**Supplementary Materials**

A picture containing text

Description automatically generated

Supplementary Figure 1. An example image of selected samples from TCGA-LUAD and LIDC-IDRI. (a) an example for high dose CT image from TCGA-LUAD; (b) an example for low dose CT image from LIDC-IDRI.

Chart, histogram

Description automatically generated

Supplementary Figure 2. The distribution of radiation exposure for selected samples in the NSCLC Radiogenomics

A picture containing diagram

Description automatically generated

Supplementary Figure 3. Example of low dose CT denoising. (a-1) The original full dose CT image; (b-1) high-noise image; (c-1) Image denoised by encoder-decoder network (\*Training at 100 epochs); (d-1) Image denoised by CGAN; (e-1) Image denoised by cycle GAN; (f-1) Image denoised by cycle GAN (ablation study); (a-2) to (f-2) Zoomed ROIs for (a-1) to (f-1).

Graphical user interface

Description automatically generated with low confidence

Supplementary Figure 4. CDF of CCC by Using Cycle GAN. (a) CDF of CCCs based on denoised low-noise by using Cycle GAN trained for different numbers of epochs; (b) CDF of CCCs based on denoised high-noise by using Cycle GAN trained for different numbers of epochs; (c) CDF of CCCs based on denoised low-noise by using Cycle GAN trained without strategy for different numbers of epochs; (d) CDF of CCCs based on denoised high-noise by using Cycle GAN trained without strategy for different numbers of epochs.

Graphical user interface, application, website

Description automatically generated

Supplementary Figure 5. Example of RIDER denoising. (a-1) One original image from RIDER; (b-1) Image denoised by encoder-decoder network (Training at 100 epochs); (c-1) Image denoised by CGAN (Training at 100 epochs); (d-1) Image denoised by simulation data trained Cycle GAN (Training at 100 epochs); (e-1) Image denoised by real data trained Cycle GAN (Training at 100 epochs); (a-2) to (e-2) Zoomed ROIs for (a-1) to (e-1).

Supplementary Table 1. Index of available patients for experiments in LUNG 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Training Samples | LUNG1-001 | LUNG1-029 | LUNG1-032 | LUNG1-051 | LUNG1-072 | LUNG1-098 |
| LUNG1-100 | LUNG1-105 | LUNG1-109 | LUNG1-115 | LUNG1-116 | LUNG1-119 |
| LUNG1-120 | LUNG1-121 | LUNG1-122 | LUNG1-124 | LUNG1-126 | LUNG1-127 |
| LUNG1-128 | LUNG1-130 | LUNG1-131 | LUNG1-132 | LUNG1-133 | LUNG1-134 |
| LUNG1-139 | LUNG1-141 | LUNG1-142 | LUNG1-145 | LUNG1-147 | LUNG1-148 |
| LUNG1-151 | LUNG1-152 | LUNG1-156 | LUNG1-157 | LUNG1-160 | LUNG1-161 |
| LUNG1-162 | LUNG1-163 | LUNG1-165 | LUNG1-169 |  |  |
| Testing Samples | LUNG1-171 | LUNG1-172 | LUNG1-174 | LUNG1-176 | LUNG1-178 | LUNG1-179 |
| LUNG1-189 | LUNG1-191 | LUNG1-192 | LUNG1-195 | LUNG1-196 | LUNG1-197 |
| LUNG1-199 | LUNG1-205 | LUNG1-206 | LUNG1-210 | LUNG1-211 | LUNG1-212 |
| LUNG1-213 | LUNG1-214 | LUNG1-215 | LUNG1-216 | LUNG1-217 | LUNG1-218 |
| LUNG1-220 | LUNG1-221 | LUNG1-222 | LUNG1-223 | LUNG1-224 | LUNG1-225 |
| LUNG1-226 | LUNG1-227 | LUNG1-229 | LUNG1-231 | LUNG1-233 | LUNG1-234 |
| LUNG1-235 | LUNG1-236 | LUNG1-237 | LUNG1-239 | LUNG1-243 | LUNG1-244 |
| LUNG1-245 | LUNG1-247 | LUNG1-249 | LUNG1-252 | LUNG1-253 | LUNG1-255 |
| LUNG1-256 | LUNG1-257 | LUNG1-259 | LUNG1-260 | LUNG1-262 | LUNG1-263 |
| LUNG1-264 | LUNG1-266 | LUNG1-267 | LUNG1-268 | LUNG1-269 | LUNG1-270 |
| LUNG1-271 | LUNG1-272 | LUNG1-273 | LUNG1-274 | LUNG1-275 | LUNG1-276 |
| LUNG1-277 | LUNG1-278 | LUNG1-280 | LUNG1-282 | LUNG1-283 | LUNG1-284 |
| LUNG1-285 | LUNG1-287 | LUNG1-288 | LUNG1-289 | LUNG1-290 | LUNG1-293 |
| LUNG1-294 | LUNG1-295 | LUNG1-296 | LUNG1-297 | LUNG1-298 | LUNG1-299 |
| LUNG1-300 | LUNG1-303 | LUNG1-304 | LUNG1-305 | LUNG1-306 | LUNG1-307 |
| LUNG1-309 | LUNG1-310 | LUNG1-311 | LUNG1-313 | LUNG1-315 | LUNG1-317 |
| LUNG1-318 | LUNG1-320 | LUNG1-321 | LUNG1-323 | LUNG1-326 | LUNG1-328 |
| LUNG1-331 | LUNG1-332 | LUNG1-334 | LUNG1-335 | LUNG1-337 | LUNG1-339 |
| LUNG1-340 | LUNG1-341 | LUNG1-342 | LUNG1-343 | LUNG1-345 | LUNG1-347 |
| LUNG1-349 | LUNG1-353 | LUNG1-354 |  |  |  |

Supplementary Table 2. Index of available patients for real data Cycle GAN training in LIDC-IDRI (Low Dose CT Domain)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Training Samples | LIDC-IDRI-0218 | LIDC-IDRI-0306 | LIDC-IDRI-0336 | LIDC-IDRI-0394 |
| LIDC-IDRI-0430 | LIDC-IDRI-0603 | LIDC-IDRI-0604 | LIDC-IDRI-0780 |
| LIDC-IDRI-0854 | LIDC-IDRI-0862 | LIDC-IDRI-0903 | LIDC-IDRI-0953 |

Supplementary Table 3. Index of available patients for real data Cycle GAN training in TCGA-LUAD (High Dose CT Domain)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Training Samples | TCGA-17-Z017 | TCGA-17-Z019 | TCGA-17-Z021 | TCGA-17-Z027 |
| TCGA-17-Z029 | TCGA-17-Z034 | TCGA-17-Z050 | TCGA-17-Z051 |
| TCGA-17-Z053 | TCGA-17-Z054 | TCGA-17-Z058 | TCGA-17-Z059 |
| TCGA-38-4628 | TCGA-50-6591 |  |  |

Supplementary Table 4. Index of available patients for pre-treatment survival predication in NSCLC Radiogenomics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| R01-001 | R01-035 | R01-062 | R01-086 | R01-108 | R01-133 |
| R01-002 | R01-037 | R01-063 | R01-088 | R01-110 | R01-134 |
| R01-003 | R01-039 | R01-064 | R01-090 | R01-111 | R01-135 |
| R01-004 | R01-040 | R01-066 | R01-091 | R01-112 | R01-136 |
| R01-005 | R01-041 | R01-067 | R01-092 | R01-113 | R01-138 |
| R01-006 | R01-042 | R01-068 | R01-093 | R01-114 | R01-139 |
| R01-007 | R01-043 | R01-070 | R01-094 | R01-115 | R01-141 |
| R01-013 | R01-045 | R01-071 | R01-095 | R01-116 | R01-142 |
| R01-017 | R01-046 | R01-072 | R01-096 | R01-117 | R01-144 |
| R01-018 | R01-047 | R01-073 | R01-097 | R01-118 | R01-145 |
| R01-019 | R01-048 | R01-074 | R01-098 | R01-119 | R01-146 |
| R01-021 | R01-049 | R01-075 | R01-100 | R01-120 |  |
| R01-023 | R01-050 | R01-077 | R01-101 | R01-121 |  |
| R01-025 | R01-051 | R01-079 | R01-102 | R01-122 |  |
| R01-029 | R01-052 | R01-081 | R01-103 | R01-123 |  |
| R01-030 | R01-055 | R01-082 | R01-104 | R01-124 |  |
| R01-031 | R01-056 | R01-083 | R01-105 | R01-125 |  |
| R01-033 | R01-060 | R01-084 | R01-106 | R01-128 |  |
| R01-034 | R01-061 | R01-085 | R01-107 | R01-129 |  |

Supplementary Table 5. Radiomics features calculated by using pyradiomics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| G11 | Index | Features | Index | Features |
| 1 | shape\_Elongation | 8 | shape\_Maximum3DDiameter |
| 2 | shape\_Flatness | 9 | shape\_MeshVolume |
| 3 | shape\_LeastAxisLength | 10 | shape\_MinorAxisLength |
| 4 | shape\_MajorAxisLength | 11 | shape\_Sphericity |
| 5 | shape\_Maximum2DDiameterColumn | 12 | shape\_SurfaceArea |
| 6 | shape\_Maximum2DDiameterRow | 13 | shape\_SurfaceVolumeRatio |
| 7 | shape\_Maximum2DDiameterSlice |  |  |
|  | Index | Features | Index | Features |
| G22 | 1 | firstorder\_10Percentile | 46 | glrlm\_LongRunLowGrayLevelEmphasis |
| 2 | firstorder\_90Percentile | 47 | glrlm\_LowGrayLevelRunEmphasis |
| 3 | firstorder\_Energy | 48 | glrlm\_RunEntropy |
| 4 | firstorder\_Entropy | 49 | glrlm\_RunLengthNonUniformity |
| 5 | firstorder\_InterquartileRange | 50 | glrlm\_RunLengthNonUniformityNormalized |
| 6 | firstorder\_Kurtosis | 51 | glrlm\_RunPercentage |
| 7 | firstorder\_Maximum | 52 | glrlm\_RunVariance |
| 8 | firstorder\_Mean | 53 | glrlm\_ShortRunEmphasis |
| 9 | firstorder\_MeanAbsoluteDeviation | 54 | glrlm\_ShortRunHighGrayLevelEmphasis |
| 10 | firstorder\_Median | 55 | glrlm\_ShortRunLowGrayLevelEmphasis |
| 11 | firstorder\_Minimum | 56 | glszm\_GrayLevelNonUniformity |
| 12 | firstorder\_Range | 57 | glszm\_GrayLevelNonUniformityNormalized |
| 13 | firstorder\_RobustMeanAbsoluteDeviation | 58 | glszm\_GrayLevelVariance |
| 14 | firstorder\_RootMeanSquared | 59 | glszm\_HighGrayLevelZoneEmphasis |
| 15 | firstorder\_Skewness | 60 | glszm\_LargeAreaEmphasis |
| 16 | firstorder\_Uniformity | 61 | glszm\_LargeAreaHighGrayLevelEmphasis |
| 17 | firstorder\_Variance | 62 | glszm\_LargeAreaLowGrayLevelEmphasis |
| 18 | glcm\_Autocorrelation | 63 | glszm\_LowGrayLevelZoneEmphasis |
| 19 | glcm\_JointAverage | 64 | glszm\_SizeZoneNonUniformity |
| 20 | glcm\_ClusterProminence | 65 | glszm\_SizeZoneNonUniformityNormalized |
| 21 | glcm\_ClusterShade | 66 | glszm\_SmallAreaEmphasis |
| 22 | glcm\_ClusterTendency | 67 | glszm\_SmallAreaHighGrayLevelEmphasis |
| 23 | glcm\_Contrast | 68 | glszm\_SmallAreaLowGrayLevelEmphasis |
| 24 | glcm\_Correlation | 69 | glszm\_ZoneEntropy |
| 25 | glcm\_DifferenceAverage | 70 | glszm\_ZonePercentage |
| 26 | glcm\_DifferenceEntropy | 71 | glszm\_ZoneVariance |
| 27 | glcm\_DifferenceVariance | 72 | gldm\_DependenceEntropy |
| 28 | glcm\_JointEnergy | 73 | gldm\_DependenceNonUniformity |
| 29 | glcm\_JointEntropy | 74 | gldm\_DependenceNonUniformityNormalized |
| 30 | glcm\_Imc1 | 75 | gldm\_DependenceVariance |
| 31 | glcm\_Imc2 | 76 | gldm\_GrayLevelNonUniformity |
| 32 | glcm\_Idm | 77 | gldm\_GrayLevelVariance |
| 33 | glcm\_Idmn | 78 | gldm\_HighGrayLevelEmphasis |
| 34 | glcm\_Id | 79 | gldm\_LargeDependenceEmphasis |
| 35 | glcm\_Idn | 80 | gldm\_LargeDependenceHighGrayLevelEmphasis |
| 36 | glcm\_InverseVariance | 81 | gldm\_LargeDependenceLowGrayLevelEmphasis |
| 37 | glcm\_MaximumProbability | 82 | gldm\_LowGrayLevelEmphasis |
| 38 | glcm\_SumEntropy | 83 | gldm\_SmallDependenceEmphasis |
| 39 | glcm\_SumSquares | 84 | gldm\_SmallDependenceHighGrayLevelEmphasis |
| 40 | glrlm\_GrayLevelNonUniformity | 85 | gldm\_SmallDependenceLowGrayLevelEmphasis |
| 41 | glrlm\_GrayLevelNonUniformityNormalized | 86 | ngtdm\_Busyness |
| 42 | glrlm\_GrayLevelVariance | 87 | ngtdm\_Coarseness |
| 43 | glrlm\_HighGrayLevelRunEmphasis | 88 | ngtdm\_Complexity |
| 44 | glrlm\_LongRunEmphasis | 89 | ngtdm\_Contrast |
| 45 | glrlm\_LongRunHighGrayLevelEmphasis | 90 | ngtdm\_Strength |

1. Feature Groups 1: shape features; 2. Feature Group 2: intensity histogram (first-order) features and textural (Haralick) features

Supplementary Table 6. RMSE, content loss and ratio of poor, medium, and good reproducibility radiomic features for images denoised by the Cycle GAN trained for different numbers of epochs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Training length  Noisy images | 25 Epochs | 50 Epochs | 75 Epochs | 100 Epochs |
| Low-noise Images | | | | |
| RMSE | 0.0178 | 0.0169 | 0.0172 | 0.0170 |
| Content loss | 0.0229 | 0.216 | 0.0217 | 0.0216 |
| CCCs ≥ 0.85 | 81% | 84% | 82% | 84% |
| 0.65≤CCCs<0.85 | 18% | 16% | 18% | 16% |
| CCCs<0.65 | 1% | 0% | 0% | 0% |
| High-noise Images | | | | |
| RMSE | 0.0193 | 0.0177 | 0.0175 | 0.0181 |
| Content loss | 0.256 | 0.0241 | 0.0248 | 0.0245 |
| CCCs > 0.85 | 77% | 86% | 86% | 86% |
| 0.65≤CCCs<0.85 | 22% | 14% | 14% | 14% |
| CCCs<0.65 | 1% | 0% | 0% | 0% |

Supplementary Table 7. RMSE, content loss and ratio of poor, medium, and good reproducibility radiomic features for images denoised by the Cycle GAN trained without training strategy for different numbers of epochs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Training length  Noisy images | 25 Epochs | 50 Epochs | 75 Epochs | 100 Epochs |
| Low-noise Images | | | | |
| RMSE | 0.0199 | 0.0167 | 0.167 | 0.0167 |
| Content loss | 0.0830 | 0.0258 | 0.0258 | 0.0258 |
| CCCs ≥ 0.85 | 61% | 86% | 86% | 86% |
| 0.65≤CCCs<0.85 | 36% | 13% | 13% | 13% |
| CCCs<0.65 | 4% | 1% | 1% | 1% |
| High-noise Images | | | | |
| RMSE | 0.0201 | 0.0188 | 0.0188 | 0.0188 |
| Content loss | 0.0877 | 0.0263 | 0.0263 | 0.0263 |
| CCCs > 0.85 | 74% | 86% | 87% | 84% |
| 0.65≤CCCs<0.85 | 25% | 11% | 11% | 12% |
| CCCs<0.65 | 1% | 3% | 2% | 3% |

Supplementary Table 8. Importance of features in 4-year pre-treatment survival predication models (Original Radiomics features)

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | Features | Index | Features |
| 1 | glszm\_LargeAreaLowGrayLevelEmphasis | 53 | firstorder\_InterquartileRange |
| 2 | ngtdm\_Coarseness | 54 | glcm\_Correlation |
| 3 | gldm\_GrayLevelVariance | 55 | glcm\_Idn |
| 4 | firstorder\_Entropy | 56 | glszm\_ZonePercentage |
| 5 | shape\_MinorAxisLength | 57 | gldm\_SmallDependenceEmphasis |
| 6 | glrlm\_GrayLevelNonUniformityNormalized | 58 | shape\_Maximum3DDiameter |
| 7 | glszm\_LargeAreaHighGrayLevelEmphasis | 59 | firstorder\_Skewness |
| 8 | glcm\_JointEntropy | 60 | glszm\_SmallAreaEmphasis |
| 9 | glrlm\_RunLengthNonUniformityNormalized | 61 | glszm\_SizeZoneNonUniformityNormalized |
| 10 | glszm\_LowGrayLevelZoneEmphasis | 62 | firstorder\_RobustMeanAbsoluteDeviation |
| 11 | glszm\_GrayLevelNonUniformityNormalized | 63 | glcm\_Autocorrelation |
| 12 | shape\_SurfaceVolumeRatio | 64 | glcm\_Idmn |
| 13 | glcm\_SumEntropy | 65 | glcm\_ClusterShade |
| 14 | gldm\_LargeDependenceLowGrayLevelEmphasis | 66 | glszm\_GrayLevelVariance |
| 15 | glszm\_SizeZoneNonUniformityNormalized | 67 | glrlm\_LongRunLowGrayLevelEmphasis |
| 16 | glrlm\_ShortRunEmphasis | 68 | shape\_Maximum2DDiameterColumn |
| 17 | gldm\_DependenceNonUniformityNormalized | 69 | glcm\_InverseVariance |
| 18 | gldm\_GrayLevelNonUniformity | 70 | glcm\_JointAverage |
| 19 | gldm\_LargeDependenceHighGrayLevelEmphasis | 71 | firstorder\_RootMeanSquared |
| 20 | shape\_MajorAxisLength | 72 | gldm\_SmallDependenceHighGrayLevelEmphasis |
| 21 | gldm\_LowGrayLevelEmphasis | 73 | ngtdm\_Contrast |
| 22 | glcm\_Idm | 74 | firstorder\_Minimum |
| 23 | glrlm\_LongRunHighGrayLevelEmphasis | 75 | glszm\_SmallAreaLowGrayLevelEmphasis |
| 24 | firstorder\_Maximum | 76 | glrlm\_LowGrayLevelRunEmphasis |
| 25 | shape\_Maximum2DDiameterColumn | 77 | glrlm\_ShortRunHighGrayLevelEmphasis |
| 26 | glcm\_Imc2 | 78 | firstorder\_MeanAbsoluteDeviation |
| 27 | shape\_Maximum2DDiameterSlice | 79 | glszm\_SmallAreaHighGrayLevelEmphasis |
| 28 | shape\_Sphericity | 80 | ngtdm\_Complexity |
| 29 | gldm\_DependenceEntropy | 81 | glrlm\_GrayLevelVariance |
| 30 | shape\_Elongation | 82 | shape\_SurfaceVolumeRatio |
| 31 | glrlm\_RunVariance | 83 | glrlm\_RunEntropy |
| 32 | glcm\_Imc1 | 84 | glszm\_HighGrayLevelZoneEmphasis |
| 33 | glrlm\_LongRunLowGrayLevelEmphasis | 85 | gldm\_DependenceEntropy |
| 34 | firstorder\_Variance | 86 | gldm\_SmallDependenceLowGrayLevelEmphasis |
| 35 | glrlm\_ShortRunHighGrayLevelEmphasis | 87 | glrlm\_ShortRunLowGrayLevelEmphasis |
| 36 | glrlm\_RunPercentage | 88 | gldm\_HighGrayLevelEmphasis |
| 37 | glrlm\_GrayLevelVariance | 89 | gldm\_DependenceNonUniformityNormalized |
| 38 | firstorder\_Mean | 90 | glcm\_DifferenceVariance |
| 39 | firstorder\_Minimum | 91 | firstorder\_Variance |
| 40 | glcm\_SumSquares | 92 | firstorder\_10Percentile |
| 41 | glszm\_ZonePercentage | 93 | shape\_Flatness |
| 42 | shape\_Flatness | 94 | gldm\_LowGrayLevelEmphasis |
| 43 | shape\_Maximum3DDiameter | 95 | glcm\_ClusterTendency |
| 44 | glcm\_ClusterShade | 96 | firstorder\_90Percentile |
| 45 | firstorder\_InterquartileRange | 97 | glszm\_LowGrayLevelZoneEmphasis |
| 46 | glcm\_DifferenceVariance | 98 | glcm\_DifferenceAverage |
| 47 | firstorder\_RobustMeanAbsoluteDeviation | 99 | gldm\_GrayLevelVariance |
| 48 | glcm\_Idmn | 100 | glszm\_GrayLevelNonUniformityNormalized |
| 49 | glcm\_Idn | 101 | glrlm\_HighGrayLevelRunEmphasis |
| 50 | ngtdm\_Complexity | 102 | glcm\_Contrast |
| 51 | Age | 103 | glcm\_SumSquares |
| 52 | shape\_SurfaceArea | 104 | firstorder\_InterquartileRange |

Supplementary Table 9. Importance of features in 4-year pre-treatment survival predication models (EDNs De-noised Radiomics features)

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | Features | Index | Features |
| 1 | glszm\_LargeAreaLowGrayLevelEmphasis | 53 | firstorder\_Uniformity |
| 2 | gldm\_GrayLevelVariance | 54 | shape\_SurfaceArea |
| 3 | glrlm\_GrayLevelNonUniformityNormalized | 55 | glszm\_HighGrayLevelZoneEmphasis |
| 4 | gldm\_LargeDependenceHighGrayLevelEmphasis | 56 | glrlm\_ShortRunLowGrayLevelEmphasis |
| 5 | gldm\_GrayLevelNonUniformity | 57 | glszm\_GrayLevelVariance |
| 6 | firstorder\_Entropy | 58 | glszm\_LargeAreaEmphasis |
| 7 | glcm\_JointEntropy | 59 | glcm\_Imc2 |
| 8 | ngtdm\_Coarseness | 60 | firstorder\_Mean |
| 9 | glcm\_SumEntropy | 61 | glcm\_ClusterTendency |
| 10 | glrlm\_ShortRunEmphasis | 62 | glcm\_Contrast |
| 11 | shape\_MinorAxisLength | 63 | glcm\_DifferenceEntropy |
| 12 | gldm\_DependenceEntropy | 64 | firstorder\_Median |
| 13 | glszm\_LargeAreaHighGrayLevelEmphasis | 65 | firstorder\_RobustMeanAbsoluteDeviation |
| 14 | glrlm\_LongRunHighGrayLevelEmphasis | 66 | glszm\_SmallAreaLowGrayLevelEmphasis |
| 15 | shape\_SurfaceVolumeRatio | 67 | firstorder\_10Percentile |
| 16 | glszm\_GrayLevelNonUniformityNormalized | 68 | glcm\_MaximumProbability |
| 17 | gldm\_LargeDependenceLowGrayLevelEmphasis | 69 | shape\_Maximum2DDiameterRow |
| 18 | gldm\_DependenceVariance | 70 | glcm\_JointEnergy |
| 19 | glcm\_Idm | 71 | glrlm\_LongRunEmphasis |
| 20 | glrlm\_RunLengthNonUniformityNormalized | 72 | shape\_LeastAxisLength |
| 21 | glszm\_LowGrayLevelZoneEmphasis | 73 | gldm\_SmallDependenceEmphasis |
| 22 | glrlm\_RunVariance | 74 | shape\_Maximum3DDiameter |
| 23 | firstorder\_Variance | 75 | gldm\_LargeDependenceEmphasis |
| 24 | shape\_MajorAxisLength | 76 | glszm\_ZonePercentage |
| 25 | shape\_Elongation | 77 | firstorder\_90Percentile |
| 26 | gldm\_LowGrayLevelEmphasis | 78 | glcm\_InverseVariance |
| 27 | glszm\_SmallAreaEmphasis | 79 | gldm\_SmallDependenceHighGrayLevelEmphasis |
| 28 | glrlm\_ShortRunHighGrayLevelEmphasis | 80 | firstorder\_Energy |
| 29 | glszm\_SizeZoneNonUniformityNormalized | 81 | glrlm\_RunEntropy |
| 30 | shape\_Maximum2DDiameterSlice | 82 | glcm\_DifferenceAverage |
| 31 | glszm\_ZoneEntropy | 83 | firstorder\_MeanAbsoluteDeviation |
| 32 | firstorder\_Maximum | 84 | glszm\_GrayLevelNonUniformity |
| 33 | gldm\_DependenceNonUniformityNormalized | 85 | ngtdm\_Busyness |
| 34 | glrlm\_GrayLevelVariance | 86 | glrlm\_LowGrayLevelRunEmphasis |
| 35 | glrlm\_RunPercentage | 87 | glcm\_Id |
| 36 | shape\_Maximum2DDiameterColumn | 88 | firstorder\_Skewness |
| 37 | glcm\_Imc1 | 89 | glrlm\_GrayLevelNonUniformity |
| 38 | shape\_Sphericity | 90 | glcm\_ClusterProminence |
| 39 | ngtdm\_Contrast | 91 | glcm\_Autocorrelation |
| 40 | glszm\_SmallAreaHighGrayLevelEmphasis | 92 | gldm\_HighGrayLevelEmphasis |
| 41 | firstorder\_InterquartileRange | 93 | Age |
| 42 | glcm\_Idmn | 94 | glszm\_ZoneVariance |
| 43 | glszm\_SizeZoneNonUniformity | 95 | glcm\_JointAverage |
| 44 | ngtdm\_Complexity | 96 | glcm\_ClusterShade |
| 45 | firstorder\_Kurtosis | 97 | glrlm\_RunLengthNonUniformity |
| 46 | glcm\_SumSquares | 98 | ngtdm\_Strength |
| 47 | glrlm\_LongRunLowGrayLevelEmphasis | 99 | glrlm\_HighGrayLevelRunEmphasis |
| 48 | glcm\_Idn | 100 | gldm\_DependenceNonUniformity |
| 49 | glcm\_DifferenceVariance | 101 | gldm\_SmallDependenceLowGrayLevelEmphasis |
| 50 | firstorder\_RootMeanSquared | 102 | shape\_MeshVolume |
| 51 | glcm\_Correlation | 103 | firstorder\_Range |
| 52 | shape\_Flatness | 104 | firstorder\_Uniformity |

Supplementary Table 10. Importance of features in 4-year pre-treatment survival predication models (CGAN De-noised Radiomics features)

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | Features | Index | Features |
| 1 | glszm\_LargeAreaLowGrayLevelEmphasis | 53 | glcm\_DifferenceAverage |
| 2 | glrlm\_GrayLevelNonUniformityNormalized | 54 | firstorder\_Uniformity |
| 3 | gldm\_GrayLevelVariance | 55 | glcm\_DifferenceVariance |
| 4 | firstorder\_Entropy | 56 | glszm\_SmallAreaLowGrayLevelEmphasis |
| 5 | gldm\_GrayLevelNonUniformity | 57 | glrlm\_LongRunLowGrayLevelEmphasis |
| 6 | ngtdm\_Coarseness | 58 | glcm\_MaximumProbability |
| 7 | glcm\_JointEntropy | 59 | ngtdm\_Strength |
| 8 | shape\_MinorAxisLength | 60 | glcm\_JointAverage |
| 9 | gldm\_DependenceEntropy | 61 | glcm\_Correlation |
| 10 | glszm\_LargeAreaHighGrayLevelEmphasis | 62 | gldm\_HighGrayLevelEmphasis |
| 11 | gldm\_LargeDependenceHighGrayLevelEmphasis | 63 | firstorder\_Range |
| 12 | glszm\_LowGrayLevelZoneEmphasis | 64 | glszm\_ZonePercentage |
| 13 | glrlm\_RunLengthNonUniformityNormalized | 65 | shape\_Maximum3DDiameter |
| 14 | glszm\_GrayLevelNonUniformityNormalized | 66 | shape\_Maximum2DDiameterRow |
| 15 | glrlm\_ShortRunEmphasis | 67 | glcm\_Autocorrelation |
| 16 | glcm\_SumEntropy | 68 | glcm\_ClusterProminence |
| 17 | shape\_SurfaceVolumeRatio | 69 | glszm\_SizeZoneNonUniformity |
| 18 | glrlm\_LongRunHighGrayLevelEmphasis | 70 | gldm\_SmallDependenceHighGrayLevelEmphasis |
| 19 | gldm\_LargeDependenceLowGrayLevelEmphasis | 71 | firstorder\_RobustMeanAbsoluteDeviation |
| 20 | gldm\_DependenceVariance | 72 | glcm\_DifferenceEntropy |
| 21 | glcm\_Idm | 73 | firstorder\_RootMeanSquared |
| 22 | firstorder\_Variance | 74 | glszm\_GrayLevelVariance |
| 23 | glrlm\_RunVariance | 75 | firstorder\_90Percentile |
| 24 | gldm\_LowGrayLevelEmphasis | 76 | glrlm\_GrayLevelNonUniformity |
| 25 | glszm\_SizeZoneNonUniformityNormalized | 77 | gldm\_SmallDependenceLowGrayLevelEmphasis |
| 26 | shape\_MajorAxisLength | 78 | glszm\_ZoneEntropy |
| 27 | glrlm\_ShortRunHighGrayLevelEmphasis | 79 | glrlm\_RunLengthNonUniformity |
| 28 | shape\_Elongation | 80 | glcm\_ClusterTendency |
| 29 | gldm\_DependenceNonUniformityNormalized | 81 | firstorder\_Median |
| 30 | glcm\_Imc1 | 82 | gldm\_DependenceNonUniformity |
| 31 | shape\_Maximum2DDiameterSlice | 83 | glcm\_InverseVariance |
| 32 | shape\_Maximum2DDiameterColumn | 84 | firstorder\_Skewness |
| 33 | glrlm\_RunPercentage | 85 | glszm\_ZoneVariance |
| 34 | glszm\_HighGrayLevelZoneEmphasis | 86 | firstorder\_10Percentile |
| 35 | glrlm\_GrayLevelVariance | 87 | shape\_MeshVolume |
| 36 | glrlm\_ShortRunLowGrayLevelEmphasis | 88 | firstorder\_Kurtosis |
| 37 | firstorder\_Maximum | 89 | ngtdm\_Contrast |
| 38 | shape\_Flatness | 90 | glcm\_JointEnergy |
| 39 | shape\_Sphericity | 91 | glszm\_GrayLevelNonUniformity |
| 40 | ngtdm\_Complexity | 92 | firstorder\_Mean |
| 41 | glcm\_Imc2 | 93 | firstorder\_MeanAbsoluteDeviation |
| 42 | glszm\_SmallAreaEmphasis | 94 | firstorder\_Minimum |
| 43 | glcm\_Idmn | 95 | shape\_LeastAxisLength |
| 44 | glrlm\_LongRunEmphasis | 96 | Age |
| 45 | glcm\_ClusterShade | 97 | glcm\_Contrast |
| 46 | glszm\_SmallAreaHighGrayLevelEmphasis | 98 | gldm\_SmallDependenceEmphasis |
| 47 | glszm\_LargeAreaEmphasis | 99 | glrlm\_RunEntropy |
| 48 | glcm\_Idn | 100 | ngtdm\_Busyness |
| 49 | gldm\_LargeDependenceEmphasis | 101 | glrlm\_LowGrayLevelRunEmphasis |
| 50 | firstorder\_InterquartileRange | 102 | firstorder\_Energy |
| 51 | glrlm\_HighGrayLevelRunEmphasis | 103 | glcm\_Id |
| 52 | glcm\_SumSquares | 104 | glcm\_DifferenceAverage |

Supplementary Table 11. Importance of features in 4-year pre-treatment survival predication models (Cycle-GAN De-noised Radiomics features)

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | Features | Index | Features |
| 1 | glrlm\_GrayLevelNonUniformityNormalized | 53 | shape\_Flatness |
| 2 | glszm\_LargeAreaLowGrayLevelEmphasis | 54 | shape\_LeastAxisLength |
| 3 | gldm\_GrayLevelVariance | 55 | shape\_SurfaceArea |
| 4 | firstorder\_Entropy | 56 | firstorder\_Energy |
| 5 | shape\_MinorAxisLength | 57 | glcm\_Correlation |
| 6 | ngtdm\_Coarseness | 58 | firstorder\_10Percentile |
| 7 | glcm\_JointEntropy | 59 | firstorder\_MeanAbsoluteDeviation |
| 8 | glrlm\_RunLengthNonUniformityNormalized | 60 | Age |
| 9 | gldm\_DependenceEntropy | 61 | shape\_Maximum2DDiameterRow |
| 10 | glszm\_LargeAreaHighGrayLevelEmphasis | 62 | glszm\_HighGrayLevelZoneEmphasis |
| 11 | glszm\_LowGrayLevelZoneEmphasis | 63 | firstorder\_RobustMeanAbsoluteDeviation |
| 12 | glszm\_GrayLevelNonUniformityNormalized | 64 | firstorder\_Range |
| 13 | glszm\_SizeZoneNonUniformityNormalized | 65 | glszm\_LargeAreaEmphasis |
| 14 | gldm\_LargeDependenceHighGrayLevelEmphasis | 66 | gldm\_DependenceNonUniformity |
| 15 | shape\_SurfaceVolumeRatio | 67 | glcm\_ClusterTendency |
| 16 | glszm\_ZoneEntropy | 68 | glcm\_JointEnergy |
| 17 | glrlm\_ShortRunEmphasis | 69 | firstorder\_Skewness |
| 18 | gldm\_GrayLevelNonUniformity | 70 | firstorder\_RootMeanSquared |
| 19 | shape\_Elongation | 71 | firstorder\_Median |
| 20 | gldm\_DependenceNonUniformityNormalized | 72 | glrlm\_ShortRunLowGrayLevelEmphasis |
| 21 | glrlm\_LongRunHighGrayLevelEmphasis | 73 | firstorder\_Kurtosis |
| 22 | glcm\_SumEntropy | 74 | glcm\_MaximumProbability |
| 23 | shape\_MajorAxisLength | 75 | glcm\_ClusterShade |
| 24 | glcm\_Idm | 76 | glcm\_ClusterProminence |
| 25 | firstorder\_Variance | 77 | glrlm\_RunLengthNonUniformity |
| 26 | glszm\_SizeZoneNonUniformity | 78 | glszm\_ZonePercentage |
| 27 | gldm\_LargeDependenceLowGrayLevelEmphasis | 79 | firstorder\_90Percentile |
| 28 | firstorder\_Maximum | 80 | glszm\_SmallAreaLowGrayLevelEmphasis |
| 29 | shape\_Maximum2DDiameterSlice | 81 | glszm\_ZoneVariance |
| 30 | glrlm\_RunVariance | 82 | gldm\_SmallDependenceLowGrayLevelEmphasis |
| 31 | glrlm\_ShortRunHighGrayLevelEmphasis | 83 | glcm\_InverseVariance |
| 32 | gldm\_DependenceVariance | 84 | gldm\_SmallDependenceEmphasis |
| 33 | shape\_Maximum2DDiameterColumn | 85 | glszm\_GrayLevelNonUniformity |
| 34 | glcm\_Imc1 | 86 | glcm\_JointAverage |
| 35 | shape\_Sphericity | 87 | glrlm\_RunEntropy |
| 36 | gldm\_LowGrayLevelEmphasis | 88 | glrlm\_LongRunEmphasis |
| 37 | glrlm\_RunPercentage | 89 | shape\_Maximum3DDiameter |
| 38 | glrlm\_GrayLevelVariance | 90 | glcm\_DifferenceAverage |
| 39 | glcm\_Id | 91 | ngtdm\_Busyness |
| 40 | glcm\_Idmn | 92 | glrlm\_HighGrayLevelRunEmphasis |
| 41 | glszm\_SmallAreaEmphasis | 93 | gldm\_LargeDependenceEmphasis |
| 42 | glrlm\_LongRunLowGrayLevelEmphasis | 94 | glcm\_Contrast |
| 43 | firstorder\_Uniformity | 95 | gldm\_HighGrayLevelEmphasis |
| 44 | firstorder\_Minimum | 96 | glcm\_Autocorrelation |
| 45 | glszm\_SmallAreaHighGrayLevelEmphasis | 97 | glszm\_GrayLevelVariance |
| 46 | glrlm\_LowGrayLevelRunEmphasis | 98 | shape\_MeshVolume |
| 47 | firstorder\_InterquartileRange | 99 | glcm\_DifferenceEntropy |
| 48 | ngtdm\_Complexity | 100 | glcm\_Imc2 |
| 49 | glcm\_SumSquares | 101 | ngtdm\_Contrast |
| 50 | glcm\_Idn | 102 | glrlm\_GrayLevelNonUniformity |
| 51 | ngtdm\_Strength | 103 | gldm\_SmallDependenceHighGrayLevelEmphasis |
| 52 | firstorder\_Mean | 104 | shape\_Flatness |

Supplementary Method 1: Noise

The high quality NSCLC-Radiomics collection [1] (hereafter called LUNG 1), which contains CT scans of 422 non-small cell lung cancer (NSCLC) patients, as our experimental dataset. These CT scans included annotations drawn by specialist radiation oncologists that delineate a region of interest (ROI), the gross tumor volume. ROIs were necessary to be able to compute radiomic features. The CT images for which the dose level (‘parameter exposure’ in DICOM metadata) was missing (n=200) were excluded from further analyses. We considered CT images scanned at 400 milliampere-seconds (mAs) and above as full dose CT (n=157, the index of LUNG 1 patients included in the experiments can be found in Supplementary Table 1, supplementary materials are available: https://gitlab.com/UM-CDS/low-dose-ct-denoising/-/branches). These data were be used for training (n=40, 4260 frames) and testing (n=117, 13423 frames). Conversely, we designated CT images scanned at 50 mAs as low dose CT, taking the same definition as a prior Low Dose CT Grand Challenge [2][3].

As mentioned, training of EDNs and CGANs require paired images, in our case, pairs of matching low dose and full dose CT scans. However, LUNG 1 contains no paired images, thus we simulated the noisy degradation present in low dose CT images by introducing noise using the method proposed in literature [2][3]. In these, the authors had mimicked CT scanners’ behavior by adding noise with a normal distribution into a sinogram (by Radon transform) and reconstructed the CT image from the modified sinogram to obtain simulated noisy images. We used a similar method to add noise in the original sinogram as follows:

(1)

where is the measurement along the -th ray path; is the read-out error; represents the original line integral of attenuation coefficients along the -th ray path; and, is the black scanner factor, which follows a normal distribution. The intensity of noise added to the image can be controlled through the parameter .

To simulate low dose CT images (scanned with 50mAs) from full dose CT images (scanned with 400 mAs), we first measured the noise intensity introduced in images with lower doses by scanning a Gammex 467 CT phantom (Middleton, WI, USA) using a Philips Brilliance Big Bore CT at two dose levels (400 mAs and 50 mAs)[4]. The signal-to-noise ratio (SNR) of the real phantom dataset was 19.7 dB (95%CI [17.8, 21.6]). We thus estimated that a σ value of 0.0035 best estimated the noise in 50 mAs CT images when generated from 400 mAs images. The SNR in the simulated low-noise images was 18.3 (95%CI, [16.9, 20.1]) dB, close to the real value. To assess the reproducibility of radiomic features with noise of different intensities, we added stronger noise (25 times noise power) by setting σ to 0.0068 to mimic CT images with stronger noise (referred to as simulated high-noise images hereafter). The SNR in the simulated high-noise images had thus reduced to 6.0 (95%CI, [5.9,6.1]) dB. Additionally, extraneous noise introduced by the Radon transform and inverse Radon transform was filtered from the simulated images. A comparison of noise in simulated images and in real phantom scans is shown in following Figure 1, the intensity of noise in real phantom is 17.1 dB and average noise power spectra density within whole image is 45.8 W/Hz. The intensity of noise in simulated low-noise images is 19.4 dB and average noise power spectra density within whole image is 3.6 W/Hz, intensity of noise in simulated high-noise images is 6.1 dB and average noise power spectra density within whole image is 6.0 W/Hz.

A picture containing timeline

Description automatically generated

Figure 1.A comparison of noise in simulated images and in a real phantom image. (a) real phantom image; (b) simulated low-noise image; (c) simulated high-noise image.

**Reference**

1. Aerts, H. J. W. L., Wee, L., Rios Velazquez, E., Leijenaar, R. T. H., Parmar, C., Grossmann, P., … Lambin, P. (2019). Data From NSCLC-Radiomics [Data set]. The Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2015.PF0M9REI>
2. Chen, Hu, et al. "Low-dose CT with a residual encoder-decoder convolutional neural network." IEEE transactions on medical imaging 36.12 (2017): 2524-2535.
3. McCollough, Cynthia H., et al. "Low‐dose CT for the detection and classification of metastatic liver lesions: Results of the 2016 Low Dose CT Grand Challenge." Medical physics 44.10 (2017): e339-e352.
4. Zhovannik, Ivan, et al. "Learning from scanners: Bias reduction and feature correction in radiomics." Clinical and translational radiation oncology 19 (2019): 33-38.