

AI-Driven Early Detection of Pediatric Pneumonia: Optimizing Convolutional Neural Networks for Chest X-Ray Analysis in Children Under 5

Adelaide Louise M. Naquila^{1,*}, Riana Lhei D. Almodovar¹, Jamiel Joy V. Rebuena¹, Elijah John B. Ramos¹
and Ronnel P. Agulto¹

¹ *Department of Electronics and Computer Engineering
De La Salle University - School of Innovation and Sustainability
Biñan, Laguna, Philippines*

**Corresponding Author: adelaide_naquila@dlsu.edu.ph*

Abstract: Pneumonia is one of the leading illnesses that causes death, particularly for children under 5. Artificial intelligence can help medical professionals with the early diagnosis of pneumonia. This study presents the development and evaluation of a convolutional neural network (CNN) model for binary image classification of chest X-rays of children into Pneumonia and Normal categories. The model was trained using a publicly available dataset in Kaggle, including two categories, “PNEUMONIA” and “NORMAL” images. The dataset was preprocessed using standard techniques such as resizing the image to 150 x 150 pixels, data augmentation, and normalization for enhancing generalization. A custom CNN model is built using the TensorFlow Sequential model. It consists of three convolutional blocks with increasing filter sizes, ReLU activation, batch normalization, max pooling, and dropout for regularization. The model was trained and tested under varying batch sizes, epochs, and filter/neuronal configurations to determine optimal performance settings. Evaluations revealed that a batch size of 16 in 25 epochs yielded the highest test accuracy of 92.31%, with strong F1 scores for both classes. Furthermore, architectural tuning by modifying the number of filters and neurons showed that increasing the model’s complexity significantly improved performance. In particular, the model with quadruple the base achieved a 91% test accuracy with an F1 score of 0.93 for Pneumonia and 0.87 for Normal, which had strong diagnostic reliability. The findings suggest that deeper CNN architectures can improve classification accuracy, especially in critical medical applications where high recall for positive cases is crucial. For future works, better feature extraction may be explored by integrating pre-trained models such as ResNet or EfficientNet and expanding the dataset images to accommodate other varying factors.

Key Words: pneumonia; deep learning; convolutional neural network; classification; hyperparameter tuning

1. INTRODUCTION

Pneumonia, a known illness that causes inflammation of the lungs, either both lungs or one lung of the human body, (Cleveland Clinic, 2025). Various organisms in the lungs can cause the infection, including bacteria, viruses, and fungi (Mayo Clinic, 2020). The air sacs of lungs filled with sac or pus, making it a serious lung infection, that can also range from mild or severe (John Hopkins Medicine, 2025).

Pneumonia accounted for 14% of all child fatalities in 2019, killing a number of 740,180 children under the age of five. (World Health Organization, 2022). The Philippine Statistics reports pneumonia as the fourth leading mortality with 12,892 deaths from January to May 2024 (Philippine Statistics Authority, 2024).

Symptoms of having pneumonia include coughing, shortness of breath, chest pain, high temperature, an aching body, loss of appetite, when breathing creates a wheezing sound, and an aching body (National Health Service, 2023).

For pneumonia diagnosis, there are various ways for the health professionals to examine a suspected patient. A physical examination, blood tests, pulse oximetry, sputum testing, and a chest X-ray are all included. (Mayo Clinic, 2020).

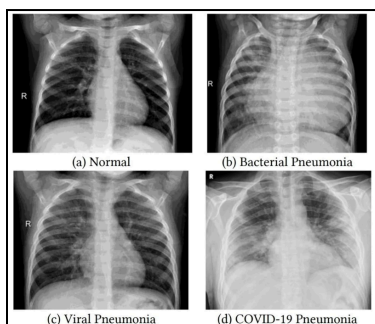


Fig. 1. Example of frontal chest X-ray image with various types of pneumonia (Thomas, 2020)

Chest X-rays are essential for diagnosing pneumonia, as they provide a visual representation of lung abnormalities, and it helps medical professionals recognize inflammatory changes and fluid accumulation, which are vital indicators of pneumonia (One Step Diagnostic, 2024).

As deep learning and artificial intelligence (AI)

continue to develop, automated medical image analysis is emerging as a useful tool to help radiologists and increase accuracy. Convolutional Neural Networks (CNNs), a known deep learning technique, which was developed for image analysis, has demonstrated remarkable performance in finding anomalies, diagnosing diseases, and identifying malignancies. CNNs improve diagnosis by automatically extracting information from medical images.

Using CNNs, the research seeks to develop an automated technique for detecting pneumonia from chest X-ray images. A dataset of X-ray images will be used to train the model, which will then identify patterns that differentiate pneumonia cases from healthy ones. The goal is to improve healthcare and assist radiologists in making better decisions, with a system that can classify images from X-rays. This study intends to decrease misdiagnosis and enhance access to high-quality healthcare, particularly in underprivileged areas, by employing deep learning for pneumonia identification.

2. METHODOLOGY

2.1 Architecture

The core CNN architecture remains consistent across all models, including the base model and its variants. This project will customize its own CNN using the TensorFlow Sequential model.

The base model will consist of the input, output, and 3 hidden layers. The input accepts 150x150-pixel grayscale images. The hidden layer is composed of 3 convolutional blocks. Each block contains a 3x3 kernel size with varying number of filters in the Conv2D layer, as well as the Batch Normalization and MaxPooling2D layers. The first block contains 32 filters, the second contains 64 filters, and the third block has 128 filters in the Conv2D layer. Every hidden layer utilizes the ReLU activation function and has a dropout layer set at the rate of 0.5. Finally, the output layer of the model utilizes the binary cross-entropy loss function, Adam optimizer, and sigmoid activation function.

Part 2 will be tuning hyperparameters related to learning. The different models for this part will have the same architecture as the base model that is trained with varying batch sizes and number of epochs as follows:

- Model 2.1: Batch size 16 and 15 epochs

- Model 2.2: Batch size 16 and 25 epochs
- Model 2.3: Batch size 16 and 50 epochs
- Model 2.4: (Base) Batch size 32 and 15 epochs
- Model 2.5: Batch size 32 and 25 epochs
- Model 2.6: Batch size 32 and 50 epochs
- Model 2.7: Batch size 64 and 15 epochs
- Model 2.8: Batch size 64 and 25 epochs
- Model 2.9: Batch size 64 and 50 epochs

Part 3 will tune hyperparameters related to network architecture – the number of filters/neurons. The first number signifies the filters in the input layer, followed by three numbers for the filters in the hidden layer, and the last number being the neurons in the fully connected layer. The configurations for each model are:

- Model 3.1: Model with HALF the base number of filter/neurons (8, 16, 32, 64, and 32)
- Model 3.2: Base Model (16, 32, 64, 128, and 64)
- Model 3.3: Model with DOUBLE the base number of filter/neurons (32, 64, 128, 256, and 128)
- Model 3.4: Model with QUADRUPLE the base number of filter/neurons (64, 128, 256, 512, and 256)

2.2 Dataset

This study utilized the Chest X-Ray Images (Pneumonia) dataset owned by Paul Timothy Mooney and hosted on Kaggle. The dataset includes 5,863 grayscale X-ray images, all labeled, from patients aged one to five at Guangzhou Women and Children's Medical Center in Guangzhou. These images come from retrospective patient cohorts. These were divided into two categories: "PNEUMONIA" and "NORMAL" (Mooney, 2018). Along with this, the dataset was organized into three subsets:

- Training Set: 5,216 images
- Validation Set: 16 images
- Test Set: 624 images

Each image was loaded and processed by resizing it to 150x150 pixels and converted to single-channel (grayscale) format for model input consistency. In order to improve model generalization and expand the dataset size, data augmentation will be performed onto the dataset.

2.3 Model Design

The base model was initially trained for 15 epochs with a batch size of 32 and the Adam optimizer set to 0.001. Each of the three convolutional blocks in the hidden layer comprises a Conv2D layer then followed by Batch Normalization and MaxPooling2D, allowing the model to learn spatial hierarchies incrementally while reducing internal covariate shift and increasing stability. Due to its non-linearity and computational simplicity, ReLU (Rectified Linear Unit) was chosen as the activation function for the convolutional and dense layers to alleviate the vanishing gradient problem and accelerate convergence.

Since overfitting is a common concern in medical imaging due to limited and often imbalanced datasets, a dropout layer was included subsequent to the dense layer to prevent this. During training, this will deactivate a fraction of neurons at random to prevent the network from becoming unduly reliant on certain paths through the network, improving generalization. The final output layer is composed of a single neuron with a sigmoid activation function, which is suited for binary classification since it produces a probability between 0 and 1, representing the likelihood of an image being classified as "PNEUMONIA" (Daniel, 2023).

The utilized dataset classifies the results on a binary classification as denoted by only NORMAL or PNEUMONIA classification. With that, the binary cross-entropy loss function was used due to the binary nature of the classification task, and this type of loss function fits well with binary results equivalent to 0 and 1 as opposed to categorical cross-entropy (Apeinans et al., 2024). A ReduceLROnPlateau learning rate scheduler and early stopping mechanism were used to prevent overfitting and accelerate convergence. At the same time, regularization was applied through dropout layers and batch normalization after each convolution.

To investigate whether the base model was already optimal, two experimental models were designed. Model 2 tuned learning-related hyperparameters by keeping the base architecture constant and varying batch sizes (16, 32, and 64) with epochs set to 15, 25, and 50, producing eight labeled variants. This setup enabled a systematic analysis of how batch size and training duration affect learning efficiency and accuracy.

Meanwhile, Model 3 explored hyperparameters related to network architecture. Using the same training configuration as the base model (batch size of 32 and 15 epochs), the number of neurons in the fully connected

layer and filters in each convolutional layer varied. Three variants were tested: half, double, and quadruple the original values. These architectural modifications aimed to assess the impact of model capacity would affect classification performance, training time, and susceptibility to overfitting.

3. RESULTS AND DISCUSSION

3.1 Base Model

After training, the base model was able to obtain a test accuracy of 88.78%. This indicates that the base model has good performance as a baseline tool for detecting pneumonia as it is capable of correctly classifying the majority of the cases on the test dataset.

Table 1. Base Model Classification Report

	Precision	Recall	F1-score	Support
Pneumonia	0.93	0.89	0.91	390
Normal	0.83	0.88	0.85	234
Accuracy			0.89	624
Macro Avg	0.88	0.89	0.88	624
Weighted Avg	0.89	0.89	0.89	624

Table 1 lists the classification report of the base model after performing model prediction. An 89% accuracy was achieved on the testing data. The base model demonstrated a macro average F1-score of 0.88 and weighted average F1-score of 0.89, indicating a strong balanced performance across both classes.

In the Pneumonia class, a high precision and a recall of 0.89 was observed, showing that most cases were correctly identified with strong confidence. The Normal class had a 0.83 precision and a 0.88 recall, suggesting that there are some misclassifications, but the majority are still correctly identified. Overall, the F1-scores of 0.91 for the Pneumonia class and 0.85 Normal class indicate a good diagnostic performance.

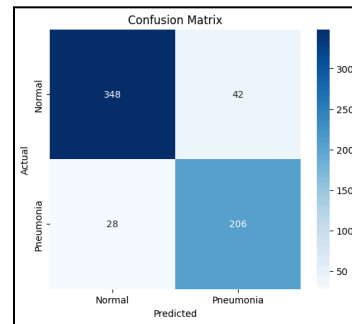


Fig. 2. Confusion Matrix of Base Model

Figure 2 shows the resulting confusion matrix of the base model. As shown, 206 Pneumonia cases and 348 Normal cases were correctly classified. 42 Normal cases were misidentified as Pneumonia, whereas 28 Pneumonia instances were incorrectly labeled as Normal.

3.2 Hyperparameters Related to Learning

Table 2. Summary of Model 2 Accuracy Results

Model	Batch Size	Epochs	Test Accuracy	Pneumonia F1	Normal F1
2.1	16	15	91.19%	0.93	0.88
2.2	16	25	92.31%	0.94	0.89
2.3	16	50	91.03%	0.93	0.87
2.4	32	15	88.78%	0.91	0.85
2.5	32	25	91.19%	0.93	0.88
2.6	32	50	90.54%	0.93	0.87
2.7	64	15	82.69%	0.85	0.80
2.8	64	25	90.22%	0.92	0.86
2.9	64	50	90.22%	0.92	0.86

Table 2 lists the summary of the performance metrics results across various combinations of batch

sizes and epochs, and they highlighted the impact of the hyperparameters used related to learning on the CNN model's performance in binary image classification of pneumonia and normal chest X-rays.

Model 2.2 achieved the highest test accuracy at 92.31% with the highest F1 score as well in the Pneumonia and Normal classification. This suggests that extending the training period allowed the model to better learn the meaningful patterns without overfitting.

Table 3. Model 2.2 Classification Report

	Precision	Recall	F1-score	Support
Pneumonia	0.92	0.96	0.94	390
Normal	0.93	0.86	0.89	234
Accuracy			0.92	624
Macro Avg	0.92	0.91	0.92	624
Weighted Avg	0.92	0.92	0.92	624

Table 3 displays the classification report for Model 2.2 which achieved a 92% accuracy. With both macro and weighted average F1-scores of 0.92, even with class imbalance, the model performed well across classes.

For the Pneumonia class, a 0.92 precision and 0.96 recall indicates a strong and confident detection. The Normal class had a precision of 0.93 and a recall of 0.86, suggesting the majority of it being correctly identified but with a slightly lower recall. F1-scores were 0.94 for Pneumonia and 0.89 for Normal, displaying overall strong and balanced performance.

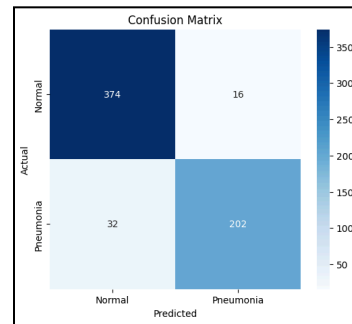


Fig. 3. Confusion Matrix of Model 2.2

Figure 3 depicts the confusion matrix of Model 2.2 after performing the model prediction. As shown, 202 Pneumonia cases and 374 Normal cases were correctly classified. Meanwhile, 16 Normal cases were incorrectly classified as Pneumonia and 32 Pneumonia instances were misidentified as Normal.

3.3 Hyperparameters Related to Network Architecture

Table 4. Summary of Model 3 Accuracy Results

Model	Number of Filters/Neurons	Test Accuracy	Pneumonia F1	Normal F1
3.1	HALF (8, 16, 32, 64, and 32)	88.14%	0.91	0.84
3.2	BASE (16, 32, 64, 128 and 64)	88.78%	0.91	0.85
3.3	DOUBLE (32, 64, 128, 256, and 128)	90.06%	0.92	0.86
3.4	QUADRUPLE (64, 128, 256, 512 and 256)	90.71%	0.93	0.87

Table 4 lists the varying impact of scaling the number of filters and neurons in the convolutional and fully connected layers while keeping all other training configurations identical to those used in the base model.

The best performance was observed in Model 3.4 with quadruple the base number of filters and neurons which resulted in an accuracy of 90.71% and an

increased F1 score as well. Increasing the number of filters and neurons significantly improves the learning capability of the model and its classification performance, particularly in distinguishing between the subtle differences in pneumonia and normal chest X-ray images.

Table 5. Model 3.4 Classification Report

	Precision	Recall	F1-score	Support
Pneumonia	0.90	0.95	0.93	390
Normal	0.92	0.83	0.87	234
Accuracy			0.91	624
Macro Avg	0.91	0.89	0.90	624
Weighted Avg	0.91	0.91	0.91	624

Table 5 shows the resulting report for Model 3.4 with quadruple the base number of filters and neurons. It demonstrated a strong overall performance as it achieved an accuracy of 91%. It shows high effectiveness in accurately classifying pneumonia cases as given by the Pneumonia class, reaching a precision of 0.90, a recall of 0.95, and an F1-score of 0.93. The model performs slightly less accurately for the Normal class with a precision of 0.92 but a lower recall of 0.83. It might occasionally misclassify normal cases as pneumonia.

Although the macro and weighted average shows a slight tendency to over-predict pneumonia, this bias is often preferred where false negatives are more dangerous than false positives like clinical screenings.

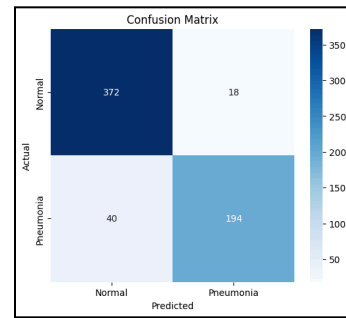


Fig. 4. Confusion Matrix of Model 3.4

Figure 4 shows the resulting confusion matrix of Model 3.4 after performing the model prediction. As shown, 194 Pneumonia cases and 372 Normal cases were correctly classified. On the other hand, only 18 Normal instances were mislabeled as Pneumonia and 40 Pneumonia cases were misdiagnosed as Normal.

3.4 Clinical Relevance Interpretability

Performance metrics are essential for assessing the clinical reliability of CNN models. Accuracy determines the overall correctness of the model's predictions. However, a high accuracy cannot be solely relied upon to guarantee reliability, especially in scenarios with imbalanced datasets (Sabour et al., 2017).

Precision represents how often the model's positive predictions are correct, while recall indicates how many actual positive cases the model was able to detect. A high precision helps minimize unnecessary anxiety caused by false positives, while a high recall is critical to prevent missed pneumonia diagnoses that could delay treatment (Cullerne Bown, 2024).

The F1-score evaluates the balance between detecting pneumonia cases and minimization of false positives and negatives (Google Developers, 2025). A high F1 score demonstrates the model's capability to classify pneumonia cases accurately while reducing false positives and negatives, proving its reliability in clinical settings.

The pediatric focus of the model helps lessen misdiagnosis which may lead to either overtreatment or dangerous delays in care, as the radiographic



presentation of pneumonia in young children can differ substantially from adults (O'Grady et al., 2014).

As pneumonia remains a leading cause of morbidity in children under five, particularly in low or middle-income countries, early and accurate diagnosis is needed to reduce this burden but access is limited in these resource-constrained areas. Fine-tuning the neural network model performance addresses this gap by being an effective decision support tool trained specifically on pediatric chest X-ray data, especially in environments with limited access to radiologists or high patient volumes.

Once accessible to the public, this model can support healthcare delivery in remote areas by optimizing resource allocation (Li et al., 2023). However, clinical integration requires interpretability and transparency to ensure patient safety and gain professional trust. The study's model can be a stepping stone to allowing healthcare professionals to validate the output of the model and be informed of its decision-making process through techniques like adding Gradient-weighted Class Activation Mapping (Grad-CAM) (Parikh & Matthew, 2023).

This integration enhances diagnostic efficiency and accuracy in this vulnerable age group while making sure that AI tools complement rather than replace clinical expertise and that they maintain the central role of clinicians in patient care. It contributes a scalable solution that can be integrated into radiology workflows and mobile health units, especially in pediatric settings where timely diagnosis is life-saving.

4. CONCLUSIONS

Deep learning architecture such as Convolutional Neural Networks can be utilized to detect Pneumonia through chest X-rays. This would significantly aid medical professionals, as Pneumonia became one of the leading causes of death, especially to children under the age of five.

This paper successfully developed and optimized a CNN model for the binary classification of chest X-ray images, specifically targeting the detection of Pneumonia versus Normal cases. Through

Sequential modeling with varying batch sizes, epochs, and the number of filters/neurons per layer, the study identified optimal training configurations that improved model performance. Notably, increasing the architectural complexity of the CNN resulted in improved test accuracy and F1 scores with the best performing model achieving 92% accuracy, an F1 score of 0.94 for Pneumonia and 0.89 for Normal. Moreover, smaller batch sizes combined with a moderate number of epochs tend to enhance the model generalization and accuracy. Conversely, larger batch sizes (especially when paired with fewer epochs) lead to poorer model performance.

Several enhancements are recommended to further improve model performance and applicability. Integrating transfer learning using pre-trained models like ResNet or EfficientNet may be considered for better feature extraction, especially for limited datasets. Additionally, expanding the dataset to include different age groups, severity levels, and other pneumonia-based complications could improve generalizability as the study is only limited to the chest X-ray images of children aged one to five years old.

5. REFERENCES

- Apeinans, I., Kodors, S., & Zarembo, I. (2024). Large Language Models: Comparison of Cross-Entropy and Binary Cross-Entropy Loss. *Human Environment. Technology*, (28), 5–8. <https://doi.org/10.17770/het2024.28.8251>
- Cleveland Clinic. (2025). *Pneumonia: Causes, symptoms, diagnosis & treatment*. Cleveland Clinic. <https://my.clevelandclinic.org/health/diseases/4471-pneumonia>
- Cullerne Bown, W. (2024). Sensitivity and specificity versus precision and recall, and related dilemmas. *Journal of Classification*, 41(3), 402–426. <https://doi.org/10.1007/s00357-024-09478-y>
- Daniel, S. (2023, July 8). *Activation function and loss function for classification and regression analysis*. Medium.

<https://medium.com/@stephy.SD99/activation-function-and-loss-function-for-classification-and-regression-analysis-f1e765cb96d>

Google Developers. (2025, March 3). *Classification: Accuracy, recall, precision, and related metrics*. Google Developers.
<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

Johns Hopkins Medicine. (2025). *Pneumonia*. Johns Hopkins Medicine.
<https://www.hopkinsmedicine.org/health/conditions-and-diseases/pneumonia>

Li, M., Jiang, Y., Zhang, Y., & Zhu, H. (2023). Medical image analysis using deep learning algorithms. *Frontiers in Public Health*, 11(1273253).
<https://doi.org/10.3389/fpubh.2023.1273253>

Mayo Clinic. (2020, June 13). *Pneumonia - Diagnosis and treatment*. Mayo Clinic.
<https://www.mayoclinic.org/diseases-conditions/pneumonia/diagnosis-treatment/drc-20354210>

Mayo Clinic. (2020, June 13). *Pneumonia - Symptoms and causes*. Mayo Clinic.
<https://www.mayoclinic.org/diseases-conditions/pneumonia/symptoms-causes/syc-20354204>

Mooney, P. (2018). *Chest X-Ray Images (Pneumonia)*. Kaggle.
<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>

National Health Service. (2023, January 12). *Pneumonia*. NHS. <https://www.nhs.uk/conditions/pneumonia/>

O'Grady, K.-A. F., Torzillo, P. J., Frawley, K., & Chang, A. B. (2014). The radiological diagnosis of pneumonia in children. *Pneumonia*, 5(S1), 38–51.
<https://doi.org/10.15172/pneu.2014.5/482>

One Step Diagnostic. (2024, May 14). *Chest X-Ray for Pneumonia: Diagnosis and Interpretation*. One Step Diagnostic.
<https://onestepdiagnostic.com/chest-x-ray-for-pneumonia-diagnosis/>

Parikh, K. V., & Mathew, T. J. (2023). COVision: convolutional neural network for the differentiation of COVID-19 from common pulmonary conditions using CT scans. *BMC Pulmonary Medicine*, 23(1).
<https://doi.org/10.1186/s12890-023-02723-x>

Philippine Statistics Authority. (2024, November 28). *2024 Causes of Death in the Philippines*. Philippine Statistics Authority.
<https://psa.gov.ph/content/2024-causes-death-philippines-provisional-30-september-2024>

Sabour, S., Abbasnezhad, O., Mozaffarian, S., & Kangavari, H. N. (2017). Accuracy and precision in medical researches; common mistakes and misinterpretations. *World J. Res. Rev.(WJRR)*, 4, 058-060.

Thomas, L. (2020, December 18). *Transfer learning exploits chest-Xray to diagnose COVID-19 pneumonia*. News-Medical.Net.
<https://www.news-medical.net/news/20201218/Transfer-learning-exploits-chest-Xray-to-diagnose-COVID-19-pneumonia.aspx>

World Health Organization. (2022, November 11). *Pneumonia in children*. World Health Organization.
<https://www.who.int/news-room/fact-sheets/detail/pneumonia>