



---

# Data-driven risk stratification and precision management of pulmonary nodules detected on chest computed tomography

---

In the format provided by the authors and unedited

---

## **Supplementary Information**

**Supplementary Table 1. Demographics of the lung screening trial participants in the primary dataset and the independent testing dataset.** Age is represented as median (25th, 75th percentiles).

**Supplementary Table 2. Weighting rule for the use of lung nodules in AI model training.** For nodules with pathological findings, their weights are set to 1, reflecting their accurate classification. For nodules without pathological findings, they are considered benign in model training, with weight varying from 1 to 0.25 as the risk level changes from 1 to 4.

**Supplementary Table 3. The number of pulmonary nodules with different risk levels in the primary dataset and the independent testing dataset.**

**Supplementary Table 4. Size distribution of the three types of nodules in the primary dataset and the independent testing dataset.** Each number is represented as median (25th, 75th percentiles, mm). Statistical analyses were performed among four categories using Kruskal-Wallis  $H$  tests, and the number of nodules in each category could be referred to Supplementary Table 3. All two-tailed  $P$  values among four categories were lower than 0.001.

**Supplementary Table 5. AUC values of Lung-RADS v2022 and C-Lung-RADS in distinguishing suspicious malignant nodules at Phase 1.** Each number is represented by mean and 95% confidence interval (CI). Statistical analyses were performed using ordinary two-way ANOVA followed by Sidak's multiple comparison tests, with two-tailed  $P$  values as follows.

**Supplementary Table 6. The FNR, FPR, PPV, and NPV in the Lung-RADS v2022 and C-Lung-RADS both based on nodule's density and size in Phase 1.** Statistical analyses were performed using chi-square tests, with two-tailed  $P$  values as follows.

**Supplementary Table 7. Demographics of participants at Phase 2.** Age is represented as median (25th, 75th percentiles). The age distribution of participants with benign and malignant nodules was compared using the Mann-Whitney  $U$  test with two-

tailed  $P$  values as follows. Statistical analysis of categorical variables was performed using the chi-square tests with two-tailed  $P$  values as follows.

**Supplementary Table 8. Distribution of the malignancy probability in single-, dual-, and multi-dimension models at Phase 2.** For the malignancy probability with asymmetrical distribution, it is represented as median (25th, 75th percentiles). To quantitatively compare the malignancy probability of the benign and the malignant, statistical analyses were performed using Mann-Whitney  $U$  tests, with two-tailed  $P$  values as follows.

**Supplementary Table 9. Quantitative metrics including AUC, accuracy, sensitivity, and specificity of the single-, dual-, and multi-dimension models in identifying the extremely high-risk nodules at Phase 2.** Each number is represented with the mean and 95% confidence interval (CI). Statistical analyses were performed using ordinary two-way ANOVA followed by Tukey's multiple comparison tests. The two-tailed adjusted  $P$  values were listed below.

**Supplementary Table 1. Demographics of the lung screening trial participants in the primary dataset and the independent testing dataset.** Age is represented as median (25th, 75th percentiles).

Parameters N(%)	Primary dataset			Independent testing dataset
	Overall	Training	Internal testing	
<b>No. of participants</b>	45064	36052	9012	14437
<b>Age (median, years)</b>	47 (38, 56)	47 (38, 55)	47 (38, 56)	57 (51, 66)
< 45	18367 (40.8%)	14660 (40.7%)	3707 (41.1%)	842 (5.8%)
≥ 45 to < 60	19237 (42.7%)	15448 (42.8%)	3789 (42.1%)	7369 (51.1%)
≥ 60	7460 (16.5%)	5944 (16.5%)	1516 (16.8%)	6226 (43.1%)
<b>Sex</b>				
<b>Male</b>	26470 (58.7%)	21171 (58.7%)	5299 (58.8%)	5118 (35.5%)
<b>Female</b>	18594 (41.3%)	14881 (41.3%)	3713 (41.2%)	9319 (64.5%)
<b>Smoking status</b>				
<b>Yes</b>	12361 (27.4%)	9919 (27.5%)	2442 (27.1%)	3309 (22.9%)
<b>No</b>	32198 (71.5%)	25741 (71.4%)	6457 (71.6%)	10866 (75.3%)
<b>Unknown</b>	505 (1.1%)	392 (1.1%)	113 (1.3%)	262 (1.8%)
<b>History of cancer</b>				
<b>Yes</b>	61 (0.1%)	44 (0.1%)	17 (0.2%)	129 (0.9%)
<b>No</b>	44255 (98.2%)	35425 (98.3%)	8830 (98.0%)	11130 (77.1%)
<b>Unknown</b>	748 (1.7%)	583 (1.6%)	165 (1.8%)	3178 (22.0%)
<b>Family history of cancer</b>				
<b>Yes</b>	5835 (13.0%)	4647 (12.9%)	1188 (13.2%)	3723 (25.8%)
<b>No</b>	38679 (85.8%)	30973 (85.9%)	7706 (85.5%)	8611 (59.6%)
<b>Unknown</b>	550 (1.2%)	432 (1.2%)	118 (1.3%)	2103 (14.6%)
<b>Family history of lung cancer</b>				
<b>Yes</b>	1754 (3.9%)	1422 (3.9%)	332 (3.7%)	259 (1.8%)
<b>No</b>	42572 (94.5%)	34050 (94.5%)	8522 (94.6%)	12075 (83.6%)
<b>Unknown</b>	738 (1.6%)	580 (1.6%)	158 (1.7%)	2103 (14.6%)
<b>Ground truth</b>				
<b>Label 1</b>	35291 (78.3%)	28253 (78.4%)	7038 (78.1%)	11344 (78.6%)
<b>Label 2</b>	6918 (15.4%)	5559 (15.4%)	1359 (15.1%)	2208 (15.3%)
<b>Label 3</b>	877 (1.9%)	685 (1.9%)	192 (2.1%)	304 (2.1%)
<b>Label 4</b>	1978 (4.4%)	1555 (4.3%)	423 (4.7%)	581 (4.0%)
<b>Pathologically confirmed lung cancer</b>				
<b>Yes</b>	1153 (2.6%)	909 (2.5%)	244 (2.7%)	139 (1.0%)
<b>No</b>	43911 (97.4%)	35143 (97.5%)	8768 (97.3%)	14298 (99.0%)

**Supplementary Table 2. Weighting rule for the use of lung nodules in AI model training.** For nodules with pathological findings, their weights are set to 1, reflecting their accurate classification. For nodules without pathological findings, they are considered benign in model training, with weight varying from 1 to 0.25 as the risk level changes from 1 to 4.

		With pathological finding		Without pathological findings
		Malignant (Label = 1)	Benign (Label = 0)	(Label = 0)
<b>Clinicians rating</b>	<b>Risk level = 1</b>	Weight = 1.00	Weight = 1.00	Weight = 1.00
	<b>Risk level = 2</b>	Weight = 1.00	Weight = 1.00	Weight = 0.75
	<b>Risk level = 3</b>	Weight = 1.00	Weight = 1.00	Weight = 0.50
	<b>Risk level = 4</b>	Weight = 1.00	Weight = 1.00	Weight = 0.25

**Supplementary Table 3. The number of pulmonary nodules with different risk levels in the primary dataset and the independent testing dataset.**

Type N(%)	Overall	Ground truth			
		Label 1	Label 2	Label 3	Label 4
<b>Primary dataset</b>					
<b>All nodules</b>	45064 (100%)	35291 (78.3%)	6918 (15.4%)	877 (1.9%)	1978 (4.4%)
<b>Solid nodules</b>	25129 (100%)	22053 (87.8%)	2416 (9.6%)	239 (0.9%)	421 (1.7%)
<b>mGGNs</b>	2215 (100%)	674 (30.4%)	749 (33.8%)	118 (5.4%)	674 (30.4%)
<b>pGGNs</b>	17720 (100%)	12564 (70.9%)	3753 (21.2%)	520 (2.9%)	883 (5.0%)
<b>Independent testing dataset</b>					
<b>All nodules</b>	14437 (100%)	11344 (78.6%)	2208 (15.3%)	304 (2.1%)	581 (4.0%)
<b>Solid nodules</b>	9807 (100%)	8447 (86.1%)	1014 (10.4%)	167 (1.7%)	179 (1.8%)
<b>mGGNs</b>	703 (100%)	251 (35.7%)	214 (30.5%)	41 (5.8%)	197 (28.0%)
<b>pGGNs</b>	3927 (100%)	2646 (67.4%)	980 (25.0%)	96 (2.4%)	205 (5.2%)

mGGN, mixed ground glass nodule; pGGN, pure ground-glass nodule

**Supplementary Table 4. Size distribution of the three types of nodules in the primary dataset and the independent testing dataset.** Each number is represented as median (25th, 75th percentiles, mm). Statistical analyses were performed among four categories using Kruskal-Wallis  $H$  tests, and the number of nodules in each category could be referred to Supplementary Table 3. All two-tailed  $P$  values among four categories were lower than 0.001.

Type	Overall	Ground truth				$P$
		Label 1	Label 2	Label 3	Label 4	
<b>Primary dataset</b>						
<b>All nodules</b>	4.90 (4.23, 5.81)	4.68 (4.13, 5.35)	6.10 (5.03, 7.55)	7.30 (5.87, 10.02)	10.22 (7.01, 14.97)	<0.001
<b>Solid nodules</b>	4.68 (4.11, 5.43)	4.56 (4.05, 5.20)	6.30 (5.12, 8.11)	8.96 (6.07, 11.93)	12.29 (8.51, 18.47)	<0.001
<b>mGGNs</b>	9.79 (6.13, 14.50)	5.82 (5.13, 7.21)	10.62 (6.73, 14.18)	10.03 (7.17, 13.82)	14.30 (10.40, 19.12)	<0.001
<b>Solid components</b>	5.78 (3.55, 10.58)	3.42 (1.30, 4.98)	6.62 (4.14, 10.64)	6.70 (4.66, 11.23)	9.80 (5.84, 15.05)	<0.001
<b>pGGNs</b>	5.10 (4.40, 6.01)	4.85 (4.24, 5.54)	5.69 (4.86, 6.67)	6.70 (5.58, 8.79)	7.43 (6.39, 9.97)	<0.001
<b>Independent testing dataset</b>						
<b>All nodules</b>	4.79 (3.96, 5.96)	4.52 (3.81, 5.39)	5.88 (4.84, 7.59)	7.47 (5.56, 10.82)	10.90 (7.40, 15.81)	<0.001
<b>Solid nodules</b>	4.62 (3.88, 5.60)	4.48 (3.79, 5.31)	5.80 (4.84, 7.25)	7.08 (5.17, 10.63)	12.71 (9.63, 20.15)	<0.001
<b>mGGNs</b>	9.52 (6.36, 14.13)	6.66 (5.37, 11.70)	9.07 (6.41, 12.55)	10.70 (8.77, 12.49)	13.64 (9.12, 16.83)	<0.001
<b>Solid components</b>	5.49 (3.07, 10.07)	4.54 (1.50, 7.39)	4.81 (2.85, 8.03)	7.90 (5.67, 10.35)	9.02 (5.13, 14.42)	<0.001
<b>pGGNs</b>	4.93 (4.05, 6.20)	4.57 (3.85, 5.49)	5.64 (4.68, 6.97)	6.60 (5.53, 9.15)	7.85 (6.31, 10.72)	<0.001

mGGN, mixed ground glass nodule; pGGN, pure ground-glass nodule

**Supplementary Table 5. AUC values of Lung-RADS v2022 and C-Lung-RADS in distinguishing suspicious malignant nodules at Phase 1.** Each number is represented by mean and 95% confidence interval (CI). Statistical analyses were performed using ordinary two-way ANOVA followed by Sidak's multiple comparison tests, with two-tailed *P* values as follows.

	<b>Overall</b>	<b>Solid nodules</b>	<b>mGGNs</b>	<b>pGGNs</b>
<b>Internal testing dataset</b>				
<b>Lung-RADS</b>	0.761 (0.759-0.762)	0.924 (0.922-0.927)	0.697 (0.694-0.699)	0.506 (0.506-0.507)
<b>C-Lung-RADS</b>	0.899 (0.898-0.900)	0.927 (0.925-0.930)	0.710 (0.708-0.713)	0.829 (0.827-0.830)
<b><i>P</i></b>	<0.001	0.163	<0.001	<0.001
<b>Independent testing dataset</b>				
<b>Lung-RADS</b>	0.820 (0.817-0.822)	0.973 (0.972-0.973)	0.639 (0.636-0.641)	0.499 (0.499-0.499)
<b>C-Lung-RADS</b>	0.912 (0.911-0.913)	0.974 (0.973-0.974)	0.704 (0.701-0.706)	0.799 (0.796-0.802)
<b><i>P</i></b>	<0.001	0.764	<0.001	<0.001

mGGN, mixed ground glass nodule; pGGN, pure ground-glass nodule



**Supplementary Table 6. The FNR, FPR, PPV, and NPV in the Lung-RADS v2022 and C-Lung-RADS both based on nodule's density and size in Phase 1.** Statistical analyses were performed using chi-square tests, with two-tailed *P* values as follows.

Variables	C-Lung-RADS		Lung-RADS		<i>P</i>
	% (95% CI)	<i>n/N</i>	% (95% CI)	<i>n/N</i>	
<b>Internal testing dataset</b>					
<b>FNR</b>	7.4 (4.7-11.4)	18/244	41.4 (35.4-47.7)	101/244	<0.001
<b>FPR</b>	20.8 (20.0-21.6)	1823/8768	11.5 (10.8-12.2)	1007/8768	<0.001
<b>PPV</b>	11.0 (9.8-12.5)	226/2049	12.4 (10.7-14.5)	143/1150	0.233
<b>NPV</b>	99.7 (99.6-99.8)	6945/6963	98.7 (98.4-98.9)	7761/7862	<0.001
<b>Independent testing dataset</b>					
<b>FNR</b>	3.6 (1.5-8.1)	5/139	24.5 (18.1-32.2)	34/139	<0.001
<b>FPR</b>	23.7 (23.0-24.4)	3393/14298	16.4 (15.8-17.0)	2350/14298	<0.001
<b>PPV</b>	3.8 (3.2-4.5)	134/3527	4.3 (3.5-5.2)	105/2455	0.353
<b>NPV</b>	99.9 (99.9-99.9)	10905/10910	99.7 (99.6-99.8)	11948/11982	<0.001

FNR, false negative rate; FPR, false positive rate; PPV, positive predictive value; NPV, negative predictive value.

**Supplementary Table 7. Demographics of participants at Phase 2.** Age is represented as median (25th, 75th percentiles). The age distribution of participants with benign and malignant nodules was compared using the Mann-Whitney *U* test with two-tailed *P* values as follows. Statistical analysis of categorical variables was performed using the chi-square tests with two-tailed *P* values as follows.

Parameters N(%)	Overall	Benign	Malignant	<i>P</i> (Benign vs. Malignant)
<b>Training dataset</b>				
<b>No. of participants</b>	5452	4665	787	
<b>Age (median, years)</b>	50 (43, 57)	49 (42, 56)	53 (46, 62)	<0.001
< 45	1618 (29.7%)	1439 (30.8%)	179 (22.7%)	<0.001
≥ 45 to < 60	2765 (50.7%)	2401 (51.5%)	364 (46.3%)	
≥ 60	1069 (19.6%)	825 (17.7%)	244 (31.0%)	
<b>Sex</b>				
<b>Male</b>	3168 (58.1%)	2854 (61.2%)	314 (39.9%)	<0.001
<b>Female</b>	2284 (41.9%)	1811 (38.8%)	473 (60.1%)	
<b>Smoking status</b>				
<b>Yes</b>	1601 (29.4%)	1421 (30.5%)	180 (22.9%)	<0.001
<b>No</b>	3763 (69.0%)	3225 (69.1%)	538 (68.4%)	
<b>Unknown</b>	88 (1.6%)	19 (0.4%)	69 (8.7%)	
<b>History of cancer</b>				
<b>Yes</b>	8 (0.1%)	2 (0.1%)	6 (0.8%)	<0.001
<b>No</b>	5189 (95.2%)	4642 (99.5%)	547 (69.5%)	
<b>Unknown</b>	255 (4.7%)	21 (0.4%)	234(29.7%)	
<b>Family history of cancer</b>				
<b>Yes</b>	865 (15.9%)	679 (14.5%)	186 (23.6%)	<0.001
<b>No</b>	4459 (81.8%)	3969 (85.1%)	490 (62.3%)	
<b>Unknown</b>	128 (2.3%)	17 (0.4%)	111 (14.1%)	
<b>Family history of lung cancer</b>				
<b>Yes</b>	265 (4.9%)	228 (4.9%)	37 (4.7%)	<0.001
<b>No</b>	4935 (90.5%)	4419 (94.7%)	516 (65.6%)	
<b>Unknown</b>	252 (4.6%)	18 (0.4%)	234 (29.7%)	
<b>Internal testing dataset</b>				
<b>No. of participants</b>	1351	1142	209	
<b>Age (median, years)</b>	50 (42, 58)	50 (42, 57)	54 (45, 64)	<0.001
< 45	427 (31.6%)	375 (32.8%)	52 (24.9%)	<0.001
≥ 45 to < 60	628 (46.5%)	545 (47.7%)	83 (39.7%)	
≥ 60	296 (21.9%)	222 (19.5%)	74 (35.4%)	
<b>Sex</b>				
<b>Male</b>	774 (57.3%)	685 (60.0%)	89 (42.6%)	<0.001
<b>Female</b>	577 (42.7%)	457 (40.0%)	120 (57.4%)	

<b>Smoking status</b>				
Yes	380 (28.1%)	331 (29%)	49 (23.4%)	<0.001
No	946 (70.0%)	809 (70.8%)	137 (65.6%)	
Unknown	25 (1.9%)	2 (0.2%)	23 (11.0%)	
<b>History of cancer</b>				
Yes	6 (0.4%)	1 (0.1%)	5 (2.4%)	<0.001
No	1283 (95.0%)	1139 (99.7%)	144 (68.9%)	
Unknown	62 (4.6%)	2 (0.2%)	60 (28.7%)	
<b>Family history of cancer</b>				
Yes	203 (15%)	173 (15.1%)	30 (14.3%)	<0.001
No	1117 (82.7%)	967 (84.7%)	150 (71.8%)	
Unknown	31 (2.3%)	2 (0.2%)	29 (13.9%)	
<b>Family history of lung cancer</b>				
Yes	50 (3.7%)	47 (4.1%)	3 (1.4%)	<0.001
No	1239 (91.7%)	1093 (95.7%)	146 (69.9%)	
Unknown	62 (4.6%)	2 (0.2%)	60 (28.7%)	
<b>Independent testing dataset</b>				
No. of participants	1951	1812	139	
Age (median, years)	62 (54, 68)	62 (54, 68)	62.0 (54, 67)	0.192
< 45	49 (2.5%)	43 (2.4%)	6 (4.3%)	0.233
≥ 45 to < 60	820 (42.0%)	768 (42.4%)	52 (37.4%)	
≥ 60	1082 (55.5%)	1001 (55.2%)	81 (58.3%)	
<b>Sex</b>				
Male	727 (37.3%)	685 (37.8%)	42 (30.2%)	0.075
Female	1224 (62.7%)	1127 (62.2%)	97 (69.8%)	
<b>Smoking status</b>				
Yes	477 (24.5%)	449 (24.8%)	28 (20.2%)	0.439
No	1440 (73.8%)	1331 (73.4%)	109 (78.4%)	
Unknown	34 (1.7%)	32 (1.8%)	2 (1.4%)	
<b>History of cancer</b>				
Yes	22 (1.1%)	22 (1.2%)	0 (0)	0.005
No	1548 (79.4%)	1450 (80%)	98 (70.5%)	
Unknown	381 (19.5%)	340 (18.8%)	41 (29.5%)	
<b>Family history of cancer</b>				
Yes	1407 (72.1%)	1368 (75.5%)	39 (28.1%)	<0.001
No	518 (26.6%)	441 (24.3%)	77 (55.4%)	
Unknown	26 (1.3%)	3 (0.2%)	23 (16.5%)	
<b>Family history of lung cancer</b>				
Yes	22 (1.1%)	18 (1.0%)	4 (2.9%)	<0.001
No	1903 (97.6%)	1791 (98.8%)	112 (80.6%)	
Unknown	26 (1.3%)	3 (0.2%)	23 (16.5%)	

**Supplementary Table 8. Distribution of the malignancy probability in single-, dual-, and multi-dimension models at Phase 2.** For the malignancy probability with asymmetrical distribution, it is represented as median (25th, 75th percentiles). To quantitatively compare the malignancy probability of the benign and the malignant, statistical analyses were performed using Mann-Whitney  $U$  tests, with two-tailed  $P$  values as follows.

<b>Datasets</b>	<b>Benign</b>	<b>Malignant</b>	<b><math>P</math>(Benign vs. Malignant)</b>
<b>Training dataset</b>			
<b>Single-dimension model (Imaging)</b>	0.236 (0.150, 0.290)	0.653 (0.483, 0.742)	<0.001
<b>Dual-dimension model (Imaging + Clinical)</b>	0.282 (0.197, 0.353)	0.713 (0.550, 0.802)	<0.001
<b>Multi-dimension model (Imaging + Clinical + Follow-up)</b>	0.288 (0.205, 0.363)	0.753 (0.570, 0.870)	<0.001
<b>Internal testing dataset</b>			
<b>Single-dimension model</b>	0.232 (0.154, 0.290)	0.706 (0.530, 0.783)	< 0.001
<b>Dual-dimension model</b>	0.283 (0.196, 0.353)	0.752 (0.579, 0.832)	< 0.001
<b>Multi-dimension model</b>	0.290 (0.203, 0.358)	0.808 (0.653, 0.911)	< 0.001
<b>Independent testing dataset</b>			
<b>Single-dimension model</b>	0.195 (0.130, 0.290)	0.656 (0.440, 0.767)	<0.001
<b>Dual-dimension model</b>	0.256 (0.185, 0.353)	0.713 (0.502, 0.830)	<0.001
<b>Multi-dimension model</b>	0.259 (0.189, 0.354)	0.713 (0.535, 0.834)	<0.001

**Supplementary Table 9. Quantitative metrics including AUC, accuracy, sensitivity, and specificity of the single-, dual-, and multi-dimension models in identifying the extremely high-risk nodules at Phase 2.** Each number is represented with the mean and 95% confidence interval (CI). Statistical analyses were performed using ordinary two-way ANOVA followed by Tukey’s multiple comparison tests. The two-tailed adjusted *P* values were listed below.

<b>Datasets</b>	<b>Single-dimension</b>	<b>Dual-dimension</b>	<b>Multi-dimension</b>	<i>P</i> (Single- vs. Dual- dimension)	<i>P</i> (Single- vs. Multi- dimension)	<i>P</i> (Dual- vs. Multi- dimension)
<b>Internal testing dataset</b>						
<b>AUC</b>	0.881 (0.880-0.882)	0.882 (0.881-0.883)	0.918 (0.918-0.919)	0.451	< 0.001	< 0.001
<b>Accuracy</b>	0.832 (0.831-0.832)	0.828 (0.827-0.828)	0.829 (0.829-0.830)	< 0.001	< 0.001	0.025
<b>Sensitivity</b>	79.6% (79.4%-79.7%)	82.4% (82.2%-82.5%)	85.1% (85.0%-85.3%)	< 0.001	< 0.001	< 0.001
<b>Specificity</b>	83.4% (83.4%-0.835)	82.8% (82.8%-82.9%)	82.8% (82.8%-82.9%)	< 0.001	< 0.001	1.000
<b>Independent testing dataset</b>						
<b>AUC</b>	0.924 (0.923-0.926)	0.926 (0.924-0.927)	0.927 (0.926-0.928)	0.876	0.565	0.857
<b>Accuracy</b>	0.882 (0.881-0.882)	0.877 (0.876-0.877)	0.877 (0.876-0.877)	0.067	0.068	1.000
<b>Sensitivity</b>	64.3% (63.6%-65.1%)	78.3% (77.8%-78.9%)	85.6% (85.1%-86.1%)	<0.001	<0.001	<0.001
<b>Specificity</b>	88.4% (88.3%-88.5%)	87.8% (87.7%-87.8%)	87.7% (87.6%-87.7%)	0.012	0.004	0.946

AUC, area under the curve.