

Detection of Pneumonia in Children through Chest Radiographs using Artificial Intelligence in a Low-Resource Setting: A Pilot Study

Taofeeq Oluwatosin Togunwa ^{1,2,*}, Abdulhammed Opeyemi Babatunde ^{1,2}, Oluwatosin Ebunoluwa Fatade ³, Richard Olatunji ³, Godwin Ogbole ³, Adegoke Falade ⁴

¹ College of Medicine, University of Ibadan, Ibadan, Nigeria

² College Research and Innovation Hub, Ibadan, Nigeria

³ Department of Radiology, University College Hospital, Ibadan, Nigeria

⁴ Department of Paediatrics, University College Hospital, Ibadan, Nigeria

* Corresponding author

E-mail: togunwataofeeq@gmail.com

Abstract

Background: Pneumonia is a leading cause of death among children under 5 years in low- and-middle-income-countries (LMICs), causing an estimated 700,000 deaths annually. This burden is compounded by limited diagnostic imaging expertise. Artificial intelligence (AI) has potential to improve pneumonia diagnosis from chest radiographs (CXRs) through enhanced accuracy and faster diagnostic time. However, most AI models lack validation on prospective clinical data from LMICs, limiting their real-world applicability. This study aims to develop and validate an AI model for childhood pneumonia detection using Nigerian CXR data.

Methods: In a multi-center cross-sectional study in Ibadan, Nigeria, CXRs were prospectively collected from University College Hospital (a tertiary hospital) and Rainbow-Scans (a private diagnostic center) radiology departments via cluster sampling (November 2023–August 2024). An AI model was developed on open-source paediatric CXR dataset from the USA, to classify the local prospective CXRs as either normal or pneumonia. Two blinded radiologists provided consensus classification as the reference standard. The model's accuracy, precision, recall, F1-score, and area-under-the-curve (AUC) were evaluated.

Results: The AI model was developed on 5,232 open-source paediatric CXRs, divided into training (1,349 normal, 3,883 pneumonia) and internal test (234 normal, 390 pneumonia) sets, and externally tested on 190 radiologist-labeled Nigerian CXRs (93 normal, 97 pneumonia). The model achieved 86% accuracy, 0.83 precision, 0.98 recall, 0.79 F1-score, and 0.93 AUC on the internal test, and 58% accuracy, 0.62 precision, 0.48 recall, 0.68 F1-score, and 0.65 AUC on the external test.

Conclusion: This study illustrates AI’s potential for childhood pneumonia diagnosis but reveals challenges when applied across diverse healthcare environments, as revealed by discrepancies between internal and external evaluations. This performance gap likely stems from differences in imaging protocols/equipment between LMICs and high-income settings. Hence, public health priority should be developing robust, locally relevant datasets in Africa to facilitate sustainable and independent AI development within African healthcare.

Keywords: childhood pneumonia; artificial intelligence; chest radiographs; low-and-middle-income countries; diagnostic accuracy

Author Summary

Pneumonia is a leading cause of death in children under five, especially in low-resource settings like Nigeria, where access to diagnostic tools and expertise is limited. Our study explores how artificial intelligence (AI) can help address this gap by detecting pneumonia from chest X-rays. We trained an AI model using a large dataset of children’s X-rays from the United States and tested it on images collected in Nigeria.

While the AI model performed well on the U.S. data, its accuracy dropped significantly when tested on the Nigerian X-rays. This reveals how differences in imaging techniques and equipment between countries can affect the performance of such models. It highlights the need for AI systems to be adapted to local contexts to ensure they are reliable and effective in real-world settings.

Our findings underline the importance of creating high-quality, locally relevant datasets in Africa to support the development of AI tools that address the unique challenges of the region. By

investing in such efforts, we can improve access to life-saving technologies, particularly for vulnerable populations in resource-limited healthcare systems.

Introduction

Globally, pneumonia remains a leading cause of under-5 mortality.¹ In 2015, approximately 700,000 under-5 children died from pneumonia worldwide, mostly in low and middle income countries (LMICs).² Sub-Saharan Africa (SSA) continues to bear a significant portion of this burden, with Nigeria accounting for 169,000 under-5 community-acquired pneumonia deaths in 2021, the highest absolute number for any single country.³ Nigeria is not on track to achieve the Global Action Plan for Pneumonia and Diarrhoea targets to reduce mortality to 3 per 1,000 by 2025.⁴ This highlights the need for urgent, effective, evidence-based interventions.

Childhood pneumonia in LMICs is largely driven by viral pathogens like respiratory syncytial virus (RSV). However, bacterial pathogens contribute remarkably to moderate to severe disease, particularly in children with underlying health conditions.⁵ Diagnosis in sub-Saharan Africa (SSA) heavily relies on clinical judgment.⁶ Additional danger signs like feeding difficulties, convulsions and hypothermia categorize severe cases. While most children recover fully, around 3-5% experience pulmonary or systemic complications, contributing to significant morbidity and mortality.⁷ The economic impact in Nigeria, though not precisely quantified, likely mirrors findings from Uganda where almost 40% of households face catastrophic health expenditures due to childhood pneumonia treatment costs.⁸

Standard diagnostic protocols for childhood pneumonia in LMICs typically involve clinical assessment supplemented by laboratory and microbiological testing. However, imaging becomes essential if outpatient treatment fails, or in cases of hospital admission or hospital-acquired pneumonia.⁹ Chest radiographs are the primary imaging method for childhood pneumonia; nonetheless, significant variability in interpretation among radiologists is a common limitation of this technique.⁹ Moreover, paediatric radiology in LMICs faces additional challenges: scarcity of specialized equipment, radiation safety concerns, and a shortage of trained personnel.^{10, 11} Nigeria faces a significant shortage of healthcare resources, with a doctor-to-population ratio of 0.17 per 1,000, among the lowest in Africa.¹² Additionally, there are only about 250 to 300 radiologists, roughly one for every 658,000 people.¹² Furthermore, the country suffers from immigration of healthcare workers to developed countries.¹³ This healthcare exodus deprives African countries of crucial expertise, particularly in specialties like radiology. The resulting disparities highlight the urgent need for innovative interventions to enhance early pneumonia diagnosis in children.

Artificial intelligence (AI) has made significant strides in clinical medicine and healthcare over the past decade. This era of digital medicine heralds technological advances, especially in diagnostic tools, enhancing care quality and response times.¹⁴ Convolutional neural networks (CNNs) are a widely used AI tool in image processing and classification; renowned for their pattern detection capabilities.¹⁵ First implemented by Fukushima in 1980 and formalized by LeCun in 1998, CNNs use convolutions, mathematical operations that allow the network to identify patterns by integrating small image regions. CNNs have shown promise in diagnosing diseases such as tuberculosis, pneumoconiosis, pneumonia, and COVID-19.¹⁶ Despite the emergence of newer visual transformer models¹⁷, CNNs remain valuable due to their extensive use, simplicity, and strong research foundation.

Ayan, Karabulut ¹⁸ developed an AI paediatric pneumonia detection system using an ensemble of CNN models (VGG-16, VGG-19, ResNet-50, Inception-V3, MobileNet, SqueezeNet, Xception). Utilizing the ensemble method outperformed individual models, achieving an AUC of 95.21% and sensitivity of 97.76% on test data. Similar efforts by Labhane, Pansare ¹⁹ developed four models (basic CNN, VGG16, VGG19, InceptionV3), and used transfer learning techniques for paediatric pneumonia detection.

Furthermore, a recent systematic review by Field, Tam ²⁰ evaluated the efficacy of AI in classifying paediatric pneumonia on chest radiographs. Their results revealed an ensemble AI achieved 96.3% sensitivity, DenseNet201 showed 94% specificity and 95% accuracy, and VGG16 had a 96.2% AUC. Some AI models nearly reached 100% diagnostic accuracy. The review emphasized AI's potential to enhance diagnosis but called for comparisons with radiologists and further clinical validation.

Many of these models have scarcely been tested in clinical settings, and similar models have not been developed in Africa for childhood pneumonia detection. Therefore, this study aims to develop an AI model, using transfer learning with CNNs, to detect pneumonia in under-5 children, utilizing chest radiographs from patients in Ibadan, Nigeria.

Results

Data

The training set and internal test (validation) set by Kermany, Zhang ²¹, include 5,232 chest radiographs and 624 chest radiographs, respectively. The training set comprises 1,349 normal chest

radiographs and 3,883 pneumonia chest radiographs, while the internal test set contains 234 normal chest radiographs and 390 pneumonia chest radiographs. These details are depicted in Figure 1.

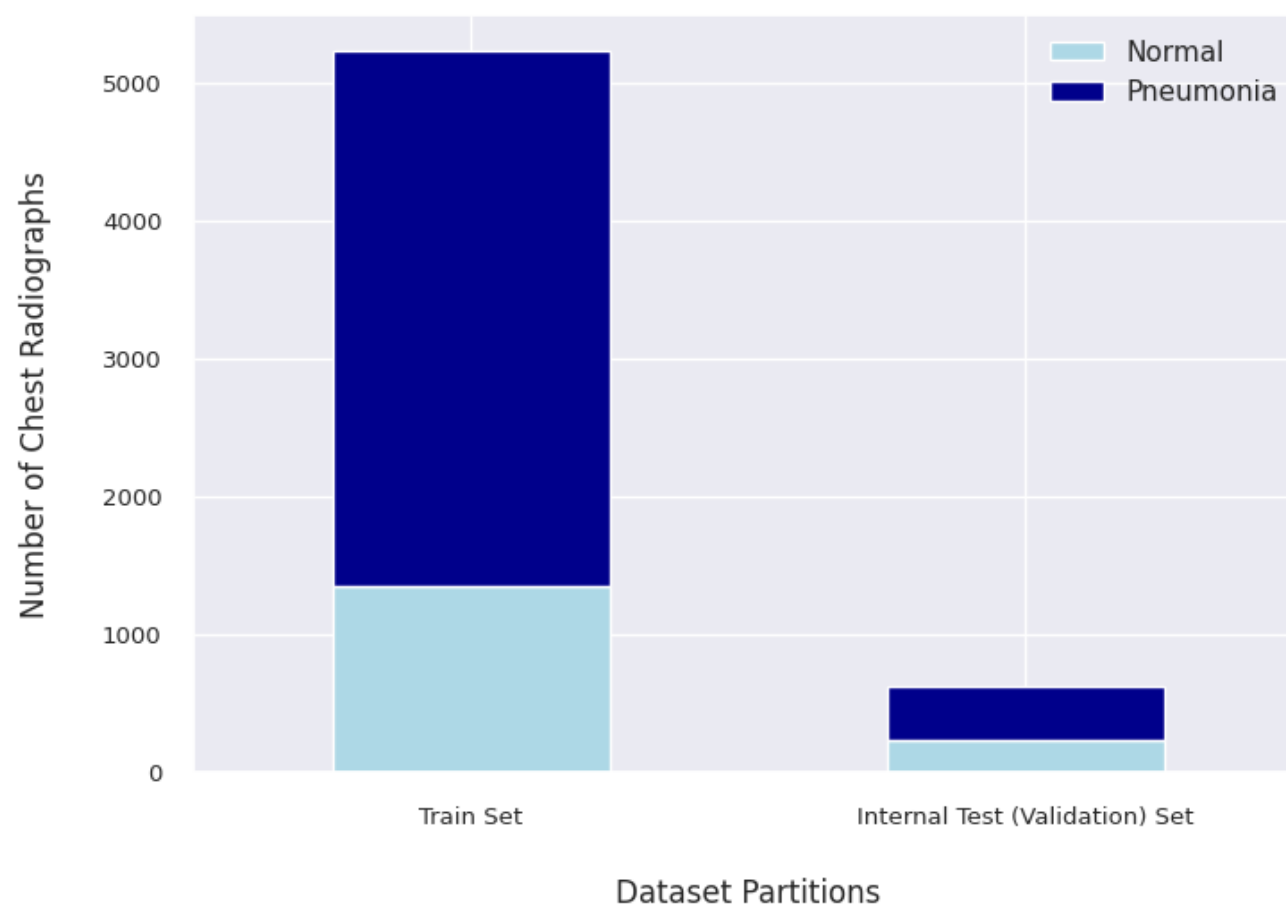


Figure 1. Distribution of Radiographs in the Training and Internal Test Sets

The training set includes 1,349 normal and 3,883 pneumonia chest radiographs, while the internal test set consists of 234 normal and 390 pneumonia chest radiographs

The external test set consists of 190 chest radiographs, selected according to the inclusion and exclusion criteria. Of these, 93 were normal, and 97 showed pneumonia. This distribution is illustrated in Figure 2.

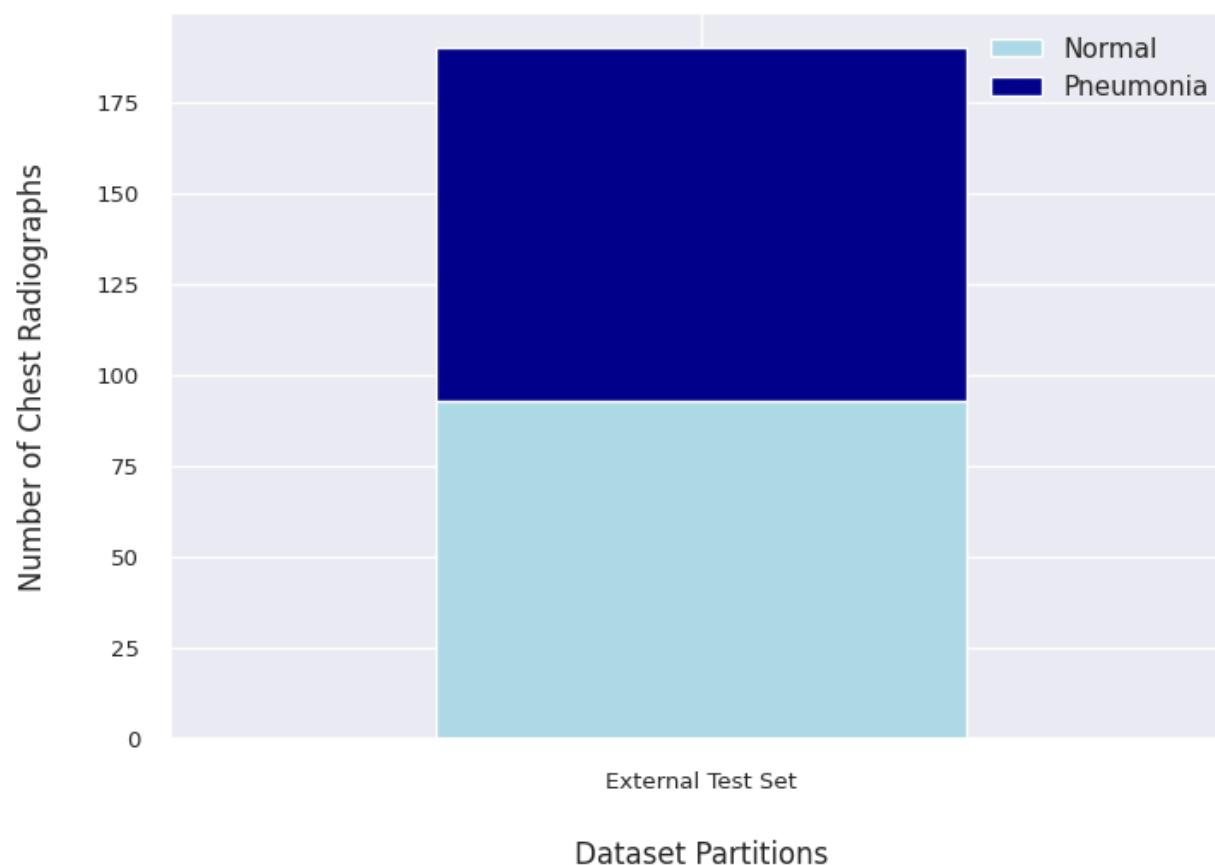


Figure 2. Distribution of the External Test Set

The test set includes 93 normal chest radiographs and 97 pneumonia chest radiographs

Model Performance

Training and Validation (Internal Test). The base model and the fine-tuned model are evaluated across accuracy, precision, recall, F1 score, and area under the curve (AUC).²² The initial hyperparameter optimization using Optuna²³ for the base model identified the optimal settings as a maximum pooling strategy, a dropout rate of 0.25, and an initial learning rate of 4.23×10^{-3} .

Subsequently, the learning rate for the fine-tuning process was found to be optimal at 2.12×10^{-4} , lower than the initial value. The model was initially trained for 6 epochs and then fine-tuned for another 8 epochs. The results from the training and internal testing (validation) of the model are presented in Figure 3.

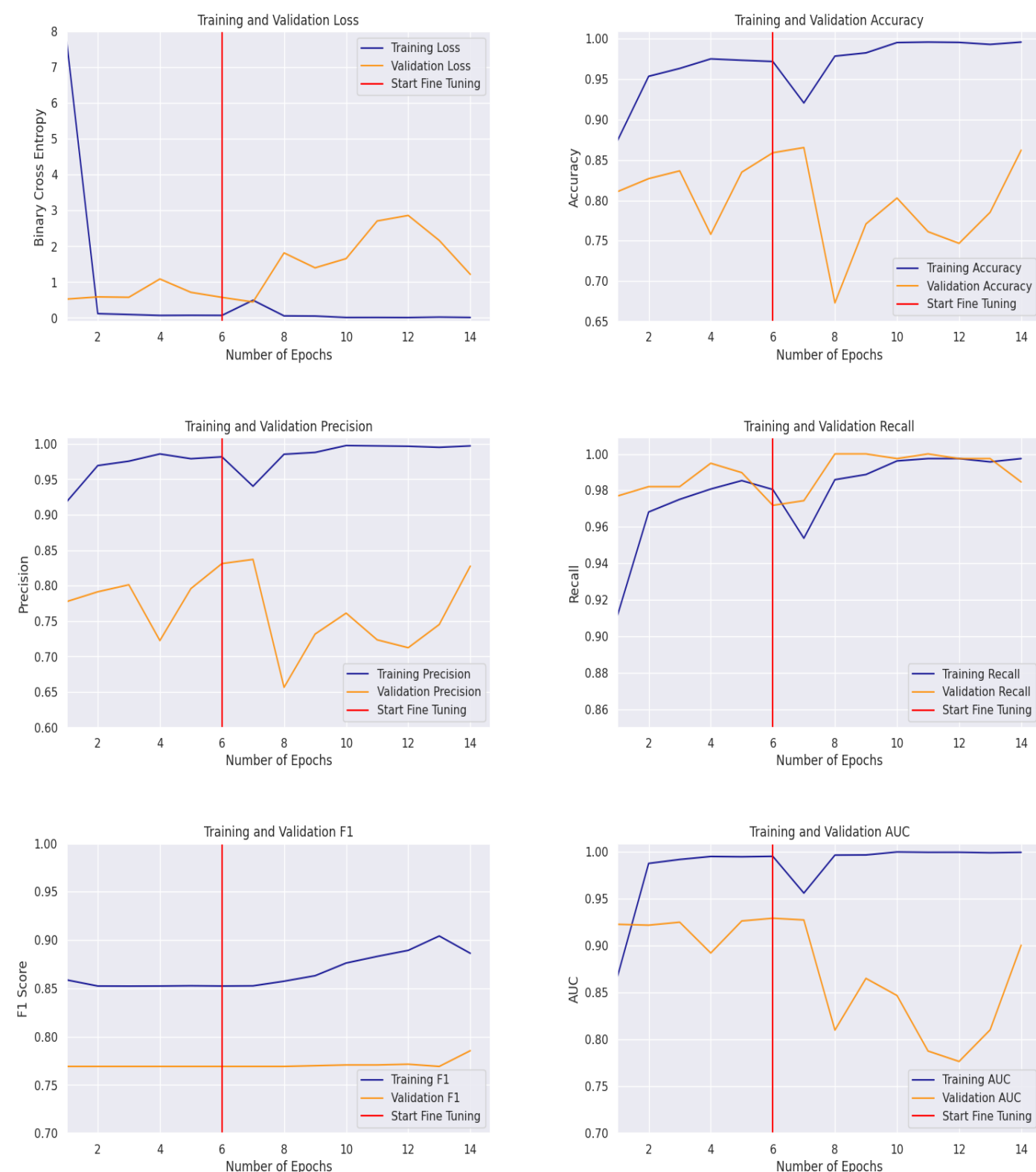


Figure 3. Training and Validation Metrics Over Epochs: Loss, Accuracy, Precision, Recall, F1, and AUC

Metrics show the model's performance trends before and after fine-tuning, across training and validation sets, providing insights into model behaviour

The training loss drops sharply after the first epoch and stabilizes at a low value of 0.07. The validation loss also begins low but fluctuates, increasing slightly after the fourth epoch before returning to 0.57. Following fine-tuning, the training loss remains consistently low at 0.013. However, the validation loss rises briefly between the tenth and twelfth epochs, before a stable decline to 1.22—still, a higher value than before fine-tuning.

The accuracy plot reveals a steady improvement in training accuracy across epochs, eventually reaching close to 0.95 pre-fine tuning. Validation accuracy shows an overall upward trend, though it fluctuates significantly—starting at 0.81, dropping sharply in the fourth epoch, and recovering to 0.86 by the sixth epoch. Following fine-tuning, the training accuracy remains near-perfect at 1.00, confirming the model's strong performance on the training set. However, validation accuracy remains volatile but stabilizes at 0.86 by the fourteenth epoch.

The model demonstrates strong precision during training and validation, with pre-fine-tuning scores of 0.98 and 0.83, respectively. After fine-tuning, the precision remains consistent, reaching 1.00 for training and maintaining 0.83 for validation. The lower validation precision suggests a tendency toward false positives. Recall values are stable, with pre-fine-tuning scores of 0.98 for

training and 0.97 for validation, improving to 1.00 and 0.98 after fine-tuning, indicating the model effectively captures positive cases.

The F1 score, which balances precision and recall, was initially 0.85 for the training set and 0.77 for the validation set. After fine-tuning, these values improved slightly to 0.89 and 0.79, respectively. The AUC plot reflects a similar trend, with a training AUC of 0.99 and a validation AUC of 0.93 before fine-tuning, and 0.99 and 0.90 after.

External Test. The base model and the fine-tuned model are evaluated on the external test set. The results are presented in Table 1, comparing both performances across various metrics with 95% confidence intervals.

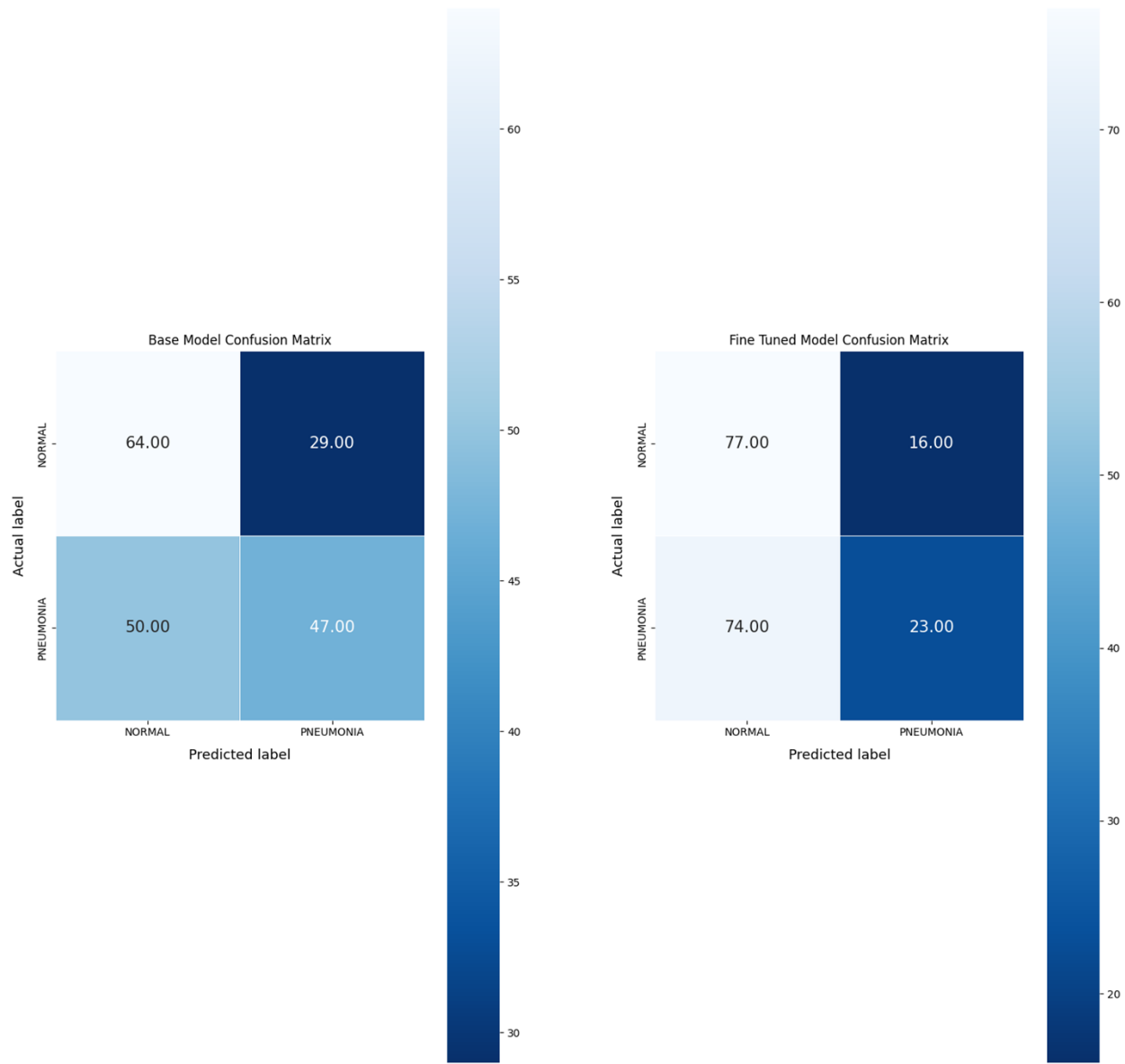
	Loss	Accuracy	F1	Precision	Recall	AUC
Base	1.10	0.58	0.68	0.62	0.48	0.65
Fine Tuned	3.40	0.53	0.67	0.59	0.24	0.57

Table 1. Comparison of Model Performance Metrics on External Test Set

The base model demonstrates a lower loss (1.10) and higher accuracy (0.58) compared to the fine-tuned model, which has a loss of 3.40 and accuracy of 0.53. The F1 score is similar for both models, with the base at 0.68 and the fine-tuned at 0.67, indicating comparable balance between precision and recall. However, precision is slightly better in the base model (0.62) versus the fine-tuned model (0.59), while recall significantly drops in the fine-tuned model (0.24) compared to the base model (0.48). The area under the curve (AUC) also shows a decline from 0.65 in the base model to 0.57 in the fine-tuned model.

201 Further error analysis is presented through the comparison of the confusion matrices of the base
202 and fine-tuned models (Figure 4) to assess how each model balances sensitivity and specificity,
203 particularly in distinguishing pneumonia cases.

204



205

206

Figure 4. Comparison of Confusion Matrices for Base Model and Fine-Tuned Model Classification Performance

The base model correctly identified 64 out of 93 normal cases and 47 out of 97 pneumonia cases. It misclassified 29 normal radiographs as pneumonia and 50 pneumonia radiographs as normal. The overall performance of the base model demonstrated relatively balanced prediction accuracy between both categories, though it struggled more with distinguishing pneumonia cases.

After fine-tuning, the model exhibited an improvement in identifying normal cases, correctly classifying 77 out of 93 (82.8%) normal radiographs. False positives, where normal cases were misclassified as pneumonia, decreased from 29 to 16. However, the fine-tuned model's performance in identifying pneumonia cases declined. It correctly classified only 23 out of 97 pneumonia cases (23.7%), misclassifying 74 pneumonia radiographs as normal, resulting in a notable increase in false negatives.

While the fine-tuned model improved its accuracy in identifying normal cases, it showed reduced effectiveness in detecting pneumonia, leading to an offset between sensitivity and specificity compared to the base model.

Discussion

This study demonstrates the potential of using artificial intelligence (AI) to detect childhood pneumonia from chest radiographs in resource-limited settings like Nigeria. The model, based on the VGG-19 architecture, was able to train effectively, minimizing loss and achieving reasonable performance on internal test data. However, when tested on the external dataset from Nigeria, both

the base and fine-tuned models struggled, suggesting limitations in generalizing to local image data. It is likely that differences in image acquisition processes and subtle variations, such as image resolution, between the US and Nigerian datasets may have influenced the model's performance. Image resolution is a critical factor in medical imaging, as it directly affects the visibility of fine details necessary for accurate diagnosis. Differences in pixel spacing, compression, or preprocessing between datasets can introduce domain shifts that impact model generalizability.²⁴

²⁵ Studies have demonstrated that higher-resolution images improve classification performance in deep learning medical imaging tasks, while lower resolutions may lead to the loss of critical diagnostic information and reduced model accuracy.^{26, 27}

Additionally, fine-tuning did not seem to improve the model's performance, potentially due to catastrophic forgetting—a phenomenon where previously learned knowledge is overwritten when new, limited data is introduced.²⁸ In this case, since the VGG-19 model was pre-trained on the ImageNet dataset, which contains over 1 million images, fine-tuning on a much smaller dataset of only a few thousand images may have caused the model to 'forget' some of the general features it had learned from the larger dataset, leading to the inability to generalize well on an external dataset.²⁹

One of the main strengths of this study is its pioneering approach, as it is the first to model an AI architecture for pneumonia detection in Nigeria using a large open-access dataset and testing on local imaging data. The use of transfer learning also aligns with the realities of low-resource settings, where computational and data resources are often constrained.³⁰ Another strength lies in the rigorous comparison between base and fine-tuned models, offering insights into the limitations of model adaptation in different environments.

However, the study has some weaknesses. The reliance on a single CNN architecture (VGG-19) may have limited the model's ability to generalize, particularly when faced with local variations in image quality and acquisition protocols. Additionally, the relatively small external test dataset may have constrained the evaluation of the model's robustness. Lastly, the study did not explore alternative ensemble models or newer architectures like vision transformers (VT), which could potentially perform better in heterogeneous datasets.^{31, 32}

Compared to studies like Mabrouk, Díaz Redondo³³, which used an ensemble approach to achieve an F1-score of over 0.9, this study's use of a single model may have contributed to the lower overall performance. Gummadi, Vootla³⁴, who similarly employed three different architectures (VGG-16, VGG-19, and Inception ResNet V2), reported significantly higher accuracy (95.82%) and other metrics. However, these studies did not assess how well the models generalize on out-of-domain or racially different data, an aspect this study has highlighted.

Nonetheless, recent authors have continued to emphasize the problem of models performing well on internal data but failing on external dataset.³⁵ Glocker, Jones³⁶ highlighted that models built on top of the chest radiography foundation model consistently underperformed compared to the CheXpert model, resulting in differences in relative performance and raising concerns about subgroup disparities. Our study reinforces the idea that while a model may perform well during internal validation, differences in regional data may lead to significant performance drops when the model is applied to new populations.

Particularly, AI models trained on datasets from high-resource settings may not generalize well to local contexts in low-resource settings, such as Nigeria. Differences in radiomics between regions, as well as variances in imaging techniques and quality, likely contribute to this limitation. For clinicians and policymakers, this highlights the importance of localized model development and

validation. Clinicians in resource-limited settings might need AI tools specifically adapted to local data to achieve reliable diagnostic support. Policymakers should prioritize the development of localized, robust, and high-quality imaging databases in Nigeria. These databases would support the development of AI models tailored to the specific needs of the population and improve diagnostic accuracy in these settings.

Several questions remain unanswered. First, the precise nature of the radiomic differences between Nigerian and US datasets requires further investigation. This would help in understanding how to better tailor AI models to local data. Additionally, the question of whether ensemble models or newer architectures, such as vision transformers (VT)³², could better generalize to diverse datasets needs exploration. Future research should consider using ensemble models combining multiple CNN architectures or a mix of CNNs and VTs, or even exclusive ensembles of VTs, to better capture global and local image features. This should focus on evaluating the interoperability and generalization performance of newer AI architectures on diverse datasets, including those from data-sparse, low-resource regions, to draw more robust conclusions.

Conclusions

This study aimed to develop an AI model using transfer learning with CNNs to detect pneumonia in under-5 children, utilizing chest radiographs from patients in Ibadan, Nigeria. The results reveal a significant disparity in the model's performance when applied to internal versus external testing datasets, highlighting the challenges of generalizing AI models across diverse populations. This emphasizes the importance of tailoring AI solutions to specific local contexts and the critical need for high-quality, representative imaging databases in Nigeria to support independent AI development in medical imaging. Future research should also focus on building more robust and

generalizable AI models to enhance their reliability and applicability across diverse populations. Addressing these challenges is essential for the successful adoption of AI technologies in clinical practice, ultimately improving diagnostic capabilities and patient care in the region.

Methodology

This study was conducted in accordance with the recommendations outlined in the 2024 Update of the Checklist for Artificial Intelligence in Medical Imaging (CLAIM).³⁷

Study Design

This is a multicenter cross-sectional study that evaluated the performance of an AI model compared to a human radiologist in differentiating radiographs of under-5 children with pneumonia (variable 1) from those without pneumonia (variable 2). The study includes both retrospective and prospective elements: the AI model is trained and internally tested on a publicly available dataset and externally tested on data collected prospectively from the study centers.

This study was designed to develop and test an AI model for diagnosing pneumonia in under-5 children, using transfer learning with pre-trained convolutional neural networks (CNNs). The goals are to explore the model's potential effectiveness in a real-world clinical setting, assessing how well it can classify radiographs of children with pneumonia and those without, addressing a significant clinical need in resource-limited environments. This exploratory aspect is crucial given the limited prior clinical validation of similar models, especially in low-resource environments like Nigeria. This allows for an investigation into the model's feasibility to provide support for radiologists in diagnosing pneumonia.

Study Setting

The study was carried out at the University College Hospital (UCH), Ibadan and Rainbow Scans Diagnostic Center, Ibadan, from November 2023 to August 2024. UCH is the oldest and largest tertiary hospital in Nigeria where several cases of childhood pneumonia are managed. UCH has board certified radiologists who possess decades of experience in paediatric imaging. Additionally, Rainbow Scans is a private medical diagnostic center also located in Ibadan; offers a range of radiological services and expertise in paediatric conditions

Data Sources

The training data consisted of a paediatric pneumonia chest radiograph dataset, which was sourced from the publicly available repository created by Kermany, Zhang²¹ on the Kaggle platform. The dataset comprised chest radiographs of paediatric patients from University of California San Diego, USA, that were categorized as either normal or having pneumonia.

The test data comprised chest radiographs of paediatric patients aged 3 months to 5 years from the Radiology Department at the University College Hospital, Ibadan, Nigeria. This is as radiological features of pneumonia in patients below 3 months often overlap with congenital or neonatal conditions, while those older than 5 years may present with patterns resembling adult pneumonia, complicating the analysis. Additional radiographs were collected from the Rainbow Scans Diagnostic Center, Ibadan, Nigeria.

Inclusion Criteria. The chest radiographs with pneumonia must (1) be a confirmed radiological diagnosis of pneumonia by the consultant radiologist (2) have no complication on

radiograph (such as pleural effusion, pneumatocoele etc). The variable group without pneumonia will include chest radiograph of under-5 children with normal findings.

Exclusion Criteria. The exclusion criteria include: (1) the chest radiographs of children older than 5 years or younger than 3 months, (2) the chest radiographs with complications of pneumonia, and (3) chest radiographs diagnosed with respiratory conditions other than pneumonia.

Data De-identification and Preprocessing. The chest radiographs were accessed using the RadiAnt DICOM Viewer³⁸ (version 2023.1, 64-bit). During this process, personal information was removed to protect patient privacy. Specifically, all identifiers, including patient names, ages, hospital numbers, and gender, were eliminated from the images. Further preprocessing was performed using the default preprocessing function specific to the model. Details of this process will be discussed in the model section.

Image Acquisition Protocol

The anteroposterior projection is the alternative when posteroanterior projection cannot be done in under-5 children. The child is made to stand or sit upright or held upright depending on the age and level of cooperation. Hands are held above the head to remove the scapula from the lung fields. Immobilization devices such as the Pigg-o-stat are not available, and the accompanying adult puts on a lead jacket to protect them from scatter radiation. Images are taken when the child is calm (not crying or restless). Based on the child's age and size, the kilovolt peak (kVp) of 60-70 is used as the least setting and higher values for larger children, up to 90-125 kVp. The milliampereseconds (mAs) is adjusted based on the child's size and body habitus. The ALARA principle (As Low As Reasonably Achievable) is adopted. Shielding, especially of the gonads, is done when necessary.

Reference Standard

The reference standard for this study was established through radiographic evaluation, determined by consensus between experienced clinicians, specifically the Chief Resident of Radiology and a Consultant Radiologist at the University College Hospital, Ibadan. Each chest radiograph is examined for typical features consistent with pneumonia, such as focal or diffuse alveolar opacities, consolidation, or interstitial infiltrates.¹ Only radiographs with clearly confirmed pneumonia and no radiological complications (such as pleural effusion or pneumatocele) are included in the pneumonia group. Normal chest radiographs are classified as those without signs of pneumonia or any other pathology.

The rationale for selecting this reference standard lies in the clinical expertise of the radiologists involved. Using a dual assessment from both a Consultant Radiologist and the Chief Resident enhances the accuracy and reliability of the diagnoses, reducing the likelihood of misclassification. While this expert-driven standard ensures high accuracy, there is a potential limitation regarding variability in interpretation. In these cases of initial disagreement, an in-house consensus review meeting is conducted to resolve discrepancies.

Data Partitions

The dataset was partitioned into training, internal test (validation), and external test sets. The training set comprised 5,232 chest radiographs, accounting for 86.54% of the total dataset. The internal validation set contained 624 chest radiographs, representing 10.32% of the dataset, and was used for model optimization. The external test set, prospectively collected, included 190 chest radiographs, making up 3.14% of the dataset. To prevent overlap between partitions, radiographs in the external test set were disjointed at the patient level, ensuring that no patient appeared in

more than one partition. This approach mitigates the risk of data leakage and maintains the independence of the external test set for unbiased performance evaluation.

Testing Data

Sample Size Calculation and Sampling Technique. The sample size was calculated with the Leslie Fischer's formula;

$$\frac{Z^2 pq}{d^2}$$

z = appropriate value from the normal distribution for 95% confidence, 1.96

p = prevalence of pneumonia in under-5 children, 12.7% in southern Nigeria.³⁹

d = desired precision, 0.05.

$q = 1 - p$

Sample size (N) = $\frac{1.96^2 \times 0.127 \times (1 - 0.127)}{0.05^2}$

Hence, the estimated test sample size was 170. Using a 1:1 ratio for two variables considered, 85 chest radiographs each for subjects with pneumonia and without pneumonia (normal) are required. A cluster sampling was used to select chest radiographs of subjects with pneumonia and without pneumonia (normal) using the monthly register of chest radiographs of under-5 patients done by the Radiology Department, University College Hospital, and Rainbow Scans Diagnostics Centre. The eligible chest radiographs were selected from the dataset into clusters of pneumonia and (normal) no pneumonia. The chest radiographs were collected from November 2023 to August 2024, when the sample size was attained.

399 ***Model Development***

400 The model in this study is based on the VGG19 architecture⁴⁰ (Figure 5) and employs transfer
 401 learning, where all layers are frozen, and then the architecture is adapted for binary classification.
 402 The inputs to the model are RGB images of size 224x224x3 with a batch size of 64, preprocessed
 403 using the “preprocess_input” function from the TensorFlow implementation of VGG19. The base
 404 VGG19 model is loaded with pre-trained ImageNet weights and includes all layers up to the final
 405 convolutional block, with no additional fully connected layers i.e the default classification head
 406 was not included. The model is then set up with an additional fully connected layer of 512 units
 407 with ReLU activation, and then a drop out layer before a custom binary classification head.



Figure 5. Detailed Architecture of the Model

The neural network architecture includes an input layer for 224x224x3 preprocessed images. It utilizes a VGG19 base model with frozen weights up to the final convolutional block. Features then pass through a dense layer with ReLU units, followed by a dropout layer. The output layer, a single neuron with sigmoid activation, provides binary classification probabilities.

Model Training

The fully connected layers on top of the base model are initialized with Glorot uniform initialization⁴¹ (the default in TensorFlow). The model is then trained with the Adam optimizer⁴², with the learning rate tuned between $1e-5$ and $1e-2$. Early stopping is employed with a patience of 5 epochs, and training is limited to 20 epochs. Hyperparameter optimization is automated through Optuna²³, and spanned 25 trials. Dropout rates, pooling strategies, and learning rates are adjusted during this process to identify the best configuration. The model with the highest validation accuracy is selected. Validation accuracy serves as a benchmark for the model's expected performance on real-world, unseen data.

Model Fine-tuning

For fine-tuning, the last 7 layers of the VGG19 model are set to be trainable, while the remaining layers remain frozen. The model is similarly configured for binary classification, using a sigmoid output to predict the likelihood of the target class.⁴³ Training is carried out with the Adam optimizer, with the learning rate fine-tuned between the initial best rate and one-hundredth of that value. Early stopping is applied with a patience of 7 epochs, and training was limited to a maximum of 50 epochs. Binary cross-entropy served as the loss function. Hyperparameter optimization is automated using Optuna, which conducted 50 trials to find the best parameter combinations. The learning rate is fine-tuned during this process, and similarly, the model with the highest validation accuracy was selected as the optimal version.

Evaluation

The evaluation metrics include accuracy, precision, recall, F1 score, and area under the curve (AUC).²² Two evaluation rounds are conducted. First, the initial model using only transfer learning is assessed across the train, internal test, and external test sets. Next, the fine-tuned model is

evaluated using the same metrics and datasets. A comparison of the results from both rounds follows.

Acknowledgement

We sincerely appreciate the support provided by the faculty at the Duke Center for Policy Impact in Global Health, Duke University, for their funding and mentorship. We also extend our gratitude to the leadership of the University of Ibadan Medical Students Association (UIMSA) and the College of Medicine, University of Ibadan, for their invaluable guidance and collaboration. We also acknowledge the support of the Radiology Department, University College Hospital and Rainbow Diagnostics (worthy of mention, Mrs Maryam) for granting access to the radiographs database.

Funding

The study was funded by the UIMSA & Duke Center for Policy Impact in Global Health Mentored Student Research on Children's Health Disparities in Nigeria.

Conflict of Interest

Authors have no conflict of interest.

456

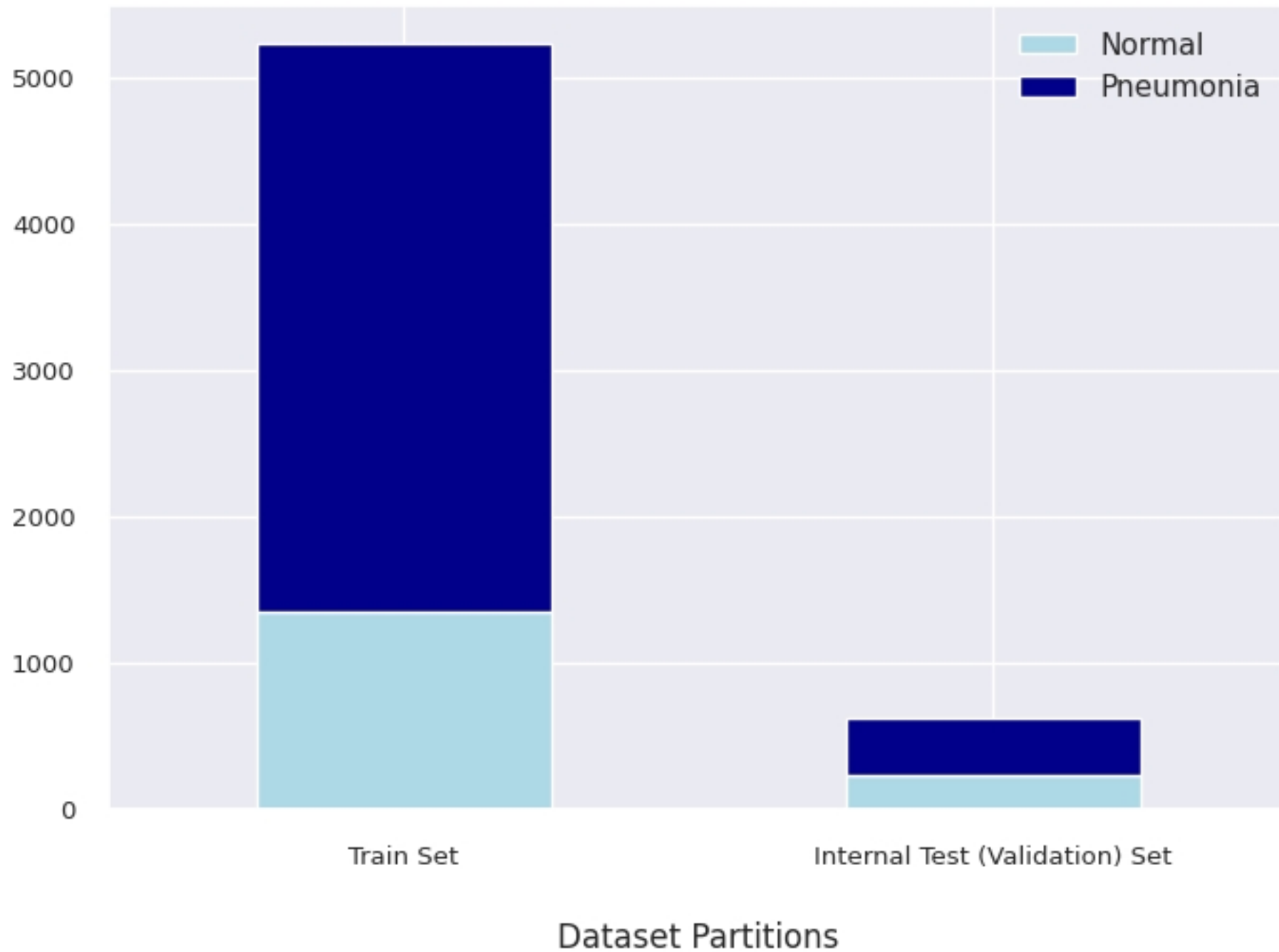
457 **References**

- 458 1. Yadav KK, Awasthi S. Childhood Pneumonia: What's Unchanged, and What's New?
459 Indian Journal of Pediatrics. 2023;90(7):693-9.
- 460 2. Marangu D, Zar HJ. Childhood pneumonia in sub-Saharan Africa: still a challenge. J Pan
461 Afr Thorac Soc. 2021;2(1):1-3.
- 462 3. Villavicencio F, Perin J, Eilerts-Spinelli H, Yeung D, Prieto-Merino D, Hug L, et al.
463 Global, regional, and national causes of death in children and adolescents younger than 20 years:
464 an open data portal with estimates for 2000-21. The Lancet Global health. 2024;12(1):e16-e7.
- 465 4. Shittu F, Agwai IC, Falade AG, Bakare AA, Graham H, Iuliano A, et al. Health system
466 challenges for improved childhood pneumonia case management in Lagos and Jigawa, Nigeria.
467 Pediatric Pulmonology. 2020;55(S1):S78-S90.
- 468 5. von Mollendorf C, Berger D, Gwee A, Duke T, Graham SM, Russell FM, et al. Aetiology
469 of childhood pneumonia in low- and middle-income countries in the era of vaccination: a
470 systematic review. Journal of global health. 2022;12:10009.
- 471 6. Organization WH. Revised WHO classification and treatment of pneumonia in children
472 at health facilities: evidence summaries. 2014.
- 473 7. Crame E, Shields MD, McCrossan P. Paediatric pneumonia: a guide to diagnosis,
474 investigation and treatment. Paediatrics and Child Health. 2021.
- 475 8. Ekirapa-Kiracho E, De Broucker G, Ssebagereka A, Mutebi A, Apolot RR, Patenaude B,
476 et al. The economic burden of pneumonia in children under five in Uganda. Vaccine: X.
477 2021;8:100095.
- 478 9. Alexopoulou E, Prountzos S, Raissaki M, Mazioti A, Caro-Dominguez P, Hirsch FW, et
479 al. Imaging of Acute Complications of Community-Acquired Pneumonia in the Paediatric
480 Population—From Chest Radiography to MRI. Children. 2024;11(1):122.
- 481 10. Andronikou S, McHugh K, Abdurahman N, Khoury B, Mngomezulu V, Brant WE, et al.
482 Paediatric radiology seen from Africa. Part I: providing diagnostic imaging to a young
483 population. Pediatric Radiology. 2011;41(7):811-25.
- 484 11. Olatunji RB, Akinmoladun JA, Atalabi OM. Capacity for paediatric radiology in Nigeria:
485 a survey of radiologists. Pediatric Radiology. 2021;51(4):587-91.
- 486 12. Idowu BM, Okedere TA. Diagnostic radiology in Nigeria: a country report. Journal of
487 Global Radiology. 2020;6(1).
- 488 13. Iyawe EP, Idowu BM, Omoleye OJ. Corrigendum: Radiology subspecialisation in Africa:
489 A review of the current status. SA Journal of Radiology. 2021;25(1).
- 490 14. Chang A. The Role of Artificial Intelligence in Digital Health. In: Meyers A, editor.
491 Digital Health Entrepreneurship. Cham: Springer International Publishing; 2023. p. 75-85.
- 492 15. Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image
493 understanding: a survey. Evolutionary Intelligence. 2022;15(1):1-22.
- 494 16. Akhter Y, Singh R, Vatsa M. AI-based radiodiagnosis using chest X-rays: A review.
495 Frontiers in Big Data. 2023;6.
- 496 17. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, et al. Transformers in
497 medical imaging: A survey. Medical Image Analysis. 2023;88:102802.
- 498 18. Ayan E, Karabulut B, Ünver HM. Diagnosis of Pediatric Pneumonia with Ensemble of
499 Deep Convolutional Neural Networks in Chest X-Ray Images. Arabian Journal for Science and
500 Engineering. 2022;47(2):2123-39.

19. Labhane G, Pansare R, Maheshwari S, Tiwari R, Shukla A, editors. Detection of Pediatric Pneumonia from Chest X-Ray Images using CNN and Transfer Learning. 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE); 2020 7-8 Feb. 2020.
20. Field EL, Tam W, Moore N, McEntee M. Efficacy of Artificial Intelligence in the Categorisation of Paediatric Pneumonia on Chest Radiographs: A Systematic Review. *Children*. 2023;10(3):576.
21. Kermany D, Zhang K, Goldbaum M. Labeled optical coherence tomography (oct) and chest x-ray images for classification. Mendeley Data, v2 <https://doi.org/10.17632/rscbjbr9sj> <https://nihcc.app.box.com/v/ChestXray-NIHCC>. 2018.
22. Naidu G, Zuva T, Sibanda EM, editors. A Review of Evaluation Metrics in Machine Learning Algorithms 2023; Cham: Springer International Publishing.
23. Akiba T, Sano S, Yanase T, Ohta T, Koyama M, editors. Optuna: A next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining; 2019.
24. Xu Y, Quan R, Xu W, Huang Y, Chen X, Liu F. Advances in Medical Image Segmentation: A Comprehensive Review of Traditional, Deep Learning and Hybrid Approaches. *Bioengineering*. 2024;11(10):1034.
25. Yan W, Wang Y, Gu S, Huang L, Yan F, Xia L, et al., editors. The domain shift problem of medical image segmentation and vendor-adaptation by Unet-GAN. Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22; 2019: Springer.
26. Thambawita V, Strümke I, Hicks SA, Halvorsen P, Parasa S, Riegler MA. Impact of Image Resolution on Deep Learning Performance in Endoscopy Image Classification: An Experimental Study Using a Large Dataset of Endoscopic Images. *Diagnostics*. 2021;11(12):2183.
27. Wollek A, Hyska S, Sabel B, Ingrisich M, Lasser T. Exploring the Impact of Image Resolution on Chest X-ray Classification Performance. arXiv preprint arXiv:230606051. 2023.
28. Aleixo EL, Colonna JG, Cristo M, Fernandes E. Catastrophic forgetting in deep learning: A comprehensive taxonomy. arXiv preprint arXiv:231210549. 2023.
29. Morid MA, Borjali A, Del Fiol G. A scoping review of transfer learning research on medical image analysis using ImageNet. *Computers in Biology and Medicine*. 2021;128:104115.
30. Iman M, Arabnia HR, Rasheed K. A Review of Deep Transfer Learning and Recent Advancements. *Technologies*. 2023;11(2):40.
31. Ganaie MA, Hu M, Malik AK, Tanveer M, Suganthan PN. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*. 2022;115:105151.
32. Azad R, Kazerouni A, Heidari M, Aghdam EK, Molaei A, Jia Y, et al. Advances in medical image analysis with vision Transformers: A comprehensive review. *Medical Image Analysis*. 2024;91:103000.
33. Mabrouk A, Díaz Redondo RP, Dahou A, Abd Elaziz M, Kayed M. Pneumonia Detection on Chest X-ray Images Using Ensemble of Deep Convolutional Neural Networks. *Applied Sciences*. 2022;12(13):6448.
34. Gummadi SD, Vootla Y, Ghosh A, Kartheek PN, Kandimalla AK, editors. Transfer Learning based Detection of Pneumonia from Chest X-Ray Images. 2021 13th International Conference on Computational Intelligence and Communication Networks (CICN); 2021 22-23 Sept. 2021.

35. Maleki F, Ovens K. Generalizability of Machine Learning Models: Quantitative Evaluation of Three Methodological Pitfalls. 2023;5(1):e220028.
36. Glocker B, Jones C, Roschewitz M, Winzeck S. Risk of Bias in Chest Radiography Deep Learning Foundation Models. Radiology: Artificial Intelligence. 2023;5(6):e230060.
37. Tejani AS, Klontzas ME, Gatti AA, Mongan JT, Moy L, Park SH, et al. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. Radiology: Artificial Intelligence. 2024;6(4):e240300.
38. Medixant. RadiAnt DICOM Viewer 2021. 2023.1:[Available from: <https://www.radiantviewer.com>.
39. Igweonu-Nwakile C, Adejimi A, Roberts A, Oluwole E, Oridota O, Oyeleye O, et al. Prevalence of Pneumonia and Its Determinants among Under-five Children attending a Primary Health Care Clinic in Amuwo Odofin Local Government Area, Lagos, Nigeria. Journal of Community Medicine and Primary Health Care. 2023;35(1):40-9.
40. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014.
41. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Yee Whye T, Mike T, editors. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; Proceedings of Machine Learning Research: PMLR; 2010. p. 249--56.
42. Kingma DP. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.
43. Pratiwi H, Windarto AP, Susliansyah S, Aria RR, Susilowati S, Rahayu LK, et al. Sigmoid Activation Function in Selecting the Best Model of Artificial Neural Networks. Journal of Physics: Conference Series. 2020;1471(1):012010.

Number of Chest Radiographs



Figure

Number of Chest Radiographs

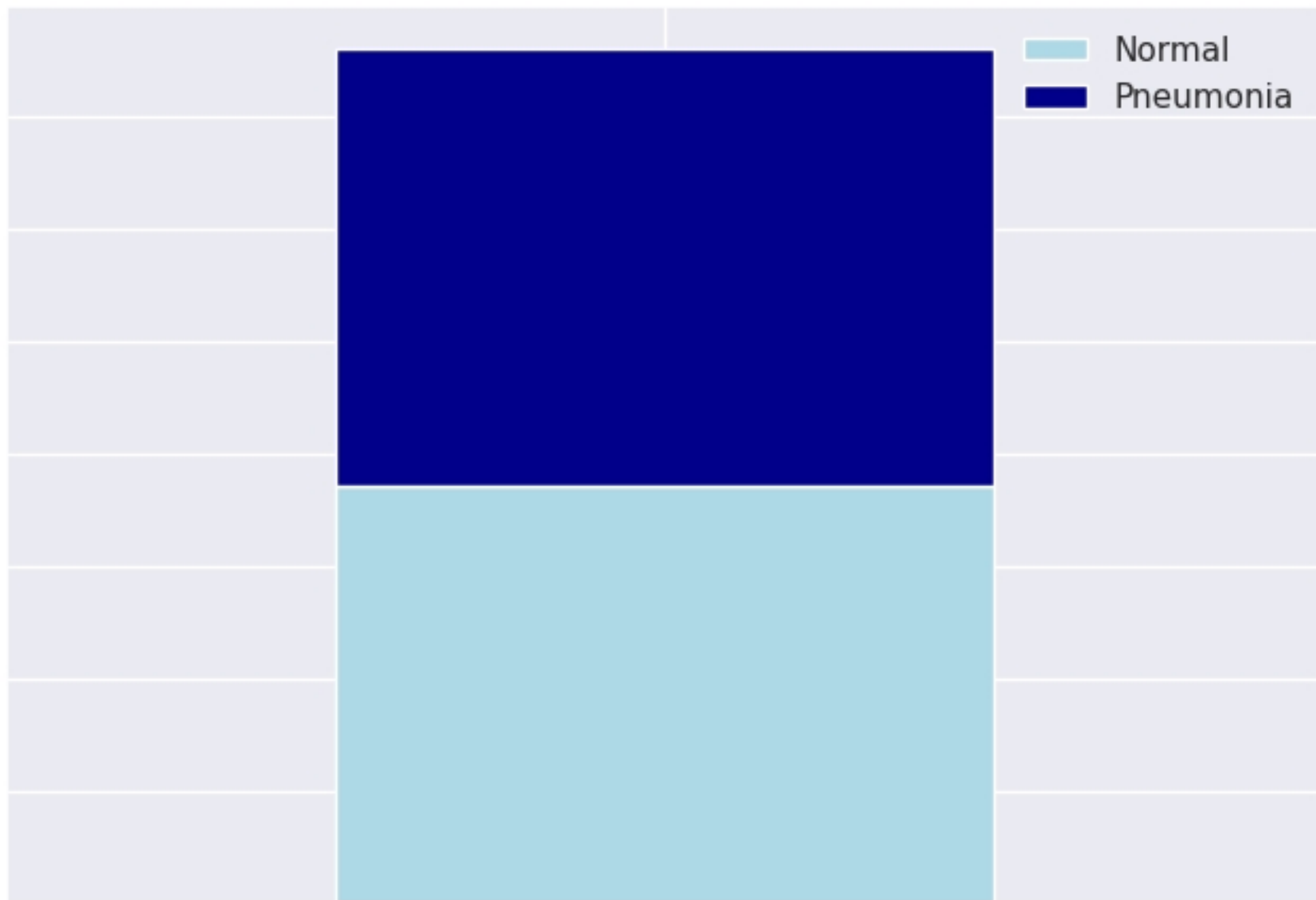
Normal
Pneumonia

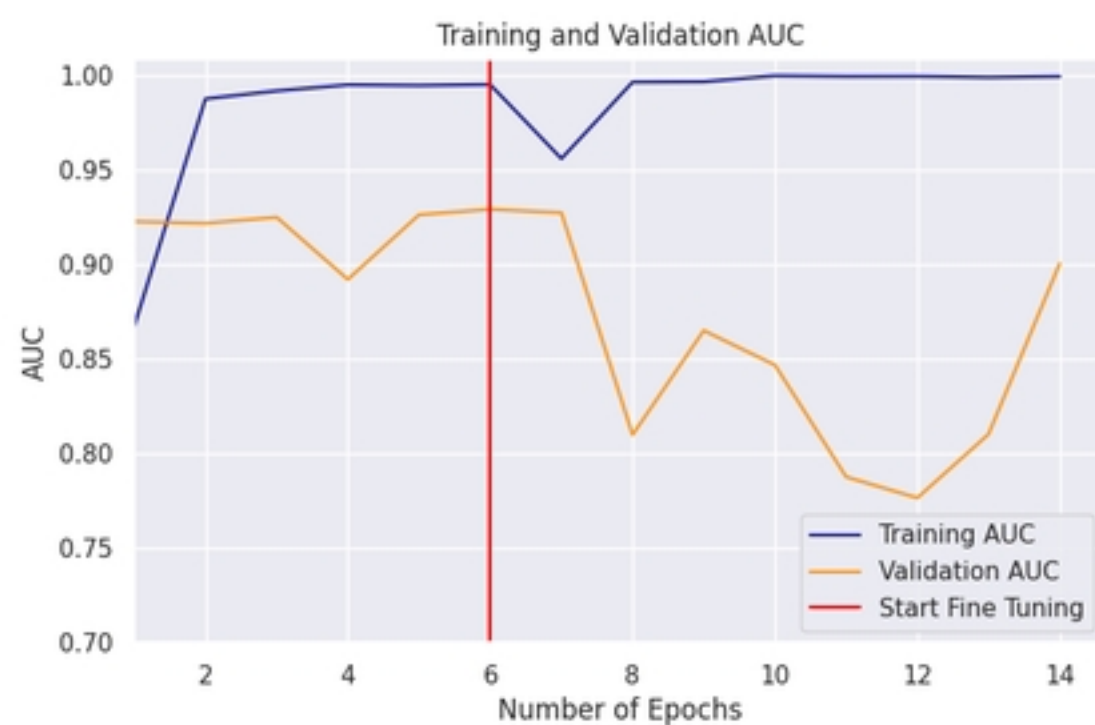
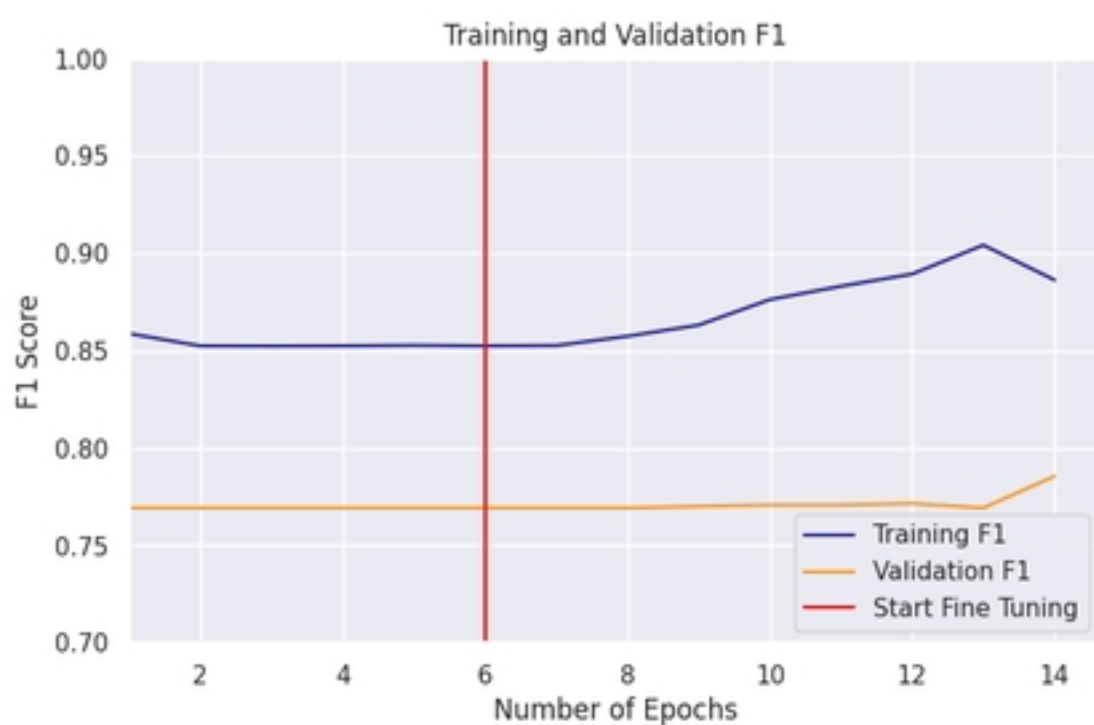
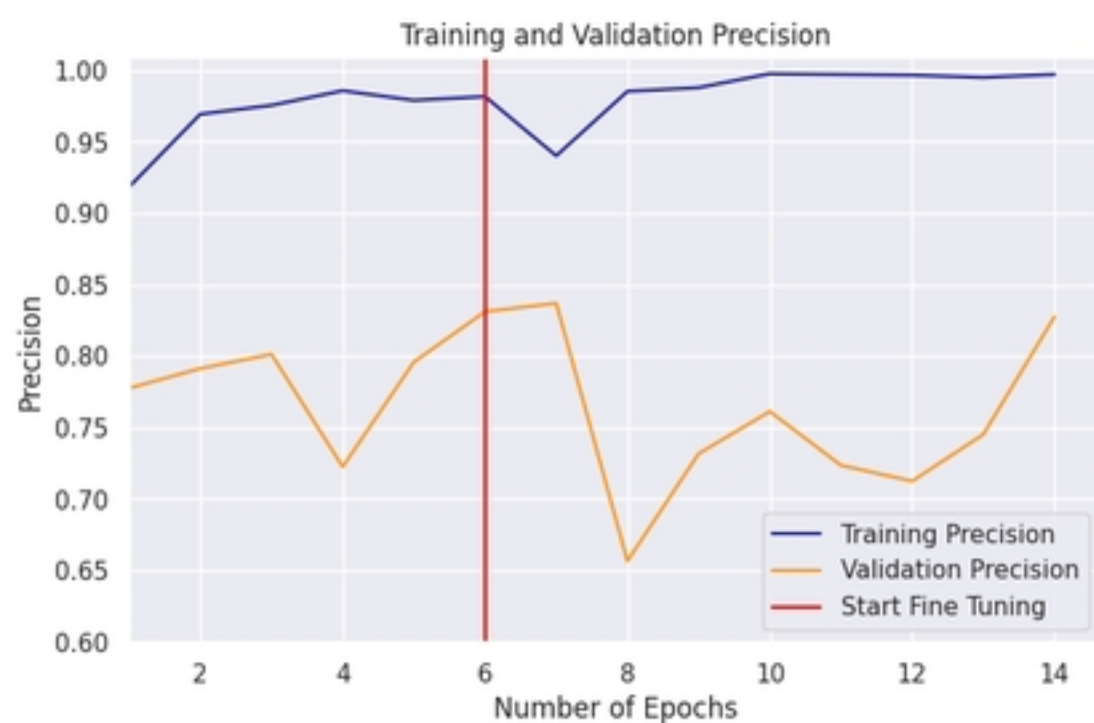
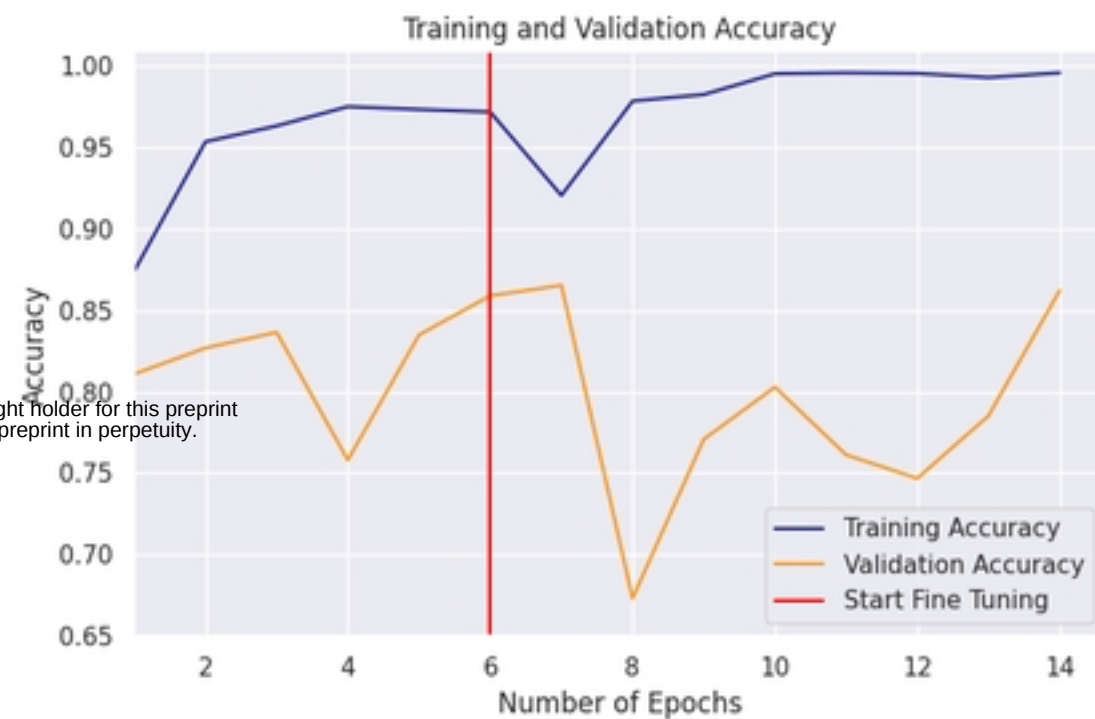
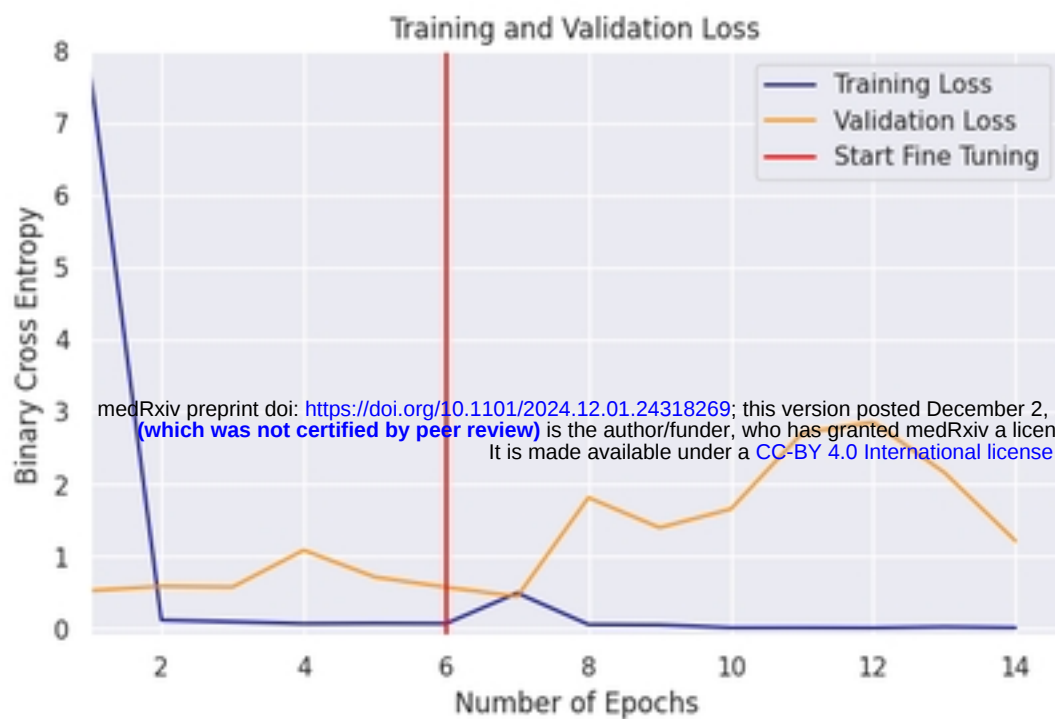
175
150
125
100
75
50
25
0

External Test Set

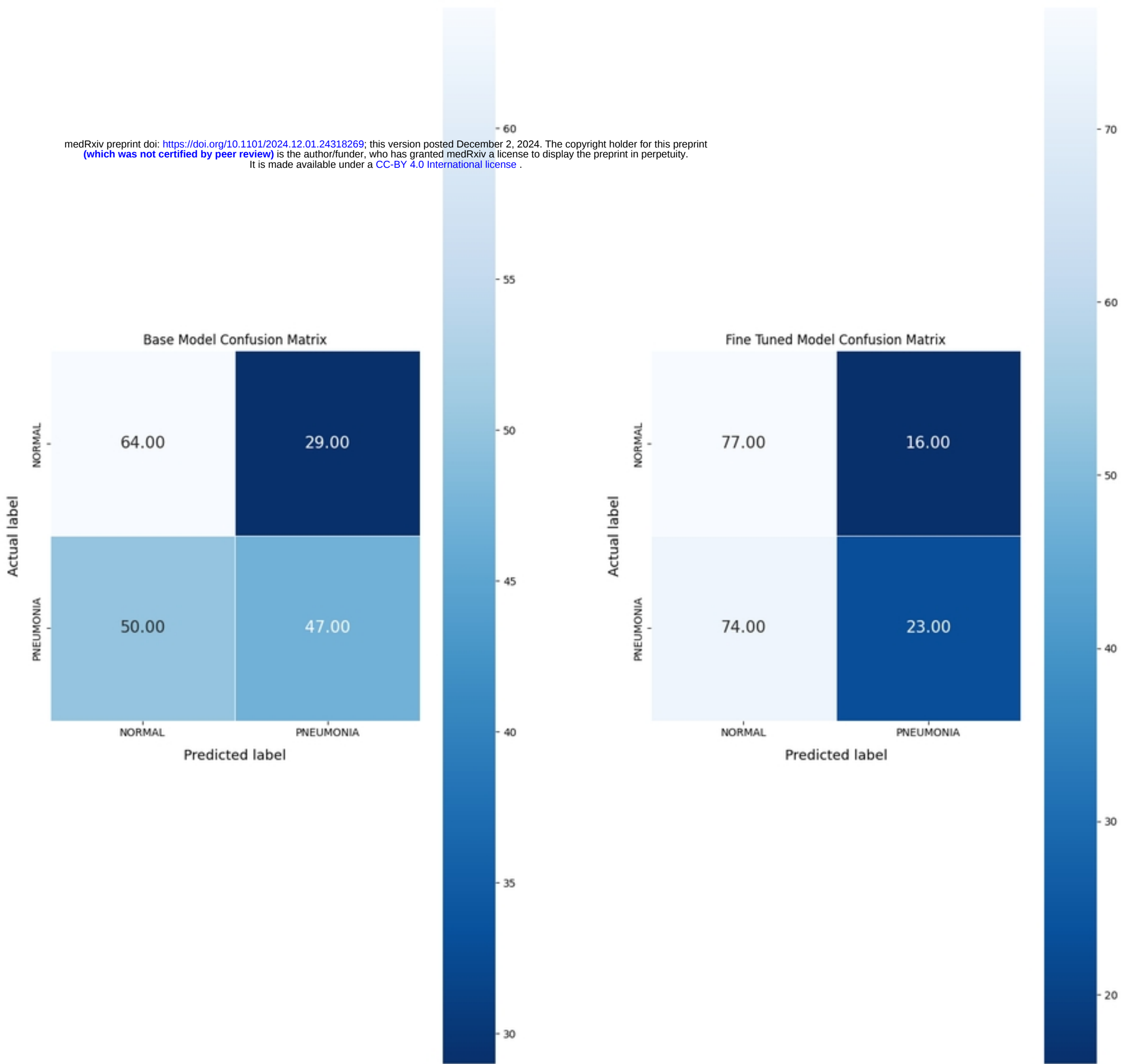
Dataset Partitions

Figure

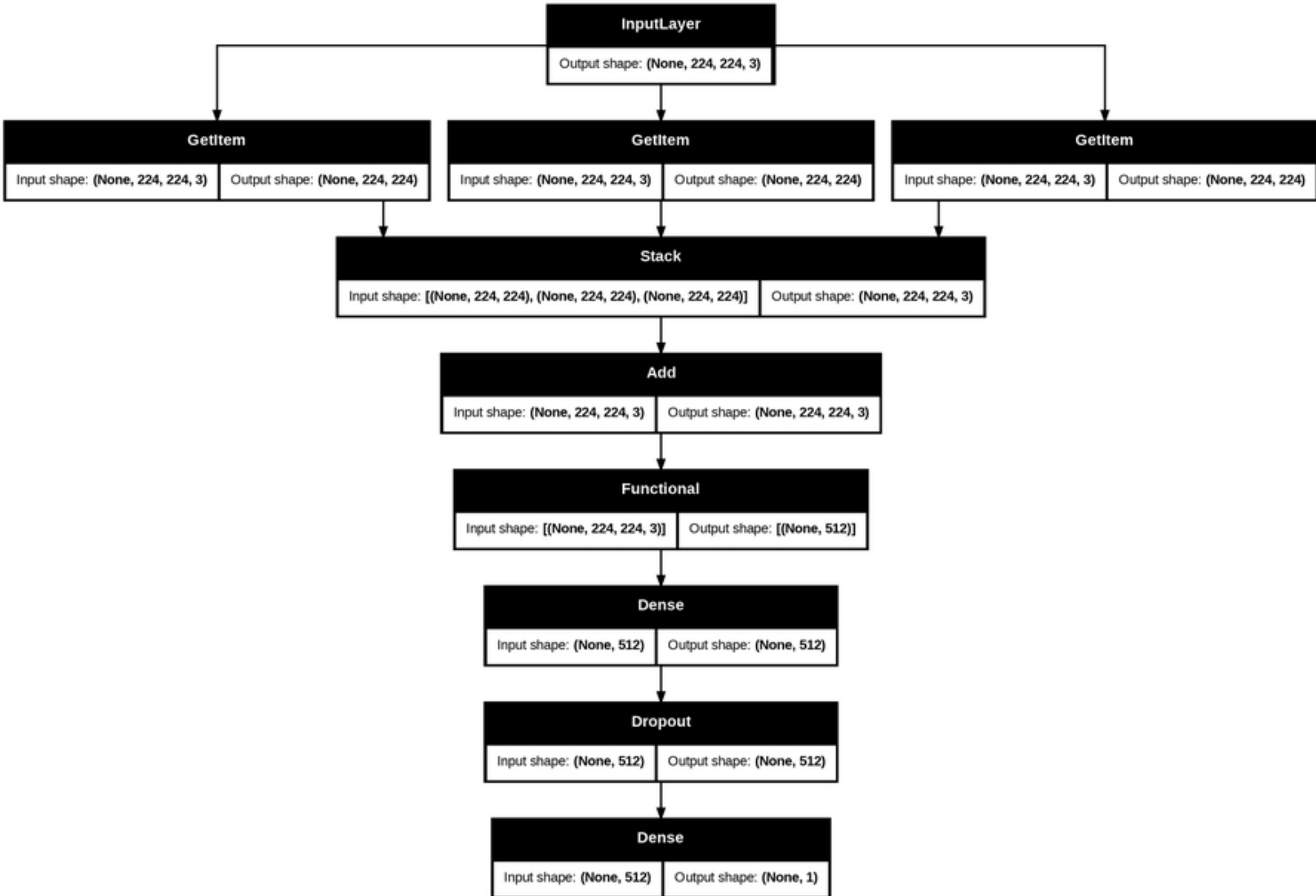




Figure



Figure



Figure