

Supplementary

Table S1 Extracted radiomics features

| First order (N=18) | Shape (N=14) | GLDM (N=14) | GLRLM (N=16) | GLSZM (N=16) | GLCM (N=22) |
|--------------------------------|----------------------------|---|--------------------------------------|--------------------------------------|---------------------|
| Interquartile Range | Voxel Volume | Gray Level Variance | Short Run Low Gray Level Emphasis | Gray Level Variance | Joint Average |
| Skewness | Maximum 3D Diameter | High Gray Level Emphasis | Gray Level Variance | Zone Variance | Joint Entropy |
| Uniformity | Mesh Volume | Dependence Entropy | Low Gray Level Run Emphasis | Gray Level Non-Uniformity Normalized | Cluster Shade |
| Median | Major Axis Length | Dependence Non-Uniformity | Gray Level Non-Uniformity Normalized | Size Zone Non-Uniformity Normalized | Maximum Probability |
| Energy | Sphericity | Gray Level Non-Uniformity | Run Variance | Size Zone Non-Uniformity | Idmn |
| Robust Mean Absolute Deviation | Least Axis Length | Small Dependence Emphasis | Gray Level Non-Uniformity | Gray Level Non-Uniformity | Joint Energy |
| Mean Absolute Deviation | Elongation | Small Dependence High Gray Level Emphasis | Long Run Emphasis | Large Area Emphasis | Contrast |
| Total Energy | Surface Volume Ratio | Dependence Non-Uniformity Normalized | Short Run High Gray Level Emphasis | Small Area High Gray Level Emphasis | Difference Entropy |
| Maximum | Maximum 2D Diameter Slice | Large Dependence Emphasis | Run Length Non-Uniformity | Zone Percentage | Inverse Variance |
| Root Mean Squared | Flatness | Large Dependence Low Gray Level Emphasis | Short Run Emphasis | Large Area Low Gray Level Emphasis | Difference Variance |
| 90 Percentile | Surface Area | Dependence Variance | Long Run High Gray Level Emphasis | Large Area High Gray Level Emphasis | Idn |
| Minimum | Minor Axis Length | Large Dependence High Gray Level Emphasis | Run Percentage | High Gray Level Zone Emphasis | Idm |
| Entropy | Maximum 2D Diameter Column | Small Dependence Low Gray Level Emphasis | Long Run Low Gray Level Emphasis | Small Area Emphasis | Correlation |
| Range | Maximum 2D Diameter Row | Low Gray Level Emphasis | Run Entropy | Low Gray Level Zone Emphasis | Autocorrelation |
| Variance | | | High Gray Level Run Emphasis | Zone Entropy | Sum Entropy |
| 10 Percentile | | | Run Length Non-Uniformity Normalized | Small Area Low Gray Level Emphasis | Sum Squares |
| Kurtosis | | | | | Cluster Prominence |
| Mean | | | | | Imc2 |
| | | | | | Imc1 |
| | | | | | Difference Average |
| | | | | | Id |
| | | | | | Cluster Tendency |

Table S2 Histopathological diagnosis of enrolled nodules

| Pathological diagnosis | All nodules (N=1171) | Solid nodules (N=548) | Subsolid nodules (N=623) |
|--------------------------|----------------------|-----------------------|--------------------------|
| Malignant nodules | | | |
| Adenocarcinomas | 741 (97.37) | 223 (92.15) | 518 (99.81) |
| Squamous carcinomas | 15 (1.97) | 14 (5.78) | 1 (0.19) |
| Other types | 5 (0.66) | 5 (2.07) | 0 (0.00) |
| In total | 761 (100.00) | 242 (100.00) | 519 (100.00) |
| Benign nodules | | | |
| Inflammatory nodules | 255 (62.20) | 182 (59.48) | 73 (70.19) |
| Benign tumors | 82 (20.00) | 80 (26.14) | 2 (1.92) |
| Other types | 73 (17.80) | 44 (14.38) | 29 (27.89) |
| In total | 410 (100.00) | 306 (100.00) | 104 (100.00) |

Table S3 Details of missing data for CEA, CYFRA21-1 and NSE

| | CEA | CYFRA21-1 | NSE |
|----------------------------------|------|-----------|-------|
| Malignant nodules (N=761) | | | |
| Effective cases | 562 | 537 | 513 |
| Missing cases | 199 | 224 | 248 |
| Percentage of missing | 0.26 | 0.29 | 0.33 |
| Median, ng/ml | 1.80 | 1.87 | 13.27 |
| Benign nodules (n=410) | | | |
| Effective cases | 241 | 228 | 227 |
| Missing cases | 169 | 182 | 183 |
| Percentage of missing | 0.41 | 0.44 | 0.45 |
| Median, ng/ml | 1.71 | 1.86 | 12.06 |

CEA, carcinoembryonic antigen; CYFRA21-1, cytokeratin 19 fragment; NSE, neuron specific enolase.

Table S4 Selected features for all nodules, nodules ≤10 mm, solid and subsolid nodules.

| Feature | Coefficient | Count |
|--|--------------|-------|
| All nodules (features, n=74) | | |
| texture | -0.1523835 | 1 |
| age | 0.06533589 | 2 |
| wavelet_lll_firstorder_root mean squared | -0.04389031 | 3 |
| spiculation | 0.042131495 | 4 |
| original_shape_major axis length | -0.04090707 | 5 |
| diameter | 0.036909364 | 6 |
| shape | 0.033710454 | 7 |
| history of malignancy | 0.03174518 | 8 |
| log_3mm_glszm_size zone nonuniformity | 0.0308566 | 9 |
| wavelet_hlh_glrjm_long run low gray level emphasis | 0.030578818 | 10 |
| wavelet_hll_glcm_maximum probability | -0.030241398 | 11 |
| wavelet_lhh_glcm_idmn | 0.027672501 | 12 |
| sex | 0.024286829 | 13 |
| log_2mm_glcm_cluster prominence | -0.023833996 | 14 |
| wavelet_hlh_glcm_idmn | -0.022335792 | 15 |
| wavelet_lhh_glcm_correlation | -0.022032931 | 16 |
| family history of lung cancer | 0.02008666 | 17 |
| wavelet_hlh_firstorder_median | 0.0199771 | 18 |
| location | 0.019024547 | 19 |
| neuron specific enolase (NSE) | 0.017194185 | 20 |
| carcinoembryonic antigen (CEA) | 0.0162424 | 21 |
| aprotrombin time | -0.016034134 | 22 |
| log_3mm_glszm_small area emphasis | 0.015090621 | 23 |
| wavelet_hlh_glcm_inverse variance | -0.01478048 | 24 |
| wavelet_hlh_glcm_cluster prominence | -0.013684566 | 25 |
| wavelet_hhh_firstorder_skewness | -0.013468539 | 26 |
| log_4mm_gldm_small dependence high gray level emphasis | 0.013222083 | 27 |
| original_gldm_small dependence low gray level emphasis | 0.011755479 | 28 |
| log_1mm_firstorder_robust mean absolute deviation | -0.01169028 | 29 |
| wavelet_lhh_glrjm_short run low gray level emphasis | -0.011426748 | 30 |
| wavelet_hhh_firstorder_median | -0.011410688 | 31 |
| wavelet_lhh_glszm_large area high gray level emphasis | 0.011297441 | 32 |
| wavelet_lll_glcm_correlation | 0.011281525 | 33 |
| wavelet_lhh_gldm_small dependence low gray level emphasis | -0.011063043 | 34 |
| wavelet_hll_firstorder_skewness | -0.010512063 | 35 |
| blood platelet | 0.009993129 | 36 |
| wavelet_lll_firstorder_skewness | 0.009779177 | 37 |
| wavelet_hlh_firstorder_skewness | 0.009215213 | 38 |
| log_4mm_glszm_size zone nonuniformity normalized | -0.008668284 | 39 |
| wavelet_lhh_firstorder_mean | 0.008643762 | 40 |
| log_5mm_gldm_large dependence high gray level emphasis | 0.008336597 | 41 |
| log_3mm_glcm_imc1 | 0.008259805 | 42 |
| wavelet_lhh_glszm_small area low gray level emphasis | -0.008223074 | 43 |
| log_5mm_glszm_gray level variance | 0.00729779 | 44 |
| log_1mm_gldm_dependence nonuniformity normalized | -0.006937272 | 45 |
| log_4mm_glszm_small area low gray level emphasis | 0.00678647 | 46 |
| log_5mm_glcm_idn | 0.006554504 | 47 |
| prothrombin time | 0.006366947 | 48 |
| log_5mm_glszm_size zone nonuniformity | 0.00632784 | 49 |
| wavelet_lhh_glcm_cluster tendency | -0.005792676 | 50 |
| wavelet_hhh_gldm_small dependence low gray level emphasis | -0.005367769 | 51 |
| wavelet_lhh_glszm_large area high gray level emphasis | -0.005241822 | 52 |
| log_3mm_gldm_large dependence low gray level emphasis | -0.004694483 | 53 |
| log_3mm_gldm_small dependence low gray level emphasis | -0.004678963 | 54 |
| wavelet_hhl_glcm_correlation | -0.004455136 | 55 |
| wavelet_lhh_glcm_joint average | 0.004391401 | 56 |
| log_2mm_glszm_low gray level zone emphasis | -0.004090064 | 57 |
| log_2mm_glcm_cluster shade | 0.003950946 | 58 |
| wavelet_lll_firstorder_10 percentile | -0.003925614 | 59 |
| original_firstorder_kurtosis | 0.003857596 | 60 |
| wavelet_hlh_glcm_correlation | -0.003388095 | 61 |
| log_5mm_glcm_difference variance | 0.002327749 | 62 |
| wavelet_lhh_glcm_cluster shade | -0.002186119 | 63 |
| log_2mm_glszm_large area emphasis | 0.002185869 | 64 |
| wavelet_lhh_firstorder_median | -0.001794971 | 65 |
| log_1mm_glszm_large area low gray level emphasis | 0.000774499 | 66 |
| log_2mm_glszm_small area low gray level emphasis | -0.000730023 | 67 |
| lobulation | 0.000719328 | 68 |
| wavelet_hlh_firstorder_mean | -0.000579703 | 69 |
| log_5mm_glszm_large area low gray level emphasis | 0.000499966 | 70 |
| log_5mm_glszm_zone percentage | 0.00046101 | 71 |
| log_5mm_firstorder_kurtosis | 0.000458805 | 72 |
| log_5mm_firstorder_maximum | 0.000431416 | 73 |
| log_3mm_glszm_small area low gray level emphasis | -0.000344796 | 74 |
| nodules ≤ 10 mm (features, n=67) | | |
| wavelet_lll_firstorder_root mean squared | -0.085387975 | 1 |
| wavelet_lll_firstorder_10 percentile | -0.052760538 | 2 |
| wavelet_lhl_firstorder_skewness | 0.04368638 | 3 |
| wavelet_lll_glcm_autocorrelation | -0.037189778 | 4 |
| wavelet_lhh_glcm_cluster tendency | -0.036917366 | 5 |
| log_4mm_glszm_size zone nonuniformity normalized | -0.03664406 | 6 |
| wavelet_hlh_glcm_idn | -0.03649864 | 7 |
| cytokeratin 19 fragment (cyfra21_1) | 0.03539686 | 8 |
| original_firstorder_skewness | 0.030780079 | 9 |
| neuron specific enolase (NSE) | 0.03052495 | 10 |
| wavelet_lhh_glcm_correlation | -0.029156856 | 11 |
| log_5mm_glszm_large area low gray level emphasis | 0.028894316 | 12 |
| texture | -0.028710542 | 13 |
| diameter | 0.027427517 | 14 |
| spiculation | 0.027001955 | 15 |
| wavelet_lhh_gldm_dependence variance | 0.025864117 | 16 |
| log_4mm_glszm_small area low gray level emphasis | 0.025831908 | 17 |
| wavelet_lhl_firstorder_kurtosis | -0.024995528 | 18 |
| lobulation | 0.02439137 | 19 |
| log_1mm_glcm_idmn | 0.024272965 | 20 |
| carcinoembryonic antigen (CEA) | 0.024164213 | 21 |
| wavelet_hhl_firstorder_mean | 0.023521949 | 22 |
| wavelet_lll_glcm_idmn | 0.023353273 | 23 |
| history of malignancy | 0.021469418 | 24 |
| wavelet_hlh_glcm_idmn | -0.01970135 | 25 |
| wavelet_hlh_firstorder_median | 0.019359384 | 26 |
| wavelet_hlh_glcm_correlation | -0.018328678 | 27 |
| log_1mm_gldm_large dependence low gray level emphasis | -0.018173037 | 28 |
| wavelet_hhl_firstorder_skewness | -0.016232854 | 29 |
| wavelet_hhh_firstorder_median | -0.015945809 | 30 |
| wavelet_hlh_glcm_cluster shade | 0.015767435 | 31 |
| log_2mm_firstorder_kurtosis | -0.014968184 | 32 |
| wavelet_lhh_firstorder_skewness | 0.014899718 | 33 |
| wavelet_lhh_gldm_large dependence low gray level emphasis | -0.013667205 | 34 |
| log_5mm_glcm_idn | 0.013578288 | 35 |
| wavelet_lhh_firstorder_skewness | -0.013431707 | 36 |
| log_4mm_gldm_large dependence high gray level emphasis | 0.012531875 | 37 |
| log_3mm_glszm_low gray level zone emphasis | -0.012381748 | 38 |
| log_5mm_glcm_cluster prominence | 0.012375816 | 39 |
| original_shape_maximum 2d diameter slice | -0.0119441 | 40 |
| log_2mm_gldm_large dependence low gray level emphasis | 0.011709016 | 41 |
| prothrombin time | 0.011531907 | 42 |
| log_3mm_glcm_imc1 | 0.010953448 | 43 |
| wavelet_hlh_glcm_autocorrelation | -0.010753098 | 44 |
| wavelet_lhh_glszm_zone entropy | 0.010242574 | 45 |
| log_4mm_glcm_inverse variance | 0.009940656 | 46 |
| log_4mm_glcm_cluster prominence | 0.009320633 | 47 |
| wavelet_hlh_gldm_large dependence high gray level emphasis | -0.009075509 | 48 |
| wavelet_lhh_firstorder_interquartile range | -0.008467706 | 49 |
| age | 0.007606886 | 50 |
| red blood cell | -0.007256369 | 51 |
| wavelet_hhl_glcm_correlation | -0.007253644 | 52 |
| log_4mm_firstorder_kurtosis | 0.007172 | 53 |
| wavelet_hll_firstorder_skewness | -0.007003864 | 54 |
| wavelet_hhh_gldm_large dependence high gray level emphasis | 0.004918731 | 55 |
| log_5mm_glszm_small area low gray level emphasis | -0.004511342 | 56 |
| original_shape_elongation | -0.004177442 | 57 |
| wavelet_lhh_gldm_large dependence high gray level emphasis | 0.003418303 | 58 |
| wavelet_hhl_glszm_large area high gray level emphasis | 0.003053799 | 59 |
| original_shape_flatness | -0.002683324 | 60 |
| log_2mm_glszm_gray level variance | -0.002040724 | 61 |
| log_1mm_glcm_correlation | 0.001979248 | 62 |
| wavelet_hll_glszm_large area low gray level emphasis | 0.001149169 | 63 |
| log_2mm_glszm_gray level nonuniformity | 0.00089789 | 64 |
| wavelet_lhh_gldm_gray level nonuniformity normalized | -0.000575396 | 65 |
| log_3mm_firstorder_kurtosis | -0.000409742 | 66 |
| original_glcm_cluster shade | 6.95E-06 | 67 |
| Solid nodules (features, n=51) | | |
| age | 0.08456472 | 1 |
| spiculation | 0.078593165 | 2 |
| sex | 0.061542835 | 3 |
| wavelet_lhh_glcm_correlation | -0.055488173 | 4 |
| shape | 0.052257538 | 5 |
| wavelet_lhh_glrjm_short run low gray level emphasis | -0.04648701 | 6 |
| log_5mm_glcm_inverse variance | -0.044164747 | 7 |
| smoking | 0.04239887 | 8 |
| history of malignancy | 0.037123434 | 9 |
| wavelet_hll_glcm_maximum probability | -0.031502098 | 10 |
| original_firstorder_10 percentile | -0.0312687 | 11 |
| log_1mm_firstorder_interquartile range | -0.027092805 | 12 |
| wavelet_hhl_firstorder_median | -0.027003296 | 13 |
| wavelet_hlh_firstorder_median | 0.024898052 | 14 |
| carcinoembryonic antigen (CEA) | 0.024303196 | 15 |
| red blood cell | -0.021868914 | 16 |
| original_shape_sphericity | 0.021541847 | 17 |
| aprotrombin time | -0.020654099 | 18 |
| wavelet_hhh_glcm_cluster shade | -0.019607673 | 19 |
| cytokeratin 19 fragment (cyfra21_1) | -0.018995605 | 20 |
| wavelet_hhh_glcm_cluster prominence | -0.01898479 | 21 |
| log_5mm_glszm_large area low gray level emphasis | -0.018560542 | 22 |
| wavelet_hhh_glszm_large area emphasis | 0.018066432 | 23 |
| wavelet_lhh_glszm_large area low gray level emphasis | 0.01795587 | 24 |
| wavelet_lll_firstorder_root mean squared | -0.017841883 | 25 |
| log_1mm_glszm_large area low gray level emphasis | -0.01751925 | 26 |
| wavelet_lhh_firstorder_median | -0.017030738 | 27 |
| wavelet_lhh_gldm_large dependence high gray level emphasis | -0.014039104 | 28 |
| wavelet_lll_glcm_correlation | 0.013814446 | 29 |
| log_5mm_glszm_gray level nonuniformity normalized | -0.010168748 | 30 |
| log_5mm_firstorder_maximum | 0.010164706 | 31 |
| original_gldm_low gray level emphasis | 0.009151111 | 32 |
| original_shape_elongation | 0.009042126 | 33 |
| log_1mm_gldm_large dependence low gray level emphasis | -0.008320518 | 34 |
| wavelet_hll_firstorder_skewness | -0.008298235 | 35 |
| log_3mm_glszm_gray level nonuniformity | 0.008272954 | 36 |
| log_1mm_firstorder_skewness | -0.007213732 | 37 |
| neuron specific enolase (NSE) | 0.007036805 | 38 |
| wavelet_hlh_firstorder_skewness | 0.006946954 | 39 |
| log_2mm_glcm_maximum probability | -0.006592018 | 40 |
| wavelet_hlh_glcm_cluster prominence | -0.005666115 | 41 |
| log_5mm_glszm_gray level variance | 0.005162543 | 42 |
| log_3mm_firstorder_skewness | 0.005006426 | 43 |
| log_2mm_glszm_large area low gray level emphasis | -0.004477511 | 44 |
| log_5mm_glszm_zone percentage | 0.004358722 | 45 |
| diameter | 0.003045135 | 46 |
| white blood cell | 0.001779955 | 47 |
| wavelet_lll_firstorder_uniformity | -0.000929418 | 48 |
| log_4mm_gldm_small dependence high gray level emphasis | 0.000884861 | 49 |
| wavelet_lhh_firstorder_mean | 0.000590627 | 50 |
| wavelet_lhh_glcm_cluster prominence | 0.000421362 | 51 |
| Subsolid nodules (features, n=3) | | |
| diameter | 0.071469665 | 1 |
| original_glcm_joint entropy | 0.061700333 | 2 |
| wavelet_HHH_gldm_dependence entropy | 0.001622507 | 3 |

Table S5 Performance of established models on training dataset

| Group | Methods | Accuracy | F1 Score | Recall | Precision | Sensitivity | Specificity | AUC |
|----------------------|---------------|----------|----------|--------|-----------|-------------|-------------|------|
| All nodules | XGBOOST | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Random Forest | 0.92 | 0.94 | 0.99 | 0.90 | 0.99 | 0.79 | 0.98 |
| | SVM | 0.90 | 0.92 | 0.97 | 0.88 | 0.97 | 0.76 | 0.95 |
| Nodules ≤ 10 mm | XGBOOST | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Random Forest | 0.98 | 0.98 | 1.00 | 0.97 | 1.00 | 0.97 | 0.99 |
| | SVM | 0.93 | 0.94 | 0.97 | 0.91 | 0.97 | 0.88 | 0.98 |
| Solid nodules | XGBOOST | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Random Forest | 0.98 | 0.98 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 |
| | SVM | 0.88 | 0.87 | 0.90 | 0.85 | 0.90 | 0.87 | 0.94 |
| Subsolid nodules | XGBOOST | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Random Forest | 0.80 | 0.81 | 0.83 | 0.79 | 0.83 | 0.78 | 0.93 |
| | Logistic | 0.73 | 0.72 | 0.70 | 0.74 | 0.70 | 0.76 | 0.78 |

AUC, area under the ROC curve.