Convolutional neural networks for rapid diagnosis and lesion detection of pediatric mycoplasma pneumoniae pneumonia using chest X-rays

Jiaming Deng[1,2,3#], Danqing Yu[1#], Boyuan Peng[1,2,3#], Dongmei Yu[3,4#], Bangxun Mao[1], Jiaqi Yang[5], Peiwu Qin[1,2,3], Xingru Huang[3], Chenggang Yan[3], Zhendong Luo[6]*, Jiansong Ji[1]*, Zhenglin Chen[1,2,3]*

1. Zhejiang Key Laboratory of Imaging and Interventional Medicine, Zhejiang Engineering Research Center of Interventional Medicine Engineering and Biotechnology, The Fifth Affiliated Hospital of Wenzhou Medical University, Lishui 323000, China
2. Institute of Biopharmaceutical and Health Engineering, Shenzhen International Graduate School, Tsinghua University, Shenzhen, Guangdong, China
3. School of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang Province, 310018, China
4. School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, Shandong 264209, China
5. State Key Laboratory of Urban Water Resources and Environment, School of Civil & Environmental Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China
6. Department of Radiology, The University of Hong Kong - Shenzhen Hospital, Shenzhen 518055, China

#These authors contributed equally to this work.

*Corresponding authors E-mail: lzhend@163.com (P. Luo); jjstcty@sina.com (P. Ji); chenzlin1992@163.com (D. Chen)

Abstract

Mycoplasma pneumoniae pneumonia (MPP) presents significant diagnostic challenges in pediatric healthcare, particularly in regions like China that with large population density. To ease the burden on radiologists, we utilized convolutional neural networks (CNN), structured as state-of-the-art computational efficient architecture, for faster MPP detection and pediatric pneumonia diagnosis. Our model, trained on 3,345 chest X-ray (CXR) images, including 833 MPP cases, distinguishes MPP from viral, bacterial, and normal cases. The model achieved an accuracy of 88.20% and an AUROC of 0.9218 across all classes, with a specific accuracy of 97.64% for MPP. We also integrated explainability techniques to help radiologists localize lesions in CXR images, with extra consideration on the deployment targeted for mobile devices.

1. Introduction

Mycoplasma pneumoniae pneumonia (MPP) is a common community-acquired infection affecting primarily children and young adults, marked by pulmonary inflammation involving various lung sites. Studies[1–4] show that MPP accounts for 10-40% of community-acquired pneumonia (CAP) cases in children over 5 years old in China. Since last September, a rise in MP infections has put significant strain on pediatric and respiratory clinics nationwide, due to MPP's distinct symptoms and diagnostic features[5].

The clinical diagnosis of MPP relies on a combination of microbiological, serological, and imaging tests[6]. While MP culture is considered the "gold standard"[7], its slow growth and specific requirements make it impractical for routine use. Other tests, such as MP nucleic acid and antibody tests, offer faster results but suffer from low sensitivity and specificity, with positive rates around 60%. Antibody tests, in particular, can take five to seven days to yield results[8].

The clinical symptoms of MPP typically include fever and cough, often accompanied by headache, runny nose, sore throat, or earache. Fever is usually moderate to high, with persistent fever signaling more severe illness. Coughs are

often severe, resembling whooping cough, and wheezing may occur, especially in younger children. Early lung involvement can be subtle, but as the disease progresses, abnormal breath sounds and rales may develop. While imaging is crucial for assessing severity and prognosis, misinterpretation by inexperienced physicians is common, and analyzing chest X-rays (CXR) is time-consuming. Given the success of deep convolutional neural networks (CNNs) in detecting various pneumonias[9–17], deep learning offers a promising way to improve the efficiency and accuracy of MPP diagnosis.

Deep learning, particularly convolutional neural networks (CNNs), has proven highly effective in tasks like image recognition and medical diagnosis. CNNs, designed specifically for visual data, use convolutional layers to extract key features, followed by pooling layers to simplify the data while retaining crucial patterns[18–20]. This structure allows CNNs to excel in complex visual tasks. Building on this, we utilize CNNs alongside explainability techniques to enhance rapid localization and diagnosis of pneumonia lesions.

We collected CXR images of children aged 0-12 diagnosed with MPP from multiple medical institutions across China, adding 833 Mycoplasma-CXR images and 169 normal-CXR images to a publicly available dataset [21] containing 669 normal, 858 bacterial, and 816 viral CXR images. The dataset was split into training, validation, and testing sets in an 8:1:1 ratio. Using this combined dataset, we developed a ConvNeXt-Tiny model to predict MPP in children and highlight suspected lesion areas in CXR images. Our best model achieved 88.20% accuracy, an AUC of 0.9218, and an F1 score of 0.8824 on the test set. To improve usability for physicians and radiologists, we designed a mobile application named PneumoniaApp for integrating model and desision visualization tools

The remainder of our paper is structured as follows: Section 2 reviews AI-assisted diagnostic methods, Section 3 details our methodology, and Section 4 presents performance comparisons and visualization results, including a demo of the PneumoniaApp user interface.

2. Related Works

Recent advancements in pneumonia detection using medical imaging techniques, such as chest X-rays (CXR) and computed tomography (CT), have been largely driven by deep-learning-based artificial intelligence. In this section, we review relevant studies prior to pediatric MPP diagnosis and explain the motivation behind our research. Stephen et al. [22] introduced a straightforward model that effectively performs classification tasks using deep neural networks. This model, designed for pneumonia image classification, employs CNNs to extract relevant features through neuron convolution. Similarly, Sharma et al. [23] proposed various CNN architectures to automatically differentiate between normal and pneumonia CXR images. Kundu et al. designed an ensemble model that combines three classification networks—GoogLeNet, ResNet-18, and DenseNet-121—using weighted averaging to enhance classification accuracy. However, these studies fail to account for differences in radiographic appearances of pneumonia caused by various pathogens, which are crucial for clinical diagnosis. Additionally, notable discrepancies exist between CXR features in adults and children, potentially leading to biases in the models. Given that pneumonia is more lethal in children, there is an urgent need for computer-aided diagnostic methods specifically tailored to pediatric pneumonia to distinguish between different pneumonia types and normal cases.

Several researchers have conducted targeted studies on pneumonia in children or differentiated pneumonia pathogens. Liz et al. [24] utilized two datasets: the X-ray Pediatric Pneumonia (XrPP) dataset from Ben-Gurion University and a publicly available collection of pediatric CXR images. These datasets comprise 950 annotated CXR images of children aged 1 to 16 years, designed to train models for pediatric pneumonia diagnosis. They evaluated six different CNN architectures and built an ensemble model for this purpose. Arun et al. [25] used the pediatric pneumonia dataset from Kermany et al. [21], featuring CXR images of children aged 1 to 5 years. They preprocessed the images with resizing and normalization, applying geometric transformations (rotation, scaling, and flipping) for data augmentation to reduce overfitting. They employed pre-trained deep CNN architectures with channel

attention modules after the last convolutional block and utilized kernel principal component analysis to reduce the dimensionality of features extracted. A stacking classifier combining various machine learning algorithms, including SVC, logistic regression, KNN, Nu-SVC, and XGBClassifier, was used for the final classification. However, these studies overlooked MPP, a common form of pediatric pneumonia, and used overly complex methods that diverged from state-of-the-art deep learning practices. In contrast, Serener et al. [26] proposed using deep learning techniques and CNNs to distinguish between MPP and viral pneumonia, such as COVID-19, in CT images. They compared the performance of various convolutional network architectures like ResNet-50 and DenseNet-121 on CT datasets. While CT scans offer greater sensitivity and resolution for pneumonia diagnosis, they come with higher costs and increased radiation exposure risks. Consequently, methods relying on CT images often face challenges in practical applications for pediatric pneumonia diagnosis, particularly in resource-limited medical settings.

Few of the previously mentioned deep learning-based pneumonia diagnostic methods employed interpretable techniques [27,28], and there is a lack of thorough analysis of their results. This indicates that while interpretability is vital for understanding model decision-making and optimizing performance and reliability in clinical pneumonia diagnosis, the current literature inadequately addresses the specific challenges of implementing and deploying deep learning in real-world settings. This represents the second motivation for our research: utilizing interpretability techniques to enhance lesion localization in pediatric MPP.

Moreover, many studies rely on deep CNN architectures such as GoogLeNet and DenseNet-121, which demand substantial computational resources, especially on mobile devices. The frequent use of ensemble learning techniques further complicates mobile deployment. Therefore, our third motivation is to adopt state-of-the-art deep visual architectures designed for computational efficiency, facilitating the effective deployment of deep learning-based mobile applications in practical scenarios.

3. Materials and Methods

In this section, we present our approach to building the dataset and developing pediatric pneumonia classification models

3.1 Data collection

We compiled our pediatric pneumonia CXR dataset from two Chinese hospitals, The University of Hong Kong - Shenzhen Hospital and Lishui Central Hospital, as well as a publicly available online dataset [25]. From the hospitals, we collected CXR images of children (aged 0-12) diagnosed with MPP. For additional cases of normal, bacterial, and viral pneumonia, we used images from the public dataset. To maintain balance across categories, we randomly sampled cases from each class and merged them with our hospital data. The final dataset includes four categories: normal, bacterial, viral, and Mycoplasma pneumonia. After shuffling, we split the dataset into training (80%), validation (10%), and testing (10%) sets. Table 1 shows the distribution of these categories across the dataset splits.

	Normal	Bacterial	Virus	Mycoplasma	total
Training	670	686	652	666	2674
Validation	83	85	81	83	332
Testing	85	87	83	84	339
Total	838(169)	858	816	833	3345

Table 1. Category distribution in the constructed Pneumonia dataset. The number within parentheses for the NORMAL category represents the number of normal samples collected from the hospitals.

3.2 Data Preprocessing

To reduce the impact of irrelevant factors, such as acquisition devices and storage formats, on the learning process, which could hinder the model's ability to differentiate between pneumonia types, we applied several preprocessing steps. First, to standardize pixel value interpretation across images with varying bit depths, we

scaled the pixel values of each image to a range of 0 to 1 using the scaling transformation (Eq. (1))

$$x_{ij}^{'} = \frac{x_{ij} - \min(X)}{\max(X) - \min(X)} \quad (1)$$

In this equation, x_{ij} and $x_{ij}^{'}$ represent the pixel values at position ij in the pixel matrix, and $X$is the collection of pixel values in the image. We then resized all images to 224x224 pixels using bilinear interpolation and duplicated the grayscale images across the channel dimension to match the model input requirements. Contrast enhancement was applied using the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique to emphasize details, which is especially effective for X-rays due to their continuous exposure method [29]. Finally, Z-score standardization (Eq. (2)) was applied to ensure uniform pixel value distribution across images.

$$x_{ijc}^{'} = \frac{x_{ijc} - \mu_c}{\sigma_c} \quad (2)$$

Where i, j has the same meaning as in Eq. (1), c represents the channel dimension, μ_c denotes the mean of the dataset on the channel, and σ_c represents the standard deviation of the dataset on the channel.

3.3 Data augmentation

Data augmentation expands the dataset by applying random, non-destructive transformations to the images, helping to mitigate overfitting due to the limited dataset size and variability introduced from different sources. We applied automatic augmentation to the training set using the torchvision API, with transformations selected to account for potential perturbations from mobile phone camera captures. For validation and testing, fixed (non-random) transformations were used to ensure

consistent evaluation across models. Figures 1 and 2 show the augmented CXR images and the data preprocessing pipeline, respectively.

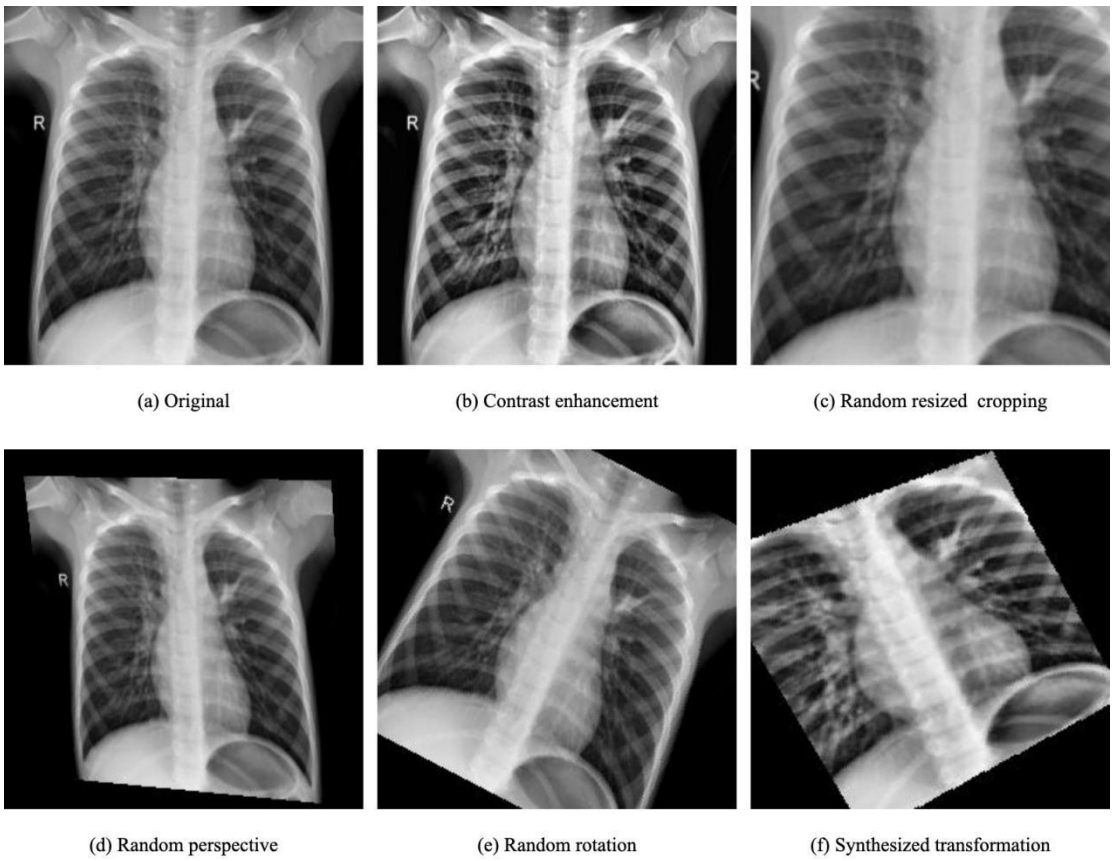


Figure 1. Example of original and augmented CXR images

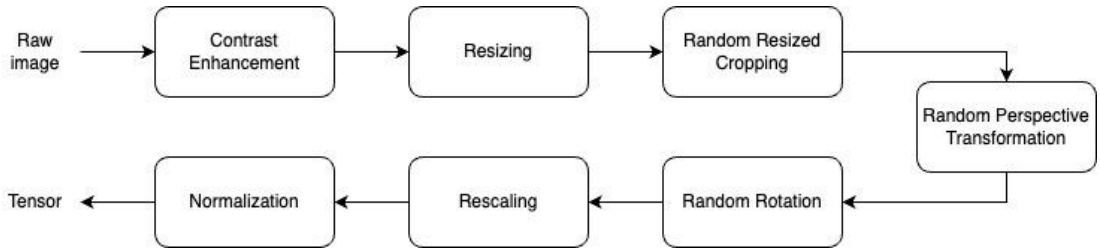


Figure 2. Preprocessing and augmentation pipeline for training data.

3.4 Model Architecture

We used convolutional neural networks (CNNs) to train classifiers for predicting pediatric pneumonia from CXR images. CNNs [31] are feedforward neural networks designed to automatically extract image features through convolutional layers, eliminating the need for manual feature engineering. They are widely used in AI-

assisted diagnosis tasks like medical image classification, object detection, and 3D reconstruction.

ResNet, a modern deep convolutional network, introduces residual connections to overcome the vanishing gradient issue, enabling the training of deeper networks. ResNet has shown strong performance in feature extraction, particularly after its success on the ImageNet dataset, and is widely used for deep feature extraction. In Section 4.2, we used a lightweight version, ResNet-18, to train models on the pediatric pneumonia CXR dataset. ResNet-18 comprises 8 basic blocks, along with input and output layers, where each block includes two convolutional modules with minor differences. Figure 3 shows the architecture of ResNet-18.

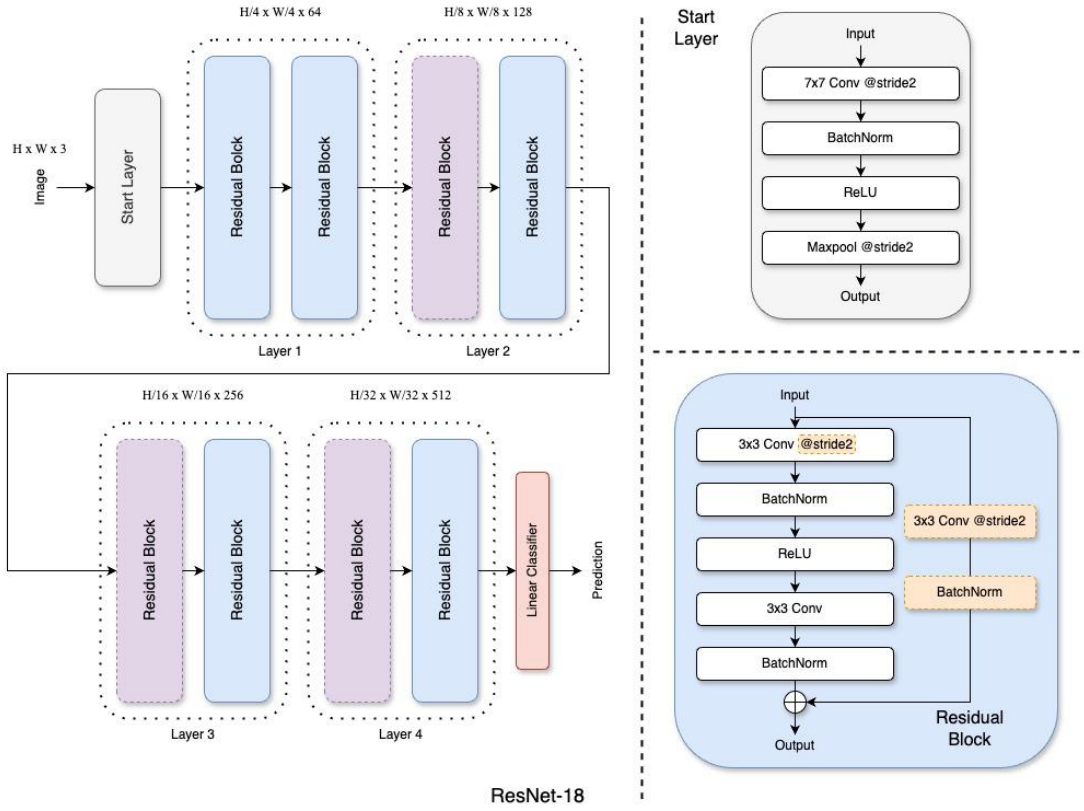


Figure 3. The architecture of ResNet-18: Layer 1 contains two consecutive Residual blocks, followed by three layers with a downsampling Residual block and a regular Residual block. The dashed boxes in the Residual block illustration highlight operations specific to downsampling blocks, marked with purple dashed boxes in the ResNet-18 pipeline.

To compare the feature extraction capabilities of various classification network backbones on pediatric pneumonia CXR images, we selected additional lightweight architectures based on prior studies [32]. Along with ResNet, we included EfficientNet[39], RegNet [33], ConvNeXt [34], and Swin Transformer [35]. Given the limited computational resources and storage, we prioritized efficient architectures to ensure good inference performance on mobile devices.

Swin Transformer is a notable example of vision transformers, which have gained popularity since the introduction of ViT [36,37], incorporating the Transformer architecture [38] into vision tasks and achieving state-of-the-art results. We chose the lightweight SwinV2-Tiny variant to train our pediatric pneumonia detection model. Figure 4 illustrates the architecture of SwinTransformer-Tiny.

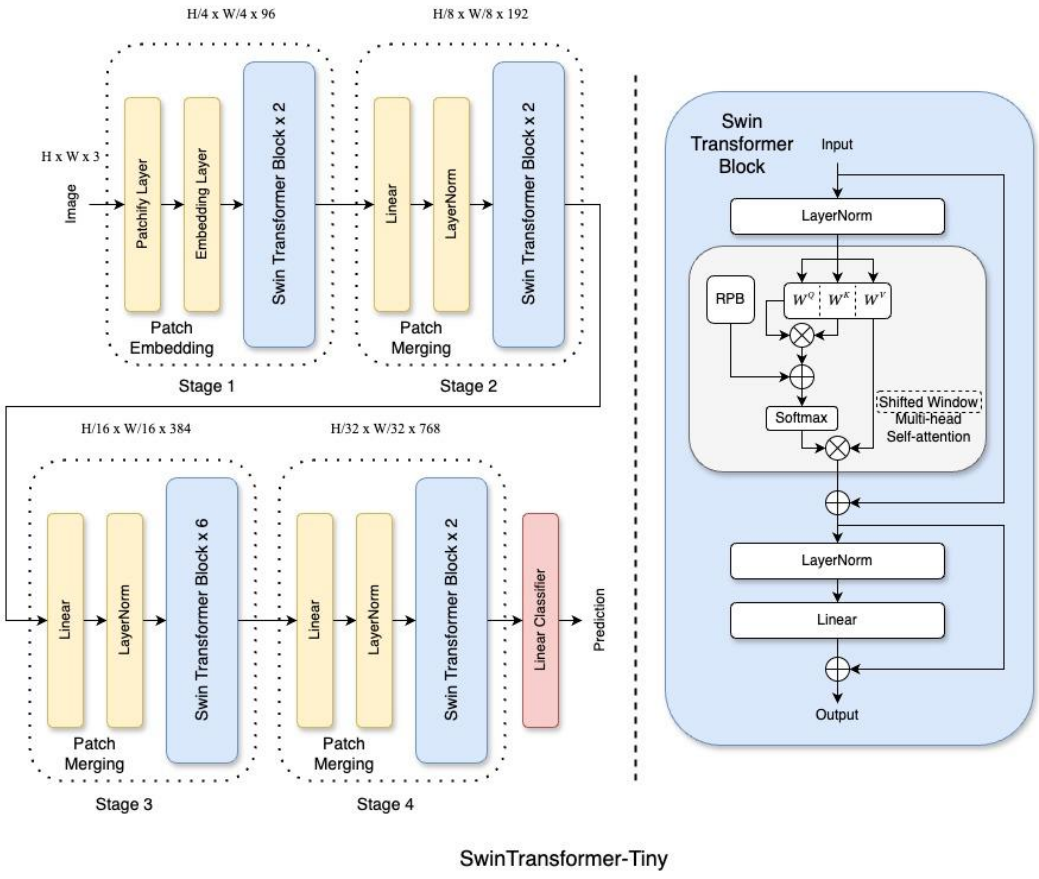


Figure 4. Architecture of SwinTransformer-Tiny. Stage 1 involves patchifying the image and embedding tokens through an embedding layer, followed by two Swin Transformer blocks. The first block applies standard multi-head self-attention, and the

second uses shifted window multi-head self-attention. In subsequent stages, feature maps are merged with linear layers and normalization modules before entering more Swin Transformer blocks. The attention mechanism alternates between standard and shifted window versions in each stage.

ConvNeXt is a modern CNN architecture that integrates design features like the inverted bottleneck, GELU activation, and stage ratios similar to the Swin Transformer. ConvNeXt has outperformed visual Transformers in various recognition tasks. For our study, we selected the lightweight variant, ConvNeXt-Tiny. The architecture of ConvNeXt-Tiny is shown in Figure 5.

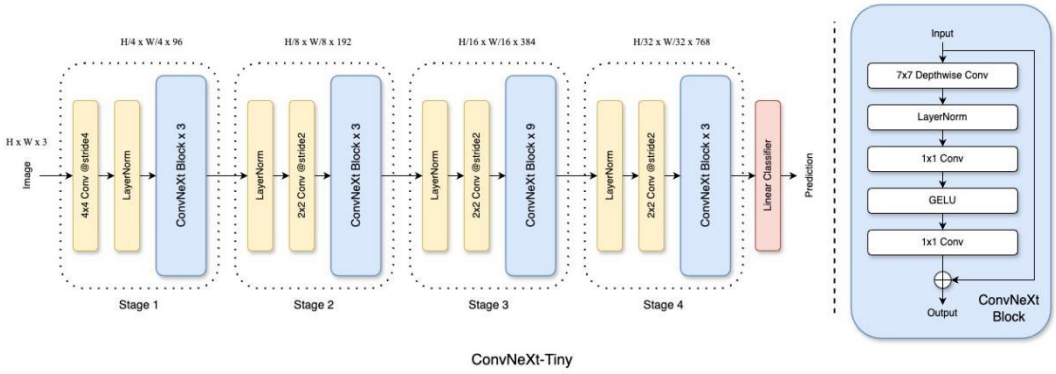


Figure 5. Architecture of ConvNeXt-Tiny. The model consists of four stages. In Stage 1, the input image is patchified through non-overlapping 4x4 convolution layers, followed by layer normalization. In the following three stages, feature maps are normalized, downsampled with 2x2 convolutions (stride 2), and processed through ConvNeXt blocks. Depthwise convolution uses group convolution, followed by 1x1 convolutions.

3.5 Class Activation Map

Class Activation Map (CAM) is a technique for visualizing features in deep convolutional (also adaptable for transformer-based) networks. It provides visual explanations for individual inputs by using a weighted combination of activation maps from convolutional layers. CAMs are valuable for understanding the mechanisms and decision-making processes of deep networks. The CAM is defined

by Eq. (3), where c represents the target class, A_k denotes the feature map, α_k is the weight coefficient for the feature map, and N_l is the number of feature maps in layer 1 . The $ReLU$ activation function used in CAM is described in Eq. (4).

$$L_{CAM}^{c} = ReLU(\sum_{k=1}^{N_l} \alpha_k A_k) \quad (3)$$

$$ReLU(x) = \max(0, x) \quad (4)$$

CAMs generate matrices that match the size of the original image, with values ranging from 0 to 1. These matrices are usually displayed as grayscale images, where each pixel value reflects the significance of the corresponding feature in the original image for the model's decision-making.

We use CAMs to highlight key feature regions in the original images for our pediatric pneumonia classification model. Section 4.3 presents qualitative results from different CAM techniques.

3.6 Transfer Learning

In deep learning, larger datasets generally improve model performance. However, due to limited data for pediatric MPP and pneumonia classification, expanding the dataset is not feasible. To address this, we use transfer learning by applying pre-trained weights from large-scale datasets like ImageNet-1K. This approach allows our models to benefit from previously learned features, resulting in faster convergence and reduced overfitting.

For our 4-class classification task, we convert the model's output into confidence scores for each class using the softmax function, as defined in Eq. (5). In this equation, $\mathbf{x}$ is a vector of n dimensions, with $x_i$and $x_j$representing the i-th and j-th components of $\mathbf{x}$, respectively. The class corresponding to the highest component in the probability distribution vector is regarded as the model's classification result.

$$softmax(\mathbf{x})_{\mathrm{i}} = \frac{e^{x_i}}{\sum_{j=0}^{n} e^{x_j}} \quad (5)$$

4. Results and Discussion

In this section, we explore three main aspects: (1) the impact of different model architectures on the performance of pediatric pneumonia classification, evaluated through metrics including accuracy, recall, precision, AUC, and F1 score; (2) the influence of two CAM techniques, Grad-CAM[40] and Score-CAM[41], on the visualization results of model outputs. (3) comparison of the Score-CAM visualizations and radiologists' annotation on the pneumonia legion.

4.1 Transfer Learning Performance

4.1.1 Experiment settings

Given the limited computing resources on mobile devices, we prioritize the efficiency of model architectures to facilitate mobile deployment and enhance performance on lower-end phones. As discussed in Section 3.2, we leverage insights from [32] to evaluate factors such as prediction accuracy on ImageNet, availability of pre-trained weights, transferability, model parameters, and computational demands. Based on these criteria, we select five lightweight architectures: ResNet-18, RegNetX-400mf, EfficientNet-B0, SwinV2-Tiny, and ConvNeXt-Tiny. The experiments in this section compare the transfer learning performance of these architectures on the pediatric pneumonia dataset.

For image preprocessing and data augmentation, we follow these steps: 1) apply CLAHE with a clip limit of 2.0 and grid size of (8, 8); 2) resize images to maintain their original aspect ratio, with a width of 256 pixels; 3) perform random cropping and resizing, with a crop ratio between 0.4 and 0.8, then resize to 224x224 using bilinear interpolation; 4) apply random perspective transformation with a distortion scale of 0.4 (60% probability); 5) rotate images randomly between -45° and 45°; 6) rescale pixel values to [0, 1]; 7) normalize pixel values with a mean of [0.485, 0.456, 0.406] and standard deviation of [0.229, 0.224, 0.225]. The full preprocessing and augmentation pipeline is shown in Figure 2.

For each architecture, we use backpropagation on the training set to update model parameters and tune hyperparameters based on validation performance. After building the models with PyTorch, we initialize them with pre-trained "ImageNet-1K-V1"

weights from Torchvision. The original fully connected (FC) layer, which outputs a 1000x1 tensor, is replaced with a randomly initialized FC layer that outputs a 4x1 tensor for classifying the four CXR categories. We apply the Adam optimizer with an initial learning rate of 1E-4 and a weight decay of 5E-3. Additionally, a linear learning rate scheduler reduces the learning rate by 0.1 every 50 epochs. We use an early-stop strategy to prevent overfitting during training. In our implementation, the model state is saved and updated whenever the validation loss reaches a new minimum.

Our experiments were conducted on a single RTX 2080Ti GPU in a Windows 10 environment. After multiple rounds of training, validation, and hyperparameter tuning, we identified the optimal hyperparameter settings. The model was trained for up to 100 epochs, using an early stopping strategy to prevent overfitting. The model state was saved whenever the validation loss reached a new minimum. The model with the best validation performance after 100 epochs was selected for comparison across different architectures. Figure 6 shows the validation loss curve for the five models during the first 50 epochs, as no further improvements were observed beyond this point.

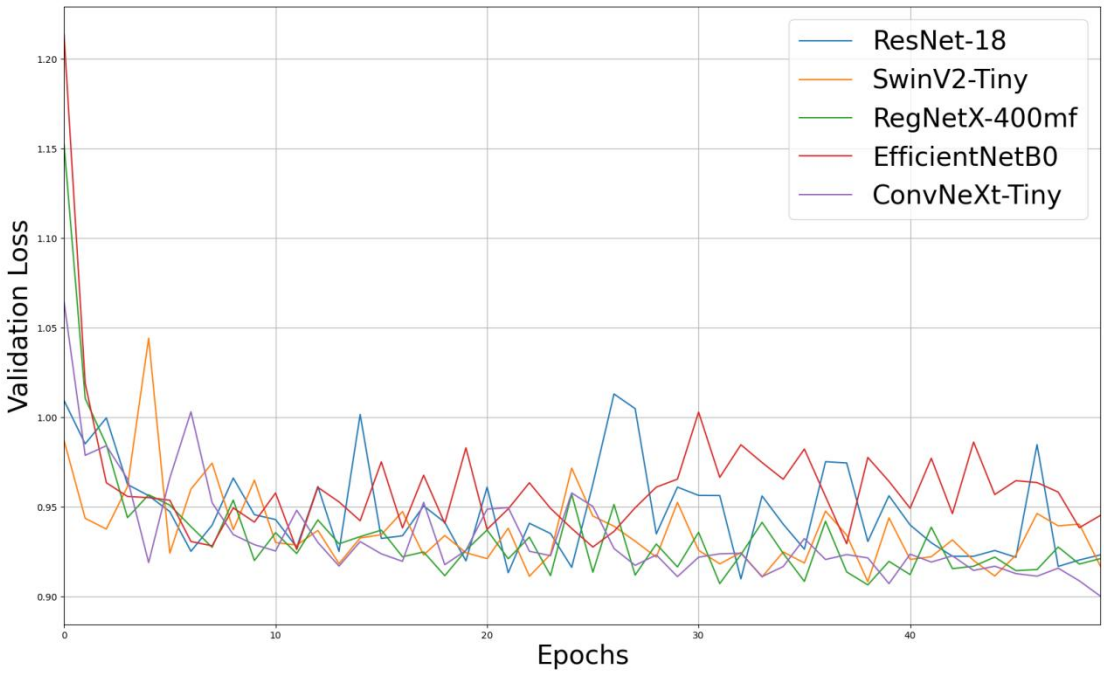


Figure 6: Validation loss evolution for ResNet-18, SwinV2-Tiny, RegNetX-400mf, EfficientNetB0, and ConvNeXt-Tiny models during the first 50 epochs.

4.1.2 Metrics and Performance

We assess the model's performance on the test dataset using several key metrics, including accuracy, recall, precision, AUC, and F1 score. For binary classification, where samples are categorized as true positive (TP), false negative (FN), false positive (FP), and true negative (TN), these metrics are defined in Eq. (6) (7) (8) (9).

$$recall = \frac{\mathrm{TP}}{TP+FN} \tag{6}$$

$$precision = \frac{\mathrm{TP}}{TP+FP} \tag{7}$$

$$accuracy = \frac{\mathrm{TP+TN}}{TP+TN+FP+FN} \tag{8}$$

$$F1\ score = \frac{2*\mathrm{precision}*\mathrm{recall}}{recall+precision} = \frac{2\mathrm{TP}}{2TP+FP+FN} \tag{9}$$

To compute the Area Under the Curve (AUC), where P represents the set of all positive samples N represents the set of all negative samples, and rank_s represents the rank of sample s when all samples are sorted in ascending order according to the model's predicted score, AUC can be calculated as:

To calculate the Area Under the Curve (AUC), let P represent the set of positive samples, N the set of negative samples, and rank_s the rank of sample s when all samples are sorted in ascending order by the model's predicted score. AUC is computed as shown in Eq (10).

$$AUC = \frac{\sum_{s\in P} \mathrm{rank}_s - \frac{|P|(|P|+1)}{2}}{|P||N|} \tag{10}$$

In our multi-class pediatric pneumonia classification task, we use macro-averaged metrics, including recall, precision, accuracy, AUC, and F1 score. These metrics provide an overall assessment of the model's performance across all classes.

The performance metrics results obtained on the test dataset for the optimal models of each architecture (Table 2.) show that the ConvNeXt-Tiny model achieves a prediction accuracy of 88.20%, an AUC of 0.9218, and an F1 score of 0.8824, which is significantly superior to the other four architectures. Additionally, the ConvNeXt-Tiny model's parameter size and computational demand on devices are considered to be more computational efficient for mobile deployment in our case.

Architecture	Acc.	Recall	Precision	Auc	F1 score	Number of Parameters	Computational cost (GFLOPs)
ResNet-18	82.01%	0.8219	0.8385	0.8811	0.8201	11.7M	1.81
RegNetX-400mf	82.89%	0.8294	0.8424	0.8862	0.8335	5.5M	0.41
EfficientNet-B0	82.01%	0.8205	0.8386	0.8803	0.8261	5.3M	0.39
SwinV2-Tiny	84.07%	0.8400	0.8473	0.8933	0.8400	28.4M	5.94
ConvNeXt-Tiny	**88.20%**	**0.8830**	**0.8844**	**0.9218**	**0.8824**	28.6M	4.46

Table 2. Performance of ResNet-18, RegNetX-400mf, EfficientNet-B0, SwinV2-Tiny, and ConvNeXt-Tiny optimal models. Boldface for best performer in the comparison.

4.1.3 Ablation Study on Data Augmentation Strategy

In this subsection, we perform an ablation study to assess the necessity of random data augmentation (described in Section 3.3) for the transfer learning of our models. We focus on two main concerns: 1) whether the selected data augmentation strategies enhance model performance on the test dataset, and 2) whether they help mitigate overfitting on the training dataset. The experimental settings and results are presented in Table 3.

training settings	training auc	test acc	test auc	over-fitting indicator
no random	0.9916	84.96%	0.8993	0.0923

augmentation				
+ cropping	0.9682	86.43%(+1.46%)	0.9104 (+0.0111)	0.0578 (-0.0345)
+ rotation	0.9712	84.37%(-0.59%)	0.8965 (-0.0028)	0.0747 (- 0.0176)
+ perspective transformation	0.9590	**89.09%(+4.13%)**	**0.9272 (+0.0279)**	0.0318 (-0.0605)
all augmentation	0.9407	88.20%(+3.23%)	0.9218 (+0.0225)	**0.0189 (-0.0734)**

Table 3. Ablation study results.

In the experiments, we maintained consistent training settings while using the ConvNeXt-Tiny model to assess various data augmentation procedures. To ensure uniformity in data input size, we replaced random cropping with central cropping in the no-random cropping setting. We measured model overfitting by comparing the AUC of the training and test sets. The results in Table 3. indicate that each random augmentation operation effectively reduces overfitting. Notably, random cropping improved accuracy by 1.46% and AUC by 0.0111, while random view transformation enhanced accuracy by 4.13% and AUC by 0.0279. A model incorporating all three random augmentation transformations (applied to all training process in Section 4.1.2) achieved comparable accuracy and AUC to one using only random viewpoint transformation, but with a smaller training-test AUC gap. Given the differences between our dataset and the real population, we believe this model demonstrates superior generalization to unseen samples.

4.2 Model Interpretability Study

4.2.1 CAM technique selection

We applied the CAM techniques discussed in Section 3.2 to visualize the features extracted by our optimal models, aiming to highlight lesion regions. Specifically, we used two popular CAM variants: the gradient-based Grad-CAM and the score-based Score-CAM. For comparison, alongside ConvNeXt-Tiny, we also visualized features from the lightweight RegNetX-400mf model. As shown, RegNetX-400mf achieved a prediction accuracy of 82.89%, an AUC of 0.8862, and an F1 score of 0.8335 on the test dataset—lower than the ConvNeXt-Tiny model, which operates at 4.46 GFLOPs.

We found that the visualized feature maps (Figures 7 and 8) from both models offer insights into how they classify pneumonia, but with distinct characteristics. In the RegNetX-400mf model, both CAM techniques produced similar results, with diffuse high-heat regions covering nearly the entire lung area in the CXR images. In contrast, ConvNeXt's high-heat regions were more focused. This suggests that the limited parameter size and low computational load of RegNetX-400mf may restrict its ability to learn detailed features from the pediatric pneumonia dataset. In short, models like RegNetX-400mf may lack the complexity needed to accurately localize lesions in CXRs.

Compared to RegNetX-400mf, the ConvNeXt-Tiny model shows notable differences in feature maps across CAM techniques (Figure 8). In Grad-CAM visualizations, high-heat regions often focus on irrelevant areas (such as for the Mycoplasma class), suggesting that gradient-based methods may misrepresent the model's decision process in this task. In contrast, Score-CAM visualizations more accurately highlight suspected lesion areas related to pneumonia.

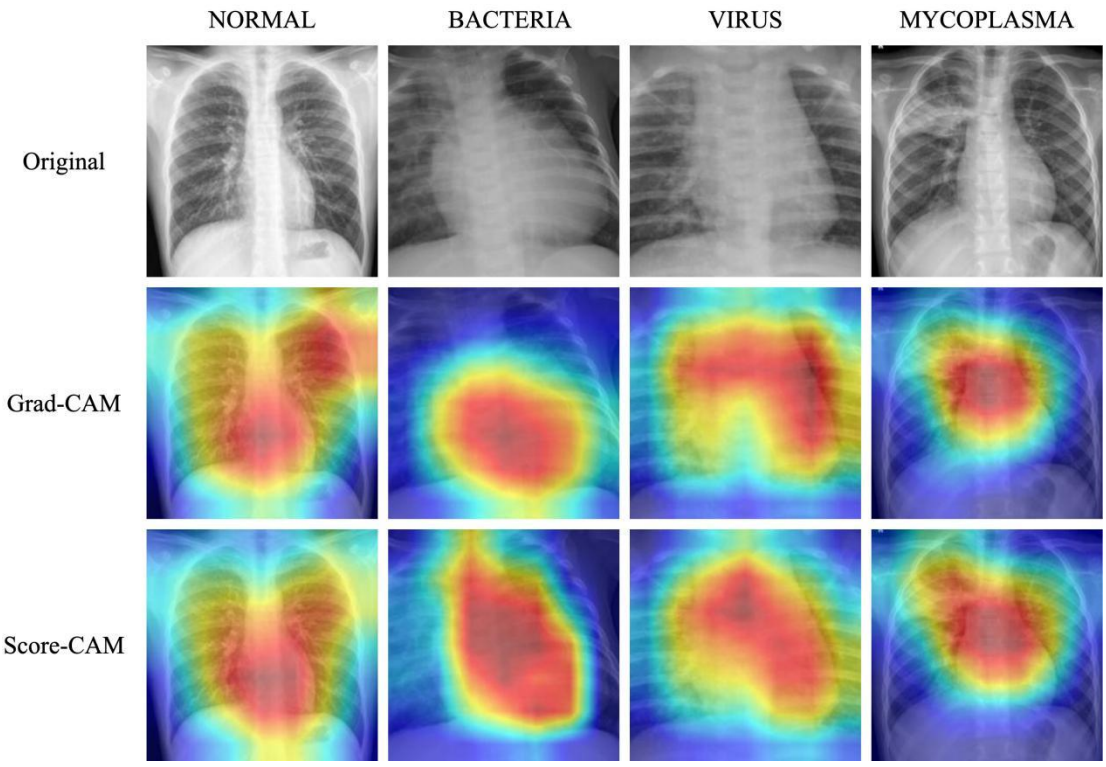


Figure 7. The feature visualization results of the RegNetX-400mf model.

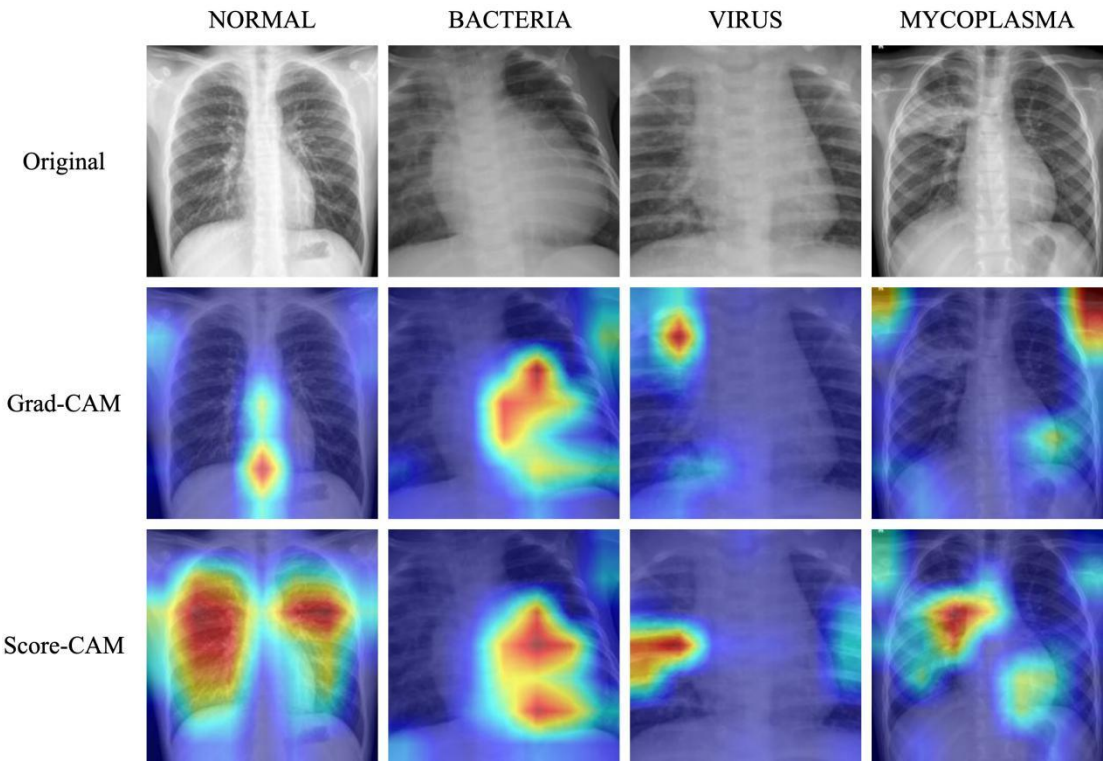


Figure 8. The feature visualization results of the ConvNeXt-Tiny model.

4.2.2 Clinical Validation

To validate the effectiveness of our method and the accuracy of CAM-based lesion localization, we compare the results with radiologist-annotated CXRs. Using the ConvNeXt-Tiny model trained in Section 4.1.1 and Score-CAM, we generate visual hints for MPP lesions on the test dataset. The comparison is shown in Figure 9.

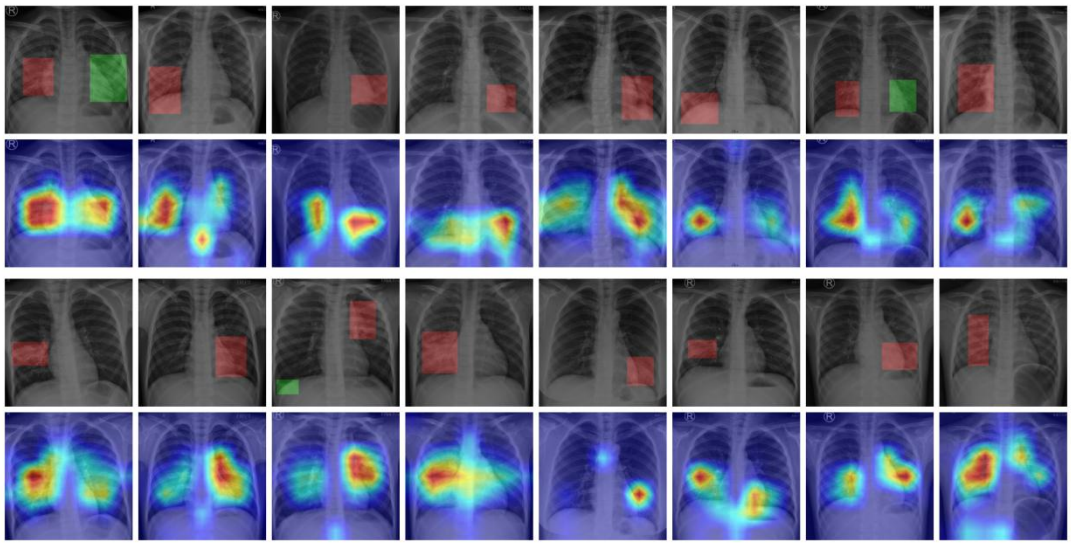
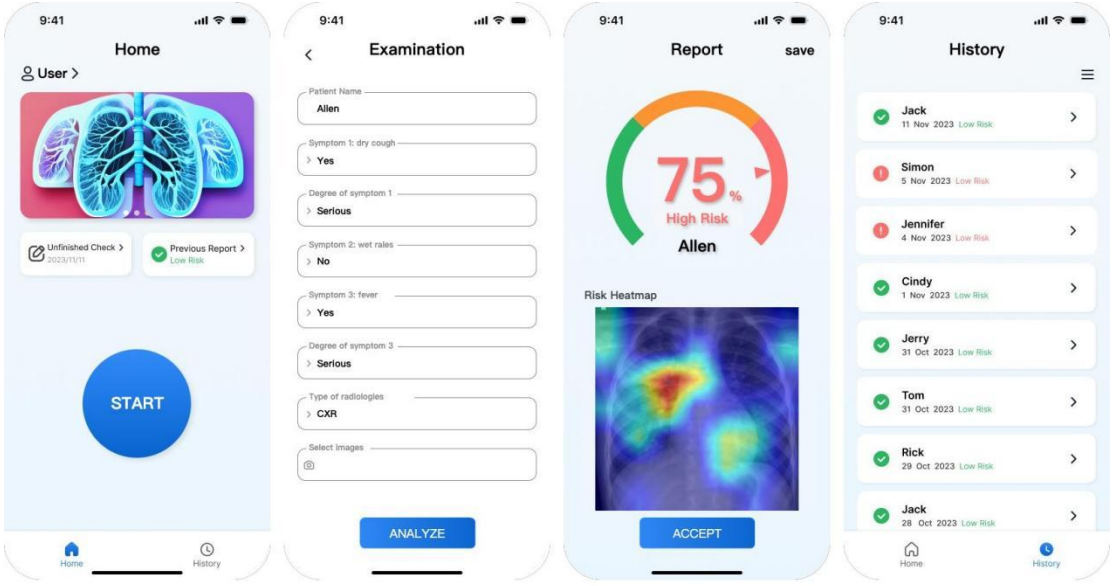


Figure 9. Comparison of Score-CAM results and radiologists' annotations. The first and third rows show the CXR annotations for MPP lesions, while the second and fourth rows display the corresponding Score-CAM results.

The ConvNeXt-Tiny model with Score-CAM effectively localized suspected MPP lesion areas in CXRs. Although the boundaries of the high-heat regions in the heatmaps are somewhat vague, the accuracy is sufficient to serve as a visual aid for detecting MPP lesions. This approach can assist respiratory physicians and radiologists in faster identification of MPP cases during daily diagnosis.

4.3 Mobile App Demo

To facilitate the application of our pediatric pneumonia prediction model in real-world scenarios, we have developed an Android application prototype named "PneumoniaAPP." The user interface of PneumoniaApp is illustrated in Figure 10. Notably, we display the softmax (Eq. (5)) mapped model output of correspondence pneumonia class as a risk score (in (c) Report page), indicating the degree of confidence for decision making of the model.



(a) Home page (b) Image analysis (c) Report page (d) History page

Figure 10. PneumoniaApp demo user interface example.

5. Conclusion

Pediatric Mycoplasma pneumoniae pneumonia (MPP) presents periodic outbreaks, posing diagnostic and treatment challenges globally. In this study, we applied advanced AI techniques to accelerate the diagnosis of various pediatric pneumonia types and the detection of MPP lesions in chest X-rays (CXRs). Using a combined CXR dataset, we trained a pneumonia classification model, employing data augmentation and early stopping to mitigate overfitting. We developed a ConvNeXt-Tiny model for pneumonia classification, comparing its performance with other lightweight architectures. The best model achieved 88.20% accuracy and an AUC of 0.9218. We also used the Score-CAM technique to assist physicians in lesion localization. To support real-world use, we created a mobile app, PneumoniaApp, integrating these techniques while ensuring user privacy.

Compared to previous studies using deep learning for pneumonia diagnosis, our approach offers several unique advantages: (1) We address Mycoplasma pneumoniae, a critical yet often overlooked pathogen, reducing the risk of misdiagnosis. (2) We focus on children aged 0-12, the most affected group in outpatient settings. (3) Our research emphasizes mobile deployment, ensuring that the models can be implemented in an Android app with less performance concerns.

Data Sharing

De-identified reader data can be made available upon request to the corresponding author for academic research purposes only. The data will be shared after signing a data use agreement, The Al software is commercially available.

Declaration of competing interest

The authors declare that they have no competing interests. The authors claim that none of the material in the paper has been published or is under consideration for publication elsewhere. All authors have seen the manuscript and approved to submit to your journal.

Ethical declaration

The study protocol was thoroughly reviewed and approved by the Ethical Committee of the Tsinghua Shenzhen International Graduate School, Tsinghua University (approval number 2023-F036). All data processing and analyzing methods were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from all subjects and/or their legal guardian(s) prior to participation in the study.

Acknowledgments

We thank the support from the National Natural Science Foundation of China 31970752,32350410397; Science, Technology, Innovation Commission of ShenzhenMunicipality, JCYJ20220530143014032, JCYJ20230807113017035, WDZC20200820173710001; Shenzhen Science and Technology Program, JCYJ20230807113017035; Shenzhen Medical Research Funds, D2301002; Department of Chemical Engineering-iBHE special cooperation joint fund project, DCE-iBHE-2022-3; Tsinghua Shenzhen International Graduate School Crossdisciplinary Research and Innovation Fund Research Plan, JC2022009; and Bureau of Planning, Land and Resources of Shenzhen Municipality (2022) 207.

References:

[1] L.-W. Gao, J. Yin, Y. Hu, X. Liu, X. Feng, J.-X. He, J. Liu, Y. Guo, B.-P. Xu, K.-L. Shen, The epidemiology of paediatric Mycoplasma pneumoniae pneumonia in North China: 2006 to 2016, Epidemiol Infect 147 (2019) e192.

[2] Z. Song, G. Jia, G. Luo, C. Han, B. Zhang, X. Wang, Global research trends of Mycoplasma pneumoniae pneumonia in children: A bibliometric analysis, Front Pediatr 11 (2023).

[3] J. Gao, B. Yue, H. Li, R. Chen, C. Wu, M. Xiao, Epidemiology and clinical features of segmental/lobar pattern Mycoplasma pneumoniae pneumonia: A ten-year retrospective clinical study, Exp Ther Med 10 (2015) 2337–2344.

[4] Y. Lu, Y. Wang, C. Hao, W. Ji, Z. Chen, W. Jiang, Y. Yan, W. Gu, Clinical characteristics of pneumonia caused by Mycoplasma pneumoniae in children of different ages, Int J Clin Exp Pathol 11 (2018) 855.

[5] C. Yan, G.-H. Xue, H.-Q. Zhao, Y.-L. Feng, J.-H. Cui, J. Yuan, Current status of Mycoplasma pneumoniae infection in China, World Journal of Pediatrics (2024) 1–4.

[6] A.M. Fletcher, S. Bhattacharyya, Infectious myelopathies, CONTINUUM: Lifelong Learning in Neurology 30 (2024) 133–159.

[7] X.-B. Zhang, W. He, Y.-H. Gui, Q. Lu, Y. Yin, J.-H. Zhang, X.-Y. Dong, Y.-W. Wang, Y.-Z. Ye, H. Xu, Current Mycoplasma pneumoniae epidemic among children in Shanghai: unusual pneumonia caused by usual pathogen, World Journal of Pediatrics (2024) 1–6.

[8] T.T.N. Dung, V.V. Phat, C. Vinh, N.P.H. Lan, N.L.N. Phuong, L.T.Q. Ngan, G. Thwaites, L. Thwaites, M. Rabaa, A.T.K. Nguyen, Development and validation of multiplex real-time PCR for simultaneous detection of six bacterial pathogens causing lower respiratory tract infections and antimicrobial resistance genes, BMC Infect Dis 24 (2024) 164.

[9] A.K. Singh, A. Kumar, V. Kumar, S. Prakash, COVID-19 Detection using adopted convolutional neural networks and high-performance computing, Multimed Tools Appl 83 (2024) 593–608.

[10] P. Dwivedi, S. Padhi, S. Chakraborty, S.C. Raikwar, Severity wise COVID-19 X-ray image augmentation and classification using structure similarity, Multimed Tools Appl (2023) 1–22.

[11] A. Sadeghi, M. Sadeghi, A. Sharifpour, M. Fakhar, Z. Zakariaei, M. Sadeghi, M. Rokni, A. Zakariaei, E.S. Banimostafavi, F. Hajati, Potential diagnostic application of a novel deep learning-based approach for COVID-19, Sci Rep 14 (2024) 280.

[12] A. Shoeibi, M. Khodatars, M. Jafari, N. Ghassemi, D. Sadeghi, P. Moridian, A. Khadem, R. Alizadehsani, S. Hussain, A. Zare, Automated detection and forecasting of covid-19 using deep learning techniques: A review, Neurocomputing (2024) 127317.

[13] R. Alaufi, M. Kalkatawi, F. Abukhodair, Challenges of deep learning diagnosis for COVID-19 from chest imaging, Multimed Tools Appl 83 (2024) 14337–14361.

[14] A.S. Moosavi, A. Mahboobi, F. Arabzadeh, N. Ramezani, H.S. Moosavi, G. Mehrpoor, Segmentation and classification of lungs CT-scan for detecting COVID-19 abnormalities by deep learning technique: U-Net model, J Family Med Prim Care 13 (2024) 691–698.

[15] J. Chen, Y. Li, L. Guo, X. Zhou, Y. Zhu, Q. He, H. Han, Q. Feng, Machine learning techniques for CT imaging diagnosis of novel coronavirus pneumonia: a review, Neural Comput Appl 36 (2024) 181–199.

[16] N.S. Alghamdi, M. Zakariah, H. Karamti, A deep CNN-based acoustic model for the identification of lung diseases utilizing extracted MFCC features from respiratory sounds,

Multimed Tools Appl (2024) 1–33.
[17] K.M. Abubeker, S. Baskar, P. Yadav, Internet of Things Assisted Wireless Body Area Network Enabled Biosensor Framework for Detecting Ventilator and Hospital-Acquired Pneumonia, (2024).
[18] M.A.A. Al-qaness, J. Zhu, D. AL-Alimi, A. Dahou, S.H. Alsamhi, M. Abd Elaziz, A.A. Ewees, Chest X-ray Images for Lung Disease Detection Using Deep Learning Techniques: A Comprehensive Survey, Archives of Computational Methods in Engineering (2024) 1–35.
[19] A.A. Nafea, S.A. Alameri, R.R. Majeed, M.A. Khalaf, M.M. AL-Ani, A Short Review on Supervised Machine Learning and Deep Learning Techniques in Computer Vision, Babylonian Journal of Machine Learning 2024 (2024) 48–55.
[20] S. Kumar, H. Kumar, G. Kumar, S.P. Singh, A. Bijalwan, M. Diwakar, A methodical exploration of imaging modalities from dataset to detection through machine learning paradigms in prominent lung disease diagnosis: a review, BMC Med Imaging 24 (2024) 30.
[21] D.S. Kermany, M. Goldbaum, W. Cai, C.C.S. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, Identifying medical diagnoses and treatable diseases by image-based deep learning, Cell 172 (2018) 1122–1131.
[22] O. Stephen, M. Sain, U.J. Maduh, D.-U. Jeong, An efficient deep learning approach to pneumonia classification in healthcare, J Healthc Eng 2019 (2019).
[23] H. Sharma, J.S. Jain, P. Bansal, S. Gupta, Feature extraction and classification of chest x-ray images using cnn to detect pneumonia, in: 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, 2020: pp. 227–231.
[24] H. Liz, M. Sánchez-Montañés, A. Tagarro, S. Domínguez-Rodríguez, R. Dagan, D. Camacho, Ensembles of convolutional neural network models for pediatric pneumonia diagnosis, Future Generation Computer Systems 122 (2021) 220–233.
[25] C.R. Asswin, D.K. KS, A. Dora, V. Ravi, V. Sowmya, E.A. Gopalakrishnan, K.P. Soman, Transfer learning approach for pediatric pneumonia diagnosis using channel attention deep CNN architectures, Eng Appl Artif Intell 123 (2023) 106416.
[26] A. Serener, S. Serte, Deep learning for mycoplasma pneumonia discrimination from pneumonias like COVID-19, in: 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), IEEE, 2020: pp. 1–5.
[27] O. Stephen, M. Sain, U.J. Maduh, D.-U. Jeong, An efficient deep learning approach to pneumonia classification in healthcare, J Healthc Eng 2019 (2019).
[28] H. Sharma, J.S. Jain, P. Bansal, S. Gupta, Feature extraction and classification of chest x-ray images using cnn to detect pneumonia, in: 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, 2020: pp. 227–231.
[29] A.M. Reza, Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement, J VLSI Signal Process Syst Signal Image Video Technol 38 (2004) 35–44.
[30] S. Dewan, C. Shi, V. Gurbaxani, Investigating the risk–return relationship of information technology investment: Firm-level empirical analysis, Manage Sci 53 (2007) 1829–1842.
[31] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, IEEE Trans Neural Netw Learn Syst 33 (2021) 6999–7019.
[32] M. Goldblum, H. Souri, R. Ni, M. Shu, V. Prabhu, G. Somepalli, P. Chattopadhyay, M. Ibrahim, A. Bardes, J. Hoffman, Battle of the backbones: A large-scale comparison of

pretrained models across computer vision tasks, Adv Neural Inf Process Syst 36 (2024).

[33] N. Schneider, F. Piewak, C. Stiller, U. Franke, RegNet: Multimodal sensor registration using deep neural networks, in: 2017 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2017: pp. 1803–1810.

[34] Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11976-11986).

[35] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: pp. 10012–10022.

[36] Y. Xu, Q. Zhang, J. Zhang, D. Tao, Vitae: Vision transformer advanced by exploring intrinsic inductive bias, Adv Neural Inf Process Syst 34 (2021) 28522–28535.

[37] Y. Fang, S. Yang, S. Wang, Y. Ge, Y. Shan, X. Wang, Unleashing vanilla vision transformer with masked image modeling for object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: pp. 6244–6253.

[38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv Neural Inf Process Syst 30 (2017).

[39] Tan, M. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946.

[40] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626).

[41] Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., ... & Hu, X. (2020). Score-CAM: Score-weighted visual explanations for convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 24-25).

[42] Peng, B., Chen, J., Githinji, P. B., Gul, I., Ye, Q., Chen, M., ... & Chen, Z. (2024). Practical Guidelines for Cell Segmentation Models Under Optical Aberrations in Microscopy. Computational and Structural Biotechnology Journal.

[43] Fan, D., Yuan, X., Wu, W., Zhu, R., Yang, X., Liao, Y., ... & Qin, P. (2022). Self-shrinking soft demoulding for complex high-aspect-ratio microchannels. Nature Communications, 13(1), 5083.

[44] Chen, Z., Li, F., Zhang, L., Lei, Z., Yang, C., Xiao, C., ... & Qin, P. (2023). Temperature tolerant all-solid-state touch panel with high stretchablity, transparency and self-healing ability. Chemical Engineering Journal, 451, 138672.

[45] Wang, C., Qin, P., Lv, D., Wan, J., Sun, S., & Ma, H. (2020). Characterization of anisotropy of the porous anodic alumina by the Mueller matrix imaging method. Optics express, 28(5), 6740-6754.

[46] Sarma, A. V., Anbanandam, A., Kelm, A., Mehra-Chaudhary, R., Wei, Y., Qin, P., ... & Van Doren, S. R. (2012). Solution NMR of a 463-residue phosphohexomutase: domain 4 mobility, substates, and phosphoryl transfer defect. Biochemistry, 51(3), 807-819.

[47] Xie, Y., Yang, H., Yuan, X., He, Q., Zhang, R., Zhu, Q., ... & Yan, C. (2021). Stroke prediction from electrocardiograms by deep neural network. Multimedia Tools and Applications, 80, 17291-17297.

[48] Chen, X., Miller, A., Cao, S., Gan, Y., Zhang, J., He, Q., ... & Du, K. (2020). Rapid

escherichia coli trapping and retrieval from bodily fluids via a three-dimensional bead-stacked nanodevice. ACS applied materials & interfaces, 12(7), 7888-7896.