Supplementary Material

Figure



**Figure S1.** Proportional diagram of various radiomic features

**Materials and methods**

**Image Preprocessing and Radiomics Feature Extraction**

**Table S1. Detailed classification of radiomics features.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type** |  |  |  |  |  |
| First Order | EnergyTotal EnergyEntropyMinimum | 10th percentile90th percentileMaximumMean | MedianInterquartile RangeRangeMean Absolute Deviation | Robust Mean Absolute DeviatioRoot Mean SquaredStandard Deviation | SkewnessKurtosisVarianceUniformity |
| Shape  | MeshSurfacePixelSurfacePerimeterPerimeterSurfaceRatioSphericitySphericalDisproportion | MaximumDiameterMajorAxisLengthMinorAxisLengthElongationCompactness1Compactness2 | MeshVolumeVoxelVolumeSurfaceAreaSurfaceVolumeRatioSphericity | SphericalDisproportionMaximum3DDiameterMaximum2DDiameterSliceMaximum2DDiameterColumnMaximum2DDiameterRow | MajorAxisLengthMinorAxisLengthLeastAxisLengthElongationFlatness |
| Grey Level Co-occurrence Matrix (GLCM) | AutocorrelationClusterProminenceClusterShadeClusterTendencyContrast | CorrelationDifferenceAverageDifferenceEntropyDifferenceVarianceId | IdmIdmnIdnImc1Imc2 | InverseVarianceJointAverageJointEnergyJointEntropyMaximumProbability | SumEntropySumSquares |
| Gray-level dependence matrix (GLDM) | DependenceEntropyDependenceNonUniformityDependenceNonUniformityNormalizedDependenceVarianceGrayLevelNonUniformity | GrayLevelVarianceHighGrayLevelEmphasisLargeDependenceEmphasisLargeDependenceHighGrayLevelEmphasisLargeDependenceLowGrayLevelEmphasis | LowGrayLevelEmphasisSmallDependenceEmphasisSmallDependenceHighGrayLevelEmphasisSmallDependenceLowGrayLevelEmphasis |  |  |
| Gray-level run length matrix (GLRLM) | GrayLevelNonUniformityGrayLevelNonUniformityNormalizedGrayLevelVarianceHighGrayLevelRunEmphasisLongRunEmphasis | LongRunHighGrayLevelEmphasisLongRunLowGrayLevelEmphasisLowGrayLevelRunEmphasisRunEntropyRunLengthNonUniformity | RunLengthNonUniformityNormalizedRunPercentageRunVarianceShortRunEmphasisShortRunHighGrayLevelEmphasis | ShortRunLowGrayLevelEmphasis |  |
| Gray-level size zone matrix (GLSZM) | GrayLevelNonUniformityGrayLevelNonUniformityNormalizedGrayLevelVarianceHighGrayLevelZoneEmphasisLargeAreaEmphasis | LargeAreaHighGrayLevelEmphasisLargeAreaLowGrayLevelEmphasisLowGrayLevelZoneEmphasisSizeZoneNonUniformitySizeZoneNonUniformityNormalized | SmallAreaEmphasisSmallAreaHighGrayLevelEmphasisSmallAreaLowGrayLevelEmphasisZoneEntropyZonePercentage | ZoneVariance |  |
| Neighboring gray tone difference matrix (NGTDM) | BusynessCoarsenessComplexityContrastStrength |  |  |  |  |
| Wavelet | HHHHLH | HHLHLL | LHHLLH | LHLLLL |  |

Note: Detailed meanings and formulas for each feature can be found at the URL (https://pyradiomics.readthedocs.io/en/latest/

Table S2. Comparison of Performance between LightGBM Model and Other Models

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | LR | SVM | KNN | RandomForest | ExtraTrees | XGBoost | MLP |
| LightGBM | Train | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| test | 0.075 | 0.407 | 0.290 | 0.634 | 0.271 | 0.227 | 0.072 |

**Results**

**Model Establishment**

After feature extraction is completed using Python code, we further use Python code to check for missing values. Through the inspection of the code, no missing values were found. This study extracts 833 radiomics features from each white and gray matter, including 14 shape features, 18 first-order features, 22 Gray-Level Co-occurrence Matrix (GLCM) features, 14 Gray-level dependence matrix (GLDM) features, 16 Gray level size zone matrix (GLSZM) features, 16 Gray level run-length matrix (GLRLM) features, 5 Neighboring gray tone difference matrix (NGTDM) features, and 728 wavelet features.

#code

import pandas as pd

# Load the CSV file into a DataFrame

file\_path = 'your\_file.csv' # Replace with the path to your CSV file

df = pd.read\_csv(file\_path)

# Check for missing values in each column

missing\_values = df.isnull().sum()

# Print the results

print("Missing value check:")

for column, count in missing\_values.items():

 if count > 0:

 print(f"{column} has {count} missing values")

 else:

 print(f"{column} has no missing values")

**Lasso model**

The specific parameters of the Lasso model can be referenced at https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.Lasso.html?highlight=lasso#sklearn.linear\_model.Lasso.

Rad\_score = 0.4845-0.00001 × WM\_original\_glszm\_GrayLevelNonUniformity

-0.0210 × WM\_original\_shape\_LeastAxisLength

+0.0010 ×WM\_wavelet\_HLH\_glszm\_SmallAreaHighGrayLevelEmphasis

+0.0030 × WM\_wavelet\_HLL\_firstorder\_Maximum

+0.0087× WM\_wavelet\_HLL\_firstorder\_Skewness

-0.0021 × WM\_wavelet\_LLH\_firstorder\_RootMeanSquared

+0.0038 × WM\_wavelet\_LLL\_firstorder\_10Percentile

-0.0025 × WM\_wavelet\_LLL\_ngtdm\_Contrast

-0.0094 × GM\_original\_glrlm\_RunEntropy

-0.0003 × GM\_wavelet\_HHH\_glszm\_SizeZoneNonUniformity

-0.0087 ×GM\_wavelet\_HLH\_glrlm\_RunEntropy

-0.0149 ×GM\_wavelet\_HLH\_glrlm\_RunVariance

-0.0141 × GM\_wavelet\_LHH\_glrlm\_RunEntropy

-0.0063 ×GM\_wavelet\_LHH\_glszm\_SizeZoneNonUniformity

-0.0004 ×GM\_wavelet\_LHL\_glrlm\_RunLengthNonUniformity

+0.0201 × GM\_wavelet\_LLL\_glcm\_ClusterShade