Supplementary Figure 1. The distribution of radiation exposure for selected samples in the (a) NSCLC Radiogenomics and (b) LIDC-IDRI datasets.

Chart, histogram, waterfall chart

Description automatically generated

Supplementary Figure 2. The architecture of of our multiple instance learning network

Diagram

Description automatically generated

The parameter settings about this network

|  |  |  |  |
| --- | --- | --- | --- |
| Layer | Type | Input Size | Output size |
| 1 | Input | -- | 103\*12 |
| 2 | FC + LReLU(Drop=0.5) | 103\*12 | 256\*12 |
| 3 | FC + LReLU(Drop=0.5) | 256\*12 | 128\*12 |
| 4 | FC + LReLU(Drop=0.5) | 128\*12 | 64\*12 |
| 5 | Attention Layer | 64\*12 | 64\*1 |
| 6 | FC+ LReLU | 64\*1 | 32\*1 |
| 7 | Output+ softmax | 32\*1 | 1\*1 |

FC means Full connected layer; LReLU means the activation function of this layer is LReLU; Drop means dropout rate of this layer is 0.5; softmax means the activation function of output layer is softmax.

Supplementary Table 1. Index of available patients for Lung cancer diagnosis in LIDC-IDRI

|  |  |  |  |
| --- | --- | --- | --- |
| LIDC-IDRI-0068 | LIDC-IDRI-0072 | LIDC-IDRI-0088 | LIDC-IDRI-0090 |
| LIDC-IDRI-0091 | LIDC-IDRI-0118 | LIDC-IDRI-0124 | LIDC-IDRI-0129 |
| LIDC-IDRI-0135 | LIDC-IDRI-0137 | LIDC-IDRI-0138 | LIDC-IDRI-0149 |
| LIDC-IDRI-0159 | LIDC-IDRI-0161 | LIDC-IDRI-0162 | LIDC-IDRI-0163 |
| LIDC-IDRI-0164 | LIDC-IDRI-0165 | LIDC-IDRI-0166 | LIDC-IDRI-0167 |
| LIDC-IDRI-0168 | LIDC-IDRI-0169 | LIDC-IDRI-0171 | LIDC-IDRI-0175 |
| LIDC-IDRI-0179 | LIDC-IDRI-0180 | LIDC-IDRI-0181 | LIDC-IDRI-0182 |
| LIDC-IDRI-0183 | LIDC-IDRI-0184 | LIDC-IDRI-0185 | LIDC-IDRI-0186 |
| LIDC-IDRI-0187 | LIDC-IDRI-0188 | LIDC-IDRI-0190 | LIDC-IDRI-0191 |
| LIDC-IDRI-0192 | LIDC-IDRI-0193 | LIDC-IDRI-0194 | LIDC-IDRI-0200 |
| LIDC-IDRI-0203 | LIDC-IDRI-0207 | LIDC-IDRI-0210 | LIDC-IDRI-0220 |
| LIDC-IDRI-0223 | LIDC-IDRI-0233 | LIDC-IDRI-0234 | LIDC-IDRI-0236 |
| LIDC-IDRI-0237 | LIDC-IDRI-0242 | LIDC-IDRI-0244 | LIDC-IDRI-0246 |
| LIDC-IDRI-0247 | LIDC-IDRI-0249 | LIDC-IDRI-0250 | LIDC-IDRI-0252 |
| LIDC-IDRI-0254 | LIDC-IDRI-0255 | LIDC-IDRI-0256 | LIDC-IDRI-0257 |
| LIDC-IDRI-0258 | LIDC-IDRI-0260 | LIDC-IDRI-0264 | LIDC-IDRI-0265 |
| LIDC-IDRI-0266 | LIDC-IDRI-0267 | LIDC-IDRI-0271 | LIDC-IDRI-0273 |
| LIDC-IDRI-0274 | LIDC-IDRI-0276 | LIDC-IDRI-0277 | LIDC-IDRI-0283 |
| LIDC-IDRI-0285 | LIDC-IDRI-0286 | LIDC-IDRI-0289 | LIDC-IDRI-0290 |
| LIDC-IDRI-0314 | LIDC-IDRI-0325 | LIDC-IDRI-0332 | LIDC-IDRI-0377 |
| LIDC-IDRI-0385 | LIDC-IDRI-0399 | LIDC-IDRI-0405 | LIDC-IDRI-0454 |
| LIDC-IDRI-0470 | LIDC-IDRI-0493 | LIDC-IDRI-0510 | LIDC-IDRI-0543 |
| LIDC-IDRI-0559 | LIDC-IDRI-0562 | LIDC-IDRI-0568 | LIDC-IDRI-0580 |
| LIDC-IDRI-0610 | LIDC-IDRI-0624 | LIDC-IDRI-0766 | LIDC-IDRI-0771 |
| LIDC-IDRI-0772 | LIDC-IDRI-0811 | LIDC-IDRI-0875 | LIDC-IDRI-0893 |
| LIDC-IDRI-0905 | LIDC-IDRI-0921 | LIDC-IDRI-0924 | LIDC-IDRI-0939 |
| LIDC-IDRI-0965 | LIDC-IDRI-0994 | LIDC-IDRI-1002 | LIDC-IDRI-1004 |
| LIDC-IDRI-1011 | LIDC-IDRI-0211 |  |  |

Supplementary Table 2. 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 3. Radiomics features calculated by using pyradiomics

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

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