THE STANFORD JOURNAL
AN INTERNATIONAL JOURNAL
THE STANFORD JOURNAL ISSN: 3048-8923
(AN INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH)
Volume–2, Issues– 3 July- September: 2025

TRANSFER LEARNING FOR LOW-DATA ENVIRONMENTS: A COMPARATIVE STUDY ON CHEST X-RAY PNEUMONIA CLASSIFICATION

Ananya Rohatgi
Student
Department of Computer Science
SRM Institute of Science and Technology, Ghaziabad
Email: av9984@srmist.edu.in

ABSTRACT

Deep learning models have achieved state-of-the-art performance in medical imaging tasks, but often require large, annotated datasets—an expensive and time-consuming resource in healthcare. Transfer learning provides a solution by leveraging features from models pretrained on large-scale datasets such as ImageNet. This paper presents a CPU-friendly comparative study of three pre-trained convolutional neural network (CNN) architectures MobileNetV2, ResNet50, and DenseNet121—on pneumonia detection from chest X-ray images under low-data conditions. Experiments were conducted on subsets of the Kaggle Chest X-Ray Pneumonia dataset with label budgets of 25, 50, and 100 images per class. MobileNetV2 demonstrated the best balance between performance and computational efficiency, with up to 0.8987 AUC and 0.846 F1-score using only 100 training images per class. The results confirm the viability of lightweight transfer learning models in low-data medical imaging environments.

Keywords—*Transfer Learning, Medical Imaging, Low-Data Regimes, MobileNetV2, ResNet50, DenseNet121, Pneumonia Detection.*

I. INTRODUCTION

Deep learning has emerged as a powerful tool for medical image classification, especially for radiological analysis. However, its success often hinges on the availability of large, annotated datasets. Medical imaging data is challenging to acquire due to privacy restrictions, annotation complexity, and the requirement for expert-level labeling.

Transfer learning addresses this issue by adapting pretrained models, often trained on large-scale datasets like ImageNet, to smaller, domain-specific datasets. In low-data environments, transfer learning significantly improves convergence speed, feature extraction quality, and generalization performance.

This study investigates the extent to which pretrained CNN architectures can perform on pneumonia detection from chest X-rays with extremely limited data. We compare MobileNetV2, ResNet50, and DenseNet121, focusing on three data budgets and providing computationally efficient training methods suited for CPU-only scenarios.

II. LITERATURE REVIEW

Transfer learning has emerged as a pivotal approach in medical image analysis, overcoming the challenge of limited annotated datasets by leveraging pretrained models trained on large-scale datasets such as ImageNet. Because annotating medical images is expensive and time-consuming, transfer learning improves model generalization and accelerates convergence by transferring learned feature representations from source domains to specific medical tasks.

Recent advances have emphasized multistage transfer learning, where intermediate fine-tuning steps bridge the gap between natural image datasets and specialized medical imagery, improving robustness and domain adaptation. This strategy allows models to better capture complex hierarchical and modality-specific features critical for medical diagnosis.

Popular architectures used as pretrained backbones include AlexNet, VGGNet, ResNet, DenseNet, and MobileNet variants, which have shown promise when fine-tuned on various medical imaging modalities such as X-rays, CT scans, and MRIs. These CNN-based models remain dominant, though recent works have started exploring Vision Transformers (ViTs) for medical tasks.

Studies have also highlighted key benefits of transfer learning in this context:

- Data efficiency, enabling learning from small datasets with reduced risk of overfitting
- Improved model convergence and training speed
- Generalization across multiple imaging modalities
- Extraction of complex features necessary to delineate subtle pathological changes

Despite these successes, challenges remain, including domain mismatch, class imbalance, and the need for interpretability. Your work, focusing on low-data regimes for pneumonia detection from chest X-rays using lightweight pretrained CNNs, aligns well with this body of research.

III. METHODOLOGY

A. Dataset

We used the publicly available Kaggle Chest X-Ray Pneumonia dataset , containing images labeled as NORMAL or PNEUMONIA. The dataset includes train/, val/, and test/ splits with class balance maintained in each subset.

B. Data Sampling Strategy

To simulate low-data environments, we sampled 25, 50, and 100 images per class from the training set. Each subset selection was stratified by class with reproducible randomness via fixed seeds.

C. Preprocessing and Augmentation

Images were resized to 160×160 pixels, normalized with ImageNet mean/std, and

augmented with random horizontal flips and ±6° rotations.

D. Model Architectures

1. MobileNetV2 – Lightweight depthwise separable convolutions.
2. ResNet50 – 50-layer deep residual network.
3. DenseNet121 – Dense connectivity between convolutional layers.

All models were pretrained on ImageNet and adapted for binary classification.

E. Transfer Learning Setup

- Feature Extract Mode: Freeze backbone, train only the classifier head.
- Optimizer: AdamW, learning rate = 1e-3, weight decay = 1e-5.
- Loss Function: Cross-entropy loss.
- Early Stopping: Patience of 3 epochs.
- Device: CPU-only to simulate low-resource deployment.

IV. EXPERIMENTAL SETUP

- Framework: PyTorch 2.0, torchvision models with pretrained weights.
- Batch Size: 16.
- Epochs: Up to 8 (with early stopping).
- Evaluation Metrics: AUC, accuracy, F1-score, precision, recall.
- Grad-CAM: Implemented for qualitative interpretability (failed for some CPU runs due to gradient hook issues).

V. RESULTS

Table I presents the test set metrics for all models and budgets.

Table I — Test Performance Comparison

Model	Budget	AUC	ACC	F1	Precision	Recall
MobileNetV2	25	0.8563	0.7548	0.8259	0.7423	0.9308
MobileNetV2	50	**0.8987**	0.7853	**0.8463**	0.7656	**0.9461**
MobileNetV2	100	0.8766	0.7949	0.8498	0.7835	0.9282
ResNet50	25	0.8102	0.6603	0.6558	**0.8938**	0.5179
ResNet50	50	0.8407	0.7724	0.8156	0.8263	0.8051
ResNet50	100	0.8741	**0.8141**	0.8596	0.8142	0.9103
DenseNet121	25	0.7727	0.7372	0.8080	0.7435	0.8846
DenseNet121	50	0.8548	0.7837	0.8244	**0.8364**	0.8128
DenseNet121	100	0.8903	0.7901	0.8461	0.7809	0.9231

Fig. 1. Confusion matrix for DenseNet121 model trained with 100 samples per class, illustrating true positive, false positive, true negative, and false negative classification counts.

VI. DISCUSSION

The results show:

- MobileNetV2 consistently performed well with small budgets, especially at 50 images per class, achieving the highest AUC and recall.
- ResNet50 excelled in precision but required more data to generalize effectively.
- DenseNet121 showed competitive F1-scores for larger budgets but underperformed in extreme low-data cases (25 images/class).
- The strong recall for MobileNetV2 indicates suitability for medical triage tasks where missing positive cases is costly.
- Grad-CAM generation failed in CPU-only mode for some runs, indicating a limitation in visual interpretability in low-resource environments.

VII. CONCLUSION

This study demonstrates that transfer learning with lightweight architectures like MobileNetV2 can deliver high diagnostic accuracy in low-data medical imaging tasks. Under constraints of limited compute and small annotated datasets, MobileNetV2 offers an optimal trade-off between performance and efficiency.

Future work will explore:

- Fine-tuning backbone layers for further gains.
- Semi-supervised and synthetic data augmentation strategies.
- Improving interpretability methods for CPU-only deployments.

THE STANFORD JOURNAL ISSN: 3048-8923
(AN INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH)
Volume–2, Issues– 3 July- September: 2025

VIII. REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, Dec. 2017.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241.

[4] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.

[5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[7] P. Mooney, "Chest X-Ray Images (Pneumonia)," Kaggle, 2018. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. [Accessed: Aug. 11, 2025].