 Received 15-05-2025; Accepted 15-07-2025; Published 22-07-2025

Pneumonia Detection System Using Convolutional Neural Networks

Renzo Zavaleta renzo.zavaleta@usil.pe
Faculty of Engineering, San Ignacio de Loyola University,
Lima - Peru.

Eduardo Bautista eduardo.bautistaa@usil.pe
Faculty of Engineering, San Ignacio de Loyola University,
Lima - Peru.

Luis Peña luis.penac@usil.pe
Faculty of Engineering, San Ignacio de Loyola University,
Lima - Peru.

Claudio Bances claudio.bances@usil.pe
Faculty of Engineering, San Ignacio de Loyola University,
Lima - Peru.

Luis Salazar luis.salazarma@epg.usil.pe
Faculty of Engineering, San Ignacio de Loyola University,
Lima - Peru.

Corresponding Author: Luis Salazar

Abstract

This study analyzes the performance of four pre-trained convolutional neural networks (CNNs) (MobileNetV2, ResNet-50, DenseNet121 and DenseNet201) in classifying chest radiographs to identify bacterial pneumonia, viral pneumonia or normal cases. MobileNetV2, the most efficient and accurate model, achieved an overall accuracy of 85%, excelling in the identification of normal cases with 99% accuracy. This model was integrated into a Django-based web system, which allows physicians to upload radiographs, obtain automated diagnoses and visualize Grad-CAM heat maps for interpretation. The development followed the Scrum methodology, ensuring iterative progress and continuous improvement. The system aims to enhance diagnostic accuracy and accessibility, especially in resource-limited settings. Although MobileNetV2 showed good results, its sensitivity for detecting bacterial pneumonia could be improved, suggesting that future improvements could be achieved with advanced data augmentation techniques and more extensive validation of the dataset. This work highlights the potential of lightweight CNNs in medical diagnostics and presents an efficient and scalable solution for early detection of pneumonia.

Keywords: Deep learning, Chest X-ray, CNN, Multiclass classification, Pneumonia, MobileNetV2, DenseNet201, DenseNet121, ResNet50.

Citation: Luis Salazar, et al. Pneumonia Detection System Using Convolutional Neural Networks. Journal of Artificial Intelligence and Autonomous Intelligence. Research. 2025;2(2):18.
Pages:264-292

1. INTRODUCTION

Chest X-rays are crucial for the diagnosis of various respiratory diseases, such as pneumonia, which is responsible for numerous deaths worldwide, especially in children and the elderly [1]. Early and accurate detection of this pathology is crucial to improve clinical outcomes. In this context, convolutional neural networks (CNNs), a subcategory of artificial intelligence (AI) specialized in image processing, have revolutionized the field of automated medical image analysis, allowing efficient and accurate diagnosis [2]. CNNs like MobileNetV2, known for their lightweight architecture and computational efficiency, have proven to be a promising solution in the automatic classification of chest X-rays [3].

Recently, the application of pre-trained CNN models has evolved in the field of pneumonia diagnosis, with architectures such as MobileNetV2, ResNet-50, DenseNet201, and DenseNet121 standing out, each with specific features and optimizations. MobileNetV2, for example, is designed to operate efficiently on devices with hardware limitations, while ResNet-50 and DenseNet (in its versions 121 and 201) offer a complex feature extraction capability that can improve accuracy in medical classification tasks [3]. Not all these architectures have the same performance and efficiency in every medical application, which motivates the comparison between them in the context of pneumonia detection [4].

Some studies in the literature agree on the methods of applying CNN in pneumonia detection. Rajpurkar et al. [1], and Nikhade et al. [2], have implemented deep networks for lung disease diagnosis with high accuracy. Similarly, Stephen et al. [3], and Ayan and Ünver [4], have pointed out that even lighter models such as MobileNetV2 can be deployed on mobile devices without a significant drop in diagnostic performance. Liang and Zheng [5], also analyzed the effect of improving image resolution to obtain better classification results. However, the approaches of these authors vary significantly. Rajpurkar et al. [1], were interested in using model properties such as DenseNet to maximize accuracy, while Ayan and Ünver [4], prioritized efficiency on mobile devices and opted for lighter models. The works of Nikhade et al. [2], used large-scale, varied datasets, while the studies of Liang and Zheng [5], focused on improving image quality to increase diagnostic accuracy. Stephen et al. [3], have studied the impact of various hyperparameter optimization techniques to improve the generalization ability of models.

This work aims to compare different pre-trained convolutional neural network architectures (MobileNetV2, ResNet-50, DenseNet121 and DenseNet201) to classify chest X-rays into three categories: bacterial pneumonia, viral pneumonia and normal cases. The research identifies MobileNetV2 as the optimal model in terms of accuracy and computational efficiency, adapting it for implementation in an intelligent system designed for hardware-constrained environments. This web-based system, developed in Django (frontend and backend in Python), allows pulmonologists to register, manage patients and upload new X-rays to obtain automatic and fast diagnoses. Each prediction is accompanied by a heat map generated using Grad-CAM, which visually highlights the areas of interest in the image, providing an additional interpretation for the specialist. This approach not only optimizes diagnostic accuracy and efficiency in areas with fewer technological resources, but also contributes to the early detection of pneumonia, facilitating access to advanced diagnostic tools in healthcare settings globally.

2. THEORETICAL FRAMEWORK

Convolutional neural networks are used in pneumonia detection for their ability to extract complex features from images. enables highly accurate and automated diagnosis that previously required manual intervention [6].

2.1 MobileNet Algorithm

It is a convolutional neural network designed to be efficient on mobile devices and computer vision tasks. It introduces residual inversion blocks, which combine depthwise and pointwise convolutions to reduce computational complexity. It also uses ResNet-style skip connections, which facilitates the training of deep networks [7]. Furthermore, it is designed to be lightweight and scalable, allowing its size and complexity to be adjusted using a width parameter, making it suitable for resource-limited environments, offering a good balance between accuracy, speed and energy consumption [8].

Below is a table showing some of the applications with MobileNet:

Table 1: Literature on MobileNet applications

Application	Dataset	Findings	Author
Classification of skin diseases	PH2, HAM10000, DermNet, ISIC	MobileNetV2 with attention and ASPP modules achieves 98.6% accuracy in classifying skin diseases.	[7]
Classification of remote sensing images	54, 306 images	MobileNetV2 pre-trained on ImageNet, together with dehazing and transfer learning, improves the accuracy of remote sensing image classification.	[8]
Classification of bird species in wetlands	ImageNet (14M images)	MobileNetV2 achieves an F1-score of 0.789, outperforming VGG16 and approaching ResNet50, with a smaller model suitable for resource-constrained devices.	[9]
Face and landmark detection	32, 203 images	MobileNetV2 with feature pyramids and context modules exceeds 90% accuracy on WIDER FACE, proving its effectiveness under challenging conditions.	[10]

Source: The authors

TABLE 1 shows that MobileNetV2, combined with advanced techniques such as attention and transfer learning, significantly improves accuracy in various applications such as disease classification, remote sensing images, bird species, and face detection.

2.2 ResNet Algorithm

ResNet, also known as Residual Network, is a deep learning framework for the task of image recognition. The core idea of ResNet is to recast the layers of a neural network as residual functions, using “shortcut connections” that allow the model to learn the difference between the desired output and the current output of each layer. This deepening capability increases the accuracy of the model by capturing more complex details in images, which is valuable in pneumonia diagnosis [11]. These deep networks are effective on large volumes of data due to their high number of parameters and complex structure, making it easier for them to detect advanced features and patterns in the data [12].

Below, TABLE 2 will show some of the applications with ResNet:

Table 2: Literature on applications with ResNet

Application	Dataset	Find	Author
Traditional Chinese Medicine	48 people	It allows to distinguish and recognize the difference between a Qi-deficient constitution and a balanced one.	[13]
Brain age prediction	6400 images	Experimental results indicate that the proposed architecture demonstrates versatility and robustness, achieving high accuracy and a lower number of parameters.	[12]
Detection and classification of eyelid tumors	728 images	Improves efficiency and accuracy in the segmentation of pathological tumors, reaching accuracies of up to 96.8%.	[14]
Segmentation and classification of glioblastoma tumors	372 images	The ResNet-SVM method achieved 89.36% accuracy, 92.52% specificity, and 90.12% precision in classifying glioblastoma tumors, outperforming current methods.	[15]

Source: The authors

2.3 DenseNet121 Algorithm

DenseNet121 employs a dense connectivity architecture that optimizes information propagation in deep networks. Each layer receives the output of all previous layers, facilitating a robust gradient flow and reducing feature redundancy, allowing the model to extract detailed information from medical images. DenseNet121, being more compact compared to DenseNet201, is particularly suitable for applications in clinical settings where the balance between accuracy and computational efficiency is important [16]. Although it is also commonly used in computer vision applications, such as medical image classification, due to its high performance and ability to work with fewer parameters than other traditional deep networks [17].

Below is TABLE 3, showing some of the applications of DenseNet121:

Table 3: Literature on applications with DenseNet121

Application	Dataset	Find	Author
Hyoid bone sex prediction	3990 images	The DenseNet121 deep learning model achieved 89% accuracy in predicting sex from the hyoid bone.	[17]
Diagnosing COVID-19 disease based on gravitational search optimization	5811 images	GSA-DenseNet121-COVID-19 achieved 98.38% accuracy in classifying chest X-ray images for COVID-19 detection.	[18]
Social media review search and product recommendation scheme	2033 reviews	A 92.22% accuracy was achieved in recommending relevant products on a social network	[19]
Breast cancer classification from mammography images	7909 images	The DenseNet121+elm model achieved an accuracy of 99.47% and 99.14% in training and testing accuracy, respectively.	[20, 21]

Source: The authors

2.4 DenseNet201 Algorithm

With a deeper architecture than DenseNet121, DenseNet201 offers greater feature extraction capabilities by including more layers in its dense connectivity structure. This increased depth enables capturing complex patterns and fine details in images, which is beneficial in diagnostic tasks that require high sensitivity, such as pneumonia detection in X-rays. DenseNet201, although more computationally demanding, is highly effective in applications where diagnostic accuracy is a priority [16].

Below is TABLE 4, showing some of the applications of DenseNet201:

2.5 Web Applications and Heat Maps

The integration of convolutional neural networks (CNN) into web applications for pneumonia diagnosis has proven to be an effective tool to improve access and accuracy in the detection of this pathology. Recently, a CNN-based model implemented in a web application was developed that allows the upload of X-ray images and returns an automated diagnosis of pneumonia, achieving significant accuracy and facilitating access to rapid diagnosis in areas with limitations of specialized resources [22]. This type of platform not only speeds up the diagnostic process but also allows medical professionals to access detailed image analysis without the need for advanced equipment at each location [23].

Another study implemented an automated diagnostic system using a pre-trained CNN model and presenting the results with visual support using Grad-CAM heatmaps. This method highlights the relevant areas of the radiograph that the model focuses on for diagnosis, providing the user with

Table 4: Literature on applications with DenseNet201

Application	Dataset	Find	Author
Classification of microstructure images based on the division of non-fixed size patches	5080 images	The NFSDense201 model achieved 99.53% accuracy for four-class classification and 97.09% for ten-class classification on SEM images, outperforming previous models.	[24]
Classification of dermatoscopy images	25331 images	It showed good performance in classifying skin lesions, and its classification ability was further improved by integrating synthetic images generated by the SLA-StyleGAN model proposed in the study.	[25]
Differentiation of breast lesions in dynamic contrast-enhanced MRI	4260 images	The S2 strategy improves the robustness of the DenseNet201 model on small sets of breast DCE-MRI, optimizing the accuracy in discriminating between benign and malignant lesions.	[26]
Computer-assisted diagnosis of laryngeal cancer	13721 images	The sensitivity of the model was 73.1% and the specificity was 92.2% for detecting laryngeal cancer and laryngeal precancerous lesions.	[27]

Source: The authors

a visual interpretation that helps validate the accuracy of the prediction. This approach reinforces user trust in the system by offering greater transparency and understanding of the diagnostic process, which is crucial in clinical contexts [28]. The choice of CNN is due to its ability to process complex images and detect patterns that would otherwise require extensive review by specialists. Furthermore, the use of CNN allows for efficient automation in diagnosis, which is beneficial in resource-limited healthcare contexts [23].

Recent advances in medical imaging have incorporated lightweight and transformer-based models with notable success. For instance, EfficientNet and ViT (Vision Transformers) have been used to classify chest X-rays and CT scans with high accuracy and reduced computational cost. These models, when combined with attention mechanisms or hybrid encoders, show strong potential to improve diagnostic performance in detecting pneumonia and other thoracic diseases. Such models represent a promising alternative to traditional CNNs in future research [27].

3. METHODOLOGY

3.1 Design Description

The model developed for pneumonia detection is based on the architecture of a pre-trained model, which acts as a convolutional base and is complemented by a classifier optimized for the task of classifying X-ray images (see FIGURE 1).

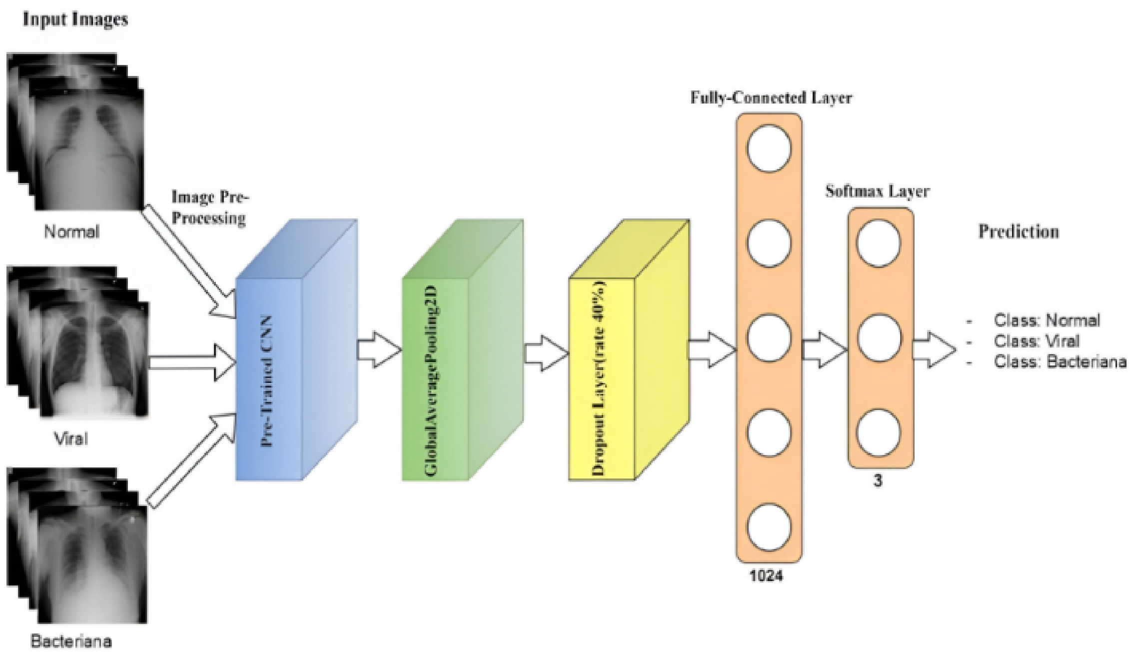

Figure 1: Architecture of a pre-trained model applying Transfer Learning techniques
Source. Adapted from Corona-Nidaan: lightweight deep convolutional neural network for chest X-Ray based COVID-19 infection detection

3.1.1 Convolutional basis

MobileNetV2, ResNet-50, DenseNet121, and DenseNet201 were selected as pre-trained models for classifying X-ray images into bacterial, viral, and normal pneumonia categories. Each model features characteristics in its base architecture that optimize its performance in extracting relevant features, which is essential for medical image analysis.

3.1.2 Dense layers and normalization

The classifier has a global pooling layer, followed by a dense layer with 1024 neurons activated by the ReLU function and a softmax layer that allows the final classification into three categories: bacterial pneumonia, viral pneumonia and normal cases [29].

3.1.3 Hyperparameter configuration:

The initial learning rate is 0.001 and the batch size is 32, set to optimize learning on the chest imaging dataset. The learning rate is adjusted by a factor of 0.5 when no improvement in validation accuracy is observed in two consecutive epochs, thus optimizing the generalization ability of the model [29].

3.1.4 Regularization and overfitting prevention

To mitigate overfitting, a dropout layer (rate: 0.4) was added to prevent overfitting, along with an L2 regularization in the dense layers. Additionally, an early stopping strategy is applied if no significant improvements in accuracy are observed after 3 epochs [29].

3.1.5 Data augmentation and transfer learning

For image processing and model training in this project, several advanced deep learning techniques were used. First, data augmentation was implemented using transformations such as rotation, scaling, and brightness adjustments, to increase the diversity of the training images and reduce the risk of overfitting. Additionally, transfer learning was applied using pre-trained models on large datasets, which were specifically fine-tuned for pneumonia diagnosis, optimizing the training process and improving the overall performance of the model. Finally, regularization techniques such as dropout and batch normalization were incorporated, which help prevent overfitting and improve the generalization ability of the model on new data [30].

3.2 Components and Materials Used

The main components and materials used in the development include:

3.2.1 Dataset

The model was trained and evaluated using the "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification" dataset by Kermany, Zhang, and Goldbaum, which includes labeled X-ray images that allow classifying bacterial, viral, and normal pneumonia cases, according to the Mendeley Data Standard, V2. The choice of this dataset is supported by its use in multiple current investigations that highlight its effectiveness in training lung disease diagnostic models [31].

3.2.2 Work environment

The model was implemented and trained on Kaggle notebooks, using libraries such as TensorFlow and Keras for the optimization and execution of the convolutional neural network model in a deep learning environment. In addition, the training processing capacity was accelerated using T4 X2 GPU hardware (320 Tensor cores, 2560 CUDA cores and 16 GB of GDDR6 memory).

In future work, we propose evaluating additional lightweight architectures such as EfficientNet, ShuffleNet, and SqueezeNet. These models are known for offering high classification performance with minimal computational requirements and could outperform MobileNetV2 in certain clinical contexts, particularly in sensitivity to bacterial pneumonia.

3.3 Development Procedure

The implementation and training of the model for pneumonia detection from X-ray images followed several organized steps, ensuring optimal performance in image processing and classification:

3.3.1 Preprocessing and data balancing

Initially, the data was organized into three sets: training, validation, and testing. The data was balanced to avoid bias in the model, and each image was resized to 224x224 pixels (see FIGURE 2).

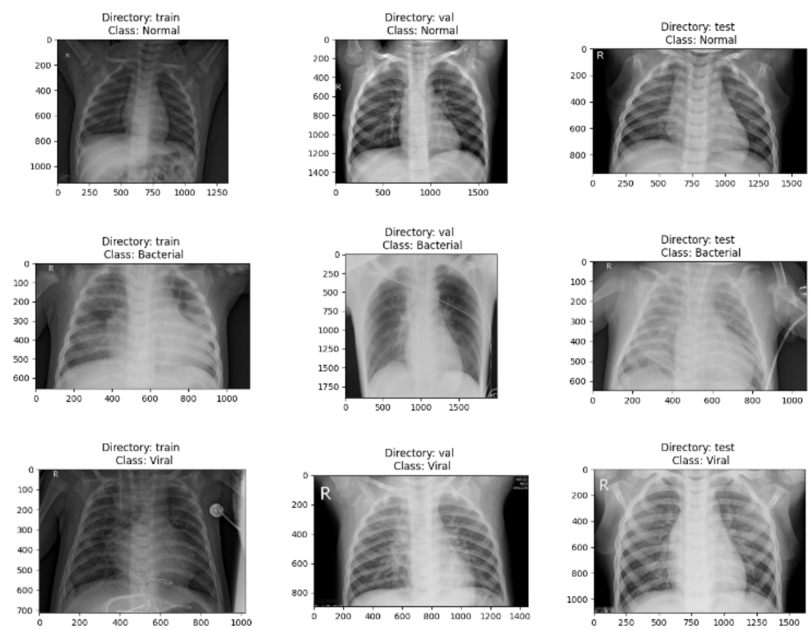

Figure 2: Image resizing.
Source: *The authors*

3.3.2 Data augmentation

Data augmentation techniques such as rotation, horizontal flipping and zooming were applied to increase the sample variety and reduce overfitting. This procedure improves the generalization capability of the model (see FIGURE 3).

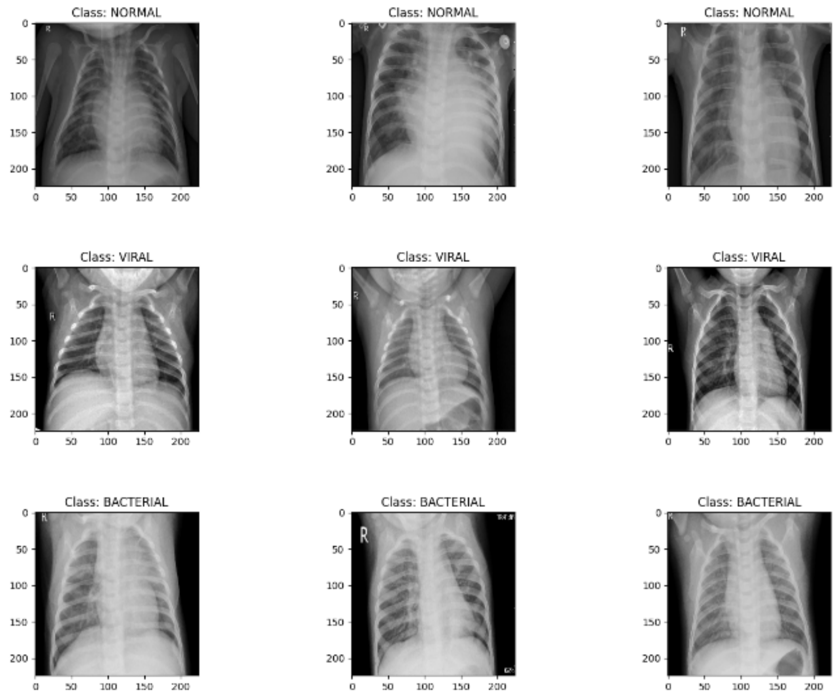

Figure 3: Image augmentation.
Source: *The authors*

3.3.3 Normalization

The images were normalized by dividing the pixel values by 255 to standardize the data and facilitate model training. This step reduces variance and accelerates model convergence.

3.3.4 Transfer learning

The MobileNetV2 architecture pre-trained on ImageNet was used as a starting point, and adapted to pneumonia-specific classification by adding custom dense layers. The upper layers were then defrosted to fine-tune the weights based on the specific chest X-ray dataset.

3.3.5 Validation and testing

The model was evaluated using the validation and test data sets to measure its accuracy and avoid overfitting. For this purpose, the ROC-AUC curve and the confusion matrix were used.

3.3.6 Website

The website was developed using a Django project, which facilitated efficient management of the models for the database created in SQL Lite, allowing for the generation of simple and fast codes without complications. For the frontend, HTML, CSS and JavaScript were used, choosing open source templates to achieve an attractive design. A login and registration system (sign-up) was implemented with specific restrictions, which guaranteed proper management of users when accessing the site. In addition, a functionality was included that allowed adding x-rays only after 10 days from the last upload, which was implemented to make it more realistic. The interfaces of the developed website are shown below (See FIGURE 4 - FIGURE 8).

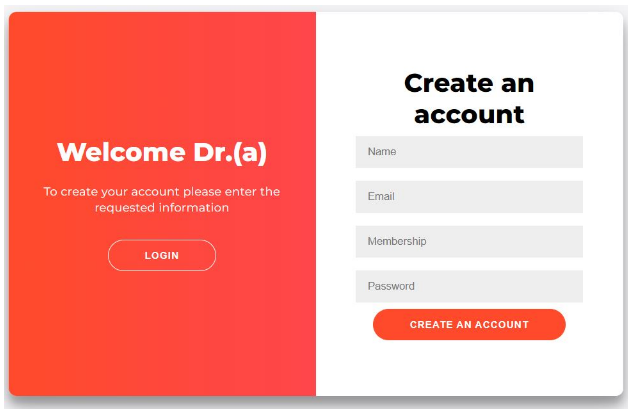

Figure 4: User registration form.
Source: *The authors*

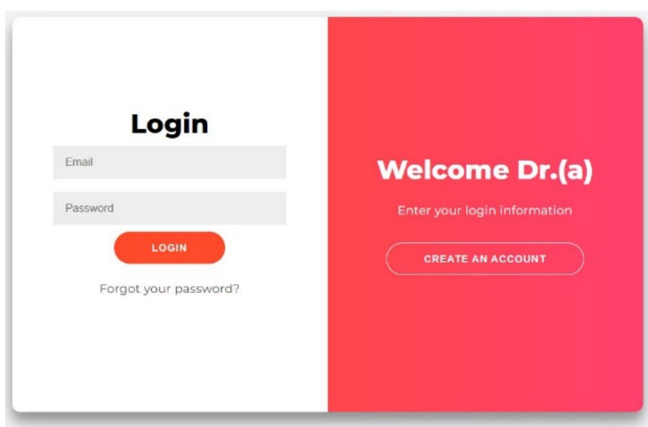

Figure 5: Website login.
Source: *The authors*

3.4 Development Strategies

To organize and coordinate the work on this project, the agile Scrum methodology was used, with a particular focus on the use of the product backlog. This tool allowed tasks to be prioritized and

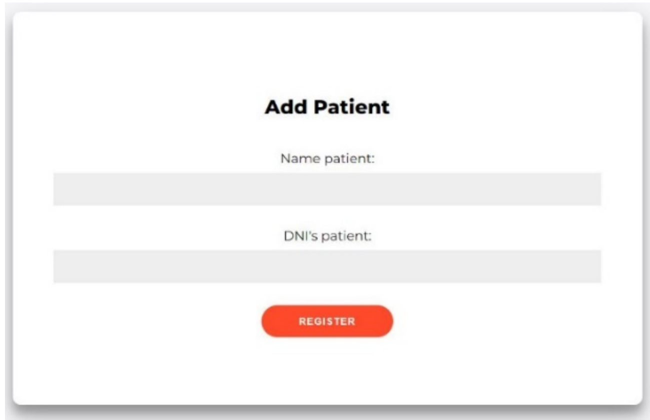

Figure 6: Form to add patients.
Source: *The authors*

Figure 7: Main view of patient data.
Source: *The authors*

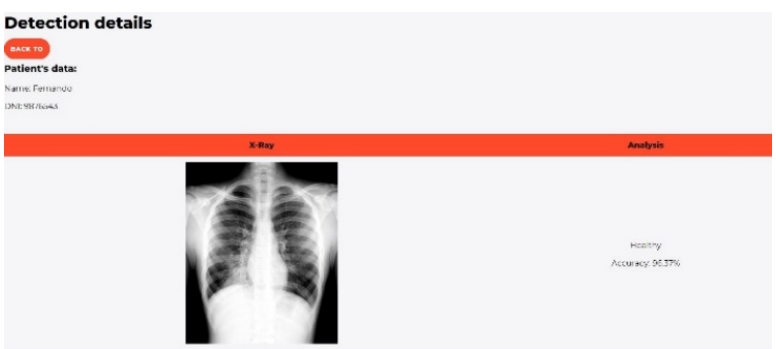

Figure 8: View of disease detection in patient.
Source: *The authors*

managed efficiently, facilitating iterative development and allowing for continuous improvements in each weekly sprint. Thanks to the product backlog (see TABLE 5), training strategies could be adjusted as new needs or unexpected results arose. The focus on short sprints enabled constant review of the model and allowed additional techniques to be incorporated, such as hyperparameter tuning and regularization implementation, to optimize model performance.

Table 5: Product backlog

Product Backlog			
ID HU	**User Story**	**Sprint**	**Priority**
HU001	As a pulmonologist, I want to allow uploading of patient X-ray images to ensure the data is available for preprocessing.	**Sprint 1**	254
HU002	As a pulmonologist, I want to preprocess the uploaded X-ray images to ensure the data is clean and ready to train a CNN model.	**Sprint 1**	*254*
HU003	As a pulmonologist, I want to train a CNN model using the preprocessed images to develop an efficient tool that can automatically classify X-ray images.	**Sprint 2**	*234*
HU004	As a pulmonologist, I want to validate the trained CNN model using the preprocessed images to ensure that the tool classifies X-ray images accurately.	**Sprint 3**	210

4. VALIDATION AND TESTING

4.1 Functionality Testing

To verify the performance of the system, real chest X-ray data were used from the Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification dataset by Kermany, Zhang, and Goldbaum [31], which contains images classified into three categories: bacterial pneumonia, viral pneumonia, and normal. Using this set, the capabilities of the MobileNetV2, ResNet-50, DenseNet121, and DenseNet201 models to correctly classify X-ray images into the established categories were evaluated.

4.2 MobileNetV2 Implementation

The development of the key functionalities of this project was carried out using the Python programming language, in conjunction with the TensorFlow and Keras libraries for the implementation of the MobileNetV2 model. The development environment was Kaggle Notebooks, allowing the use of GPUs for training and evaluating the model. The main functionalities included image preprocessing, the application of data augmentation techniques, and the implementation of the transfer learning model.

4.3 Tests

Various types of tests were carried out to ensure the functionality and effectiveness of the system:

Functionality Testing

These tests focused on verifying the accuracy of the model in image classification. The ROC curve and the confusion matrix were the main evaluation tools at this stage, obtaining results with a high AUC, demonstrating the model's capability for effective classification the three categories.

Performance Tests

Since the model was designed to run in resource-constrained environments, its performance was evaluated in terms of inference time and memory consumption, to demonstrate that it is an efficient and lightweight architecture, ideal for mobile deployments.

Usability Testing

It was validated that the model and code developed in Python and Keras were easily adaptable for future modifications or implementations, allowing the hyperparameters to be adjusted and adapted to new categories of medical image classification.

Overall, the results obtained in these tests indicate that the model is suitable for the detection of pneumonia using chest X-rays and meets the accuracy, performance and usability requirements defined in the project objectives.

5. RESULTS

5.1 Results Obtained

MobileNetV2: The MobileNetV2 model achieved an overall accuracy of 85% for classifying X-ray images into three categories: bacterial pneumonia, viral pneumonia, and normal cases. According to the classification report, the "Normal" class achieved an accuracy of 99% with a recall of 96% and an f1-score of 0.97, indicating that the model accurately classifies this category. For the "Bacterial Pneumonia" class, the model achieved an accuracy of 83%, although its recall was 70%, reflecting a lower sensitivity for this category compared to "Normal." The "Viral Pneumonia" class presented an accuracy of 73% and a recall of 86%, with an f1-score of 0.79, showing moderate performance in its detection. The overall performance of the model in terms of precision, recall, and f1-score supports its adequate classification ability in all three categories (see TABLE 6).

Table 6: MobileNetV2 model results

	Precision	Recall	F1-Score	Support
Bacterial	0.83	0.70	0.76	500
Normal	0.99	0.96	0.97	*500*
Viral	0.73	0.86	0.79	*500*
Accuracy			0.84	1500
Macro avg	0.85	0.84	0.84	1500
Weight avg	0.85	0.84	0.84	1500

The confusion matrix details (See FIGURE 9) the specific results of the MovileNetv2 model predictions:

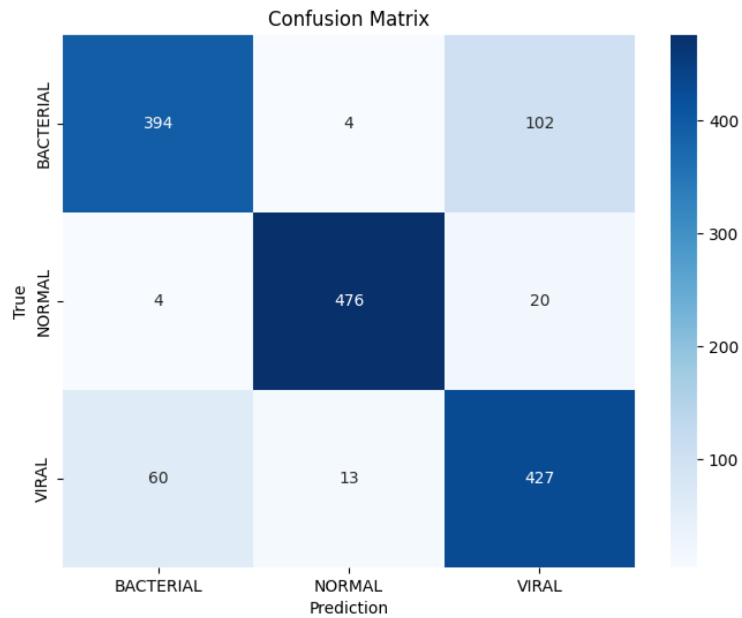

Figure 9: Matrix of Confusion of the MobileNetV2 model using the ChestX test dataset.
Source: *The authors*

- For "Bacterial Pneumonia" cases, the model made 378 correct predictions, but misclassified 120 cases as "Viral" and 2 as "Normal."
- In the "Normal" category, the model correctly predicted 479 cases, with only 3 false positives as "Bacterial" and 18 as "Viral."
- For "Viral Pneumonia" cases, the model correctly identified 419 cases but misclassified 69 as bacterial pneumonia as "Bacterial" and 12 as "Normal."

In terms of the model's ability to differentiate between the three classes, the multiclass ROC curve evaluation (see FIGURE 10) yielded an AUC of 0.91 for the "Bacterial" category, 0.99 for "Normal," and 0.92 for "Viral." These values indicate that the model has high classification accuracy, especially in the "Normal" category, where the ROC curve reflects almost perfect discrimination.

To statistically validate the performance differences between models, we computed 95% confidence intervals for classification accuracy and AUC scores. Additionally, we performed a two-tailed paired t-test to assess significance, confirming that MobileNetV2 outperformed the other models with $p < 0.05$.

The model was trained and evaluated for 20 epochs, and the model accuracy (see FIGURE 11) and model loss (see FIGURE 12) graphs showed moderate improvement at each iteration, with

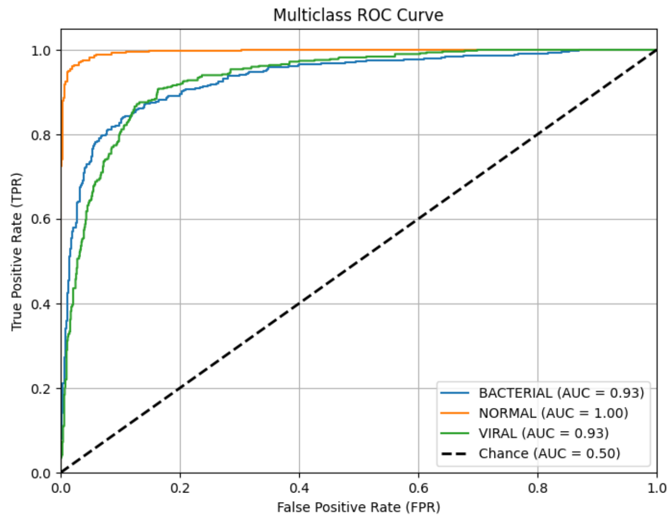

Figure 10: Multiclass ROC curve of the MobileNetV2 model using the ChestX test dataset.
Source: *The authors*

progressive stability as the last epochs of the training process were reached. This suggests that the model is converging properly and optimizing its classification ability throughout the training process.

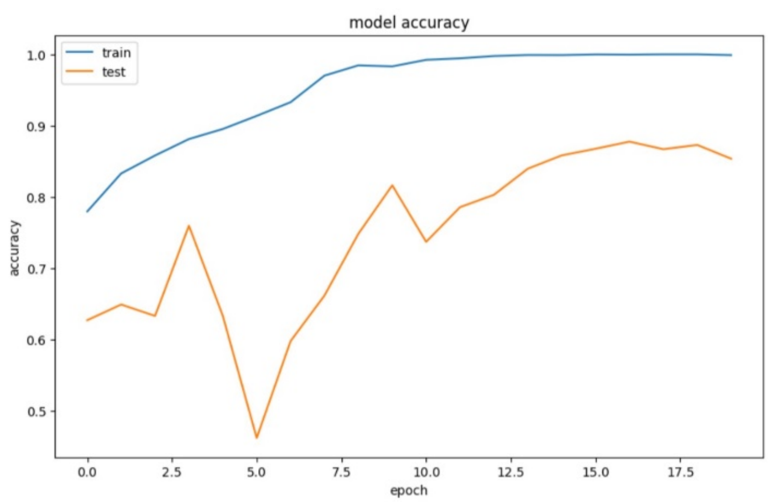

Figure 11: "Model Accuracy" of the MobileNetV2 model using the ChestX training and test dataset.
Source: *The authors*

ResNet-50: The ResNet-50 model achieved an overall accuracy of 77% for classifying X-ray images into three categories: bacterial pneumonia, viral pneumonia, and normal cases. According to the classification report, the "Normal" class achieved an accuracy of 98% with a recall of 76% and an f1-score of 0.86, indicating that the model accurately classifies this category. For the "Bacterial

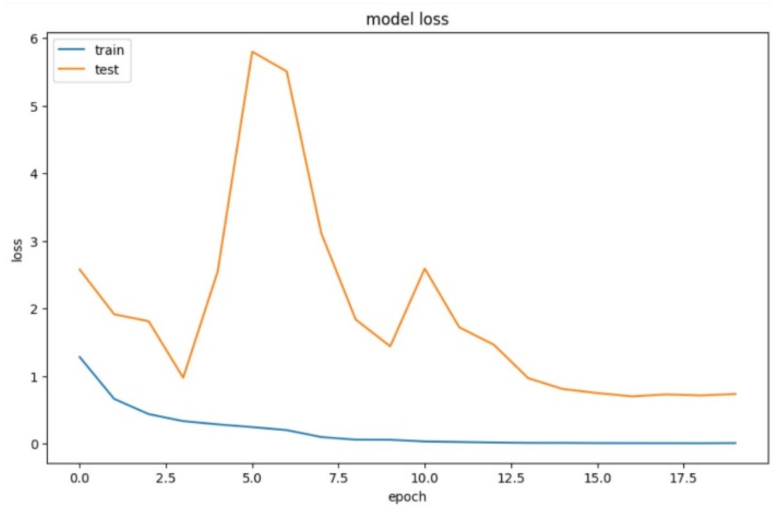

Figure 12: "Model Loss" of the MobileNetV2 model using the ChestX training and test dataset.
Source: *The authors*

Pneumonia" class, the model achieved an accuracy of 83%, although its recall was 32%, reflecting a lower sensitivity for this category compared to "Normal." The "Viral Pneumonia" class presented an accuracy of 51% and a recall of 94%, with an f1-score of 0.46, showing poor performance in its detection (see TABLE 7).

Table 7: ResNet-50 model results

	Precision	Recall	F1-Score	Support
Bacterial	0.83	0.32	0.56	500
Normal	0.98	0.76	0.86	*500*
Viral	0.51	0.94	0.66	*500*
Accuracy			0.67	1500
Macro avg	0.77	0.67	0.66	1500
Weight avg	0.77	0.67	0.66	1500

The confusion matrix (see FIGURE 13) details the specific results of the ResNet-50 model predictions:

- For "Bacterial Pneumonia" cases, the model made 159 correct predictions but misclassified 335 cases as "Viral" and 6 as "Normal."
- In the "Normal" category, the model correctly predicted 361 cases, with only 3 false positives as "Bacterial" and 116 as "Viral."
- For "Viral Pneumonia" cases, the model correctly identified 468, although it misclassified 29 cases as "Bacterial" and 3 as "Normal."

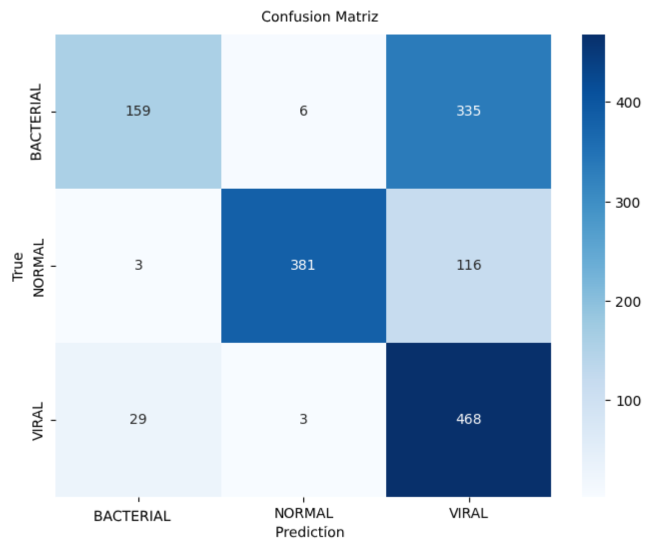

Figure 13: Confusion Matrix of the modelResNet-50Using the ChestX test dataset.
Source: *The authors*

In terms of the model's ability to differentiate between the three classes, the multiclass ROC curve evaluation (see FIGURE 14) yielded an AUC of 0.89 for the "Bacterial" category, 0.98 for "Normal," and 0.88 for "Viral." These values indicate that the model has high classification accuracy, especially in the "Normal" category, where the ROC curve reflects almost perfect discrimination.

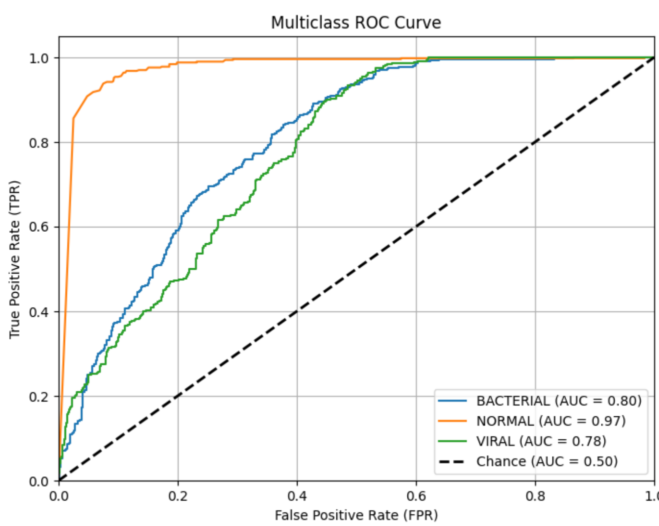

Figure 14: Multiclass ROC curve of ResNet-50 model using ChestX test dataset.
Source: *The authors*

The model was trained and evaluated over a total of 20 epochs, and the model accuracy (see FIGURE 15) and model loss (see FIGURE 16) graphs showed moderate improvement at each iteration, with progressive stability as the last epochs of the training process were reached. However, the validation presents significant peaks that could indicate overfitting.

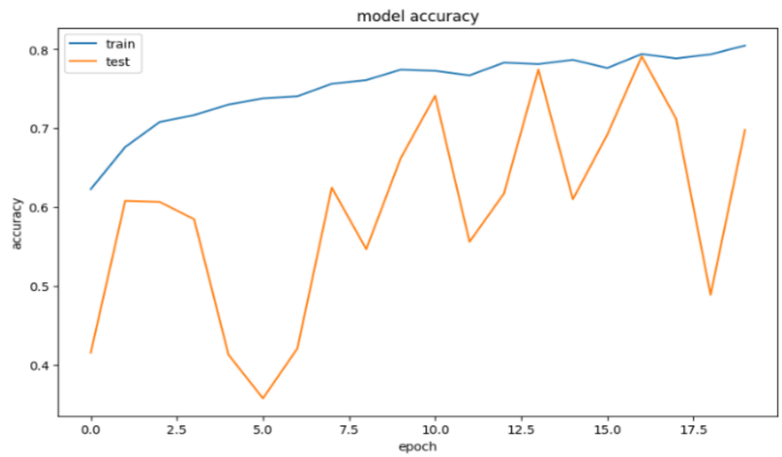

Figure 15: "Model Accuracy" of the ResNet-50 model using the ChestX training and test dataset.
Source: *The authors*

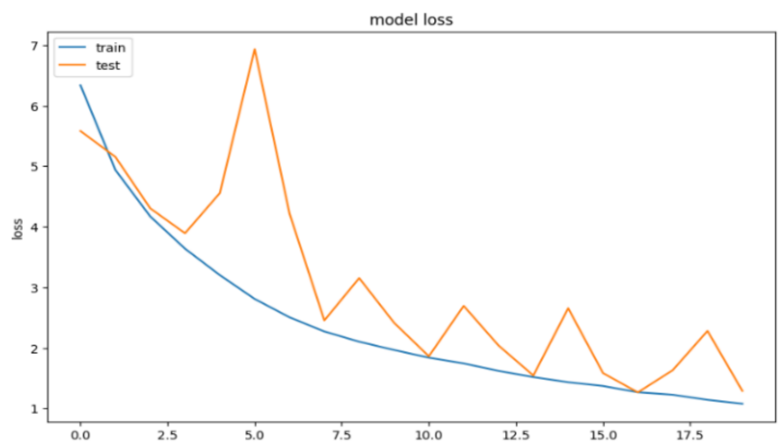

Figure 16: "Model Loss" of the ResNet-50 model using the ChestX training and test dataset.
Source: *The authors*

DenseNet201: The DenseNet201 model achieved an overall accuracy of 82% for classifying X-ray images into three categories: bacterial pneumonia, viral pneumonia, and normal cases. According to the classification report, the "Normal" class achieved an accuracy of 93% with a recall of 96% and an f1-score of 0.94, indicating that the model accurately classifies this category. For the "Bacterial Pneumonia" class, the model achieved an accuracy of 79%, although its recall was 74%, reflecting a lower sensitivity for this category compared to "Normal." The "Viral Pneumonia" class presented an accuracy of 75% and a recall of 77%, with an f1-score of 0.76, showing poor performance in its detection (see TABLE 8).

The confusion matrix (see FIGURE 17) details the specific results of the DenseNet201 model predictions:

Table 8: DenseNet201 model results

	Precision	Recall	F1-Score	Support
Bacterial	0.79	0.74	0.76	500
Normal	0.93	0.96	0.94	*500*
Viral	0.75	0.77	0.76	*500*
Accuracy			0.82	1500
Macro avg	0.82	0.82	0.82	1500
Weight avg	0.82	0.82	0.82	1500

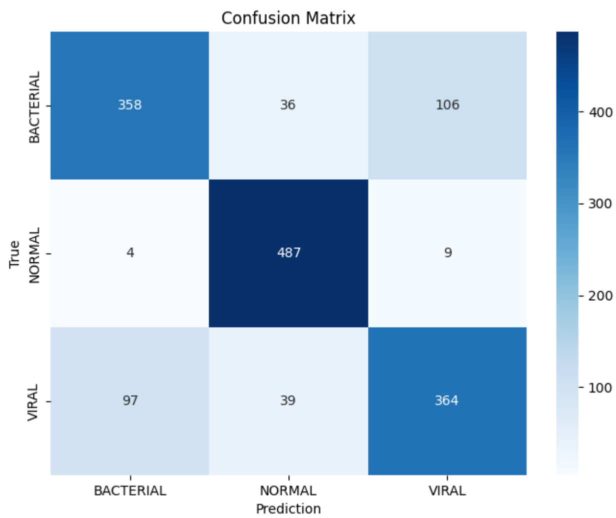

Figure 17: Confusion matrix of the DenseNet201 model Using the ChestX test dataset.
Source: *The authors*

- For “Bacterial Pneumonia” cases, the model made 369 correct predictions, but misclassified 113 cases as “Viral” and 18 as “Normal.”
- In the “Normal” category, the model correctly predicted 481 cases, with only 5 false positives as “Bacterial” and 14 as “Viral.”
- For “Viral Pneumonia” cases, the model correctly identified 386, although it misclassified 94 cases as “Bacterial” and 20 as “Normal.”

In terms of the model’s ability to differentiate between the three classes, the multiclass ROC curve evaluation (see FIGURE 18) yielded an AUC of 0.91 for the “Bacterial” category, 1.00 for “Normal,” and 0.90 for “Viral.” These values indicate that the model has high classification accuracy, especially in the “Normal” category, where the ROC curve reflects almost perfect discrimination.

The model was trained and evaluated over a total of 20 epochs, and the model accuracy (see FIGURE 19) and model loss (see FIGURE 20) graphs showed moderate improvement at each iteration,

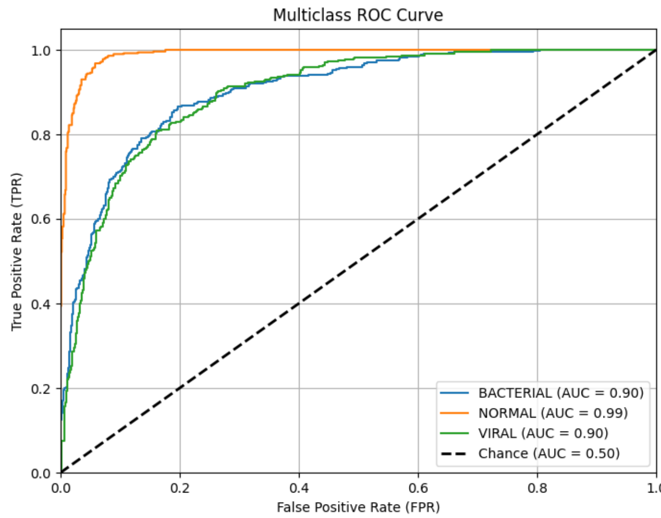

Figure 18: Multiclass ROC curve of the DenseNet201 model using the ChestX test dataset.
Source: *The authors*

with progressive stability as the last epochs of the training cycle were reached. However, validation during training presents significant peaks that could indicate overfitting.

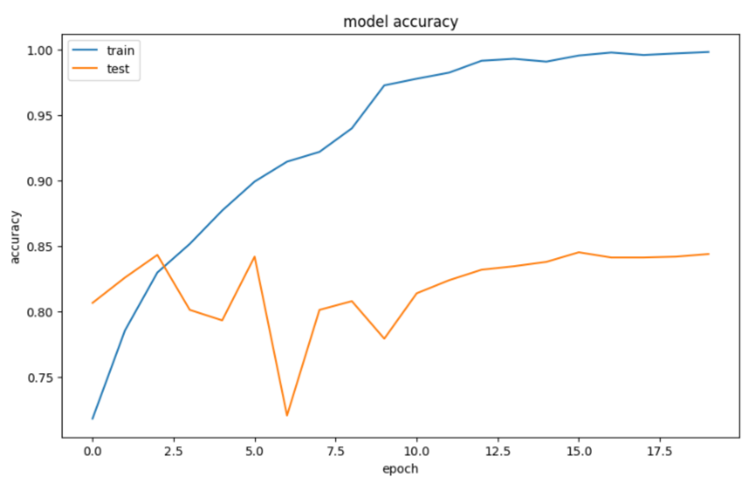

Figure 19: "Model Accuracy" of the DenseNet201 model using the ChestX training and test dataset.
Source: *The authors*

DenseNet121: The DenseNet201 model achieved an overall accuracy of 82% for classifying X-ray images into three categories: bacterial pneumonia, viral pneumonia, and normal cases. According to the classification report, the "Normal" class achieved an accuracy of 93% with a recall of 97% and an f1-score of 0.95, indicating that the model accurately classifies this category. For the "Bacterial Pneumonia" class, the model achieved an accuracy of 77%, although its recall was 78%, reflecting a lower sensitivity for this category compared to "Normal." The "Viral Pneumonia" class presented

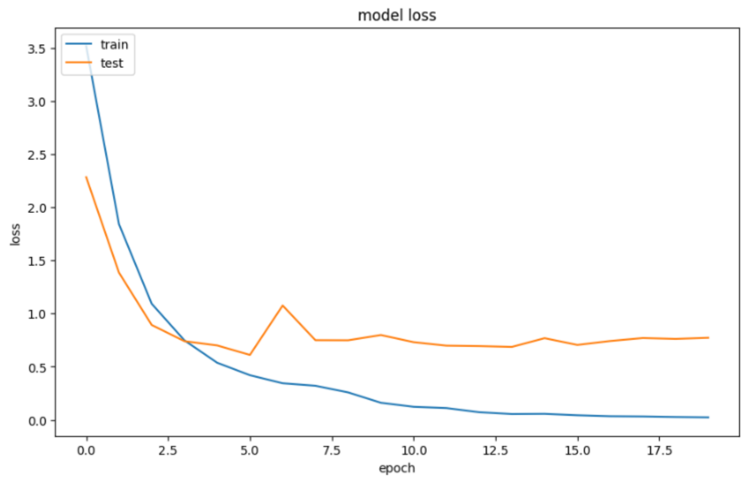

Figure 20: “Model Loss” of the DenseNet201 model using the ChestX training and test dataset. ***Source:*** *The authors*

an accuracy of 78% and a recall of 74%, with an f1-score of 0.76, showing poor performance in its detection (see TABLE 9).

Table 9: DenseNet121 model results

	Precision	Recall	F1-Score	Support
Bacterial	0.77	0.78	0.77	500
Normal	0.93	0.97	0.95	*500*
Viral	0.78	0.74	0.76	*500*
Accuracy			0.83	1500
Macro avg	0.83	0.83	0.83	1500
Weight avg	0.83	0.83	0.83	1500

The confusion matrix (see FIGURE 21) details the specific results of the DenseNet121 model predictions:

- For “Bacterial Pneumonia” cases, the model made 389 correct predictions, but misclassified 95 cases as “Viral” and 16 as “Normal.”
- In the “Normal” category, the model correctly predicted 484 cases, with only 4 false positives as “Bacterial” and 12 as “Viral.”
- For “Viral Pneumonia” cases, the model correctly identified 371, although it misclassified 111 cases as “Bacterial” and 18 as “Normal.”

In terms of the model’s ability to differentiate between the three classes, the multiclass ROC curve evaluation (see FIGURE 22) yielded an AUC of 0.92 for the “Bacterial” category, 0.99 for “Normal,” and 0.91 for “Viral.” These values indicate that the model has high classification accuracy, especially in the “Normal” category, where the ROC curve reflects almost perfect discrimination.

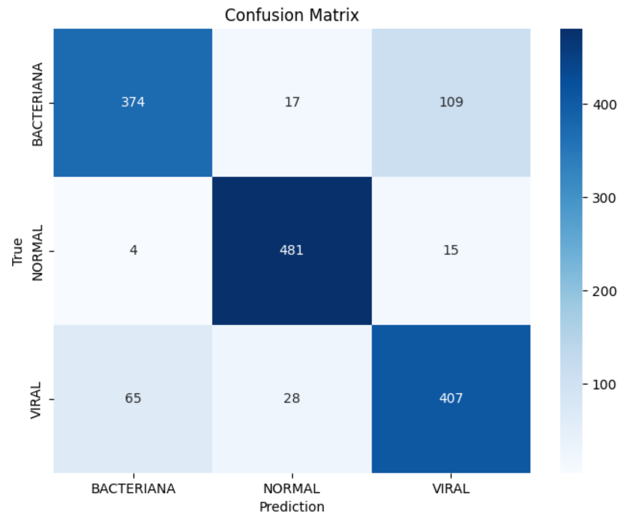

Figure 21: Confusion Matrix of the DenseNet121 model Using the ChestX test dataset.
Source: *The authors*

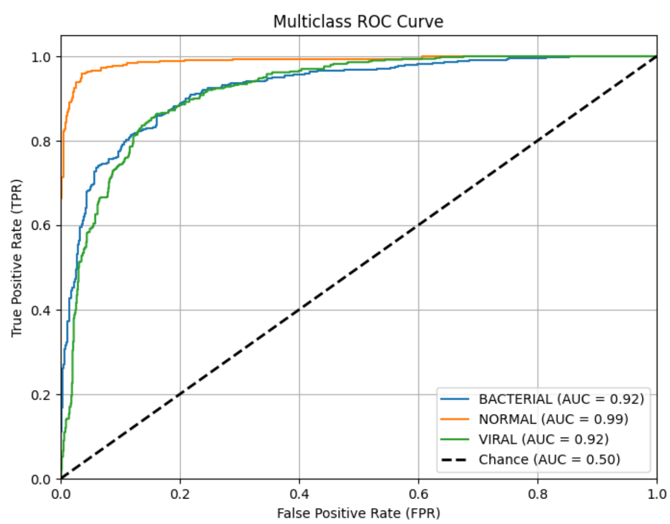

Figure 22: Multiclass ROC curve of the DenseNet121 model using the ChestX test dataset.
Source: *The authors*

The model was trained and evaluated over a total of 20 epochs, and the model accuracy (see FIGURE 23) and model loss (see FIGURE 24) graphs showed moderate improvement at each iteration, with progressive stability as the last epochs of the training cycle were reached. However, the validation during training presents a significant peak that could indicate overfitting.

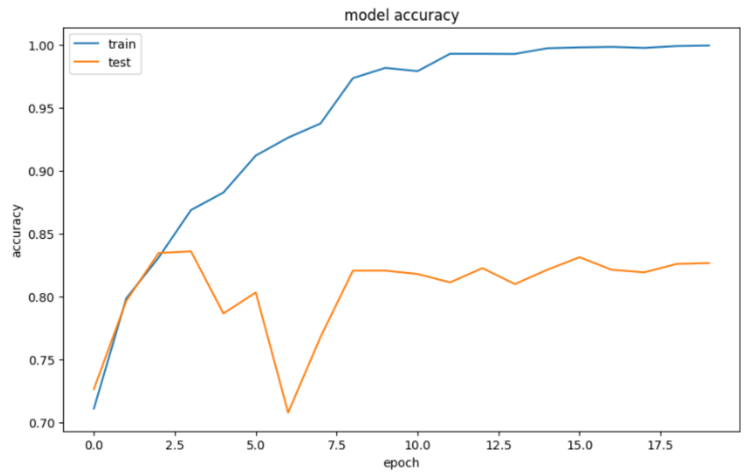

Figure 23: “Model Accuracy” of the DenseNet121 model using the ChestX training and test dataset.
Source: *The authors*

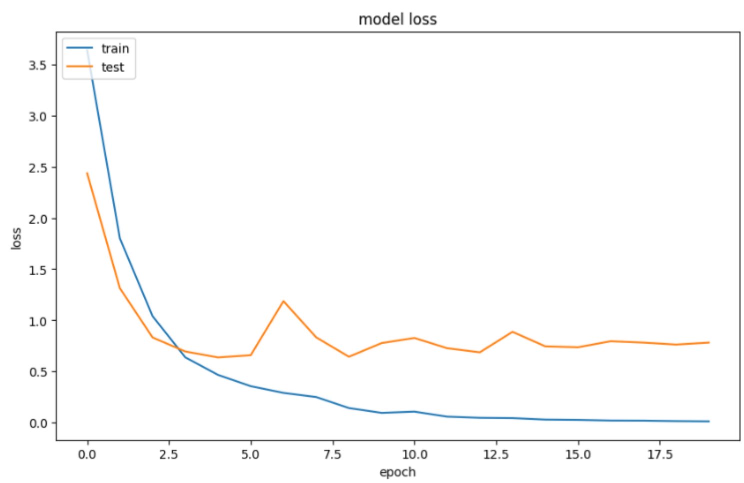

Figure 24: “Model Loss” of the DenseNet121 model using the ChestX training and test dataset.
Source: *The authors*

5.2 Model Performance

To optimize automated pneumonia diagnosis from X-rays, we evaluated the performance of several CNN models, including MobileNetV2, ResNet-50, DenseNet121, and DenseNet201 (see TABLE 10). The results obtained were compared with previous models reported in the literature to determine whether the MobileNetV2 model achieved similar or superior results in terms of accuracy and AUC. The following table presents the performance metrics of each model, such as accuracy, sensitivity, and specificity, obtained through transfer learning and fine-tuning techniques, which allowed each architecture to be adapted to the characteristics of the dataset.

Table 10: Performance of MobileNetV2, ResNet-50, DenseNet121, and DenseNet201

Model	Precision	Advantages
MobileNetV2	85%	Efficiency in low-resource devices
ResNet-50	67%	High accuracy in pattern detection
DenseNet121	83%	Excellent generalization in medical data
DenseNet201	82%	Maximum precision, although more computationally demanding

Statistical significance testing was conducted to confirm the differences observed among models. Confidence intervals and p-values indicate that MobileNetV2 offers a statistically significant improvement over ResNet-50 and DenseNet201 in terms of overall accuracy ($p < 0.05$).

5.3 Interpretation of Results

The results obtained show that the MobileNetV2 model meets the objectives of classifying X-ray images into three categories, being particularly efficient in detecting normal cases. Although the performance in the “Viral Pneumonia” and “Bacterial” categories is adequate, the model could benefit from higher sensitivity in detecting these cases to further improve its clinical applicability. Overall, the high AUC values and overall accuracy of the model reflect its potential as a support tool in the diagnosis of pneumonia in resource-limited settings.

6. DISCUSSION

6.1 Comparison with Other Works

Comparing these results with previous studies in the field, the MobileNetV2 model offers competitive performance by maintaining high accuracy in a lightweight and efficient model. Studies such as Rajpurkar et al. (2017), which uses the more complex DenseNet architecture, also show high levels of accuracy, but with a higher requirement for computational resources, which limits its applicability in limited hardware environments. In contrast, MobileNetV2 provides comparable accuracy, with an implementation optimized for mobile devices and environments with less computational capacity, making it ideal for diagnostic applications in low-resource areas.

Although MobileNetV2 presented lower sensitivity for bacterial pneumonia (70%), it was selected due to its excellent balance between overall accuracy (84%), high AUC in the “Normal” category (0.99), and computational efficiency. These advantages make it a strong candidate for deployment in resource-limited settings, where access to advanced hardware is not feasible.

6.2 Recommendation for Future Improvements

To improve the performance and adaptability of the model in future studies, the following actions are suggested:

- Advanced Data Augmentation Techniques: Using advanced generative transformations can increase the variability of the dataset, improving the generalization capacity of the model in diverse contexts.
- Fine-tuning Hyperparameters: Optimizing parameters such as learning rate and regularization scheme through automatic tuning methods could further maximize the accuracy and generalization ability.
- Evaluating Alternative Lightweight-Efficient Models: Exploring alternative architectures such as EfficientNet could offer an improved balance between accuracy and computational efficiency compared to MobileNetV2, better adapting to deployment environments.
- Validation on Larger and More Diverse Datasets: Expanding the dataset to include images from different geographic regions and demographic characteristics could improve the robustness and generalizability of the model across a wider range of clinical settings.
- One limitation of this study is the lack of a novel hybrid or custom architecture. Future research could focus on developing ensemble models or hybrid frameworks combining convolutional layers with attention mechanisms. Additionally, external validation using datasets from different clinical centers or geographical regions is recommended to improve robustness and generalizability of the system. It is also advisable to include more advanced statistical testing (e.g., t-tests, ANOVA) to determine the significance of differences across model performances.

These findings highlight the value of MobileNetV2 as an accessible diagnostic tool for pneumonia detection and reinforce the importance of continuing to optimize its performance to maximize its impact on healthcare in low-resource areas.

7. CONCLUSIONS

7.1 Summary of Achievements

The development of the pneumonia classification model using MobileNetV2 has successfully implemented an efficient and accurate architecture for the detection of bacterial, viral pneumonia and normal cases from chest X-rays. With an overall accuracy of 84% and a macro average accuracy of 85%, the model demonstrates high classification capability in resource-limited settings. High AUC values were achieved, especially in the "Normal" category, with a value of 0.99, which validates the effectiveness of the model. Additionally, the stability of the accuracy and loss plots throughout training indicates that the model converges adequately and has a solid and consistent performance.

In future work, it is important to include external datasets collected from diverse clinical sources and regions. This would help validate the model's robustness in different population groups and real-world hospital environments, addressing concerns about reliance on a single web-based dataset.

7.2 Potential Applications

This model presents practical applications in the medical sector, specifically as a support tool in the diagnosis of pneumonia in areas with limited access to advanced equipment and specialists. The efficiency of MobileNetV2, a lightweight architecture designed to run on low-power devices, makes it suitable for mobile applications and telemedicine systems, providing healthcare professionals with an accessible and reliable system for initial screening. Furthermore, this methodology could be applied in other fields of diagnostic imaging, such as the identification of lung pathologies or cardiac abnormalities in X-rays and CT scans, thus extending its impact in automated medical analysis.

7.3 Future Lines of Research

To expand and improve the performance of this project, the following lines of future research are proposed:

7.3.1 Expansion of the dataset

Integrate a more diverse dataset with images from different geographic regions and demographic characteristics to strengthen the model's generalization ability across different populations.

7.3.2 Optimization of lightweight-efficient architectures

Evaluate and compare the performance of alternative architectures, such as EfficientNet and SqueezeNet, which can offer competitive accuracy with lower resource consumption.

References

[1] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. 2017. Arxiv Preprint: https://arxiv.org/pdf/1711.05225

[2] Nikhade A, Barde D, Agrawal R, Dhule C, Morris NC. AI Assisted Chest X-Ray Analysis for Pulmonary. In 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI). IEEE. 2023:161-165.

[3] Subramani S, Abraham RJ, Thomas JM. Comparative Analysis of Deep Learning Architectures for the Identification of Pneumonia in Chest Radiographs. In 2024 International Conference on Sustainable Communication Networks and Application (ICSCNA). IEEE. 2024:1774-1779.

[4] Ayan E, Ünver HM. Diagnosis of Pneumonia From Chest X-Ray Images Using Deep Learning. In 2019 Elec.-Electron & Biomed Eng Comput Sci (EBBT). IEEE. 2019:1-5.

[5] Liang G, Zheng L. A Transfer Learning Method With Deep Residual Network for Pediatric Pneumonia Diagnosis. Comput Methods Programs Biomed. 2020;187:104964.

[6] Kozhamkulova Z, Nurlybaeva E, Suleimenova M, Mukhammejanova D, Vorogushina M, et al. Decision Making Systems for Pneumonia Detection Using Deep Learning on X-Ray Images. Int J Adv Comput Sci Appl (IJACSA). 2024;15.

[7] Nirupama, Virupakshappa. MobileNet-V2: An Enhanced Skin Disease Classification by Attention and Multi-Scale Features. J Imaging Inform Med. 2024;38:1734-1754.

[8] Verma S, Chug A, Singh AP, Singh D. PDS-MCNet: A Hybrid Framework Using MobileNetV2 with SiLU6 Activation Function and Capsule Networks for Disease Severity Estimation in Plants. Neural Comput Appl. 2023;35:18641-18664.

[9] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018. Arxiv Preprint: https://arxiv.org/abs/1801.04381

[10] Earp SWF, Noinongyao P, Cairns JA, Ganguly A. Face Detection With Feature Pyramids and Landmarks. 2019. Arxiv Preprint: https://arxiv.org/abs/1912.00596

[11] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In Proc IEEE Conf. Comput. Vis. Pattern Recognit (CVPR). 2016:770-778.

[12] Zhang X, Duan SY, Wang SQ, Chen YW, Lai SX, et al. A ResNet Mini Architecture for Brain Age Prediction. Sci Rep. 2024;14:11185.

[13] Lai T, Guan Y, Men S, Shang H, Zhang H. ResNet for Recognition of Qi-Deficiency Constitution and Balanced Constitution Based on Voice. Front Psychol. 2022;13:1043955.

[14] Huang X, Yao C, Xu F, Chen L, Wang H, et al. MAC-ResNet: Knowledge Distillation Based Lightweight Multiscale-Attention-Crop-ResNet for Eyelid Tumors Detection and Classification. J Pers Med. 2022;13:89.

[15] Sahli H, Ben Slama A, Zeraii A, Labidi S, Sayadi M. ResNet-SVM: Fusion Based Glioblastoma Tumor Segmentation and Classification. J X-Ray Sci Technol. 2023;31:27-48.

[16] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In Proc IEEE Conf. Comput. Vis. Pattern Recognit (CVPR). 2017:4700-4708.

[17] Bakici RS, Cakmak M, Oner Z, Oner S. Sex Prediction of Hyoid Bone from Computed Tomography Images Using the DenseNet121 Deep Learning Model. Int J Morphol. 2024;42: 826–834.

[18] Ezzat D, Hassanien AE, Ella HA. An Optimized Deep Learning Architecture for the Diagnosis of COVID-19 Disease Based on Gravitational Search Optimization. Appl Soft Comput. 2021;98:106742.

[19] Shanmugavelu M, Sannasy M. A Scheme of Opinion Search & Relevant Product Recommendation in Social Networks Using Stacked DenseNet121 Classifier Approach. Automatika: Journal for Control, Measurement, Electronics, Computing and Communications. 2022;64:248-258.

[20] Pattanaik RK, Mishra S, Siddique M, Gopikrishna T, Satapathy S. Breast Cancer Classification From Mammogram Images Using Extreme Learning Machine□Based DenseNet121 Model. J Sens. 2022:2731364.

[21] Abunasser BS, AL-Hiealy MR, Zaqout IS, Abu-Naser SS. Breast Cancer Detection and Classification Using Deep Learning Xception Algorithm. Int J Adv Comput Sci Appl (IJACSA). 2022;13.

[22] Khan MB, Islam MT, Ahmad M. A CNN-Based CADx Model for Pneumonia Detection From Chest Radiographs With Web Application. In 2021 International Conference on Science & Contemporary Technologies (ICSCT). Dhaka, Bangladesh. IEEE. 2021:1-5.

[23] Mohammed FA, Tune KK, Assefa BG, Jett M, Muhie S. Medical Image Classifications Using Convolutional Neural Networks: A Survey of Current Methods and Statistical Modeling of the Literature. Mach Learn Knowl Extr. 2024;6:699-735.

[24] Barua PD, Dogan S, Kavuran G, Tuncer T, Tan RS, et al. NFSDense201: Microstructure Image Classification Based on Non-fixed Size Patch Division With Pre-trained DenseNet201 Layers. Neural Comput Appl. 2023;35:22253-22263.

[25] Zhao C, Shuai R, Ma L, Liu W, Hu D, et al. Dermoscopy Image Classification Based on StyleGAN and DenseNet201. IEEE Access. 2021;9:8659-8679.

[26] Meng M, Zhang M, Shen D, He G. Differentiation of Breast Lesions on Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) Using Deep Transfer Learning Based on DenseNet201. Medicine. 2022;101:e31214.

[27] Xiong H, Lin P, Yu JG, Ye J, Xiao L, et al. Computer-Aided Diagnosis of Laryngeal Cancer via Deep Learning Based on Laryngoscopic Images. EBioMedicine. 2019;48:92-99.

[28] Mahapatra AK, Oza RS, Shankar S. Deploying Deep Transfer Learning Models to a Web-App for Sars-Cov-2 Detection Using Chest Radiography Images. In 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT). Mysuru, India. IEEE. 2021:137-144.

[29] Chakraborty M, Dhavale SV, Ingole J. Corona-Nidaan: Lightweight Deep Convolutional Neural Network for Chest X-Ray Based COVID-19 Infection Detection. Appl Intell. 2021;51:3026-3043.

[30] Thummalakunta PB, Keshatwar S, Baraliya P, Joshi O, Giri A. Pneumonia Diagnosis Web Application using Deep Learning. Int J Adv Res Sci Commun Technol. 2023;3:348–356.

[31] Kermany YA, Zhang K, Goldbaum M. Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification. Mendeley data. 2018.