# Sharpness Matters. Rethinking the Impact of Image Resolution on Medical Image Classification

**Aditya Vikas Kulkarni, MSc,** Software Developer 2, Radiology, St. Jude Children's Research Hospital
Paul Yi, MD

## Introduction

Deep learning (DL) models commonly downsample medical images to reduce computational cost, but subtle pathologies may require higher-resolution inputs. Although previous work explored resolution effects in chest x-ray (CXR) DL classification, it did not evaluate explainability or out-of-distribution (OOD) generalizability with external datasets – key features of safe and trustworthy AI. We thus ask: Does increasing image resolution for training DL CXR classifiers improve not only in-distribution accuracy, but also explainability and OOD generalizability?

## Hypothesis

Higher-resolution training improves OOD generalizability and explainability.

## Methods

We trained DenseNet-121 classifiers on the SIIM-ACR Pneumothorax dataset (n=10,675 ) and set aside 10% for holdout testing. Six resolutions (64×64 to 1024×1024) were used for model training, each tuned via an independent hyperparameter search. The final models were validated through 5-fold cross-validation and tested on held-out and external OOD (n=726; 226 pneumothoraces) test-sets . For assessing explainability, Grad-CAM saliency maps, thresholded to binary masks were compared to radiologist-performed segmentations; mean IoU and percentage overlap (%-IN) were used to measure localization. AUROCs and saliency localization were compared between models using Delong's and Wilcoxon Signed-Rank tests, respectively.

## Results

As image resolution increased, both internal and external performance improved (Fig.1A), ranging from AUROC 0.90 and 0.68 for 224x224 (internal and external, respectively) to 0.97 and 0.84 for 1024x1024 ($p < .001$, all). Generalizability was best for higher resolutions ( < 0.08 AUROC drop for 768x768 and 1024x1024 vs. ~0.2 drop for 64x64 and 128x128) [Fig.1B]. Localization quality followed similar trends, with significantly higher mIoU and %-IN for higher image resolutions in internal and external test sets (Fig. 1C-D) that were qualitatively more reliable (Fig.2).

# Conclusion

Training DL classifiers with higher resolutions enhances OOD generalizability and explainability, alongside in-distribution accuracy. While radiology DL often uses lower resolutions (e.g., 224×224), our findings support adopting higher resolutions to maximize accuracy, trust, and safety, thus advancing more reliable clinical AI deployment.

# Figure(s)

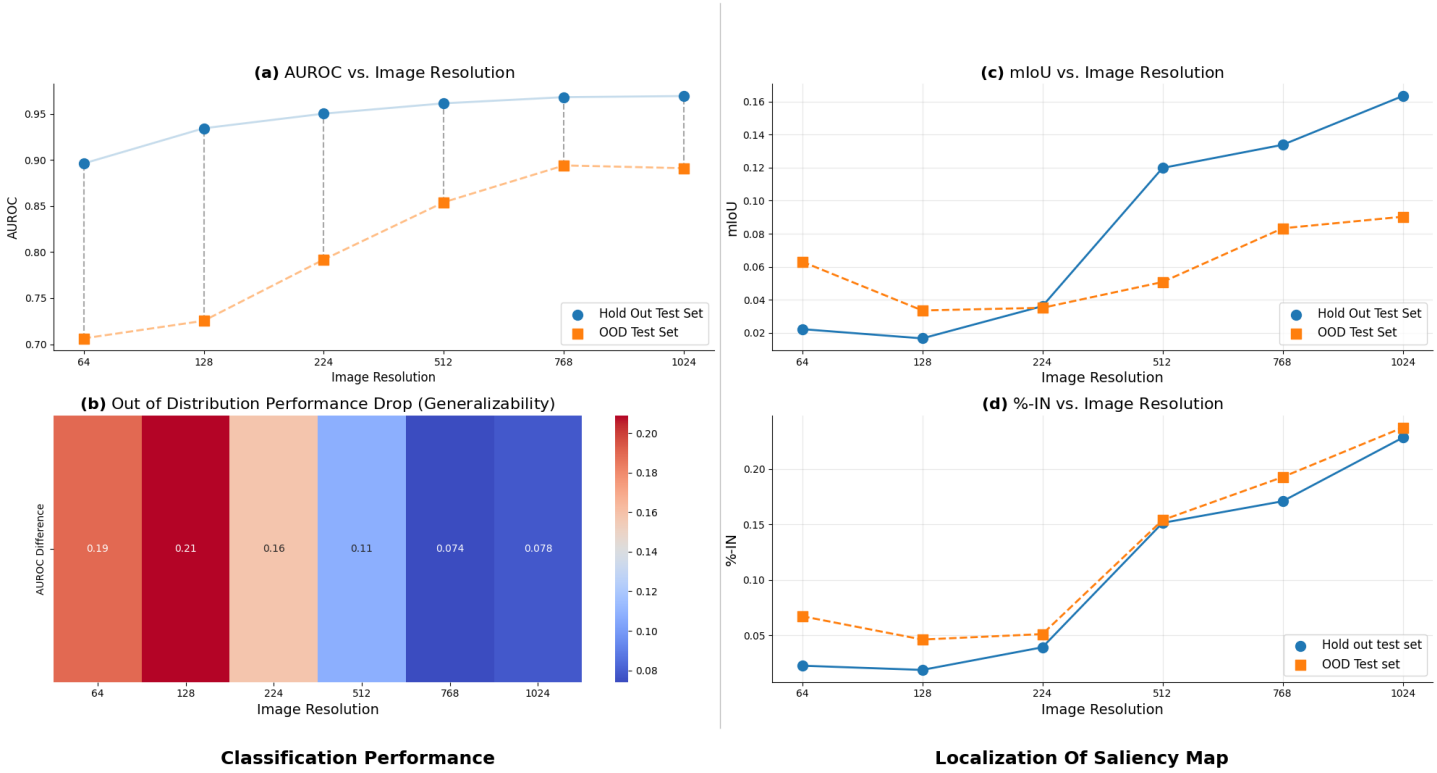


**Figure 1.** Results Overview (a) AUROC results on the hold-out and out-of-distribution (OOD) test sets, demonstrating the classification performance across resolutions. (b) A heatmap illustrating changes in disparity between in- and out-of-distribution performance, showing improved OOD robustness at higher resolutions. (c) Mean Intersection-over-Union (mIoU) across resolutions, indicating that higher resolutions yield better localization. (d) Percentage of the saliency map within the ground truth mask (%-IN), further confirming improved localization with increased resolution.

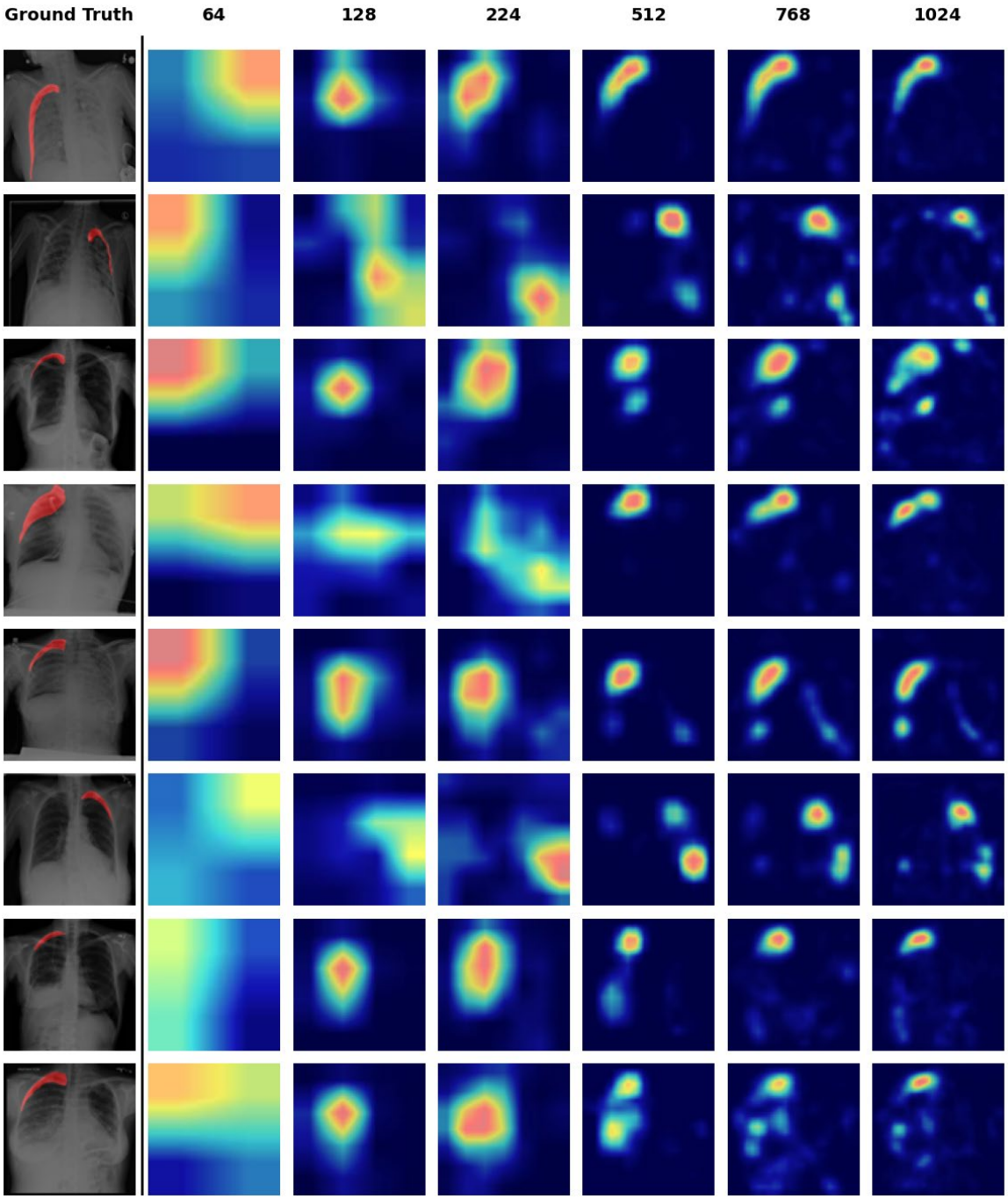


**Figure 2.** Training at higher resolution leads to more localized saliency maps. This figure illustrates saliency maps generated by models trained and evaluated at various resolutions, red colors indicate the most salient regions. Notice that increasing image resolutions result in qualitatively more reliable heatmaps that better correspond to the groundtruth masks.


## Keywords

Artificial Intelligence/Machine Learning; Emerging Technologies; Imaging Research; Quality Improvement & Quality Assurance; Security; Storage