Effect of Data Augmentation on Classification Performance of Pleural Mesothelioma Histological Subtype on CT Scans

Arina Oberoi[1], Christopher L. Valdes, MSc[1], Feng Li, MD, PhD[1], Owen Mitchell[2], Christopher Straus, MD[1], Hedy L. Kindler, MD[2] and Samuel G. Armato III, PhD[1]

1. Department of Radiology, The University of Chicago 2. Department of Medicine, The University of Chicago

2026 JULY 19–22 VANCOUVER, BC JOINT AAPM | COMP MEETING

THE UNIVERSITY OF CHICAGO Committee on Medical Physics

Abstract

Purpose: Accurate histological subtype classification of pleural mesothelioma (PM) from CT imaging remains challenging due to limited labeled data and substantial morphological variability. Data augmentation is a common strategy used to avoid overfitting when training deep learning (DL) models. This study evaluates the effect of individual image augmentation strategies on classification performance and generator reconstruction consistency in a conditional generative adversarial network (cGAN)-based PM subtype classification pipeline.

Methods: A cGAN was trained on 21 CT scans of epithelioid PM patients. Rotations (±5°, ±10°), Gaussian noise, and brightness adjustments were applied independently during training. Performance was evaluated via receiver operating characteristic (ROC) analysis on 7 epithelioid and 7 sarcomatoid held-out scans, using an area under the curve (AUC)-per-epoch framework tracking max AUC, epoch-wise trajectories, and minimum p-values.

Results: Moderate geometric and intensity-based augmentations produced small but consistent improvements in classification performance relative to the baseline, which achieved a maximum AUC of 0.80. Rotational augmentations showed the greatest improvement, with rotations of −10° achieving an AUC of 0.86. Brightening augmentations brought an AUC of 0.80, while decreases in brightness increased AUC to 0.82. Gaussian noise was the only augmentation to result in reduced performance (AUC 0.78). Across augmentations, performance gains emerged gradually across epochs, indicating improved robustness rather than early overfitting.

Conclusion: Targeted, physically plausible augmentations modestly improve mesothelioma subtype classification performance in a cGAN-based framework, while noise-driven transformations reduce generalizability. These findings suggest the importance of focusing on the selection of specific augmentation strategies and that controlled geometric transformations are more beneficial than stochastic perturbations for CT-based tumor classification.

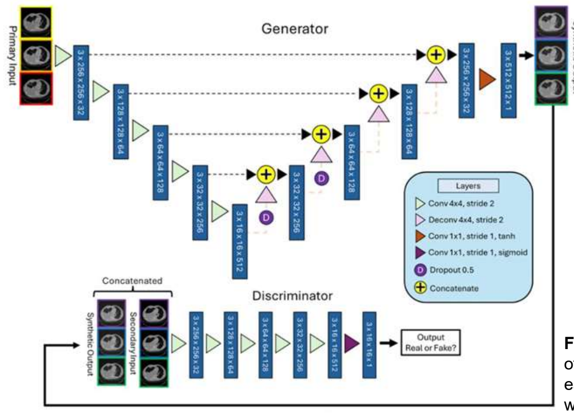
Background

- Histological subtype (epithelioid vs. sarcomatoid) drives prognosis, but CT-based classification is hard due to limited labeled data and morphological overlap.
- cGANs learn subtype-discriminative features from small datasets by combining a generator/discriminator with classification loss.
- Data augmentation combats overfitting, but which augmentation helps is rarely isolated (most pipelines stack many at once).
- This study decouples individual augmentations (rotation, noise, brightness) to test each one's isolated effect on cGAN classification performance.

Objective

- We aim to introduce a controlled framework for isolating the effects of individual data augmentations in a cGAN-based mesothelioma classification pipeline. These results would provide actionable guidance for designing augmentation strategies that can inform and improve clinical reliability in data-limited CT-based deep learning applications.

Methods



- Small subset of data: 21 epithelioid patients were used for training.
- Starting from baseline (no augmentations added during training), each augmentation was added to the training pipeline individually.
- After training, the model was tested to evaluate classification performance.

Figure 1. Network architecture of the cGAN. The shapes of each layer are depth, height, width, and channels.

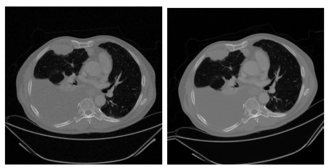


Figure 2. Sample images of augmentations. **Left:** Gaussian noise added to the image. **Right:** 10-degree counterclockwise rotation.

Generator

$L_G = L_{Adv} + \lambda_{L1} * L_{L1}$

L_{Adv} = adversarial loss

λ_{L1} = weighting factor

L_{L1} = L1 loss function

L2 Loss

$$L_2 = \sum_{i=1}^{P} (x - x')^2$$

x = real pixel value

x' = fake pixel value

Discriminator

$L_D = L_{Real} + L_{Syn}$

L_{Real} = real image loss

L_{Syn} = synthetic image loss

Figure 3. Loss functions used for testing (Generator & Discriminator) and testing (L2 Loss). Binary cross entropy (BCE) is used for adversarial learning.

Results

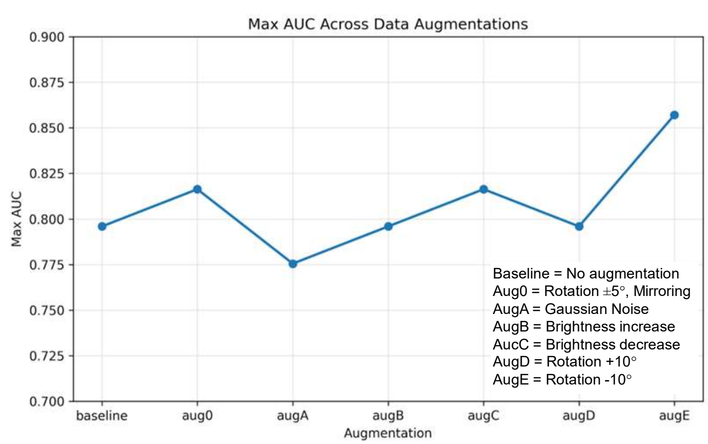


Figure 4. Maximum AUC achieved across data augmentation conditions.

Results (cont.)

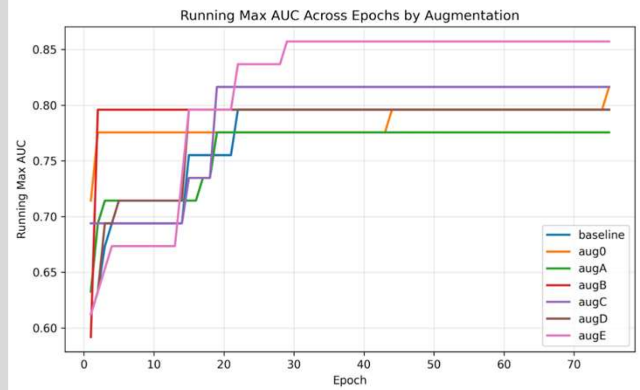


Figure 5. Running maximum AUC across training epochs by augmentation.

- Baseline achieved max AUC of 0.80.
- Rotation −10° (augE) produced the largest improvement (AUC = 0.86).
- Brightness decrease (augC) and small symmetric rotation (aug0) both improved performance to AUC = 0.82.
- Brightness increase (augB) and rotation +10° (augD) performed similarly to baseline (AUC ≈ 0.80), suggesting neutral effect.
- Gaussian noise (augA) was the only augmentation that reduced performance below baseline (AUC = 0.78), indicating stochastic perturbations may hinder rather than help generalization in this pipeline.

Summary and Discussion

- Directionality mattered for rotation, since −10° outperformed +10° by a notable margin (0.86 vs. 0.80), suggesting the model may be sensitive to orientation-specific features rather than rotation magnitude alone.
- Figure 5 shows performance gains emerged gradually across epochs for augE and augC, rather than spiking early and declining, suggesting genuine improvements in generalizability rather than an overfitting artifact.
- AugB and augD plateaued early and stayed flat, indicating these augmentations didn't meaningfully change what the model learned relative to baseline.
- Overall, geometric and intensity-based augmentations modestly but consistently outperformed the noise-based augmentation, supporting the idea that augmentation type (not just presence) drives performance gains.

Future Directions

- Expand dataset and test generalizability beyond 21 training scans.
- Combinatorial augmentation testing: test paired combinations to see whether augmentations compound or interfere with each other.
- Investigate rotation directionality: the asymmetric benefit of −10° vs. +10° suggests a bias worth characterizing to better inform anatomy-aware augmentation design.

Acknowledgements

This study was funded by the Mesothelioma Applied Research Foundation. CT scans were provided and anonymized by the Human Imaging Research Office at the University of Chicago.