Cureus Journal of Computer Science

Open Access Technical Report

Comparative Analysis of Machine Learning Techniques for Pneumonia Detection in Chest X-Ray Images

Received 06/30/2025
Review began 07/31/2025
Review ended 08/23/2025
Published 09/03/2025

© Copyright 2025
de França Albuquerque Trindade et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI:
https://doi.org/10.7759/s44389-025-08894-4

Leon de França Albuquerque Trindade [1] , Lucas Alves da Silva [1] , Manoel Leonardo Carneiro Sette Serrano [1] , Igor Gabriel Soares de Almeida Araújo [1] , Flávio Secco Fonsêca Prof. [1] , Vinícius Costa Amador [1]

1. Computer Sciences, Centro Universitário Tiradentes, Recife, BRA

Corresponding author: Vinícius Costa Amador, vinicius.amador@souunit.com.br

Abstract

Lower respiratory infections, such as pneumonia, represent a major global health problem, ranking as the world's most deadly communicable disease after COVID-19. Identifying the causative organism type (bacterial or viral) is critical, as treatments differ significantly, and bacterial pneumonia tends to develop more severe outcomes than viral cases. Imaging diagnosis is essential but relies on specialists' subjective analysis. This study aimed to develop an automated system for pneumonia identification and classification using chest X-rays to improve diagnostic accuracy and reduce analysis time. We employed the public dataset Chest X-Ray Images (Pneumonia) from Kaggle, containing 5,856 X-ray images, to train three models: Convolutional Neural Network, Random Forest, and Support Vector Machine. Preprocessing included resizing (224 × 224 pixels), normalization, and data augmentation for class balancing. The Convolutional Neural Network-based DenseNet121 model achieved the highest performance, with 81% accuracy and an area under the curve of 84.2% in multiclass classification. These results demonstrate that accessible artificial intelligence techniques can effectively support imaging diagnosis, offering clinical decision-making tools with improved speed and precision.

Categories: AI applications, Machine Learning (ML), User-interface design for medical devices and healthcare software
Keywords: pneumonia detection, artificial intelligence, x-ray, medical augmentation, human health

Introduction

According to the World Health Organization, lower respiratory tract infections are the deadliest communicable disease worldwide after COVID-19, ranking fifth among all causes of death and accounting for 2.5 million fatalities each year [1]. One of the core indicators for United Nations Sustainable Development Goal 3.2 (maternal and child health) is the proportion of children under five with pneumonia, underscoring how critical it is to combat the disease in this age group [2].

Bacterial pneumonia is most often caused by *Streptococcus pneumoniae*, the principal agent of community-acquired pneumonia, i.e., infections contracted outside the hospital [3]. Viral cases, in turn, are usually triggered by influenza A or B viruses. Bacterial infections tend to be more alarming: patients may develop more severe clinical pictures and recover more slowly, even when treated with antibiotics. Viral pneumonias typically respond faster to antiviral therapy, yet they can become life-threatening when caused by SARS-CoV-2, the virus responsible for COVID-19 [4]. Correctly identifying whether the pathogen is bacterial or viral is thus essential, both for choosing the right treatment and for anticipating the potential course of illness.

Clinical examination is combined with imaging studies for confirmation; chest radiography is favored because it delivers low radiation, is inexpensive, and is widely available. X-ray images reveal the extent and severity of infection and help clinicians monitor complications and therapeutic response. In more complex cases - when there is concern about central obstruction, cavitation, lymphadenopathy, or empyema - computed tomography (CT) provides additional clarity [5].

In this sense, a fair question remains as to whether machine learning and deep learning models, when trained solely on labeled X-ray images, are capable of accurately detecting pneumonia and distinguishing between its bacterial and viral forms based exclusively on radiographic features.

To accelerate diagnosis, machine-learning classifiers can automatically recognize imaging patterns and distinguish between the two etiologies. Diverse algorithms have been explored, including convolutional neural networks (CNN) [6], support vector machines (SVM) [7], and random forests (RF) [8]. Kaggle offers a rich pool of public datasets for training such models, attracting a community committed to innovative solutions. A notable resource is Chest X-Ray Images (Pneumonia), a curated repository of X-ray images labeled as normal, bacterial pneumonia, or viral pneumonia and validated by the scientific community [9]. Numerous studies have leveraged this and other datasets to train disease-detection models with

How to cite this article
de França Albuquerque Trindade L, Alves Da Silva L, Leonardo Carneiro Sette Serrano M, et al. (September 03, 2025) Comparative Analysis of Machine Learning Techniques for Pneumonia Detection in Chest X-Ray Images. Cureus J Comput Sci 2 : es44389-025-08894-4. DOI https://doi.org/10.7759/s44389-025-08894-4

encouraging results, while also highlighting the challenges of deploying them in real-world settings [10-12].

Thus, the objective of this study is to develop and evaluate classification models based on CNN, SVM, and RF algorithms, trained using the Chest X-Ray Images (Pneumonia) dataset, in order to accurately identify the presence of pneumonia and distinguish between its bacterial and viral etiologies. Additionally, the study aims to assess whether these models are capable of differentiating between distinct data subsets solely based on chest X-ray image patterns.

Technical Report

This study followed a retrospective observational design, focusing on the development of a predictive system based on supervised deep learning. The goal was to train an artificial intelligence model using labeled chest X-ray images (Figure *1*). The simplified step-by-step process of the work is illustrated in Figure *1*. Further methodological details are provided in the following sections.

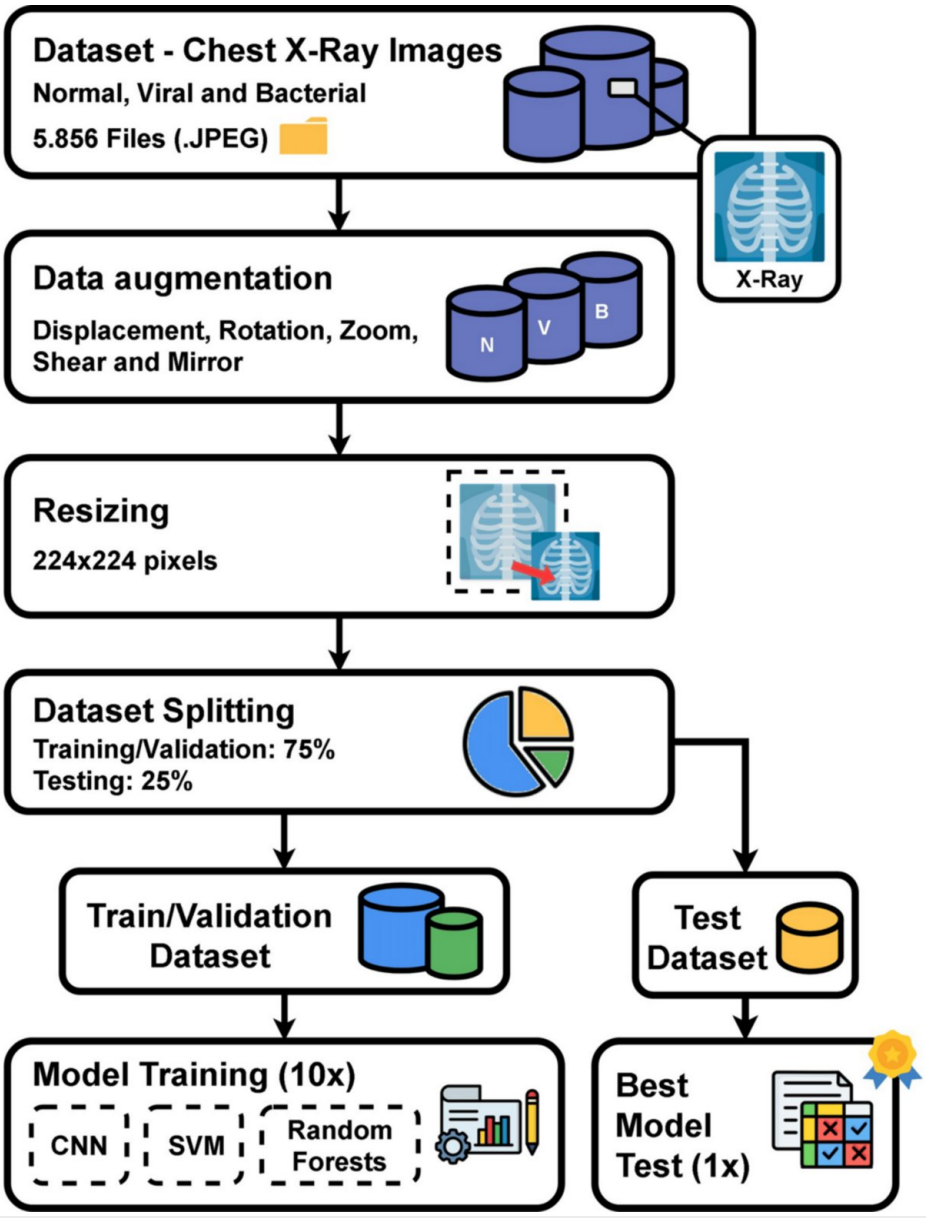

FIGURE 1: Workflow for chest X-ray image classification: data augmentation, resizing, dataset splitting, model training, and final testing with the best model

CNN, Convolutional Neural Network; SVM, Support Vector Machine

The project was implemented through a structured five-step workflow aligned with the standard data science lifecycle applied to artificial intelligence.

Dataset description and source

The dataset used in this study is a publicly available collection of human chest X-ray images, widely adopted in scientific research on pneumonia diagnosis. It was obtained from the Kaggle platform in October 2024 under a CC BY 4.0 license, permitting academic and research use.

Initially, the dataset included 5,856 JPEG-format images, divided into two main folders: NORMAL and PNEUMONIA. Although the PNEUMONIA folder contained both viral and bacterial cases, these were not in separate subfolders. The classification was inferred from the filenames, which included the keywords "virus" or "bacteria." The folder structure was divided into three subsets: training, validation, and testing (Figure 2).

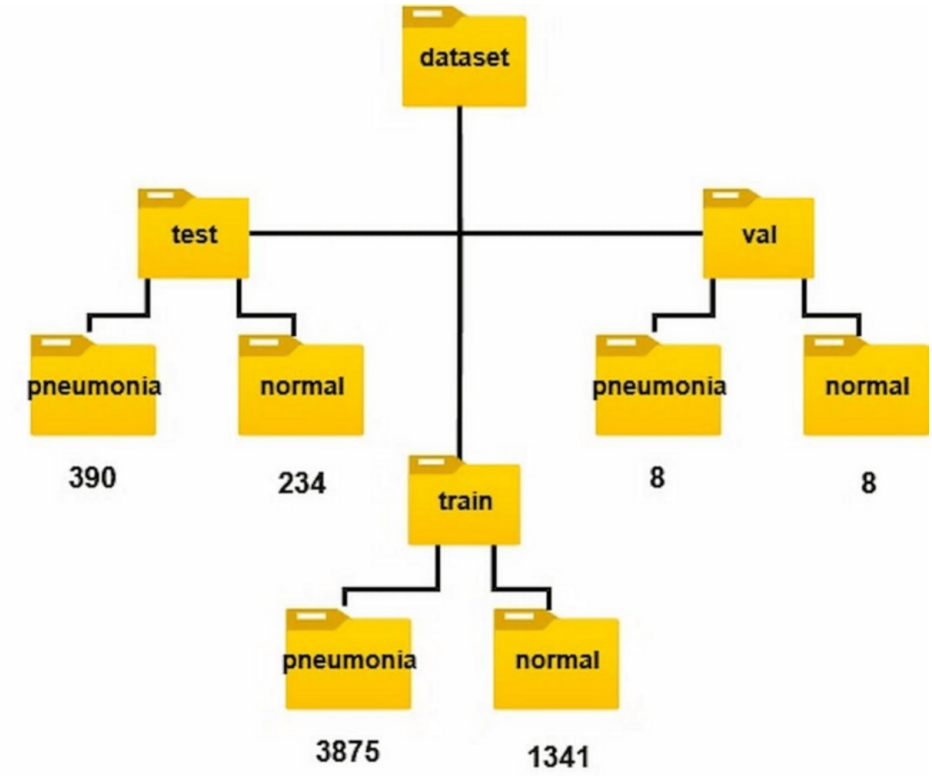

FIGURE 2: Dataset directory structure showing division into training, validation, and test sets, organized by class (normal and viral)

The images had varied resolutions and included unique identifiers in their filenames for tracking and quality control.

Preprocessing

Due to class imbalance - specifically, an empty validation folder for viral images and an underpopulated test set - adjustments were made to improve dataset representativeness and robustness. A Python script redistributed images across the validation and test sets, classifying them as normal, bacterial, or viral based on their filenames.

To mitigate class imbalance in the training set (where bacterial images outnumbered others), real-time data augmentation was applied using Keras's ImageDataGenerator. Transformations included random rotations (15°), horizontal/vertical shifts (10%), shearing (0.1), zooming (10%), and horizontal flips, with nearest-neighbor pixel fill. This enhanced model robustness by simulating clinical variability and reducing overfitting risk.

The reorganized dataset enabled more effective training strategies, such as applying the class_weight parameter to penalize misclassifications proportionally. The goal was to improve predictive performance, particularly for the underrepresented viral class, and ensure statistically representative and balanced

validation and test sets.

To align with ImageNet-pretrained weights used in the Transfer Learning approach, all images were resized to 224 × 224 pixels (the input size for DenseNet121) and normalized range.

After completing the preprocessing step on the original Chest X-Ray Images (Pneumonia) dataset [10], a new custom dataset was generated. This new dataset was subsequently used for training the three models. The data was organized into separate folders for training, validation, and testing, each containing subfolders corresponding to the classes bacterial, normal, and viral (Figure *3*). After redistribution, each validation and test subset contained 250 images per class (750 total per set) (Figure *3*).

Training was performed using the images from the train folder (4,356 images), with a batch size of 32, for up to 10 epochs, employing early stopping based on validation accuracy with a patience of five epochs. On-the-fly data augmentation and dynamic class weighting were applied to address the class imbalance, particularly in the viral category.

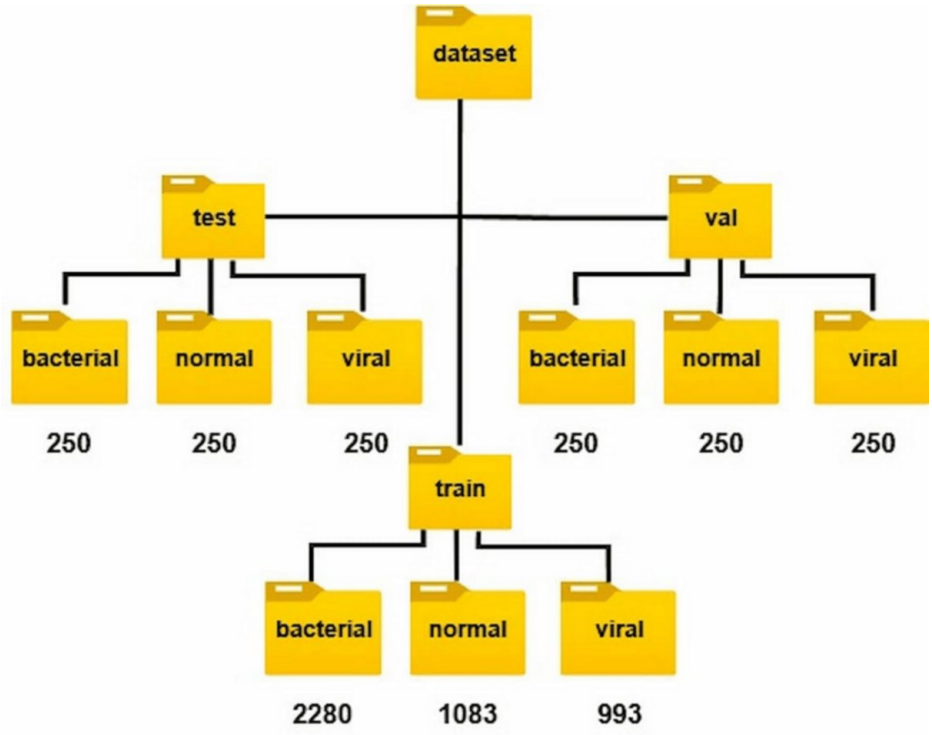

FIGURE 3: Final dataset structure with equal class distribution

Figure *4* illustrates a visual comparison between an original chest X-ray image and its augmented version, highlighting the transformations applied during the training phase.

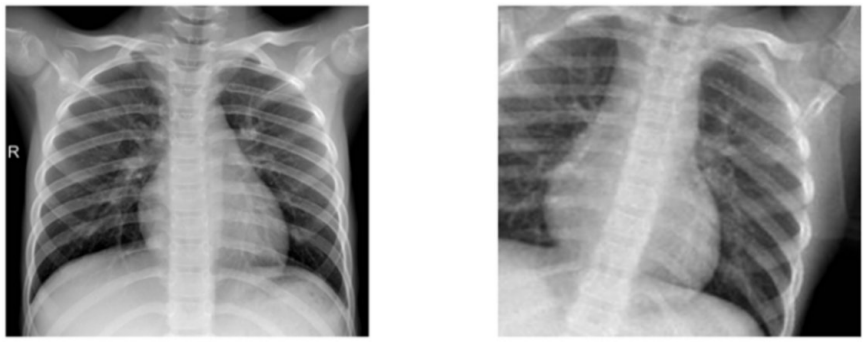

FIGURE 4: Image A: original chest X-ray (1,940 × 1,761). Image B: resized (224 × 224) with augmentation (rotation, shifting)

Model and architecture

For multiclass classification (bacterial pneumonia, viral pneumonia, and normal cases), we adopted DenseNet121, a deep convolutional neural network known for its dense connectivity that enhances gradient flow and feature reuse. Transfer Learning with ImageNet weights was used, replacing the original output layer with a custom classifier comprising a GlobalAveragePooling2D layer, a dense layer with 256 ReLU-activated units, and a softmax-activated output layer with three neurons. The model was compiled using the Adam optimizer (learning rate: 0.0001), categorical cross-entropy loss, and accuracy as the evaluation metric. Training ran for up to 10 epochs with a batch size of 32, using early stopping after five validation epochs without improvement. To address data imbalance, we applied dynamically adjusted class weights and real-time data augmentation. Images were resized and normalized as previously described. The final CNN model was saved in .h5 format for potential web integration.

The RF model leveraged CNN-extracted features combined with real-time augmentation. Different tree depths and class weights were tested, ultimately selecting 200 trees with a depth of 10 and class weights: normal = 1, bacterial = 1, and viral = 3. Ten-fold cross-validation with shuffling was applied to reduce overfitting and ensure robustness.

SVM models were also explored as multiclass classifiers. Two kernels were tested: Radial Basis Function (RBF) and Polynomial. Input data were preprocessed using traditional computer vision techniques: grayscale conversion, resizing (224 × 224), flattening, and normalization via StandardScaler.

To handle imbalance, we used class_weight='balanced'. The RBF kernel was chosen for its ability to map data into higher-dimensional spaces, while the Polynomial kernel offered an alternative for capturing complex relationships. Models were evaluated using the same training and test sets, with default hyperparameters (C=1.0, gamma='scale'). This simpler approach proved efficient and fast, making it suitable for limited-resource environments or cases needing more interpretable models.

Performance evaluation

Model performance was assessed using standard classification metrics, where true positives (TP) refer to cases correctly predicted as belonging to the positive class, true negatives (TN) are cases correctly predicted as belonging to the negative class, false positives (FP) are cases incorrectly predicted as positive when they are actually negative, and false negatives (FN) are cases incorrectly predicted as negative when they are actually positive (Table *1*).

Metrics	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-Score	$\frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$

TABLE 1: Performance metrics and corresponding formulas

TP, True Positives; TN, True Negatives; FP, False Positives; FN, False Negatives

Results

Training and Validation Results

After the methodological procedures were applied, the training phase produced distinct performance outcomes across the evaluated models. The CNN, based on the DenseNet121 architecture, achieved the highest overall performance, with an accuracy of 80.30%, sensitivity of 79.53%, specificity of 89.31%, and an area under the receiver operating characteristic curve (AUC) of 91.85%. The RF model, configured with 200 trees, reached an accuracy of 75.13%, sensitivity of 73.24%, specificity of 86.33%, and an AUC of 87.69%. The SVM with an RBF kernel obtained results comparable to the CNN, with an accuracy of 80.21%, sensitivity of 80.13%, and specificity of 90.10%. Lastly, the SVM with a Polynomial kernel demonstrated slightly lower performance, with an accuracy of 74.61%, sensitivity of 74.36%, and specificity of 87.25%, as shown in Table *2*.

Model	Accuracy	Sensitivity	Specificity	AUC
RF-100	0.750 ± 0.001	0.731 ± 0.001	0.863 ± 0.001	0.875 ± 0.001
RF-150	0.750 ± 0.001	0.731 ± 0.001	0.863 ± 0.001	0.876 ± 0.001
RF-200	0.751 ± 0.001	0.732 ± 0.001	0.863 ± 0.001	0.877 ± 0.001
CNN	0.8030 ± 0.001	0.7953 ± 0.001	0.8931 ± 0.001	0.9185 ± 0.001
SVM-POLY	0.7461 ± 0.001	0.7436 ± 0.001	0.8725 ± 0.001	0.8080 ± 0.001
SVM-RBF	0.8021 ± 0.001	0.8013 ± 0.001	0.9010 ± 0.001	0.8510 ± 0.001

TABLE 2: Comparative performance of machine learning models

AUC, Area Under the Curve; CNN, Convolutional Neural Network; POLY, Polynomial; RBF, Radial Basis Function; RF, Random Forest; SVM, Support Vector Machine

Best Model Results

The CNN with DenseNet121 architecture achieved the best overall performance. Table *3* presents its class-specific results.

Class	Accuracy (%)	Sensitivity	Specificity	AUC
Bacterial	82	0.89	0.88	0.91
Normal	82	0.82	0.91	0.92
Viral	82	0.78	0.92	0.86
Mean	82	0.83	0.903	0.89

TABLE 3: Class-wise performance metrics for the CNN model

AUC, Area Under the Curve; CNN, Convolutional Neural Network

The simulation used 250 samples per class, with true labels represented by the rows and predicted labels by the columns. The BACTERIAL class had 183 correct classifications, NORMAL had 205, and VIRAL had 176. The diagonal cells indicate correct predictions, while the off-diagonal values represent misclassifications between the classes. Figure *5* shows the confusion matrix generated based on the CNN model's performance, assuming an accuracy of 82% for each class.

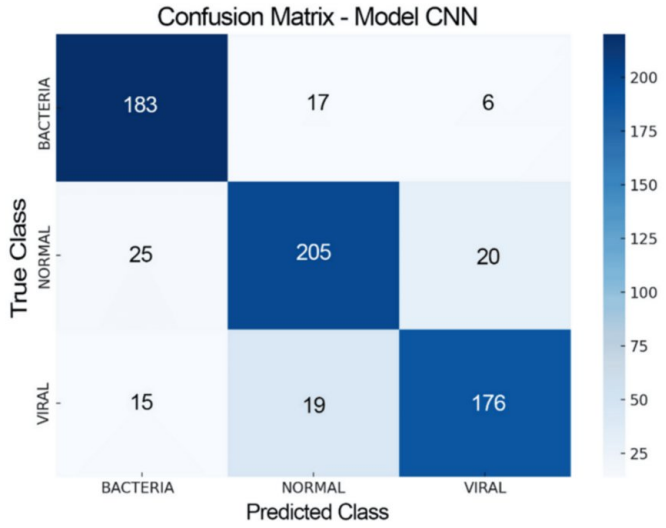

FIGURE 5: Confusion matrix – CNN model

CNN, Convolutional Neural Network

Discussion

Upon evaluating the results presented in Table *3*, it is evident that the CNN model based on the DenseNet121 architecture demonstrated solid performance when addressing the multiclass classification challenge in a scenario characterized by class imbalance. This performance aligns with existing literature, where DenseNet architectures, such as DenseNet121, are widely recognized for their effectiveness in medical image analysis by promoting feature reuse and gradient propagation. For instance, Bundea and Danciu [13] also employed DenseNet121, DenseNet169, and DenseNet201 on a pneumonia X-ray dataset from Kaggle (5,856 images), achieving a consistent accuracy of 96% with DenseNet121 variants. Similarly, Dalei and Mahapatra [14] utilized a Modified DenseNet121 in an ensemble model, reaching an accuracy of 96.4% on unseen data.

Among the three evaluated classes, the BACTERIAL class achieved the best results, which may be attributed to the larger number of training samples available for that category. The NORMAL class maintained strong performance, demonstrating stability across the assessed metrics. Although the VIRAL class still presented the lowest relative values, it showed significant improvement compared to previous approaches that did not employ balancing strategies, thus reinforcing the effectiveness of the techniques adopted in this study.

Another important aspect was the use of 10-fold cross-validation, which added greater statistical robustness to the analysis. This method enabled a more realistic assessment of the model's average performance, minimizing overfitting effects and ensuring greater reliability for future applications.

Finally, the high AUC score obtained further supports the capacity of the DenseNet121 architecture to correctly separate the classes, which is particularly relevant in clinical settings where accuracy and reduction of false negatives have a direct impact on medical decision-making and patient outcomes.

This study also assessed the performance of the RF algorithm using different configurations of the n_estimators hyperparameter, corresponding to 100, 150, and 200 trees. Analyses were based on accuracy, area under the ROC curve (AUC), sensitivity, and specificity metrics, both on average and by class: bacterial, normal, and viral.

The model with 100 trees achieved an accuracy of 75.05%, an AUC of 87.52%, an average sensitivity of 73.11%, and an average specificity of 86.29%. Compared to the model with 200 trees, the overall gains were modest. Accuracy increased by only 0.11%, AUC improved by 0.19%, and the averages for sensitivity and specificity rose by 0.18% and 0.05%, respectively. Although these improvements are positive, they remain below the 1% threshold, suggesting that the model reaches a stable performance level with the initial 100-tree configuration.

Regarding the evaluated classes (normal, bacterial, viral), the sensitivity for the bacterial class remained stable, with a slight decrease of 0.10%, dropping from 79.88% to 79.80%. For the normal class, there was a minor increase in sensitivity from 86.42% to 86.50%, while specificity rose from 93.88% to 93.96%. The most notable variation was observed in the viral class, where sensitivity increased from 53.01% to 53.42%, representing a 0.77% rise. Nevertheless, this value remains lower than those of the other classes, highlighting that detecting viral cases continues to be the model's main challenge. Specificity for this class remained unchanged at 89.02%.

In addition to CNN and tree-based algorithms, this study evaluated two configurations of the SVM algorithm, varying the type of kernel function employed. The analysis was based on average accuracy, sensitivity, and specificity metrics obtained through cross-validation.

The SVM model using the RBF kernel exhibited superior performance, with an average accuracy of 80.21%, an average sensitivity of 80.13%, and an average specificity of 90.10%. These results indicate the model's strong ability to correctly distinguish between classes, with particular emphasis on its high specificity, suggesting a lower false-positive rate. Such performance highlights the suitability of the RBF kernel in classification contexts involving complex and non-linear patterns, as observed in the analyzed images.

On the other hand, the SVM model with a polynomial kernel achieved an average accuracy of 74.61%, an average sensitivity of 74.36%, and an average specificity of 87.25%. Although these results were slightly lower than those of the RBF kernel model, the performance remains competitive, indicating that the polynomial kernel is still a viable alternative in certain data configurations.

Conclusions

The primary objective of this study was to develop and compare machine learning and deep learning models for the automatic classification of pneumonia in chest X-rays, with a focus on distinguishing between viral, bacterial, and normal cases. The results demonstrate significant advancements, particularly with the use of CNN combined with Transfer Learning. Among the models evaluated - CNN, RF, and SVM - the CNN-based architecture achieved the best performance, with an average accuracy of 81% and an AUC of 84.2%, maintaining consistent metrics across all three classes.

These findings highlight the potential of CNN-based models to be integrated into medical imaging workflows in clinical settings, contributing to improved access to early diagnosis of respiratory diseases. It is possible that, with access to more comprehensive and diverse data, as well as richer contextual information, the performance of these models could be further enhanced.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Vinícius Costa Amador, Leon de França Albuquerque Trindade, Lucas Alves da Silva, Manoel Leonardo Carneiro Sette Serrano, Igor Gabriel Soares de Almeida Araújo, Flávio Secco Fonsêca Prof.

Acquisition, analysis, or interpretation of data: Vinícius Costa Amador, Leon de França Albuquerque Trindade, Lucas Alves da Silva, Manoel Leonardo Carneiro Sette Serrano, Igor Gabriel Soares de Almeida Araújo, Flávio Secco Fonsêca Prof.

Drafting of the manuscript: Vinícius Costa Amador, Leon de França Albuquerque Trindade, Lucas Alves da Silva, Manoel Leonardo Carneiro Sette Serrano, Igor Gabriel Soares de Almeida Araújo, Flávio Secco Fonsêca Prof.

Critical review of the manuscript for important intellectual content: Vinícius Costa Amador, Leon de França Albuquerque Trindade, Lucas Alves da Silva, Manoel Leonardo Carneiro Sette Serrano, Igor Gabriel Soares de Almeida Araújo, Flávio Secco Fonsêca Prof.

Supervision: Vinícius Costa Amador, Flávio Secco Fonsêca Prof.

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. World Health Organization (WHO): The top 10 causes of death. (2024). Accessed: June 30, 2025: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.
2. World Health Organization (WHO): SDG Target 3.2: End preventable deaths of newborns and children under 5 years of age. (2025). Accessed: June 30, 2025: https://www.who.int/data/gho/data/themes/topics/sdg-target-3_2-newborn-and-child-mortality.
3. Jain S, Self WH, Wunderink RG, et al.: Community-acquired pneumonia requiring hospitalization among U.S. adults. New England Journal of Medicine. 2015, 373:415-427. 10.1056/nejmoa1500245
4. Musher DM, Thorner AR: Community-acquired pneumonia. New England Journal of Medicine. 2014, 371:1619-1628. 10.1056/nejmra1312885
5. Katz DS, Leung AN: Radiology of pneumonia. Clinics in Chest Medicine. 1999, 20:549-562. 10.1016/s0272-5231(05)70235-5
6. Breiman L: Random forests. Machine Learning. 2001, 45:5-32. 10.1023/a:1010933404324
7. Cortes C, Vapnik V: Support-vector networks. Machine learning. 1995, 20:273-297. 10.1007/bf00994018
8. LeCun Y, Bottou L, Bengio Y, Haffner P: Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998, 86:2278-2324. 10.1109/5.726791
9. Kermany D, Zhang K, Goldbaum M: Labeled optical coherence tomography (OCT) and chest X-ray images for classification. Mendeley Data. 2018, 10.17632/rscbjbr9sj.2
10. Kaggle: Chest X-Ray Images (Pneumonia). (2018). Accessed: June 30, 2025: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/data.
11. Kareem A, Liu H, Sant P: Review on pneumonia image detection: A machine learning approach. Human-Centric Intelligent Systems. 2022, 2:31-43. 10.1007/s44230-022-00002-2
12. Szepesi P, Szilágyi L: Detection of pneumonia using convolutional neural networks and deep learning. Biocybernetics and Biomedical Engineering. 2022, 42:1012-1022. 10.1016/j.bbe.2022.08.001
13. Bundea M, Danciu GM: Pneumonia image classification using DenseNet architecture. Information. 2024, 15:611. 10.3390/info15100611
14. Dalei MK, Mahapatra S: Pneumonia detection and classification from chest X-ray images using Modified DenseNet121 Deep Learning Approach. 2025 International Conference on Ambient Intelligence in Health Care (ICAIHC), Raipur Chattisgarh, India. 2025, 1-5. 10.1109/ICAIHC64101.2025.10956683