

基于多模型比较的肺炎X射线影像识别方法研究

杨冰倩^{1,2}, 张艳敏^{1,2}, 孙凤杰¹, 赵占坤¹

¹河北软件职业技术学院软件工程系, 河北 保定

²保定市人工智能 + 智慧交通领域重点实验室, 河北 保定

收稿日期: 2026年7月20日; 录用日期: 2026年8月12日; 发布日期: 2026年8月20日

摘要

目的: 比较不同卷积神经网络模型在肺炎X射线影像识别任务中的性能, 为医学影像智能识别模型的选择提供实验依据。方法: 以公开肺炎胸部X射线影像数据集(Chest X-Ray Images (Pneumonia))为研究对象, 选取AlexNet、VGG16、ResNet50、DenseNet121和MobileNetV2五种典型CNN模型进行对比实验, 采用迁移学习策略进行模型训练, 并使用Accuracy、Precision、Recall、F1-score和AUC等指标综合评价模型性能。结果: VGG16模型在肺炎识别任务中表现最优, Accuracy达90.38%, F1-score达92.59%。结论: 不同CNN模型在肺炎X射线影像识别任务中存在显著性能差异, 为临床辅助诊断系统的模型选择提供了参考依据。

关键词

肺炎识别, 卷积神经网络, X射线影像, 迁移学习, 模型比较

Research on Pneumonia X-Ray Image Recognition Method Based on Multi-Model Comparison

Bingqian Yang^{1,2}, Yanmin Zhang^{1,2}, Fengjie Sun¹, Zhankun Zhao¹

¹Department of Software Engineering, Hebei Software Institute, Baoding Hebei

²Baoding Key Laboratory of Artificial Intelligence and Intelligent Transportation, Baoding Hebei

Received: July 20, 2026; accepted: August 12, 2026; published: August 20, 2026

Abstract

Objective: To compare the performance of different convolutional neural network (CNN) models

文章引用: 杨冰倩, 张艳敏, 孙凤杰, 赵占坤. 基于多模型比较的肺炎 X 射线影像识别方法研究[J]. 软件工程与应用, 2026, 15(4): 578-585. DOI: 10.12677/sea.2026.154053

in pneumonia X-ray image recognition tasks, and to provide experimental evidence for the selection of intelligent recognition models in medical imaging. **Methods:** Using the publicly available Chest X-Ray Images (Pneumonia) dataset, five typical CNN models—AlexNet, VGG16, ResNet50, DenseNet121, and MobileNetV2—were selected for comparative experiments. Transfer learning was adopted for model training, and model performance was comprehensively evaluated using Accuracy, Precision, Recall, F1-score, and AUC. **Results:** The VGG16 model achieved the best performance in pneumonia recognition, with an Accuracy of 90.38% and an F1-score of 92.59%. **Conclusion:** Significant performance differences exist among different CNN models in pneumonia X-ray image recognition tasks, providing a reference for model selection in clinical assisted diagnosis systems.

Keywords

Pneumonia Recognition, Convolutional Neural Network, X-Ray Image, Transfer Learning, Model Comparison

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

肺炎是一种常见的呼吸系统疾病，胸部 X 射线影像(Chest X-ray, CXR)因具有检查成本低、效率高等特点，成为临床肺部疾病筛查的重要辅助工具。然而，传统影像诊断主要依赖医生经验，面对大规模医学影像数据时，容易受到工作强度和个体经验差异等因素影响。因此，利用人工智能技术辅助开展医学影像分析，提高疾病筛查效率，逐渐成为智慧医疗领域的重要研究方向。

近年来，深度学习技术在医学影像分析领域得到广泛应用。卷积神经网络(Convolutional Neural Network, CNN)能够通过多层结构自动提取图像中的特征信息，在医学图像分类、目标检测和辅助诊断等任务中表现出良好的应用效果。Krizhevsky 等提出的 AlexNet 模型推动了深度卷积神经网络在图像识别领域的发展[1]；Simonyan 等提出的 VGG 网络通过增加网络深度和采用小尺寸卷积核提升了特征表达能力[2]；He 等提出的 ResNet 通过残差连接机制有效解决了深层网络训练困难问题[3]。近年来，基于 CNN 的医学影像分析方法不断发展，研究人员利用 DenseNet、迁移学习等技术开展肺部疾病识别研究，并取得了较好的分类效果[4]。

针对肺炎 X 射线影像识别任务，现有研究主要集中于利用单一深度学习模型进行分类性能优化，而不同网络结构在医学影像任务中的适应性差异仍缺少系统比较。由于医学影像具有数据规模有限、病灶区域特征复杂等特点，不同深度网络在特征提取能力、模型复杂度以及泛化性能方面存在一定差异。因此，在统一实验条件下分析不同卷积神经网络模型对肺炎 X 射线影像识别任务的适用性，对于模型选择和医学辅助诊断系统构建具有一定参考价值。

本文以公开肺炎胸部 X 射线影像数据集(Chest X-Ray Images (Pneumonia))为研究对象，构建基于多模型比较的肺炎影像识别实验体系。选取 AlexNet、VGG16、ResNet50、DenseNet121 和 MobileNetV2 五种典型 CNN 模型进行对比实验，并采用 Accuracy、Precision、Recall、F1-score 和 AUC 等指标对模型性能进行综合评价，分析不同网络结构在肺炎 X 射线影像分类任务中的表现特点，为医学影像智能识别模型的选择提供实验依据。

2. 数据集与相关技术

2.1. 数据集介绍

本文采用 Kermamy 等人在 2018 年公开的胸部 X 射线影像数据集(Chest X-Ray Images (Pneumonia))作为实验数据, 该数据集来源于儿童患者胸部 X 射线检查影像, 主要用于肺炎辅助诊断研究[5]。相比普通自然图像数据集, 医学影像具有类别间差异小、病灶区域不明显等特点, 因此适合作为验证深度学习模型医学图像分类能力的数据来源。

该数据集包含两类胸部 X 射线影像, 分别为正常肺部影像(Normal)和肺炎影像(Pneumonia)。其中, 肺炎类别主要包括细菌性肺炎(Bacterial Pneumonia)和病毒性肺炎(Viral Pneumonia)。原始数据按照训练集(Train)、验证集(Validation)和测试集(Test)进行划分, 由于原始验证集样本数量较少, 仅包括 16 张, 本文重新对训练集和验证集进行划分, 按照 9:1 比例随机划分训练集和验证集, 测试集保持官方划分不变。具体数据分布见表 1 所示。

Table 1. Composition of the Chest X-Ray Images (Pneumonia) dataset

表 1. Chest X-Ray Images (Pneumonia)数据集组成

数据类别	Normal	Pneumonia	总数量
训练集	1072	3637	4709
验证集	277	246	523
测试集	234	390	624

2.2. 图像预处理方法

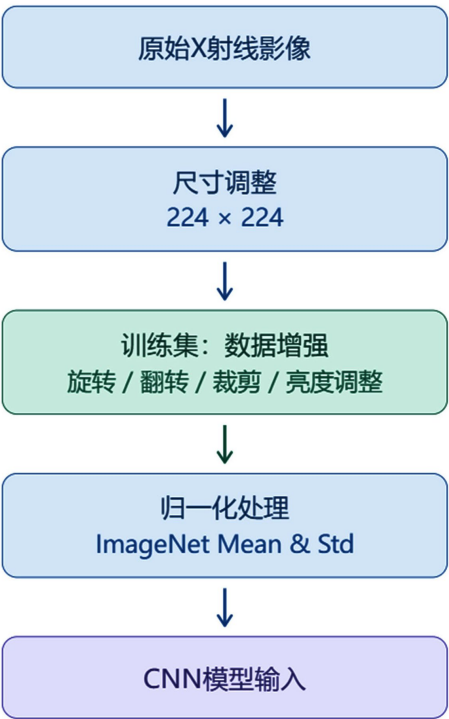


Figure 1. Flowchart of data preprocessing

图 1. 数据预处理流程图

胸部 X 射线影像具有成像设备、拍摄角度以及患者个体差异等因素造成的图像分布差异，直接将原始影像输入卷积神经网络可能会影响模型训练稳定性。为提高模型对肺炎影像特征的学习能力，本文在模型训练前对原始 X 射线图像进行统一预处理，主要包括图像尺寸调整、数据增强以及归一化处理三个步骤。数据预处理流程图如图 1 所示。

2.3. 迁移学习方法

医学影像数据通常存在样本获取困难、标注成本较高等特点，直接利用有限数据训练深层卷积神经网络容易产生过拟合问题。迁移学习(Transfer Learning)通过利用已有任务中学习到的特征知识，并将其迁移到新的目标任务中，可以有效降低模型训练对大规模标注数据的依赖，因此逐渐成为医学影像智能分析中的常用方法。

近年来，迁移学习已广泛应用于医学图像分类任务。相关研究表明，基于预训练模型的迁移学习方法能够在医学数据规模有限的情况下提升模型训练效果，并改善模型的泛化能力[6]。针对医学影像分类任务，研究人员通常采用模型结构调整、参数微调以及特征迁移等方式实现知识迁移[7]。

本文采用基于迁移学习的 CNN 模型训练策略，将 ImageNet 预训练的 AlexNet、VGG16、ResNet50、DenseNet121 和 MobileNetV2 作为骨干网络，通过替换全连接层，使模型适应肺炎 X 射线影像分类任务，并在目标数据集上进行参数优化。该方法能够充分利用预训练模型在大规模自然图像数据集上学习到的特征表示能力，提升模型在小规模医学影像数据集上的识别性能。

3. 肺炎 X 射线影像识别方法

3.1. 方法框架

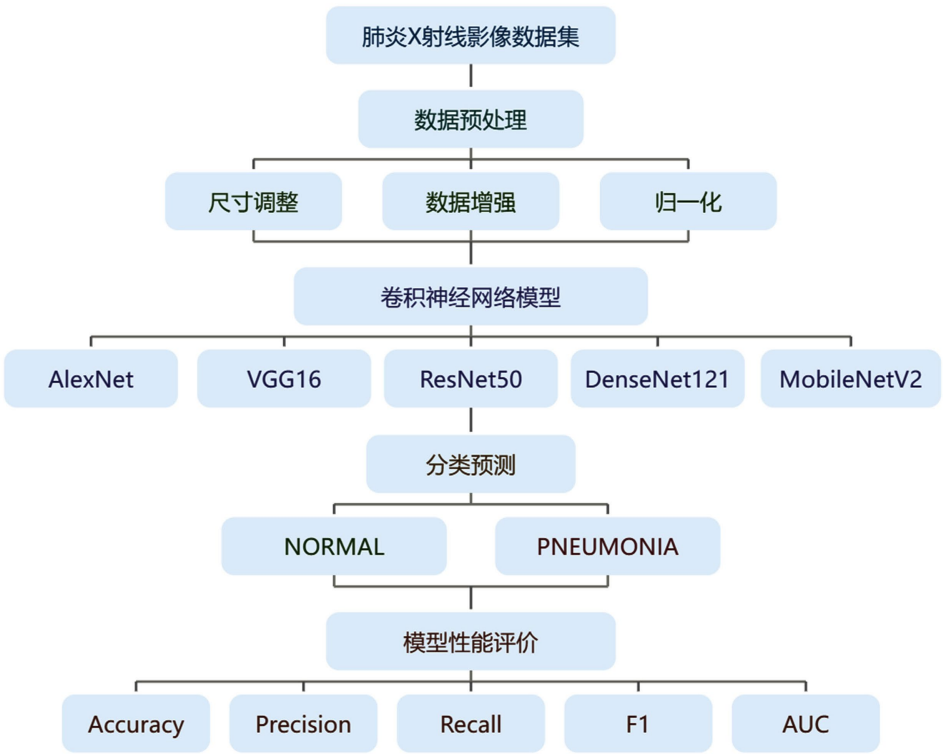


Figure 2. Framework of pneumonia X-ray image recognition based on multi-model comparison
图 2. 基于多模型比较的肺炎 X 射线影像识别框架图

针对肺炎 X 射线影像自动识别任务，本文构建了一种基于多模型比较的医学影像分类方法框架，见图 2。该方法框架主要包括数据预处理、深度学习模型构建、模型训练优化以及性能评估四个阶段。首先，对原始胸部 X 射线影像数据进行统一预处理，包括图像尺寸调整、数据增强以及灰度归一化等操作，以提高模型输入数据的一致性和多样性。其次，基于迁移学习思想，构建五种经典卷积神经网络模型，包括 AlexNet、VGG16、ResNet50、DenseNet121 和 MobileNetV2，通过统一实验条件对不同模型在肺炎影像识别任务中的性能进行比较分析。最后，通过准确率、精确率、召回率、F1-score 和 AUC 等多维度指标对模型分类性能进行综合评价，分析各模型在参数规模与模型效率方面对不同模型的适用场景。

3.2. 典型 CNN 模型构建

为全面评估不同卷积神经网络模型在肺炎 X 射线影像识别任务中的适用性，本文选取了具有代表性的五种深度学习模型进行实验比较(表 2)，包括 AlexNet、VGG16、ResNet50、DenseNet121 和 MobileNetV2。其中，AlexNet 是首个深度卷积神经网络模型；VGG16 采用连续小卷积核堆叠的深度网络结构；ResNet50 通过引入残差连接解决了深层网络训练困难问题；DenseNet121 采用密集连接机制增强特征传播能力；MobileNetV2 使用深度可分离卷积构建轻量级模型。通过上述模型组合，可以覆盖经典 CNN、残差网络、密集连接网络以及轻量级网络等不同技术路线，为比较不同模型在肺炎 X 射线影像识别中的性能差异提供实验基础。

Table 2. Comparison of typical CNN models
表 2. 典型 CNN 模型对比

模型	提出年份	核心思想	主要特点
AlexNet	2012	深层 CNN	首个深度 CNN
VGG16	2014	小卷积核堆叠	更深更宽的学习
ResNet50	2015	残差连接	梯度传播优化
DenseNet121	2017	密集连接	特征复用
MobileNetV2	2018	深度可分离卷积	轻量级模型

3.3. 模型训练与评价指标

在肺炎 X 射线影像识别任务中，模型不仅要具有较高的分类准确率，同时需要具备较强的泛化能力和稳定性。由于医学影像分类任务中不同类别的样本数量可能存在一定差异，仅依靠准确率(Accuracy)不能全面反映模型性能。

因此，本文从分类准确性、正负样本识别能力以及模型综合性能等多角度出发，选取 Accuracy、Precision、Recall、F1-score、Specificity 以及 AUC 共 6 个指标对模型进行综合评价(表 3)。其中，Accuracy 用于衡量模型整体分类的正确程度；Precision 用于评估模型预测为肺炎患者的可靠性；Recall 用于衡量模型对肺炎患者的发现能力；Specificity 用于评估模型正确识别正常样本的能力；F1-score 综合考量 Precision 和 Recall 的调和平均；AUC 用于评估模型在不同阈值下的整体分类性能。通过上述指标的综合评价，可以更全面地分析不同卷积神经网络模型在肺炎 X 射线影像识别任务中的性能差异。

4. 实验设计与结果分析

4.1. 实验环境

本文实验基于深度学习框架 PyTorch 实现模型训练和评估。实验环境配置如表 4 所示。

Table 3. Model evaluation metrics
表 3. 模型评价指标

评价指标	计算方式	评价意义
Accuracy	分类正确的样本数/总样本数	整体分类性能
Precision	预测为正类中的真正类比例	假阳性控制能力
Recall	真正类中被正确识别的比例	漏诊风险评估
Specificity	真实负类中被正确识别的比例	误诊风险评估
F1-score	Precision 与 Recall 的调和平均	综合评价模型
AUC	ROC 曲线下面积	整体辨别能力

Table 4. Experimental environment configuration
表 4. 实验环境配置

配置项	参数
操作系统	Linux
编程语言	Python
深度学习框架	PyTorch
GPU	NVIDIA GPU
CUDA 版本	CUDA 12.x
开发工具	Jupyter Notebook

4.2. 实验参数设置

通过训练和文献调研，本文采用 Adam 优化算法进行模型参数更新，通过最小化交叉熵损失函数来缩小预测结果与真实标签之间的差异。为避免模型过拟合，引入 EarlyStopping 策略，当验证集性能在连续多个 epoch 内不再提升时停止训练。主要模型训练参数配置如表 5 所示。

Table 5. Model training parameter configuration
表 5. 模型训练参数配置表

参数	设置
输入尺寸	224 × 224
Batch Size	32
优化器	Adam
初始学习率	0.0001
损失函数	CrossEntropyLoss
最大训练轮数	50
EarlyStopping	10
学习率调整	ReduceLROnPlateau

4.3. 实验结果分析

4.3.1. 量化指标比较

表 6 展示了不同模型量化指标对比。不同卷积神经网络模型在肺炎 X 射线影像识别任务中的性能存在显著差异。其中，VGG16 模型取得了最高的 Accuracy 为 90.38%，同时 F1-score 达到 92.59%，说明该模型能够较好地平衡肺炎识别与正常分类的整体性能。AlexNet 虽然网络结构较浅，但 Recall 达到 97.44%，说明其对肺炎病种具有较强的检测能力，能够有效降低漏诊风险。ResNet50 采用残差结构增强特征学习能力，在 Precision 和 Recall 之间取得了较好的平衡。DenseNet121 虽然 Recall 最高(98.46%)，但 Accuracy 较低，说明模型倾向于预测为正类，对正常样本识别存在一定不足。MobileNetV2 由于参数量限制，在计算效率方面具有优势，但模型性能相应下降，分类准确率最低。综合 Accuracy、F1-score 和 AUC 等指标，VGG16 模型在本实验条件下表现最优。

Table 6. Comparison of experimental results

表 6. 实验结果对比

模型	Accuracy	Precision	Recall	F1-score	AUC
AlexNet	0.8990	0.8776	0.9744	0.9235	0.9571
DenseNet121	0.8542	0.8188	0.9846	0.8941	0.9567
MobileNetV2	0.8157	0.7996	0.9410	0.8645	0.9242
ResNet50	0.8910	0.8833	0.9513	0.9160	0.9493
VGG16	0.9038	0.8929	0.9615	0.9259	0.9571

4.3.2. 混淆矩阵分析

在肺炎辅助诊断中，假阴性(False Negative, FN)即将患者误判为正常人的风险，在临床意义上远大于假阳性(False Positive, FP)即将正常人误判为肺炎的风险，因为 FN 意味着患者的黄金治疗窗口。基于此，我们对各模型的误分类情况进行了分析。不同模型误分类情况见表 7。

Table 7. Comparison of misclassification among different models

表 7. 不同模型误分类情况对比

模型	假阴性(FN)	假阳性(FP)	FN/FP 比值	临床风险
AlexNet	6 (最低)	85 (最高)	0.071	临床倾向性
VGG16	10	53	0.189	极高敏感度
ResNet50	15	45 (最低)	0.333	高敏感度
DenseNet121	19	49	0.388	高特异度
MobileNetV2	23	92	0.250	均衡偏向特异度

4.4. 模型选型分析

通过对不同卷积神经网络模型的综合比较，可以得出以下选型建议：(1) 若以分类准确率为首要目标，在本文实验数据集和实验参数条件下，VGG16 模型取得最佳综合性能，适合对诊断精度要求较高的应用场景。(2) 若以降低漏诊风险为首要考虑，AlexNet 具有最高的召回率，适合作为初步筛查模型。(3) 若需兼顾模型性能和计算效率，ResNet50 在 Accuracy 和模型复杂度之间取得了较好的平衡。(4) 对于资源受限的部署环境，MobileNetV2 虽然性能略低，但模型参数量最小，适合移动端或嵌入式设备部署。

5. 结论

本文针对肺炎 X 射线影像识别任务,在统一实验条件下对 AlexNet、VGG16、ResNet50、DenseNet121 和 MobileNetV2 五种典型卷积神经网络模型进行了系统的比较研究。实验结果表明,不同网络结构在肺炎影像识别任务中的性能存在明显差异,VGG16 模型在整体分类性能方面表现最优,Accuracy 和 F1-score 分别达到 90.38%和 92.59%。

本文工作为医学影像智能识别模型的选择提供了实验依据,为肺炎 X 射线影像分类任务中的模型选择提供实验参考。未来的研究工作将考虑引入更先进的模型结构和多模态数据融合方法,进一步提升肺炎影像识别的准确性和鲁棒性。

此外,本实验未对训练集类别不平衡进行专门校正,模型在正常类别上的识别性能仍有提升空间,未来可引入类别加权损失或动态阈值调整方法以进一步优化临床适用性。

基金项目

2025 年保定市科技计划项目;项目名称:医疗大数据背景下人工智能在疾病预测中的应用研究;项目编号:2511ZG018。

参考文献

- [1] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, Lake Tahoe, 3-6 December 2012, 1097-1105.
- [2] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. <https://arxiv.org/abs/1409.1556>
- [3] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016. 770-778. <https://doi.org/10.1109/cvpr.2016.90>
- [4] Rajpurkar, P., Irvin, J., Zhu, K., *et al.* (2017) CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. <https://arxiv.org/abs/1711.05225>
- [5] Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C.S., Liang, H., Baxter, S.L., *et al.* (2018) Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, **172**, 1122-1131.e9. <https://doi.org/10.1016/j.cell.2018.02.010>
- [6] 陈弘扬, 高敬阳, 赵地, 等. 深度学习与生物医学图像分析 2020 年综述[J]. 中国图象图形学报, 2021, 26(3): 456-472.
- [7] 黎英, 宋佩华. 迁移学习在医学图像分类中的研究进展[J]. 中国图象图形学报, 2022, 27(3): 721-742.