

Exploring Transfer Learning for Pneumonia Classification in Chest X-rays

A comparison of fine-tuning strategies and data augmentation strengths

Dr. Ackley Will - Department of Computing

This work was supported by a research grant from the Office of Research and Creative Scholarship at Andrews University.

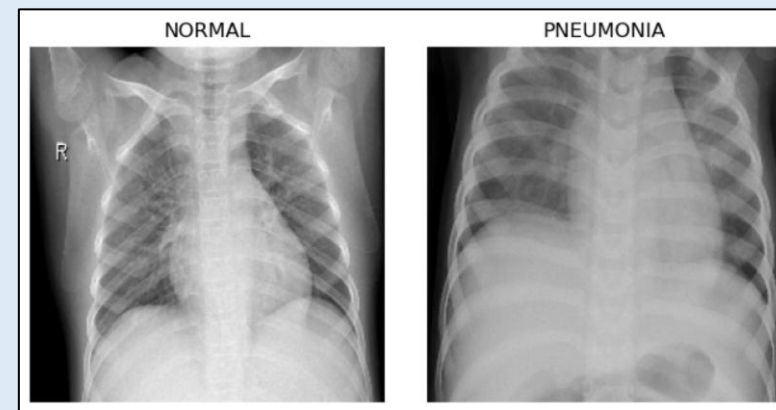
INTRODUCTION & METHODS

Background & Motivation

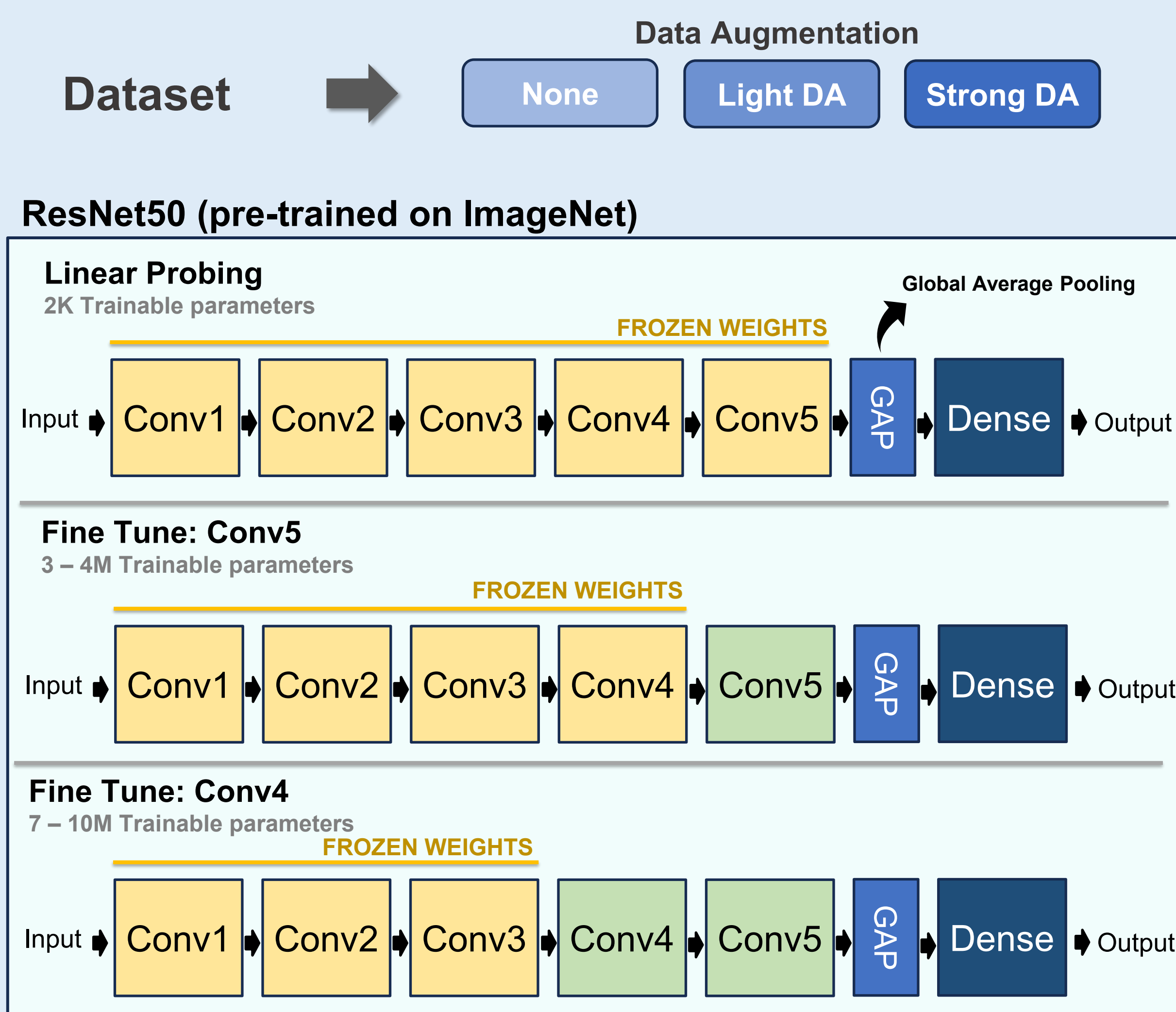
Transfer learning has become an essential approach in medical image analysis, especially when annotated data are limited. This study examines the impact of varying fine-tuning depths and data augmentation strengths on the performance of deep convolutional models for pneumonia classification in chest X-rays. A pre-trained ResNet50 backbone was adapted under three regimes: frozen (linear probing), fine-tuning of the last block, and fine-tuning of the last two blocks, using a Labeled Chest X-ray Images dataset. Fine-tuning consistently improved results, while stronger augmentation enhanced robustness and validation stability. These findings confirm that transfer learning, combined with balanced fine-tuning and augmentation, is a data-efficient approach for medical image classification.

Dataset

- Labeled Chest X-ray Images
- Two classes: Normal and Pneumonia
- Training: ~5,200 images | Test: 624 images
- Split: 80% train / 20% validation
- Images resized to 224 × 224 pixels

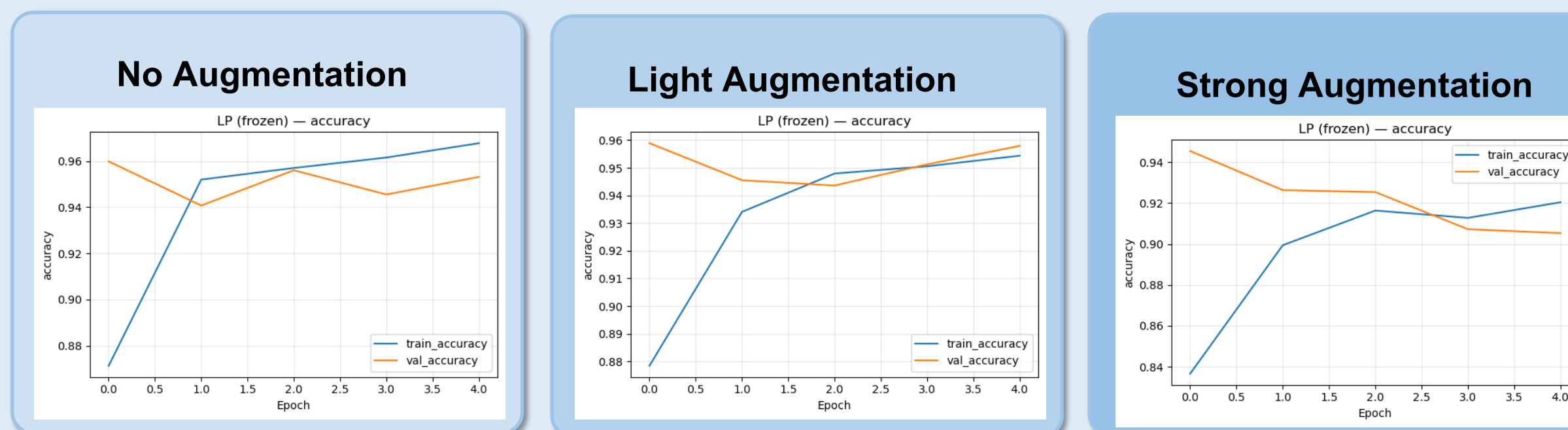


Experimental Design



RESULTS

Training and Validation Curves



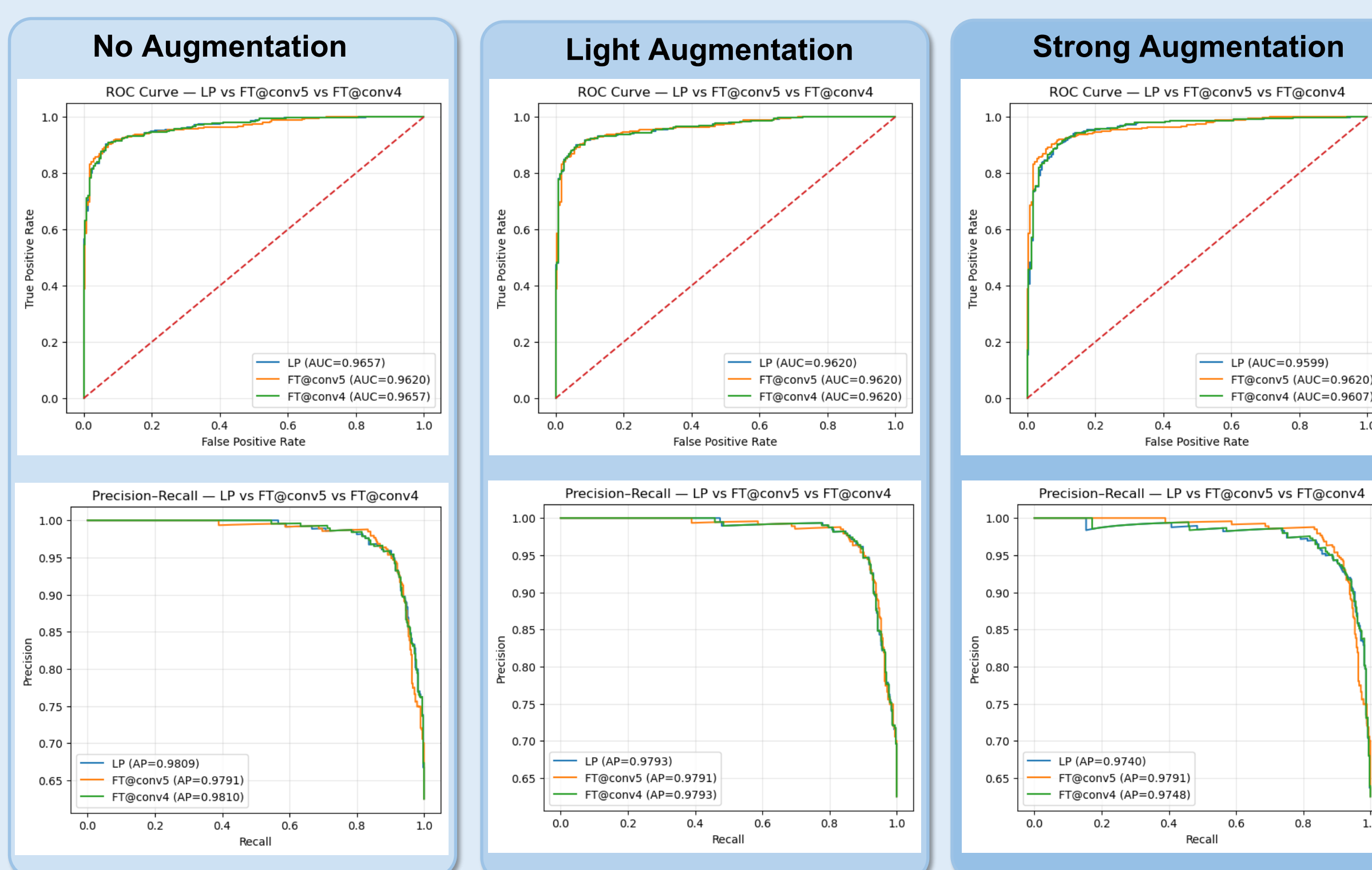
Fine-tuning yields faster and more stable convergence; stronger DA slightly delays convergence but improves generalization

Test Performance Summary

Augmentation	Model	Test Acc	AUC	AP (PR)
None	FT Conv4+5	0.8686	0.9658	0.9810
Light	FT Conv4+5	0.8510	0.9623	0.9793
Strong	FT Conv4+5	0.8910	0.9607	0.9748

Fine-tuning consistently improves accuracy. Strong DA leads to the highest test accuracy, while AUC remains stable across regimes.

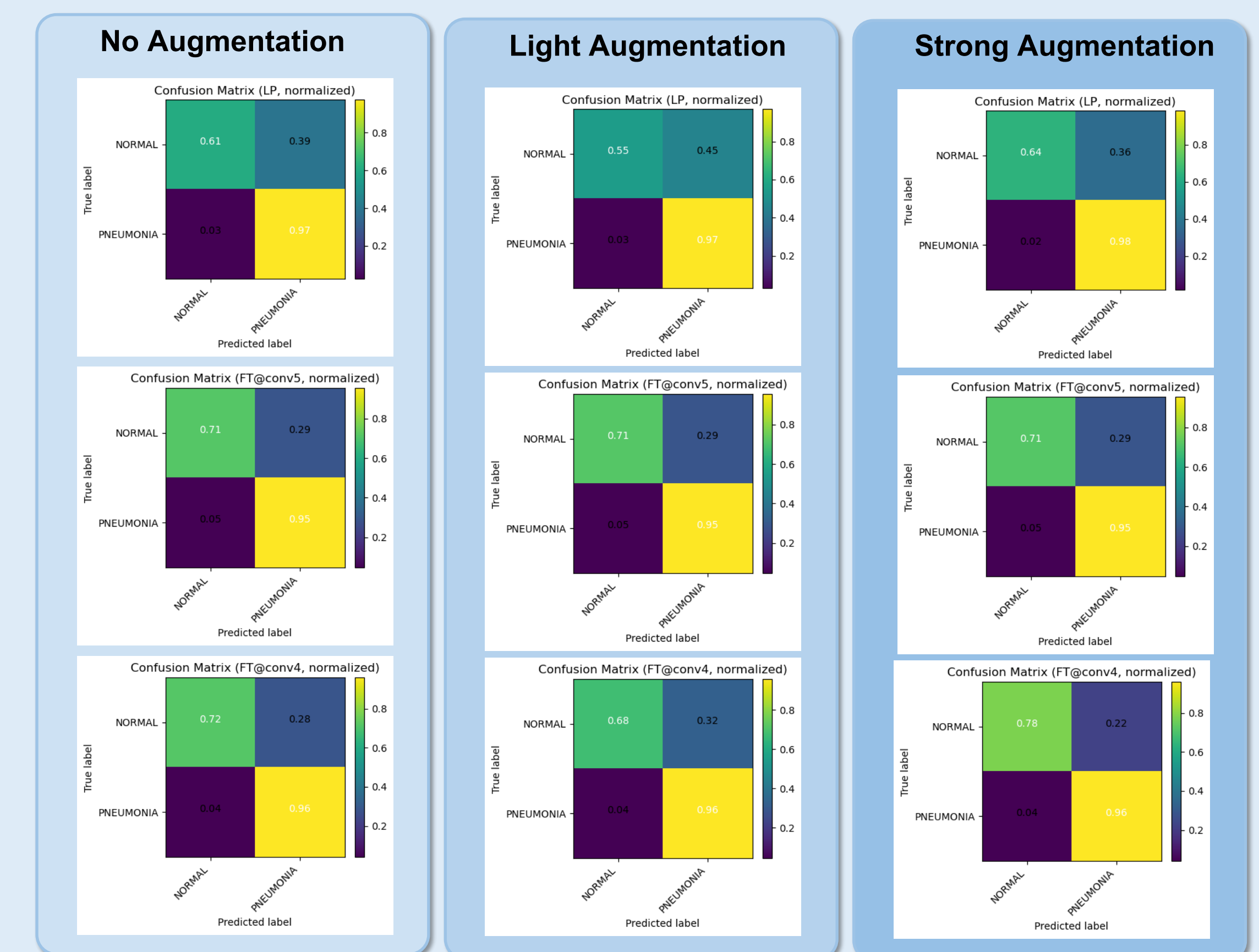
ROC and Precision–Recall Curves



All models show an AUC of approximately 0.96. precision-recall trade-offs remain consistent across augmentation strengths.

ANALYSIS & CONCLUSIONS

Confusion Matrices (Normalized)



Key Findings

- Fine-tuning improves performance across all conditions.
- Strong data augmentation increases robustness and accuracy (+2–3%).
- Light augmentation provides minimal gains.
- AUC stays consistently high (~0.96), confirming generalization.

References

- Kermany, Daniel; Zhang, Kang; Goldbaum. Michel (2018). Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images [Dataset]. Mendeley. <https://doi.org/10.17632/RSCBJBR9SJ.3>
- Sadia Showkat, Shaima Qureshi, Efficacy of Transfer Learning-based ResNet models in Chest X-ray image classification for detecting COVID-19 Pneumonia, Chemometrics and Intelligent Laboratory Systems, Volume 224, 2022, 104534, ISSN 0169-7439, <https://doi.org/10.1016/j.chemolab.2022.104534>.
- Georgiana Ingrid Stoleru, Adrian Iftene, Transfer Learning for Alzheimer's Disease Diagnosis from MRI Slices: A Comparative Study of Deep Learning Models, Procedia Computer Science, Volume 225, 2023, Pages 2614-2623, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2023.10.253>.