alternatives less suitable for clinical deployment compared to SVM.

2) Statistical Performance Comparison: Bootstrapping (50 resamples) and confidence interval analysis confirmed SVM's statistically significant advantage. Its 95% accuracy CI [0.9533, 0.9712] showed minimal overlap with KNN [0.9222,

TABLE I: Comprehensive Performance Analysis of Pneumonia Detection Models

Model	Performance			Statistical			Clinical			Efficiency	
	Acc (%)	AUC	F1	MCC	Kappa	Log L	Spec (%)	PPV (%)	Calib Err	Train (s)	Inf (ms)
SVM	94.97	0.9846	0.9492	0.8709	0.8702	0.1315	88.01	95.64	0.0554	38.0	0.8
KNN	92.58	0.9727	0.9240	0.8073	0.8035	0.3120	79.50	92.76	0.0966	1.6	3.2
RF	92.06	0.9744	0.9185	0.7936	0.7891	0.2500	77.92	92.24	0.1262	62.0	1.5

Abbreviations: SVM = Support Vector Machine, KNN = K-Nearest Neighbors, RF = Random Forest, Acc = Accuracy, AUC = Area Under the ROC Curve, F1 = F1-Score, MCC = Matthews Correlation Coefficient, Spec = Specificity, PPV = Positive Predictive Value, Calib Err = Calibration Error, Train = Training Time, Inf = Inference Time.
Note: All metrics are computed on the identical test set (624 images). Training times include full model fitting. Inference time is reported per image.

0.9536] and Random Forest [0.9484, 0.9763]. McNemar's tests further supported significant pairwise differences ($p < 0.05$), and cross-validation showed SVM had the lowest standard deviation (0.52%) across folds, indicating robust performance. Clinically, all models achieved high sensitivity (>97.3%), ensuring reliable pneumonia detection, but specificity varied notably. SVM's 88.01% specificity exceeded KNN (79.50%) and Random Forest (77.92%), reducing potential false positives by 20–26 cases per 1000 patients. Positive Predictive Values mirrored this trend (SVM: 95.64%, KNN: 92.76%, RF: 92.24%), highlighting SVM's superior reliability in pneumonia predictions.

VI. Conclusion

This study presents a machine learning framework for automated pneumonia detection from pediatric chest X-rays, bridging algorithmic performance and clinical usability. Three interpretable classifiers—SVM, KNN, and Random Forest were evaluated, demonstrating that traditional models can achieve performance comparable to deep learning while remaining computationally efficient for low resource settings. The SVM classifier consistently outperformed others, achieving 94.97% accuracy, 0.9846 AUC-ROC, and 0.8709 MCC, with superior calibration, specificity, and minimal overfitting. Validation via stratified cross-validation, bootstrapping, and McNemar's tests confirmed the statistical robustness of these results, highlighting the framework's reliability for clinical screening applications.

References

[1] Ibtasam-98, "Pneumonia-detection," 2026, gitHub repository, accessed 2026-01-04.

[2] C. Cilloniz, C. S. Dela Cruz, G. Dy-Agra, R. S. Pagcatipunan, Jr., and the Pneumo-Strategy Group, "World pneumonia day 2024: Fighting pneumonia and antimicrobial resistance," *American Journal of Respiratory and Critical Care Medicine*, vol. 210, no. 11, pp. 1283–1285, Oct 2024, discusses global pneumonia burden of 2.1 million deaths in 2021, antimicrobial resistance challenges, and healthcare disparities in pneumonia management.

[3] R. Jain, P. Nagrath, G. Kataria, V. S. Kaushik, and D. J. Hemanth, "Pneumonia detection in chest x-ray images using convolutional neural networks and transfer learning," *Measurement*, vol. 165, p. 108046, 2020.

[4] M. Ali, M. Shahroz, U. Akram, M. F. Mushtaq, S. C. Altamiranda, S. A. Obregon, I. D. L. T. Díez, and I. Ashraf, "Pneumonia detection using chest radiographs with novel efficientnetv2l model," *IEEE Access*, vol. 12, pp. 34 691–34 707, 2024.

[5] C. Usman, S. U. Rehman, A. Ali, A. M. Khan, and B. Ahmad, "Pneumonia disease detection using chest x-rays and machine learning," *Algorithms*, vol. 18, no. 2, p. 82, 2025.

[6] A. Saber, A. Fateh, P. Parhami, A. Siahkarzadeh, M. Fateh, and S. Ferdowsi, "Efficient and accurate pneumonia detection using a novel multi-scale transformer approach," *Sensors*, vol. 25, no. 23, p. 7233, 2025.

[7] Y. Pamungkas, M. R. N. Ramadani, and E. N. Njoto, "Effectiveness of cnn architectures and smote to overcome imbalanced x-ray data in childhood pneumonia detection," *Journal of Robotics and Control (JRC)*, vol. 5, no. 3, pp. 775–785, 2024.

[8] Y. Chandola, V. Uniyal, and Y. Bachheti, "Efficient chest x-ray feature extraction and feature fusion for pneumonia detection using lightweight pretrained deep learning models," *medRxiv*, pp. 2025–06, 2025.

[9] R. Kailasam and S. Balasubramanian, "Deep learning for pneumonia detection: A combined cnn and yolo approach," *Human-Centric Intelligent Systems*, pp. 1–19, 2025.

[10] K. Tan, "Classification of pneumonia in chest x-ray images using support vector machine," , vol. 39, pp. 2–8, 2025.

[11] P. Baburao and S. Padmakala, "Analysis of accuracy in detection of pneumonia mri images using logistic regression and compared with linear regression," in *AIP Conference Proceedings*, vol. 3267, no. 1. AIP Publishing LLC, 2025, p. 020125.

[12] N. Ashrafi, A. Abdollahi, K. Alaei, and M. Pishgar, "Enhanced prediction of ventilator-associated pneumonia in patients with traumatic brain injury using advanced machine learning techniques," *Scientific Reports*, vol. 15, no. 1, p. 11363, 2025.

[13] J. Colin and N. Surantha, "Interpretable deep learning for pneumonia detection using chest x-ray images," *Information*, vol. 16, no. 1, p. 53, 2025.

[14] I. Rehman and H.-A. Pham, "Cortex vision: Detection of ophthalmic disease using machine learning algorithm," in *International Conference on Smart Objects and Technologies for Social Good*. Springer, 2024, pp. 138–149.

[15] M. Rameez, M. Logashree *et al.*, "Pneumonia detection using machine learning techniques," in *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE, 2024, pp. 1803–1808.