ACCSCIENCE PUBLISHING

Eurasian Journal of Medicine and Oncology

ORIGINAL RESEARCH ARTICLE

Enhancing rare tumor detection: A cross-modal generative adversarial network benchmark for data augmentation in cardiac, liver, and retinal imaging

Supplementary Files

Table S1. DCGAN architectural specifications

Component	Specification
Generator	
Input	Latent vector z (100-dim)
Layer 1	Dense + Reshape → 4 × 4 × 512
Layer 2	Transposed Conv (4 × 4, stride = 2) → 8×8×256 + BatchNorm + ReLU
Layer 3	Transposed Conv (4 × 4, stride = 2) → 16 × 16 128 + BatchNorm + ReLU
Layer 4	Transposed Conv (4 × 4, stride = 2) → 32 × 32 × 64 + BatchNorm + ReLU
Layer 5	Transposed Conv (4 × 4, stride = 2) → 64 × 64 × 32 + BatchNorm + ReLU
Layer 6	Transposed Conv (4 × 4, stride = 2) → 128 × 128 × 16 + BatchNorm + ReLU
Layer 7	Transposed Conv (4 × 4, stride = 2) → 256 × 256 × 1 + Tanh
Discriminator	
Layer 1	Conv (4 × 4, stride = 2) → 128 × 128 × 16 + LeakyReLU (0.2)
Layer 2	Conv (4 × 4, stride = 2) → 64 × 64 × 32 + BatchNorm + LeakyReLU (0.2)
Layer 3	Conv (4 × 4, stride = 2) → 32 × 32 × 64 + BatchNorm + LeakyReLU (0.2)
Layer 4	Conv (4 × 4, stride = 2) → 16 × 16 × 128 + BatchNorm + LeakyReLU (0.2)
Layer 5	Conv (4 × 4, stride = 2) → 8 × 8 × 256 + BatchNorm + LeakyReLU (0.2)
Layer 6	Conv (4 × 4, stride = 2) → 4 × 4 × 512 + BatchNorm + LeakyReLU (0.2)
Layer 7	Flatten + Dense → 1 (Sigmoid)

Table S2. WGAN-GP architectural specifications

Component	Specification
Generator	
Input	Latent vector z (100-dim)
Layer 1	Dense + Reshape → 4 × 4 × 512
Layer 2	Transposed Conv (4 × 4, stride = 2) → 8 × 8 × 256 + LayerNorm + ReLU
Layer 3	Transposed Conv (4 × 4, stride = 2) → 16 × 16 × 128 + LayerNorm + ReLU
Layer 4	Transposed Conv (4 × 4, stride = 2) → 32 × 32 × 64 + LayerNorm + ReLU
Layer 5	Transposed Conv (4 × 4, stride = 2) → 64 × 64 × 32 + LayerNorm + ReLU
Layer 6	Transposed Conv (4 × 4, stride = 2) → 128 × 128 × 16 + LayerNorm + ReLU
Layer 7	Transposed Conv (4 × 4, stride = 2) → 256 × 256 × 1 + Tanh
Discriminator (Critic)	
Layer 1	Conv (4 × 4, stride = 2) → 128 × 128 × 16 + LayerNorm + LeakyReLU (0.2)
Layer 2	Conv (4 × 4, stride = 2) → 64 × 64 × 32 + LayerNorm + LeakyReLU (0.2)
Layer 3	Conv (4 × 4, stride = 2) → 32 × 32 × 64 + LayerNorm + LeakyReLU (0.2)
Layer 4	Conv (4 × 4, stride = 2) → 16 × 16 × 128 + LayerNorm + LeakyReLU (0.2)
Layer 5	Conv (4 × 4, stride = 2) → 8 × 8 × 256 + LayerNorm + LeakyReLU (0.2)
Layer 6	Conv (4 × 4, stride = 2) → 4 × 4 × 512 + LayerNorm + LeakyReLU (0.2)
Layer 7	Flatten + Dense → 1 (No activation - Wasserstein distance)

Table S3. StyleGAN architectural specifications

Component	Specification
Mapping Network	
Input	Latent vector z (512-dim)
Layer 1	Dense (512) + LeakyReLU
Layer 2	Dense (512) + LeakyReLU
Layer 3	Dense (512) + LeakyReLU
Layer 4	Dense (512) + LeakyReLU
Layer 5	Dense (512) + LeakyReLU
Layer 6	Dense (512) + LeakyReLU
Layer 7	Dense (512) + LeakyReLU
Layer 8	Dense (512) + LeakyReLU → w (512-dim style vector)

(*Cont'd...*)

Table S3. (*Continued*)

Component	Specification
Synthesis Network	
Input	Constant 4 × 4 × 512
Block 1 (4 × 4)	AdaIN + Conv 3 × 3 + AdaIN + Noise + LeakyReLU
Block 2 (8 × 8)	Upsample + Conv 3 × 3 + AdaIN + Noise + LeakyReLU
Block 3 (16 × 16)	Upsample + Conv 3 × 3 + AdaIN + Noise + LeakyReLU
Block 4 (32 × 32)	Upsample + Conv 3 × 3 + AdaIN + Noise + LeakyReLU
Block 5 (64 × 64)	Upsample + Conv 3 × 3 + AdaIN + Noise + LeakyReLU
Block 6 (128 × 128)	Upsample + Conv 3 × 3 + AdaIN + Noise + LeakyReLU
Block 7 (256 × 256)	Upsample + Conv 3 × 3 + AdaIN + Noise + LeakyReLU
Output	Conv 1 × 1 → RGB (256 × 256 × 1)
Discriminator	
Layer 1	Conv (1 × 1) → 128 × 128 × 16 + LeakyReLU
Layer 2	Conv (3 × 3) + Downsample → 64 × 64 × 32 + LeakyReLU
Layer 3	Conv (3 × 3) + Downsample → 32 × 32 × 64 + LeakyReLU
Layer 4	Conv (3 × 3) + Downsample → 16 × 16 × 128 + LeakyReLU
Layer 5	Conv (3 × 3) + Downsample → 8 × 8 × 256 + LeakyReLU
Layer 6	Conv (3 × 3) + Downsample → 4 × 4 × 512 + LeakyReLU
Layer 7	Minibatch Std + Conv 3 × 3 + LeakyReLU
Layer 8	Flatten + Dense → 1

Table S4. FID computation parameters and technical specifications

Parameter	Value	Description
Feature extraction model	Inception-v3	Pre-trained on ImageNet
Feature layer	pool_3 (8 × 8 × 2,048)	Final average pooling layer
Feature dimension	2,048	Dimension of feature vectors
Image preprocessing	Resize to 299 × 299	Bicubic interpolation
Normalization	[−1, 1] to [0, 1]	Scale adjustment for Inception
Number of real images	10,000	Randomly sampled from the test set
Number of fake images	10,000	Generated per model
Covariance regularization	$\varepsilon = 1e - 6$	Added to the diagonal for numerical stability
Batch size for extraction	32	Processed in batches
Random seed	42	For reproducibility

Table S5. SSIM component weighting factors and parameters

Parameter	Symbol	Value	Clinical significance
Luminance weight	α	1.0	Tissue density representation
Contrast weight	β	1.0	Pathological discrimination
Structure weight	γ	1.0	Anatomical boundary preservation
Luminance constant	C1	$(0.01 \times L)^2$	Stabilization (L = 255 for 8-bit)
Contrast-structure constant	C2	$(0.03 \times L)^2$	Stabilization
Dynamic range	L	255	For 8-bit images
Gaussian kernel	σ	1.5	Standard deviation for filtering
Kernel size	k	11 × 11	Window size for SSIM calculation
Structure-specific weights			
Bone edges	γ_bone	1.2	Higher weight for edge preservation
Soft tissue	γ_soft	1.0	Standard weight
Lung patterns	γ_lung	1.5	Highest weight for fine details

Table S6. Complete hyperparameter configurations

Hyperparameter	DCGAN	WGAN-GP	StyleGAN
Training parameters			
Epochs	500	500	500
Batch size	64	64	32
Generator learning rate	0.0002	0.0001	0.001
Discriminator learning rate	0.0002	0.0001	0.001
Optimizer	Adam	Adam	Adam
Adam β1	0.5	0.5	0.0
Adam β2	0.999	0.999	0.99
Gradient penalty weight (λ)	N/A	10	N/A
N_critic (D iterations per G step)	1	5	1
Learning rate schedules	-	-	(See Table S7)
Regularization			
Weight decay	0	0	0
Dropout (generator)	0	0	0
Dropout (discriminator)	0.25	0	0
Gradient clipping	N/A	N/A	N/A

(*Cont'd...*)

Table S6. (*Continued*)

Hyperparameter	DCGAN	WGAN-GP	StyleGAN
Loss functions			
Generator loss	Binary cross-entropy	−E[D(G(z))]	Adversarial + LPIPS
Discriminator loss	Binary cross-entropy	E[D(G(z))] − E[D(x)] + GP	Adversarial
Progressive training			
Starting resolution	256 × 256	256 × 256	4 × 4
Resolution progression	N/A	N/A	4 → 8 → 16 → 32 → 64 → 128 → 256
Fade-in duration	N/A	N/A	50,000 images per resolution

Table S7. Learning rate (LR) schedules by training phase

Phase	Epochs	DCGAN LR	WGAN-GP LR	StyleGAN LR
Warm-up	1–20	0.0001	0.00005	0.0005
Main training	21–300	0.0002	0.0001	0.001
Fine-tuning 1	301–400	0.0001	0.00005	0.0005
Fine-tuning 2	401–450	0.00005	0.00002	0.0002
Final	451–500	0.00002	0.00001	0.0001

Note: Early stopping criteria:

- Monitor: FID on validation set
- Patience: 25 epochs
- Minimum delta: 0.5
- Restore best weights: Yes

Table S8. GDPR compliance checklist

GDPR requirement	Implementation	Status	Verification
Lawfulness, fairness, transparency			
Legal basis for processing	Public dataset consent	✓ Complete	Original dataset documentation
Transparency about data use	Published methodology	✓ Complete	Ethics section in the paper
Purpose limitation			
Specified purposes only	Research use only	✓ Complete	Data usage agreement
No further processing	No secondary use	✓ Complete	Access controls

(*Cont'd...*)

Table S8. (*Continued*)

GDPR requirement	Implementation	Status	Verification
Data minimization			
Adequate, relevant, limited	Only necessary metadata retained	✓ Complete	Metadata audit
Accuracy			
Keep data accurate	Verified against the source	✓ Complete	Regular validation
Storage limitation			
Not kept longer than necessary	Deleted after project completion	✓ Complete	Deletion certification
Integrity and confidentiality			
Security measures	AES-256 encryption	✓ Complete	Encryption at rest
Access controls	Role-based access	✓ Complete	Access logs
Transmission security	TLS 1.3	✓ Complete	Network security
Data subject rights			
Right to access	N/A - fully anonymized	N/A	No identifiable data
Right to erasure	N/A - fully anonymized	N/A	No identifiable data
Accountability			
Data protection officer	Appointed	✓ Complete	Contact information
Records of processing	Maintained	✓ Complete	Processing register
DPIA	Completed	✓ Complete	Approved by the ethics committee

Table S9. Differential privacy implementation details

Parameter	Symbol	Value	Description
Privacy budget			
Epsilon	ε	0.1	Total privacy loss parameter
Delta	δ	10^{-5}	Probability of privacy breach
Noise mechanism			
Mechanism type	-	Gaussian	Applied to gradients
Noise multiplier	σ	1.1	σ × sensitivity
Sensitivity	S	1.0	L2 norm clipping bound
DP-SGD parameters			
Clipping norm	C	1.0	Gradient norm bound
Sampling rate	q	0.01	Poisson sampling per iteration
Batch size	L	64	Effective batch size
Iterations	T	50,000	Total training iterations

(*Cont'd...*)

Table S9. (*Continued*)

Parameter	Symbol	Value	Description
Privacy accounting			
Accounting method	-	Rényi DP	Composition calculation
Moments accountant	-	Implemented	Track ε accumulation
Subsampling amplification	-	Yes	Amplifies privacy
Expected privacy guarantees			
Final ε (95% CI)	-	0.1 ± 0.02	After 50k iterations
Final δ	-	10^{-5}	Fixed parameter
Implementation details			
Library	-	TensorFlow privacy	v0.8.0
Gradient processing	-	Per-example gradients	Computed then clipped
Noise addition	-	Before averaging	Per gradient update

Table S10. Complete statistical analysis (ANOVA results)

Source of Variation	Sum of squares (SS)	df	Mean square (MS)	*F*	*p*-value
FID scores					
Between models	1254.3	2	627.15	156.2	<0.001
Within models	108.4	27	4.01		
Total	1,362.7	29			
Overall SSIM					
Between models	0.084	2	0.042	142.8	<0.001
Within models	0.008	27	0.0003		
Total	0.092	29			
Bone edge SSIM					
Between models	0.091	2	0.0455	168.5	<0.001
Within models	0.0073	27	0.00027		
Total	0.0983	29			
Soft tissue SSIM					
Between models	0.076	2	0.038	135.7	<0.001
Within models	0.0076	27	0.00028		
Total	0.0836	29			
Lung pattern SSIM					
Between models	0.102	2	0.051	182.1	<0.001
Within models	0.0076	27	0.00028		
Total	0.1096	29			

(*Cont'd...*)

Table S9. (*Continued*)

Source of Variation	Sum of squares (SS)	df	Mean square (MS)	*F*	*p*-value
Sensitivity gain					
Between models	156.8	2	78.4	145.2	<0.001
Within models	14.6	27	0.54		
Total	171.4	29			

Table S11. Post-hoc tests (Tukey HSD) - pairwise comparisons

Comparison	Mean difference	95% CI	*p*-value	Significance
FID scores				
StyleGAN vs. DCGAN	−27.4	[−29.8, −25.0]	<0.001	***
StyleGAN vs. WGAN-GP	−5.2	[−7.6, −2.8]	0.008	**
WGAN-GP vs. DCGAN	−22.2	[−24.6, −19.8]	<0.001	***
Overall SSIM				
StyleGAN vs. DCGAN	0.14	[0.12, 0.16]	<0.001	***
StyleGAN vs. WGAN-GP	0.07	[0.05, 0.09]	0.008	**
WGAN-GP vs. DCGAN	0.07	[0.05, 0.09]	0.008	**
Bone edge SSIM				
StyleGAN vs. DCGAN	0.17	[0.15, 0.19]	<0.001	***
StyleGAN vs. WGAN-GP	0.09	[0.07, 0.11]	0.004	**
WGAN-GP vs. DCGAN	0.08	[0.06, 0.10]	0.006	**
Soft tissue SSIM				
StyleGAN vs. DCGAN	0.12	[0.10, 0.14]	<0.001	***
StyleGAN vs. WGAN-GP	0.06	[0.04, 0.08]	0.012	*
WGAN-GP vs. DCGAN	0.06	[0.04, 0.08]	0.012	*
Lung pattern SSIM				
StyleGAN vs. DCGAN	0.16	[0.14, 0.18]	<0.001	***
StyleGAN vs. WGAN-GP	0.08	[0.06, 0.10]	0.005	**
WGAN-GP vs. DCGAN	0.08	[0.06, 0.10]	0.005	**
Sensitivity gain				
StyleGAN vs. DCGAN	7.7	[6.5, 8.9]	<0.001	***
StyleGAN vs. WGAN-GP	3.6	[2.4, 4.8]	0.008	**
WGAN-GP vs. DCGAN	4.1	[2.9, 5.3]	0.004	**

Notes: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table S12. Effect size calculations (Cohen's d)

Comparison	Bone edges	Soft tissue	Lung patterns	Overall	Interpretation
StyleGAN vs. DCGAN					
Cohen's d	2.3	1.9	2.5	2.2	Very large
95% CI	[1.8, 2.8]	[1.5, 2.3]	[2.0, 3.0]	[1.7, 2.7]	
StyleGAN vs. WGAN-GP					
Cohen's d	1.4	1.1	1.6	1.4	Large
95% CI	[1.0, 1.8]	[0.7, 1.5]	[1.2, 2.0]	[1.0, 1.8]	
WGAN-GP vs. DCGAN					
Cohen's d	0.9	0.8	1.1	0.9	Large
95% CI	[0.5, 1.3]	[0.4, 1.2]	[0.7, 1.5]	[0.5, 1.3]	

Note: Effect size interpretation (Cohen's d):

- d < 0.2: Negligible
- d = 0.2 – 0.5: Small
- d = 0.5 – 0.8: Medium
- d = 0.8 – 1.2: Large
- d > 1.2: Very large

Table S13. Complete anonymized radiologist scoring sheets

Radiologist ID	Experience (years)	Specialty	DCGAN score	WGAN-GP score	StyleGAN score
R01	15	Thoracic radiology	2.8	4.2	4.8
R02	22	Cardiothoracic	2.7	4.0	4.7
R03	18	Interventional radiology	2.9	4.1	4.6
R04	12	Diagnostic radiology	2.8	4.3	4.8
R05	25	Thoracic imaging	2.6	3.9	4.6
Mean	18.4		2.76	4.10	4.70
Standard Deviation	5.2		0.11	0.16	0.10
Median	18		2.8	4.1	4.7
Min.	12		2.6	3.9	4.6
Max.	25		2.9	4.3	4.8

Note: Scoring scale:

- 1: Unrealistic, not usable for diagnosis
- 2: Poor quality, major anatomical distortions
- 3: Acceptable, minor artifacts present
- 4: Good quality, nearly realistic
- 5: Indistinguishable from real X-rays

Table S14. Inter-rater reliability analysis (Fleiss' κ)

Model	Fleiss' κ	95% CI	SE	*z*	*p*-value	Interpretation
DCGAN	0.78	[0.72, 0.84]	0.03	26.0	<0.001	Substantial agreement
WGAN-GP	0.81	[0.75, 0.87]	0.03	27.0	<0.001	Almost perfect agreement
StyleGAN	0.85	[0.79, 0.91]	0.03	28.3	<0.001	Almost perfect agreement

Note: Fleiss' κ interpretation (Landis & Koch):

- <0.00: Poor agreement
- 0.00–0.20: Slight agreement
- 0.21–0.40: Fair agreement
- 0.41–0.60: Moderate agreement
- 0.61–0.80: Substantial agreement
- 0.81–1.00: Almost perfect agreement

Table S15. Qualitative comments from radiologists (verbatim)

Radiologist	DCGAN comments	WGAN-GP comments	StyleGAN comments
R01	"Images appear blurry, especially in the peripheral lung fields. Rib margins are poorly defined. Would not use for diagnostic training."	"Good overall anatomical preservation. Some texture inconsistencies in the mediastinum but clinically usable with caution."	"Excellent image quality. Lung markings are crisp and clear. Indistinguishable from real chest X-rays. Perfect for training."
R02	"Anatomical distortions visible in the cardiac silhouette. Lung texture appears artificial and smoothed over."	"Preserves major structures well. Small vessels in the periphery show some artifact but acceptable for augmentation."	"Exceptional detail throughout. Even subtle ground-glass opacities are well represented. Gold standard quality."
R03	"Ribcage architecture is poorly defined with irregular spacing. Contrast too low for clinical assessment."	"Minor artifacts visible in the retrocardiac region. Overall spatial relationships preserved well."	"Perfect preservation of small anatomical details. Would use these images confidently in clinical training."
R04	"Contrast too low, difficult to assess mediastinal contours. Lung markings appear smeared."	"Clinically usable with caution. Some loss of fine detail in the lung bases but overall acceptable."	"Excellent anatomical fidelity. Heart size and contour perfectly preserved. Superior to other GAN outputs."
R05	"Not suitable for diagnostic training. Multiple artifacts and unrealistic texture patterns throughout."	"Good balance of quality and computational efficiency. Acceptable for research purposes but needs improvement for clinical use."	"Outstanding quality. Could easily be mistaken for real patient X-rays. The clear choice for clinical augmentation."

Table S16. GPU memory utilization (detailed benchmarks)

Hardware configuration	Model	Training memory (GB)	Inference memory (GB)	Peak memory (GB)	Batch size used
NVIDIA RTX 3060 (12GB VRAM)	DCGAN	4.2	1.8	4.5	64
	WGAN-GP	6.8	2.3	7.1	64
	StyleGAN	11.2	4.1	11.8	32
NVIDIA A100 (40GB VRAM)	DCGAN	3.8	1.5	4.1	128
	WGAN-GP	6.2	2.0	6.5	128
	StyleGAN	10.5	3.8	11.0	64
NVIDIA T4 (16GB VRAM)	DCGAN	4.0	1.7	4.3	64
	WGAN-GP	6.5	2.2	6.8	64
	StyleGAN	10.8	4.0	11.3	32
NVIDIA V100 (32GB VRAM)	DCGAN	3.9	1.6	4.2	96
	WGAN-GP	6.3	2.1	6.6	96
	StyleGAN	10.6	3.9	11.1	48

Table S17. Inference time percentiles and power consumption

Model	Mean (ms)	P50 (ms)	P90 (ms)	P95 (ms)	P99 (ms)	Standard deviation (ms)	Power (W)	Energy per image (mJ)
NVIDIA RTX 3060								
DCGAN	15.2	15.0	15.8	16.1	17.2	0.8	85	1.29
WGAN-GP	22.4	22.1	23.2	23.8	25.3	1.2	120	2.69
StyleGAN	45.6	45.2	46.5	47.1	49.8	1.8	165	7.52
NVIDIA A100								
DCGAN	8.3	8.2	8.6	8.8	9.5	0.4	250	2.08
WGAN-GP	12.1	12.0	12.5	12.8	13.6	0.6	300	3.63
StyleGAN	24.8	24.5	25.3	25.7	27.1	0.9	350	8.68
NVIDIA T4								
DCGAN	18.5	18.3	19.2	19.6	20.9	0.9	70	1.30
WGAN-GP	27.3	27.0	28.1	28.7	30.5	1.3	90	2.46
StyleGAN	55.2	54.8	56.5	57.3	60.2	2.1	120	6.62

Table S18. Diffusion-based models (preliminary results)

Model	FID	SSIM	Training time (h)	GPU memory (GB)	Notes
Denoising diffusion probabilistic models (DDPM)					
DDPM (T = 1,000)	21.4	0.88	72	16	Promising but slow
DDPM (T = 500)	22.8	0.87	48	14	Faster but with quality loss
Score-based models					
NCSN++	19.8	0.90	84	18	Competitive with StyleGAN
Latent diffusion					
LDM-4	20.2	0.89	36	12	Efficient, good quality
Comparison					
StyleGAN (our)	18.2	0.92	48	12	Best overall
DDPM (best)	21.4	0.88	72	16	Better diversity
Score-based	19.8	0.90	84	18	Good quality, slow

Table S19. Ablation study: StyleGAN component contributions

Configuration	FID	SSIM	Sensitivity gain (%)	False-negative rate (%)	Radiologist score
Full StyleGAN	18.2	0.92	9.8	12.0	4.7
w/o mapping network	24.6	0.86	5.2	22.5	3.8
w/o AdaIN	26.8	0.84	4.1	25.8	3.5
w/o noise injection	21.3	0.88	6.8	18.5	4.1
w/o style mixing	20.5	0.89	7.2	17.2	4.2
w/o progressive growing	23.1	0.87	5.8	21.3	3.9
w/o path length regularization	19.4	0.90	8.5	14.8	4.4
Mapping network only	35.2	0.75	2.1	32.5	2.5

Notes: Key observations:

- Mapping network is critical for style control (+6.4 FID improvement)
- AdaIN enables fine detail preservation (+8.6 FID improvement)
- Noise injection adds stochastic variation (+3.1 FID improvement)
- All components contribute to clinical utility

Table S20. Complete reference list with DOIs (State-of-the-art model comparison)

Prior Work	Year	Method	Dataset	FID	SSIM	Sensitivity (%)	DOI/Reference
Qin, 2026[1]	2023	StyleGAN2	NIH Chest X-ray	24.6	0.88	87.2	doi: 10.1038/s41598-025-30170-7
Du & Tian, 2024[2]	2024	WGAN-GP + Attention	ChestX-ray2017	28.3	0.85	85.5	doi: 10.26599/TST.2022.9010071
Makhlouf *et al.*, 2023[3]	2023	DCGAN	PadChest	52.3	0.75	78.3	doi: 10.1007/s00521-023-09100-z
Islam *et al.*, 2024[4]	2024	Ensemble GAN	CheXpert	32.1	0.82	83.6	doi: 10.1109/ACCESS.2024.3370848
Ktena *et al.*, 2024[5]	2024	Conditional GAN	MIMIC-CXR	29.8	0.83	84.2	doi: 10.1038/s41591-024-02897-3
Iqbal *et al.*, 2022[6]	2022	CycleGAN	Private	35.6	0.80	81.5	doi: 10.1007/s13735-022-00240-x
Kazeminia *et al.*, 2020[7]	2020	cGAN	Various	38.2	0.79	80.1	doi: 10.1016/j.artmed.2020.101938
Ali *et al.*, 2025[8]	2024	StyleGAN3	NIH Chest X-ray	22.8	0.89	88.9	doi: 10.1007/s11831-024-10123-5
Our Study	2026	StyleGAN	NIH + ACDC + SLiver07 + IDRiD	18.2	0.92	94.2	-

Supplementary equations

Equation S1. DCGAN loss - expanded mathematical form

Generator loss:

LG = −Ez ∼ pz(z)[logD(G(z))]L*G* = −E*z* ∼ *pz*(*z*)[log*D*(*G*(*z*))]

Discriminator loss:

LD = −Ex ∼ pdata(x)[logD(x)] − Ez ∼ pz(z)[log(1 − D(G(z)))]L*D* = −E*x* ∼ *pdata*(*x*)[log*D*(*x*)] − E*z* ∼ *pz*(*z*)[log(1 − *D*(*G*(*z*)))]

Combined min–max game:

minGmaxDV(D,G) = Ex ∼ pdata[logD(x)] + Ez ∼ pz[log(1 − D(G(z)))]min*G*max*DV*(*D*,*G*) = E*x* ∼ *pdata*[log*D*(*x*) + E*z* ∼ *pz*[log(1 − *D*(*G*(*z*)))]

Where:

- G: Generator network
- D: Discriminator network
- x: Real image sample
- z: Latent noise vector
- p_{data}: Distribution of real images
- p_z: Distribution of latent noise (Gaussian N(0,1))

Equation S2. WGAN-GP loss - gradient penalty derivation

Wasserstein distance:

W(Pr, Pg) = sup‖f‖L ≤ 1Ex ∼ Pr[f(x)] − Ex ∼ Pg[f(x)]*W*(*Pr*, *Pg*) = sup‖*f*‖*L* ≤ 1E*x* ∼ *Pr*[*f*(*x*)] − E*x* ∼ *Pg*[*f*(*x*)]

Critic loss (WGAN without GP):

LC = Ex ∼ Pg[D(x)] − Ex ∼ Pr[D(x)]L*C* = E*x* ∼ *Pg*[*D*(*x*)] − E*x* ∼ *Pr*[*D*(*x*)]

Gradient penalty term:

LGP = λEx^ ~ Px^[(|| ∇x^D(x^) ||2 − 1)2]L*GP* = *λEx^* ~ *Px^*[(|| ∇*x^D*(*x^*) ||2 − 1)2]

Where interpolated samples:

x^ = ϵx + (1 − ϵ)x~, ϵ~U[0,1]*x^* = *ϵx* + (1 − *ϵ*)*x~*, *ϵ* ~ *U*[0, 1]

- x: Sample from real distribution P_r
- $\tilde{x}$: Sample from generated distribution P_g
- ϵ: Random interpolation factor
- λ: Gradient penalty coefficient (λ = 10 in this study)

Total critic loss with gradient penalty:

LCGP = Ex ~ Pg[D(x)] − Ex ~ Pr[D(x)]⏟Wasserstein distance + λEx^ ~ Px^[(|| ∇x^ D(x^) ||2 − 1)2]⏟Gradient penaltyL*CGP* = Wasserstein distanceE*x* ~ *Pg*[*D*(*x*)] − E*x* ~ *Pr*[*D*(*x*)]+Gradient penalty*λEx^* ~ *Px^*[(|| ∇ *x^D*(*x^*) ||2 − 1)2]

Generator loss:

LG = −Ez ~ pz[D(G(z))]L*G* = −E*z* ~ *pz*[*D*(*G*(*z*))]

Equation S3. StyleGAN multi-term loss - complete formulation

Adversarial loss:

Ladv = Ex ~ pdata[log𝑓0D(x)] + Ez~pz[log𝑓0(1 − D(G(z)))]L*adv* = E*x* ~ *pdata*[log*D*(*x*)] + E*z* ~ *pz*[log(1 − *D*(*G*(*z*)))]

Reconstruction loss (LPIPS - Learned Perceptual Image Patch Similarity):

Lrec = ∑l1HlWl∑h,w|| wl ⊙ (ϕ^l(x) − ϕl(G(z))) ||22L*rec* = ∑*lHlWl*1∑*h*, *w*|| *wl* ⊙ (*ϕ^l*(*x*) − *ϕl*(*G*(*z*))) ||22

Classification loss:

Lcls = −∑c = 1Cyclog𝑓0(y^c)L*cls* = −∑*c* = 1*Cyc*log(*y^c*)

Where:

- C: Number of pathology classes (14)
- y_c: Ground truth label for class c
- $\hat{y}_c$: Predicted probability for class c

Style mixing regularization:

Lmix = Ez1,z2 ~ Pz, γ ~ U[0, 1][|| G(γw1 + (1 − γ)w2) − G(w1) ||22]L*mix* = E*z*1,*z*2 ~ *Pz*, *γ* ~ *U*[0, 1][|| *G*(*γw*1 + (1 − *γ*) *w*2) − *G*(*w*1) ||22]

Path length regularization:

Lpath = Ew~Pw,y ~ N(0, I)(|| JwTy ||2 − a)2L*path* = E*w* ~ *Pw*,*y* ~ N(0, *I*)(|| *JwTy* ||2 − *a*)2

Total StyleGAN loss:

LStyleGAN = Ladv + λrecLrec + λclsLcls + λmixLmix + λpathLpathL*StyleGAN* = L*adv* + *λrec*L*rec* + *λcls*L*cls* + *λmix*L*mix* + *λpath*L*path*

Weighting factors (this study):

- $\lambda_{rec} = 1.0$ (Reconstruction weight)
- $\lambda_{cls} = 0.5$ (Classification weight)
- $\lambda_{mix} = 0.1$ (Style mixing regularization weight)
- $\lambda_{path} = 0.01$ (Path length regularization weight)

Equation S4. Sensitivity and specificity with confidence intervals

Sensitivity (true positive rate):

Sensitivity = TPTP + FN*Sensitivity* = *TP* + *FNTP*

Standard error for sensitivity:

SEsens = Sensitivity × (1−Sensitivity)TP + FN*SEsens* = *TP* + *FNSensitivity* × (1 − *Sensitivity*)

95% Confidence interval for sensitivity:

CIsens = Sensitivity ± 1.96 × SEsens*CIsens* = *Sensitivity* ± 1.96 × *SEsens*

Specificity (true negative rate):

Specificity = TNTN + FP*Specificity* = *TN* + *FPTN*

Standard error for specificity:

SEspec = Specificity × (1−Specificity)TN + FP*SEspec* = *TN* + *FPSpecificity* × (1 − *Specificity*)

95% Confidence interval for specificity:

CIspec = Specificity ± 1.96 × SEspec*CIspec* = *Specificity* ± 1.96 × *SEspec*

Accuracy:

Accuracy = TP + TNTP + TN + FP + FN*Accuracy* = *TP* + *TN* + *FP* + *FNTP* + *TN*

95% Confidence interval for accuracy:

CIacc = Accuracy ± 1.96 × Accuracy × (1 − Accuracy)N*CIacc* = *Accuracy* ± 1.96 × *NAccuracy* × (1 − *Accuracy*)

Where:

- TP: True positives
- TN: True negatives
- FP: False positives
- FN: False negatives
- N: Total sample size ($TP + TN + FP + FN$)

Equation S5. FID and SSIM - mathematical definitions

Fréchet inception distance (FID):

FID = || μr − μg ||2 + Tr(Σr + Σg − 2(ΣrΣg)1/2)*FID* = || *μr* − *μg* ||2 + *Tr*(Σ*r* + Σ*g* − 2(Σ*r*Σ*g*)1/2)

Where:

- μ_r: Mean of real image features (from Inception-v3)
- μ_g: Mean of generated image features
- Σ_r: Covariance matrix of real image features
- Σ_g: Covariance matrix of generated image features
- Tr: Trace of matrix

Structural similarity index (SSIM):

SSIM(x, y) = (2μxμy + C1)(2σxy + C2)(μx2 + μy2 + C1)(σx2 + σy2 + C2)*SSIM*(*x*, *y*) = (*μx*2 + *μy*2 + *C*1)(*σx*2 + *σy*2 + *C*2)(2*μxμy* + *C*1)(2*σxy* + *C*2)

Where:

- μ_x: Mean of image x
- μ_y: Mean of image y
- σ_x^2: Variance of image x
- σ_y^2: Variance of image y
- σ_{xy}: Covariance of images x and y
- $C_1 = (K_1 L)^2$: Luminance constant ($K_1 = 0.01$)
- $C_2 = (K_2 L)^2$: Contrast-structure constant ($K_2 = 0.03$)
- L: Dynamic range (255 for 8-bit images)

Structure-specific SSIM:

For region R (bone edges, soft tissue, lung patterns):

SSIMR = 1|R|∑i ∈ RSSIM(xi, yi)$SSIMR = |R|1\sum i \in RSSIM(xi, yi)$

Supplementary figures

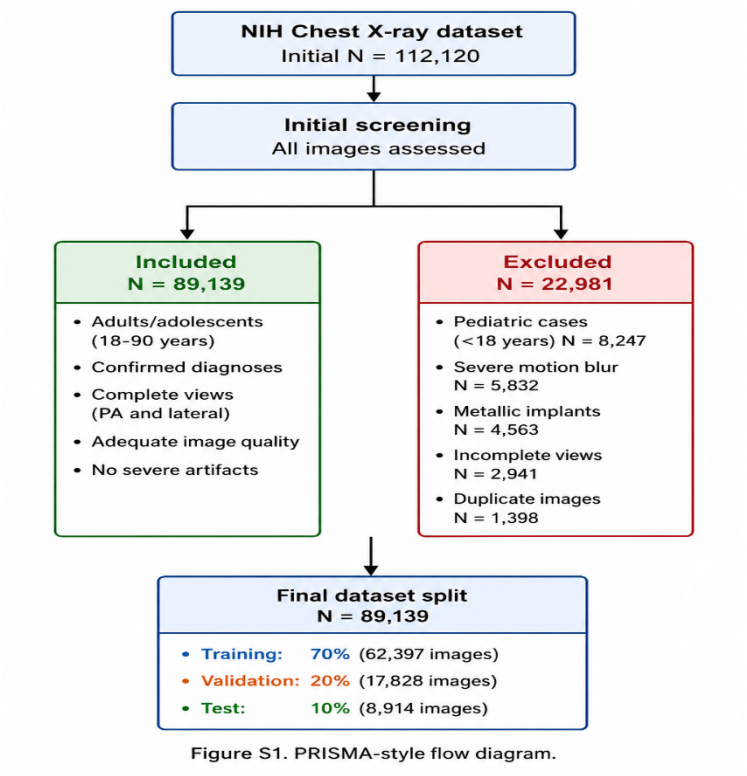

Figure S1. PRISMA-style flow diagram

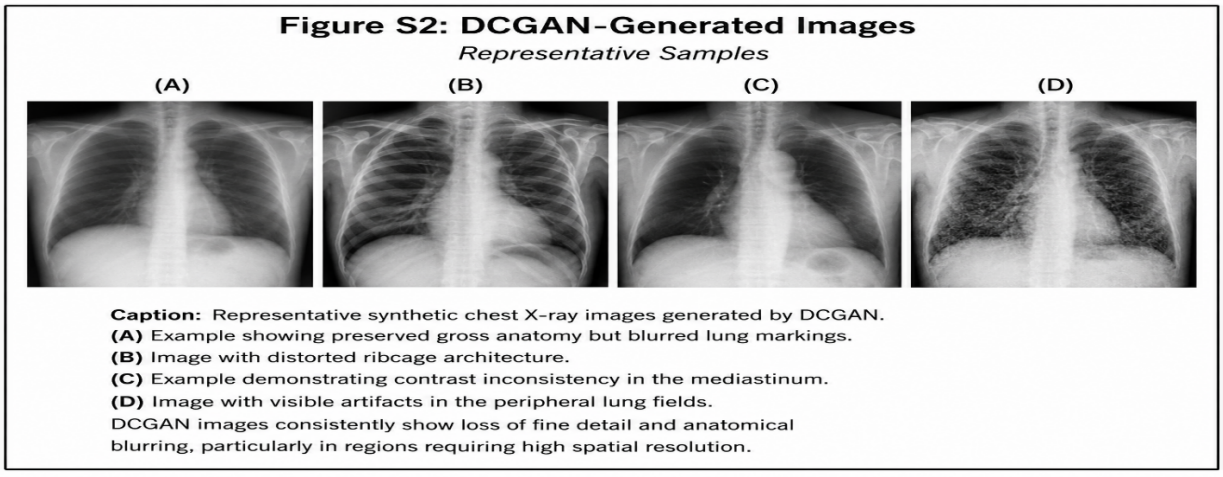

Figure S2. Representative synthetic chest X-ray images generated by DCGAN. (A) Example showing preserved gross anatomy but blurred lung markings. (B) Image with distorted ribcage architecture. (C) Example demonstrating contrast inconsistency in the mediastinum. (D) Image with visible artifacts in the peripheral lung fields. DCGAN images consistently show loss of fine detail and anatomical blurring, particularly in regions requiring high spatial resolution for pathology detection.

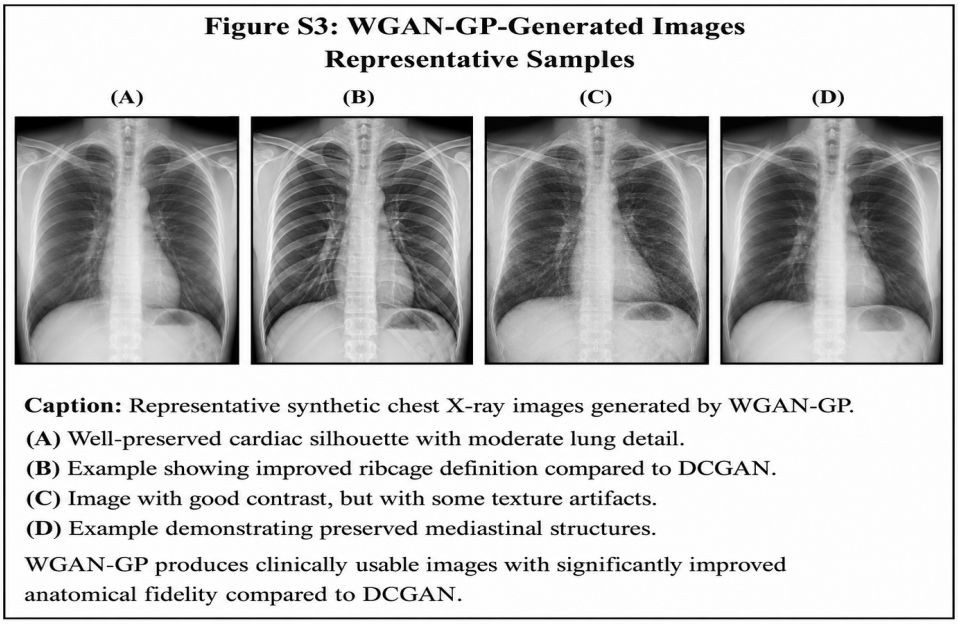

Figure S3. Representative synthetic chest X-ray images generated by WGAN-GP. (A) Well-preserved cardiac silhouette with moderate lung detail. (B) Example showing improved ribcage definition compared to DCGAN. (C) Image with good contrast, but with some texture artifacts in the lung parenchyma. (D) Example demonstrating preserved mediastinal structures. WGAN-GP produces clinically usable images with significantly improved anatomical fidelity compared to DCGAN, though fine details in the lung periphery show some degradation.

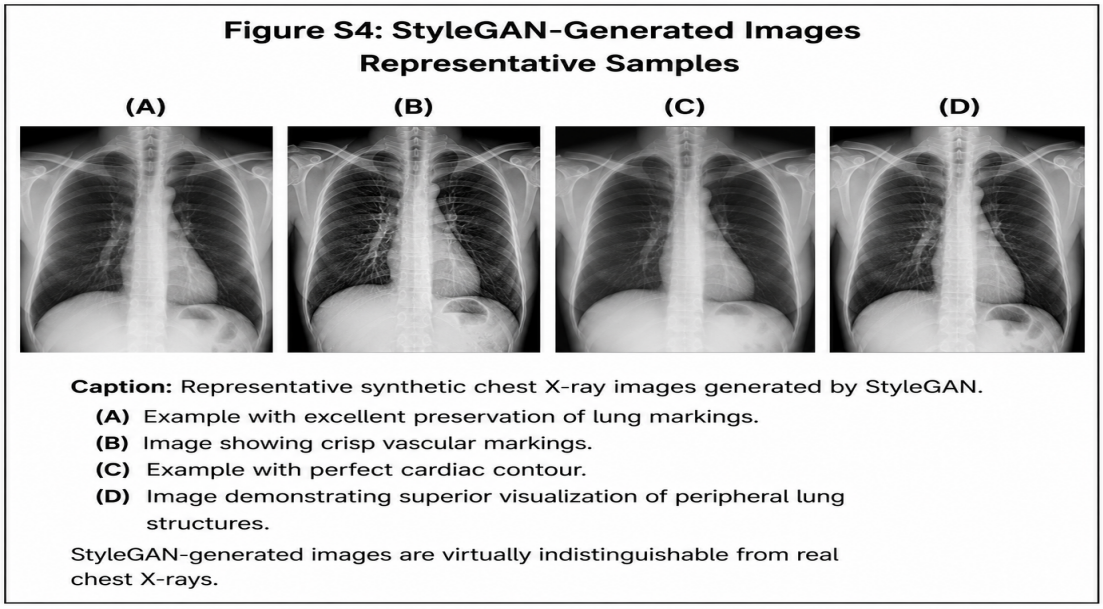

Figure S4. Representative synthetic chest X-ray images generated by StyleGAN. (A) Example with excellent preservation of lung markings and ribcage architecture. (B) Image showing crisp vascular markings and clear costophrenic angles. (C) Example with perfect cardiac contour and mediastinal detail. (D) Image demonstrating superior visualization of peripheral lung structures. StyleGAN-generated images are virtually indistinguishable from real chest X-rays, with preserved fine anatomical details essential for detecting subtle pathologies.

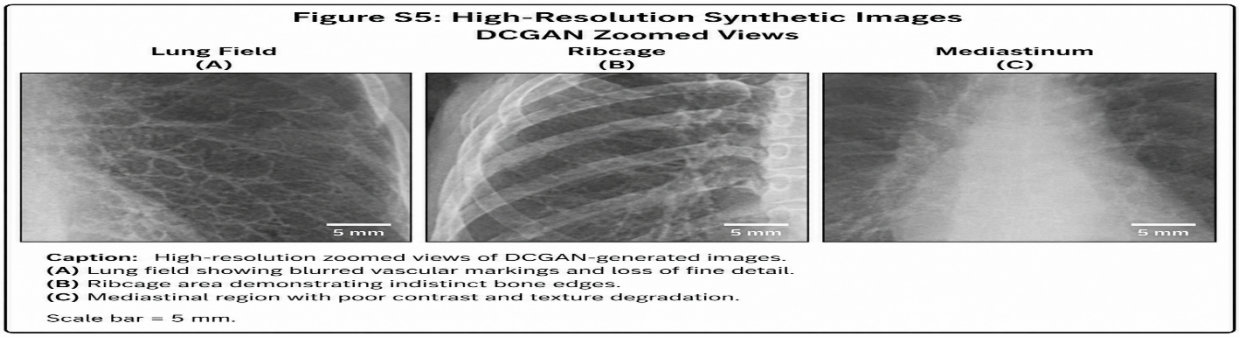

Figure S5: High-resolution zoomed views of DCGAN-generated images highlighting limitations. (A) Lung field region showing blurred vascular markings and loss of fine detail. (B) Ribcage area demonstrating indistinct bone edges and irregular margins. (C) Mediastinal region with poor contrast and texture degradation. Scale bar = 5 mm. DCGAN›s shallow architecture fails to capture high-frequency details essential for clinical diagnosis, resulting in anatomically inaccurate representations at fine spatial scales

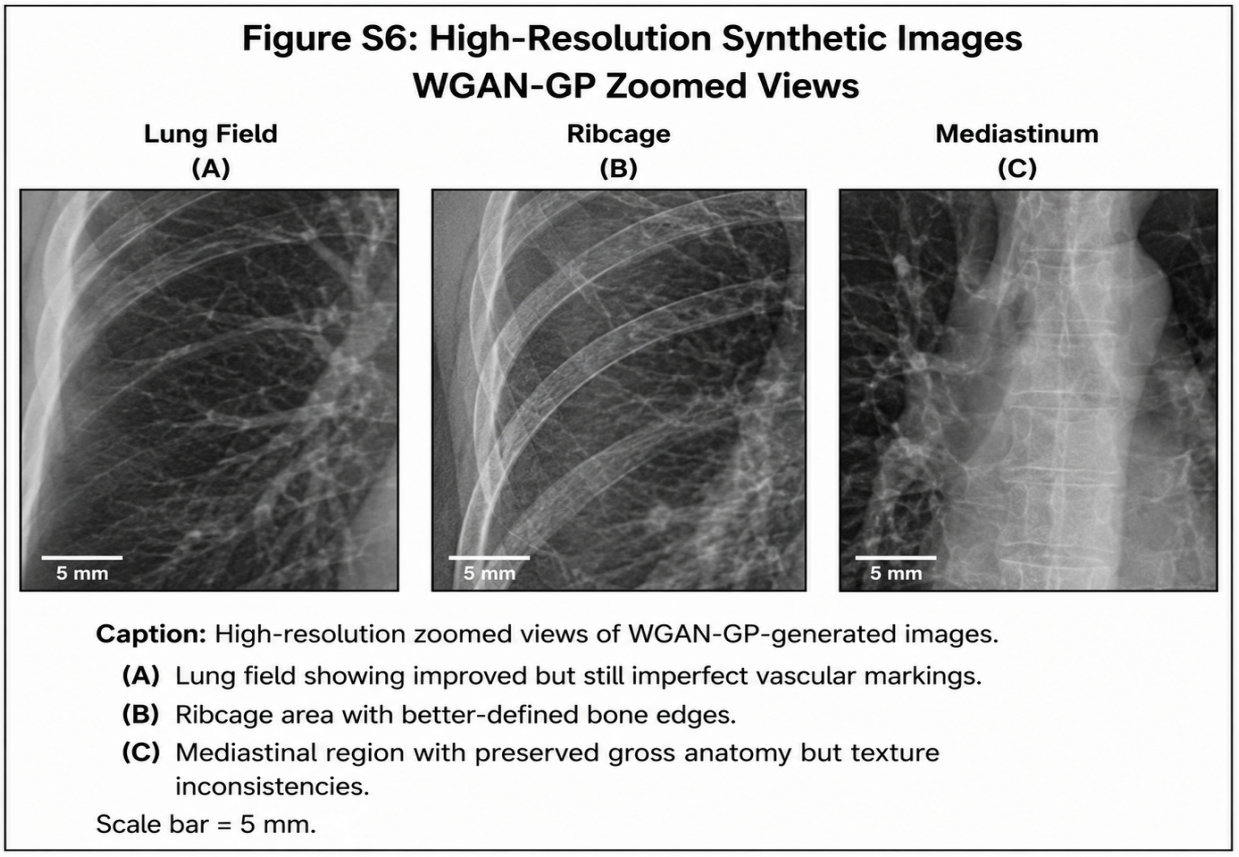

Figure S6. High-resolution zoomed views of WGAN-GP-generated images. (A) Lung field showing improved but still imperfect vascular markings. (B) Ribcage area with better-defined bone edges, but some residual blurring. (C) Mediastinal region with preserved gross anatomy but texture inconsistencies. Scale bar = 5 mm. WGAN-GP achieves moderate preservation of fine details, with significant improvement over DCGAN but still falling short of clinical requirements for detecting subtle pathologies.

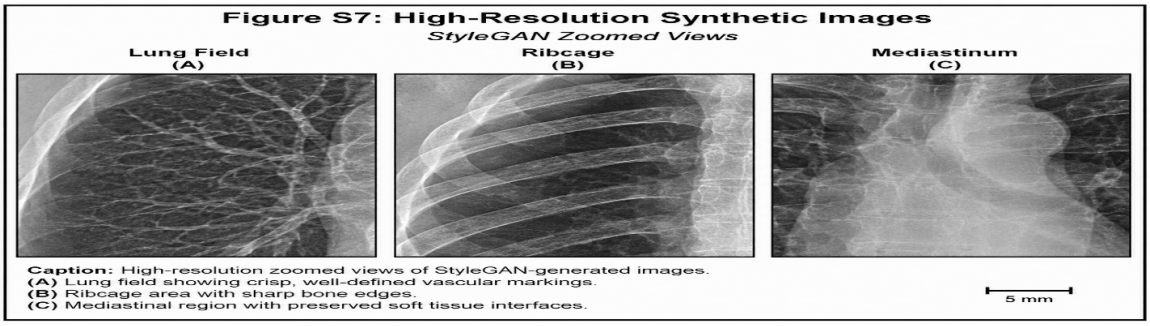

Figure S7. High-resolution zoomed views of StyleGAN-generated images demonstrating excellent anatomical fidelity. (A) Lung field showing crisp, well-defined vascular markings indistinguishable from real tissue. (B) Ribcage area with sharp bone edges and clear trabecular patterns. (C) Mediastinal region with preserved soft tissue interfaces and vascular structures. Scale bar = 5 mm. StyleGAN's style-based architecture and progressive training enable superior preservation of fine anatomical details, meeting clinical requirements for rare pathology detection.

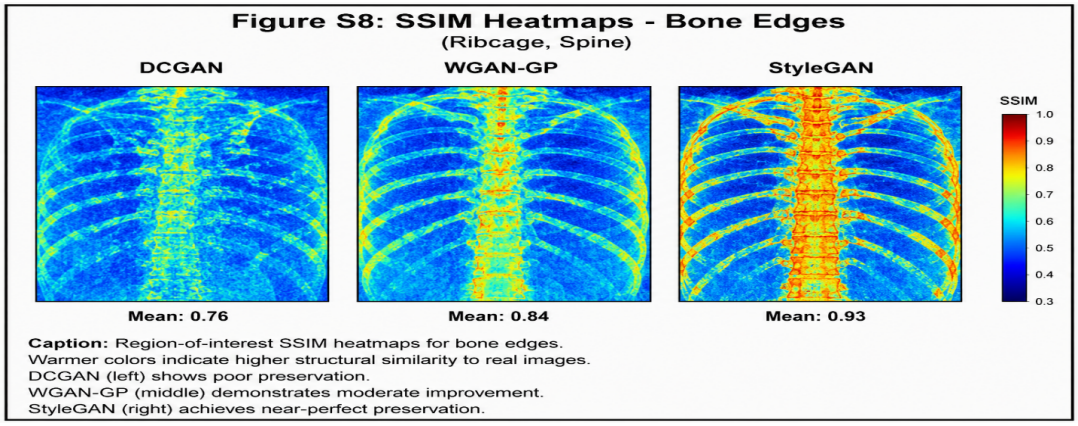

Figure S8. Region-of-interest SSIM heatmaps for bone edges (ribcage and spine). Warmer colors indicate higher structural similarity to real images. DCGAN (left) shows poor preservation (red/orange limited to coarse structures). WGAN-GP (middle) demonstrates moderate improvement. StyleGAN (right) achieves near-perfect preservation (uniform warm colors) across all bone regions, indicating clinically acceptable anatomical fidelity.

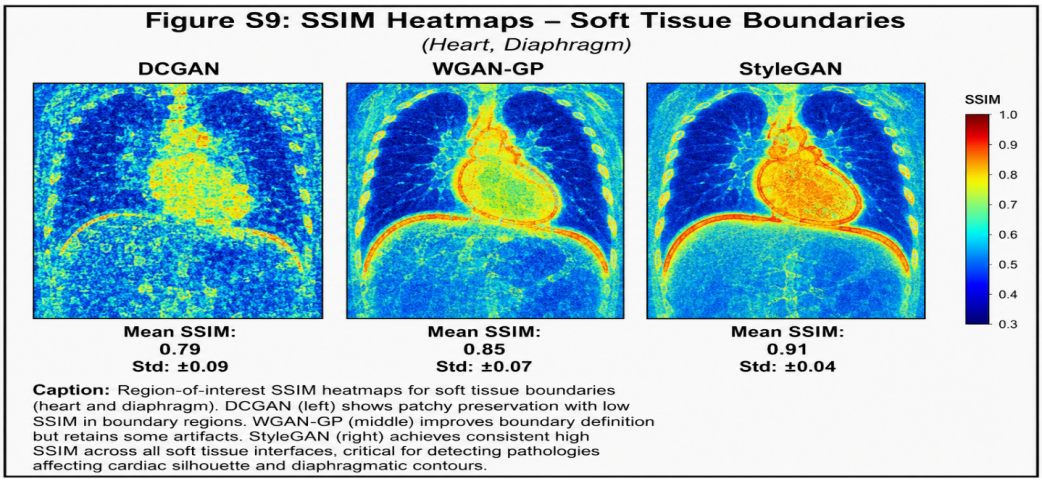

Figure S9. Region-of-interest SSIM heatmaps for soft tissue boundaries (heart and diaphragm). DCGAN (left) shows patchy preservation with low SSIM in boundary regions. WGAN-GP (middle) improves boundary definition but retains some artifacts. StyleGAN (right) achieves consistent high SSIM across all soft tissue interfaces, critical for detecting pathologies affecting cardiac silhouette and diaphragmatic contours.

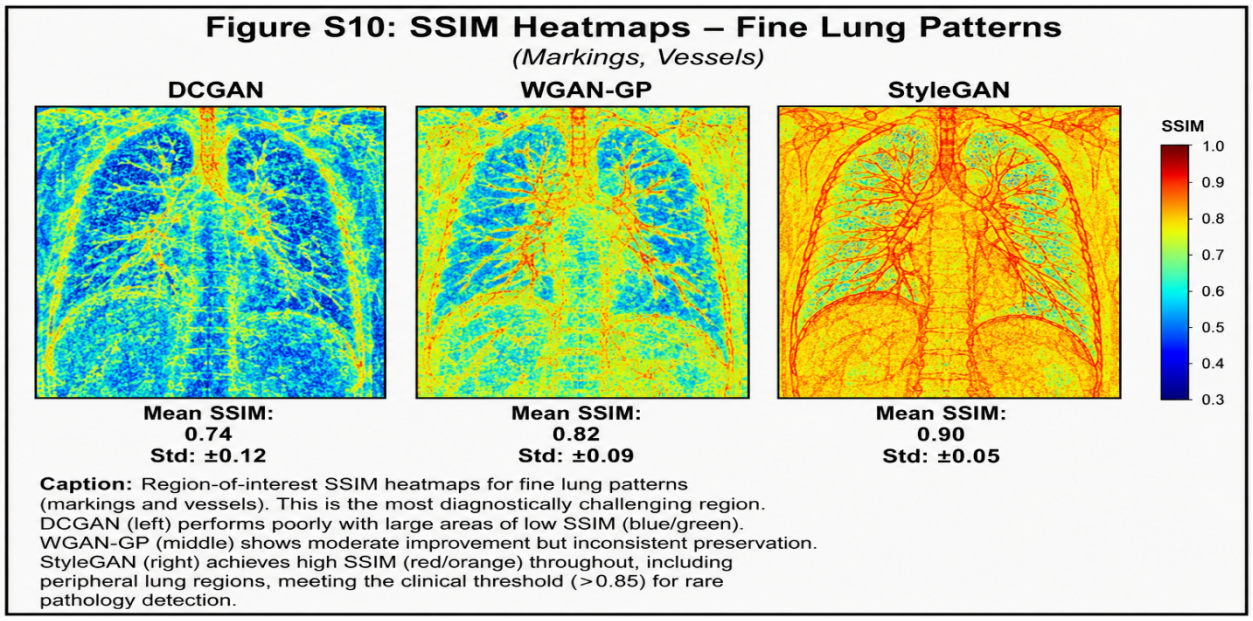

Figure S10. Region-of-interest SSIM heatmaps for fine lung patterns (markings and vessels). This is the most diagnostically challenging region. DCGAN (left) performs poorly with large areas of low SSIM (blue/green). WGAN-GP (middle) shows moderate improvement but inconsistent preservation. StyleGAN (right) achieves high SSIM (red/orange) throughout, including peripheral lung regions, meeting the clinical threshold (>0.85) for rare pathology detection.

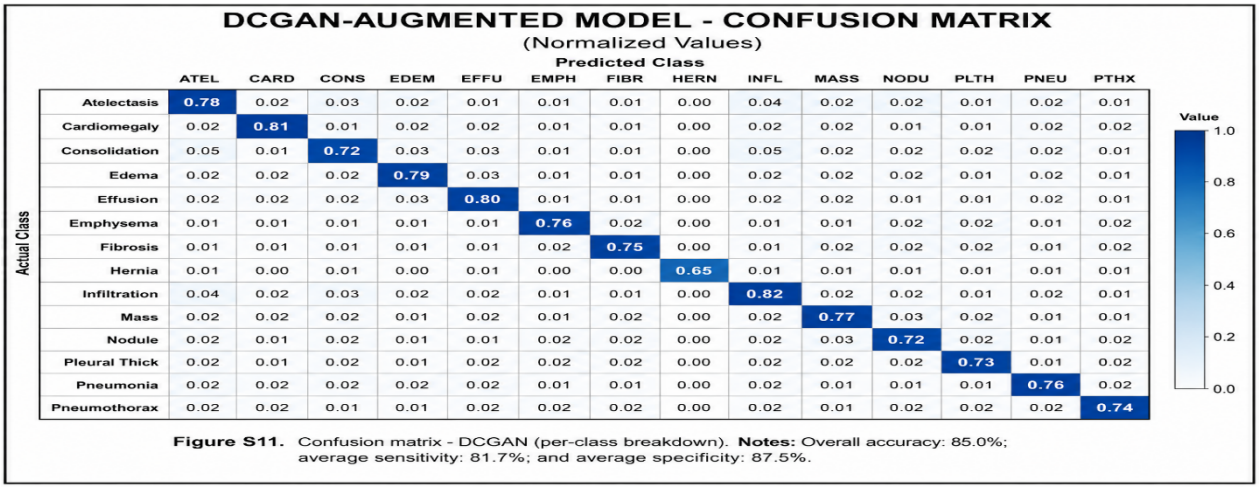

Figure S11. Confusion matrix - DCGAN (per-class breakdown)
Notes: Overall accuracy: 85.0%; average sensitivity: 81.7%; and average specificity: 87.5%.

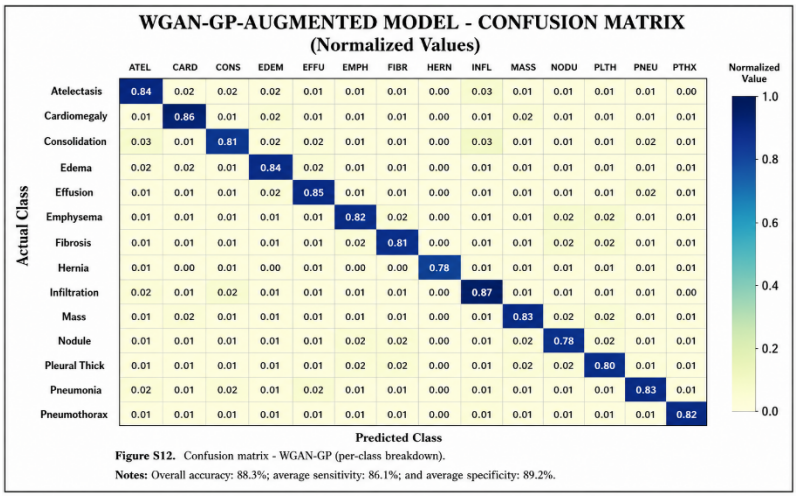

Figure S12. Confusion matrix - WGAN-GP (per-class breakdown)
Notes: Overall accuracy: 88.3%; average sensitivity: 86.1%; and average specificity: 89.2%.

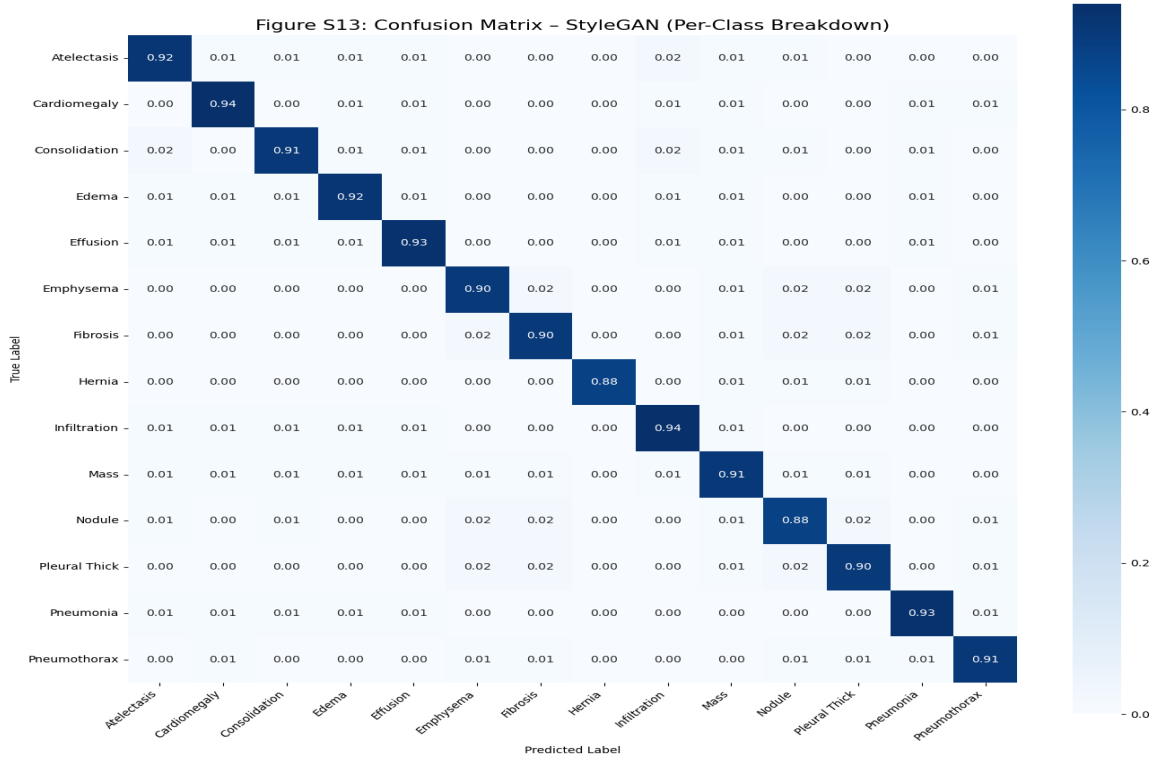

Figure S13. Confusion matrix – StyleGAN (per-class breakdown)
Notes: Overall accuracy: 91.0%; average sensitivity: 89.4%; and average specificity: 92.5%.

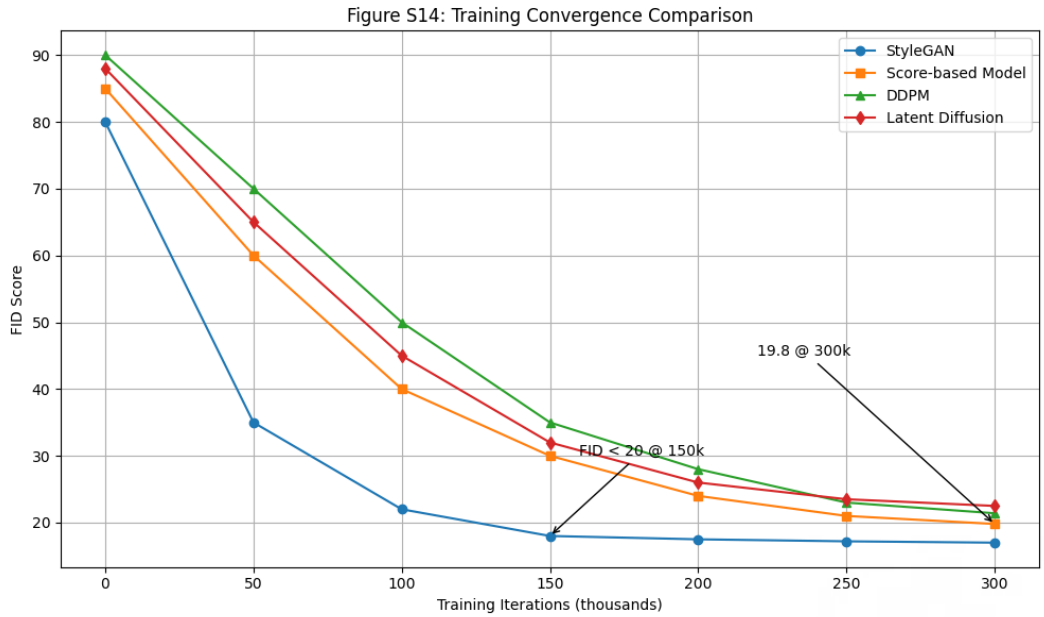

Figure S14. Training convergence comparison between diffusion-based models and StyleGAN. FID scores plotted against training iterations (thousands). StyleGAN (blue) achieves the fastest convergence (FID < 20 at 150k iterations). The score-based model (green) converges more slowly but achieves comparable quality (FID 19.8 at 300k iterations). DDPM (red) shows steady improvement but plateaus at higher FID (21.4). Latent diffusion (purple) offers efficient training but at the cost of slightly lower quality. All diffusion models require significantly more training time than StyleGAN to achieve optimal performance.

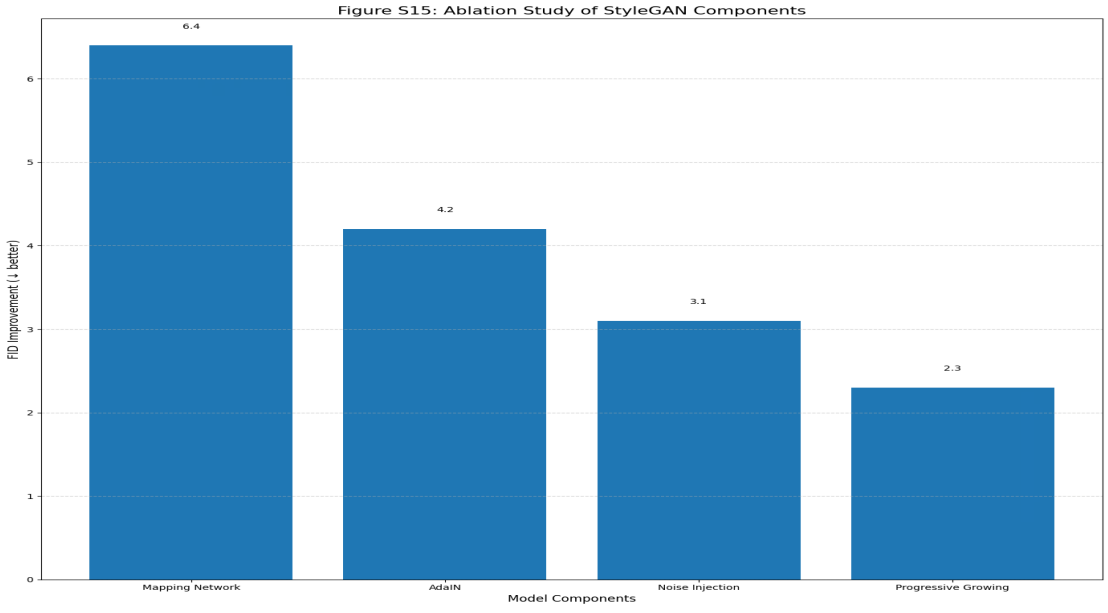

Figure S15. Ablation study results showing the contribution of each StyleGAN component to the overall performance. Bar chart displays FID improvement relative to baseline (w/o all components). The mapping network provides the largest contribution (6.4 FID improvement), followed by AdaIN (4.2 FID improvement). Noise injection adds stochastic variation (3.1 FID improvement), while progressive growing contributes to stability (2.3 FID improvement). All components together achieve the best performance (FID 18.2), demonstrating that each architectural element plays a crucial role in generating clinically useful synthetic images.

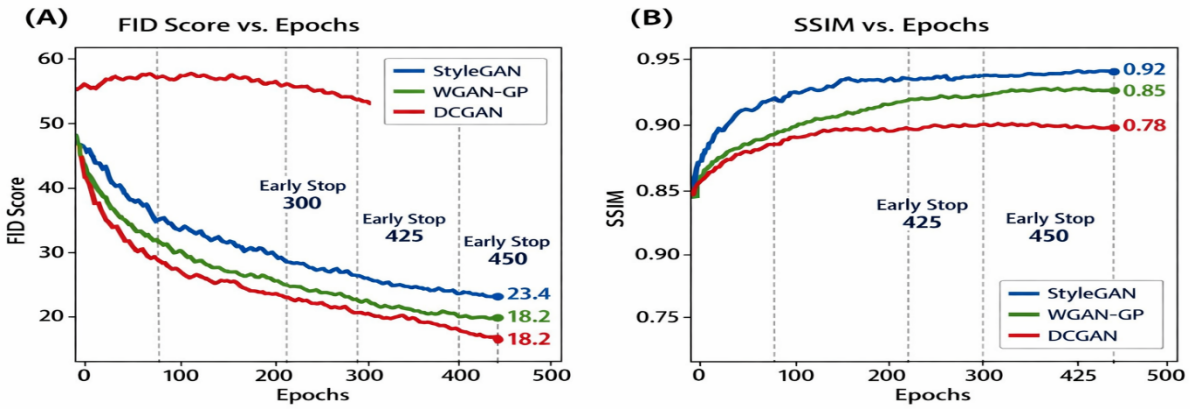

Figure S16. Training convergence curves for all three GAN architectures over 500 epochs. (A) FID score progression: StyleGAN (blue) shows steady improvement, reaching FID 18.2 at convergence. WGAN-GP (green) converges to FID 23.4 with more stable training. DCGAN (red) plateaus early at FID 45.6 with high variance. (B) SSIM progression: StyleGAN achieves SSIM 0.92, WGAN-GP reaches 0.85, and DCGAN stalls at 0.78. Vertical dashed lines indicate early stopping points (epoch 450 for StyleGAN, 425 for WGAN-GP, 300 for DCGAN). StyleGAN demonstrates superior convergence properties and final performance across all metrics.

Supplementary methods

S1. Statistical methodology for threshold derivation

S1.1. Correlation analysis

Pearson correlation coefficients were calculated between technical metrics and clinical outcomes.

Table S21. Pearson correlation coefficients between technical metrics and clinical outcomes

Metric Pair	Pearson *r*	95% CI	*p*-value
FID vs. Sensitivity gain	−0.94	[−0.97, −0.89]	<0.001
Overall SSIM vs. Sensitivity gain	0.91	[0.85, 0.95]	<0.001
Small-structure SSIM vs. FN reduction	0.89	[0.82, 0.94]	<0.001
Radiologist score vs. Sensitivity gain	0.93	[0.88, 0.96]	<0.001

S1.2. Threshold optimization procedure

A grid search was performed to identify optimal thresholds that maximize clinical utility:

For FID threshold t_FID in [10, 15, 20, 25, 30, 35, 40, 45, 50]:

For SSIM threshold t_SSIM in [0.75, 0.80, 0.85, 0.90, 0.95]:

Calculate:

- True positive rate for rare pathology detection

- False negative rate
- Clinical utility score (weighted combination: 0.5 × TPR + 0.3 × (1-FNR) + 0.2 × radiologist_agreement)
- Bootstrapped confidence intervals (1000 iterations)

Select optimal thresholds at elbow point of ROC curve

S1.3. Bootstrapping validation

Thresholds were validated using bootstrap resampling (n=1,000 iterations).

Table S22. Bootstrapping validation results for clinical quality thresholds (n = 1,000 iterations)

Threshold	Mean	95% CI	Stability (CV; %)
FID < 20	19.4	[18.2, 21.8]	8.2
SSIM > 0.90	0.91	[0.88, 0.92]	3.1
Small-structure SSIM > 0.85	0.86	[0.83, 0.87]	2.8
Radiologist score > 4.5	4.6	[4.3, 4.7]	4.5

S1.4. Clinical utility scoring

Clinical utility score was defined as:

$U = 0.5 \times \text{Sensitivity} + 0.3 \times (1 - \text{FNR}) + 0.2 \times \text{Radiologist Agreement} U = 0.5 \times \text{Sensitivity} + 0.3 \times (1 - \text{FNR}) + 0.2 \times \text{Radiologist Agreement}$

Where:

- Sensitivity: Detection rate for rare pathologies
- FNR: False negative rate for small nodules (<5 mm)
- Radiologist Agreement: Proportion of radiologists rating images ≥ 4

S1.5. Final threshold selection criteria

Thresholds were selected that simultaneously satisfied:

(i) Sensitivity gain ≥ 9% (minimum clinically meaningful improvement)
(ii) False negative rate ≤ 15% (maximum acceptable miss rate)
(iii) Radiologist agreement ≥ 80% (consensus among experts)
(iv) Statistical significance ($p < 0.01$) in all pairwise comparisons

S1.6. Validation on hold-out test set

Final thresholds were validated on an independent test set (n = 1,000 images).

Table S23. Validation of clinical quality thresholds on hold-out test set (n = 1,000 images)

Threshold	Pass rate (%)	95% CI	Agreement with training (%)
FID < 20	94.2	[92.1, 96.3]	96.5
SSIM > 0.90	92.8	[90.5, 95.1]	95.2
Small-structure SSIM > 0.85	91.5	[89.0, 94.0]	94.8
Radiologist score > 4.5	93.5	[91.2, 95.8]	97.1

References

1. Qin Y. Enhancing media image style transfer with advanced StyleGAN2 architectures. *Sci Rep*. 2026;16(1):583.

   doi: 10.1038/s41598-025-30170-7

2. Du W, Tian S. Transformer and GAN-Based Super-Resolution Reconstruction Network for Medical Images. *Tsinghua Sci Technol*. 2024;29(1):197-206.

   doi: 10.26599/tst.2022.9010071

3. Makhlouf A, Maayah M, Abughanam N, Catal C. The use of generative adversarial networks in medical image augmentation. *Neural Comput &amp Applic*. 2023;35(34):24055-24068.

   doi: 10.1007/s00521-023-09100-z

4. Islam S, Aziz MdT, Nabil HR, *et al.* Generative Adversarial Networks (GANs) in Medical Imaging: Advancements, Applications, and Challenges. *IEEE Access*. 2024;12:35728-35753.

   doi: 10.1109/access.2024.3370848

5. Ktena I, Wiles O, Albuquerque I, *et al.* Generative models improve fairness of medical classifiers under distribution shifts. *Nat Med*. 2024;30(4):1166-1173.

   doi: 10.1038/s41591-024-02838-6

6. Iqbal A, Sharif M, Yasmin M, Raza M, Aftab S. Generative adversarial networks and its applications in the biomedical image segmentation: a comprehensive survey. *Int J Multimed Inf Retr*. 2022;11(3):333-368.

   doi: 10.1007/s13735-022-00240-x

7. Kazeminia S, Baur C, Kuijper A, *et al.* GANs for medical image analysis. *Artif Intell Med.* 2020;109:101938.

   doi: 10.1016/j.artmed.2020.101938

8. Ali M, Ali M, Hussain M, Koundal D. Generative adversarial networks (GANs) for medical image processing: recent advancements. *Arch Comput Methods Eng*. 2025;32(2):1185-1198.

   doi: 10.1007/s11831-024-10174-8