1	Т	itle: From signal to knowledge: The diagnostic value of rawdata in artificial intelligence					
2		prediction of human data for the first time					
3							
4	Au	thors names					
5	Bin	gxi He <sup>1,2,3#</sup> , Yu Guo <sup>4#</sup> , Yongbei Zhu <sup>1,2,3</sup> , Lixia Tong <sup>5</sup> , Boyu Kong <sup>4</sup> , Kun Wang <sup>3</sup> , Caixia Sun <sup>1,2,3</sup> ,					
6	Hailin Li <sup>1,2,3</sup> , Feng Huang <sup>5</sup> , Liwei Wu <sup>6</sup> , Meng Wang <sup>4</sup> , Fanyang Meng <sup>4</sup> , Le Dou <sup>4</sup> , Kai Sun <sup>3</sup> , Tong						
7	Tor	g <sup>3</sup> , Zhenyu Liu <sup>3</sup> , Ziqi Wei <sup>3</sup> , Wei Mu <sup>1,2,3</sup> , Shuo Wang <sup>1,2,3</sup> , Zhenchao Tang <sup>1,2,3</sup> , Shuaitong					
8	Zha	ng <sup>1,2,3</sup> , Jingwei Wei <sup>3</sup> , Lizhi Shao <sup>3</sup> , Mengjie Fang <sup>1,2,3</sup> , Juntao Li <sup>7</sup> , Shouping Zhu <sup>7,8,9</sup> , Lili Zhou <sup>4</sup> ,					
9	Shu	o Wang <sup>4</sup> , Di Dong <sup>3*</sup> , Huimao Zhang <sup>4*</sup> , Jie Tian <sup>1,2,3*</sup>					
10							
11	#Bi	ng-Xi He, Yu Guo contributed equally to this article.					
12							
13	Au	thor affiliations					
14	1.	Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of					
15		Engineering Medicine, Beihang University, Beijing, China					
16	2.	Key Laboratory of Big Data-Based Precision Medicine, Beihang University, Ministry of					
17		Industry and Information Technology, Beijing, China					
18	3.	CAS Key Laboratory of Molecular Imaging, the State Key Laboratory of Management and					
19		Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences,					
20		Beijing, China					
21	4.	Department of Radiology, The First Hospital of Jilin University, 71 Xinmin Street,					
22		Changchun, 130021, Jilin, China.					
23	5.	Neusoft Medical Systems Co. Ltd., Shenyang, China					
24	6.	Neusoft Research of Intelligent Healthcare Technology Co. Ltd., Shenyang, China					
25	7.	School of Life Science and Technology, Xidian University & Engineering Research Center of					
26		Molecular and Neuro Imaging, Ministry of Education, Xi'an, Shaanxi 710126, China.					
27	8.	Xi'an Key Laboratory of Intelligent Sensing and Regulation of trans-Scale Life Information					
28		& International Joint Research Center for Advanced Medical Imaging and Intelligent					
29		Diagnosis and Treatment, School of Life Science and Technology, Xidian University, Xi'an,					

30 NOTE: This trapposed to guide clinical practice.

- 31 9. Innovation Center for Advanced Medical Imaging and Intelligent Medicine, Guangzhou
- 32 Institute of Technology, Xidian University, Guangzhou, Guangdong 51055, China.
- 33

34 **\*Corresponding author** 

- 35 Di Dong, PhD
- 36 The CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of
- 37 Sciences, No. 95 Zhongguancun East Road, Hai Dian District, Beijing, 100190, China.
- 38 Tel: +86 10 82618465
- **Fax:** +86 10 62527995
- 40 Email: Di.Dong@ia.ac.cn
- 41

```
42 Hui-mao Zhang, PhD
```

- 43 Department of Radiology, The First Hospital of Jilin University, Changchun, 130021, China.
- 44 **Tel:** +86 0431 85166420
- 45 Email: huimao@jlu.edu.cn
- 46

# 47 Jie Tian, PhD

- 48 FAIMBE, FIAMBE, FIEEE, FSPIE, FOSA, FIAPR, FISMRM
- 49 The CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of
- 50 Sciences, No. 95 Zhongguancun East Road, Hai Dian District, Beijing, 100190, China.
- 51 **Tel:** +86 10 82618465
- 52 **Fax:** +86 10 62527995
- 53 Email: jie.tian@ia.ac.cn
- 54
- 55 Running Head: The first signal-to-knowledge study based on human data

# 57 Abstract

Recently, image-based diagnostic technology has made encouraging and astonishing development. 58 59 Modern medical care and imaging technology are increasingly inseparable. However, the current 60 diagnosis pattern of Signal-to-Image-to-Knowledge inevitably leads to information distortion and 61 noise introduction in the procedure of image reconstruction (Signal-to-Image). Artificial 62 intelligence (AI) technologies that can mine knowledge from vast amounts of data offer 63 opportunities to disrupt established workflows. In this prospective study, for the first time, we 64 developed an AI-based Signal-to-Knowledge diagnostic scheme for lung nodule classification 65 directly from the CT rawdata (the signal). We found that the rawdata achieved almost comparable 66 performance with CT indicating that we can diagnose diseases without reconstructing images. Meanwhile, the introduction of rawdata could greatly promote the performance of CT, 67 68 demonstrating that rawdata contains some diagnostic information that CT does not have. Our 69 results break new ground and demonstrate the potential for direct Signal-to-Knowledge domain 70 analysis.

71

### 72 Introduction

73 The discovery of X-rays in 1895 ushered in a new era in the use of imaging for medical diagnostic 74 purposes. Since then, the non-invasive medical imaging technology subverts the palpation and cut-75 and-see scheme [1]. The technological advances in medical imaging have been astounding over 76 the past 120 years, and modern medical care is increasingly inseparable from imaging technology. 77 Medical imaging is essential for humans, to allow clinicians to observe from the images and 78 diagnose diseases. This process can be defined as a path of image-to-knowledge. However, 79 recently, it is found that the human ability has become a bottleneck in this path hindering the 80 accurate diagnosis and treatment of diseases [2][3].

The emergence of Artificial Intelligence (AI) technology partially solves the problem of the limited ability of humans in the diagnosis process [4][5][6][7]. AI could automatically mine the radiographic patterns that related to the occurrence and progression of diseases from the imaging data, and it has been shown to match and even surpass human abilities in many clinical applications [8][9][10][11][12]. The essential reason why AI could surpass humans may be that AI treats images as data rather than the visual image and extracts huge amounts of features for

87 analysis [13][14]. However, the medical image is compressed or filtered data to fit the human eve, 88 which may be insufficient or imperfect for diagnosis. Take computed tomography (CT) for 89 example, the CT system first collects rawdata (signal) from the patient, then the reconstruction 90 method converts rawdata to images (signal-to-image) [15]. Therefore, both AI-based and human-91 based diagnosis are processes of signal-to-image-to-knowledge. Medical images suffer from 92 information distortion in both acquisition and reconstruction processes. The current high sampling 93 frequency greatly compresses the influence of factors such as motion artifacts in the acquisition 94 process, so the main reason for the loss of resolution is concentrated in the operations such as 95 interpolation and sub-optimal statistical weighting in the reconstruction process [16]. Meanwhile, the unprocessed data size of rawdata is about 10 to 20 times larger than that of CT images (2GB 96 compare with 180MB). The huge amount of information inside the rawdata is not optimally mined 97 98 in current signal-to-image-to-knowledge process, and how to analysis rawdata is of great scientific 99 interest.

100 Skipping the image process and going directly from signal to knowledge, will hopefully bring 101 new breakthroughs in disease diagnosis. Inspired by this idea, several previous studies have talked 102 about the potential value of analysis of rawdata [17][18][19], directly from signal to knowledge. 103 De Man Q et.al. conducted a simulation experiment to detect and estimate the vessel centerline 104 from rawdata in the sinogram domain [19]. They achieved encouraging initial results showing the 105 feasibility of rawdata analysis for clinical CT analysis tasks. We have also reported our simulation 106 results about lung cancer at American Association for Cancer Research (AACR) conference [20]. 107 However, there is no study about signal-to-knowledge analysis in real clinical tasks of patients.

108 In this prospective study, for the first time, we developed an AI-based signal-to-knowledge 109 diagnostic scheme for lung nodule classification directly from the CT rawdata (The flowchart was 110 shown in Fig. 1). The value of rawdata alone (Discussion), as well as its added value to CT, are 111 studied on 276 patients. We found that the rawdata achieved almost comparable performance with 112 CT indicating that we can diagnose diseases without reconstructed images. Meanwhile, the 113 introduction of rawdata could greatly promote the performance of CT, demonstrating that rawdata 114 contains some diagnostic information that CT does not have. This research breaks the routinely 115 used circle of image-based diagnosis, which may open up a new pathway of signal-to-knowledge 116 for disease diagnosis.

117

#### 118 **Results**

#### 119 Clinical characteristics

The clinical characteristics was summarized in Table S1. A total of 276 patients were included and 120 the number of patients in the training cohort, validation cohort and test cohort were 166, 55 and 55, 121 122 respectively. Fifty percent (n = 138) patients were female and the mean of age in the entire dataset 123 was 58.48 years. Furthermore, there were 21 (8%) small cell carcinoma, 35 (13%) squamous cell 124 carcinomas and 149 (54%) adenocarcinomas. With respect to lesion location, most patients were 125 identified as right upper lobe (n = 89, 32%), followed by left lower lobe (n = 67, 24%) and left upper lobe (n = 64, 23%) in all patients. For lung cancer diagnosis, most patients (n = 225, 82%) 126 127 were evaluated as malignancy.

128

#### 129 Performance of CT model and rawdata gain model

130 This experiment explores the performance improvement that the residual fusion model (Methods) 131 based on both rawdata and CT images can bring to the model based on CT images only. For 132 further explore the repeatability and stability of this gain, we tested four different CT models 133 (Abbreviation for CTM1~CTM4; Methods) and adopted three backbone network architectures for rawdata feature extraction, namely Densenet121 (DN) [21], Resnet18 (RE) [22] and Resnext18 134 135 (RX) [23]. For each CT model (CTM), three rawdata gain models (RGM) based on different 136 backbone feature extraction networks were constructed. The performance of each RGM was 137 compared with the original CTM. The receiver operating characteristic curves (ROC) and its area 138 under the curve (AUC) of four CTMs and the corresponding RGMs based on difference backbone 139 networks is shown in Fig. 2.

For each CTM, the residual fusion models based on different backbone networks can obtain better classification performance on training, validation and test cohorts. For CTM1 model, the fusion model that produced the maximum performance improvement for the training cohort is RGM-RX1, and its AUC improvement can reach 0.051 (from 0.757 to 0.808). The fusion model that produced the maximum performance improvement for the validation cohort is RGM-RE1 and RGM-RX1, and its AUC improvement can reach 0.033 (from 0.756 to 0.789). The fusion model that produced the maximum performance improvement for the test cohort is RGM-RE1, and its

147 AUC improvement can reach 0.046 (from 0.807 to 0.853). For CTM2 model, the fusion model 148 that produced the maximum performance improvement for the training cohort is RGM-RX2, and 149 its AUC improvement can reach 0.109 (from 0.745 to 0.854). The fusion model that produced the 150 maximum performance improvement for the validation cohort is RGM-DM2, and its AUC 151 improvement can reach 0.124 (from 0.698 to 0.822). The fusion model that produced the 152 maximum performance improvement for the test cohort is RGM-DM2, and its AUC improvement 153 can reach 0.022 (from 0.760 to 0.782). For CTM3 model, the fusion model that produced the 154 maximum performance improvement for the training cohort is RGM-RX3, and its AUC 155 improvement can reach 0.083 (from 0.765 to 0.848). The fusion model that produced the 156 maximum performance improvement for the validation cohort is RGM-RX3, and its AUC improvement can reach 0.093 (from 0.760 to 0.853). The fusion model that produced the 157 158 maximum performance improvement for the test cohort is RGM-RE3, and its AUC improvement 159 can reach 0.027 (from 0.773 to 0.800). For CTM4 model, the fusion model that produced the 160 maximum performance improvement for the training cohort is RGM-RX4, and its AUC 161 improvement can reach 0.035 (from 0.832 to 0.867). The fusion model that produced the 162 maximum performance improvement for the validation cohort is RGM-DM4, and its AUC 163 improvement can reach 0.026 (from 0.756 to 0.782). The fusion model that produced the maximum performance improvement for the test cohort is RGM-RE4, and its AUC improvement 164 165 can reach 0.034 (from 0.833 to 0.867). Overall, using Resnext18 as the backbone network of 166 rawdata feature extraction can obtain the maximum average performance improvement on the 167 three cohorts.

168

#### 169 Image feature distribution of the RGMs and gain stability analysis

We performed t-SNE dimensionality reduction on all deep learning features obtained by different feature extraction networks and counted the true positives, false positives, true negatives and false negatives of each patient (**Table 1**). Besides, we assigned different colors and markers to visualize them in the same coordinate system (**Fig. 3**). It can be observed from the **Fig. 3** that the results of the various RGMs within each CTM are relatively similar, even though they use different feature extraction networks. **Table 1** also shows the same situation. The gain of the RGMS inside each CTM is approaching the same trend, such as improving malignant or benign detectable rates.

#### 177 Meanwhile, RGM-RE1, RGM-DN2, RGM-RE2, RGM-DN3, RGM-DN4, RGM-RE4, and RGM-

- 178 RX4 could achieve a significant increase in the detectable rate of one category at the expense of a
- 179 small number of the other category detectable rates.
- 180 Therefore, we calculated the optimization rate and error rate of each RGM for the CTM, and 181 also calculated the proportion of at least 2 model optimizations to all optimization samples, which 182 can reflect the stability of rawdata's gain. All results were summarized in **Table S2**.
- The results show that the analysis method incorporating the rawdata has a high optimization rate for the CTM 1~3 and is greater than the error rate, which is also reflected in the improvement of AUC. In addition, although different feature extraction networks were used to analysis the rawdata, the proportion of at least two networks that can be optimized in each CTM is about 80%. Finally, we found that 7 samples were mispredicted within 4 CTMs. For these 7 samples, the income of rawdata can correct the prediction results of the 6 CTMs, and the corrected model exists in each CTM. In summary, the gain of the rawdata for the CTM is very stable.
- 190

#### 191 Visual statistics and analysis of the RGMs

192 To better explain the prediction process of RGM, we visualized the region of most interest in the 193 RGM by using Gradient-weighted Class Activation Mapping (Grad-CAM). The predictive results 194 of RGM were most dependent on the information of the RGM-discovered suspicious areas. Fig. 4 195 illustrated the lesion masks and corresponding attention maps from different views of the rawdata. 196 From Fig. 4, we can see that the RGMs can always focus on areas of lesion for prediction 197 although the input data includes some non-lesion areas. We also calculated the average attention 198 score of each voxel in lesion and non-lesion in rawdata (Method), and the result showed that the 199 attention score of the lesion area was 1-2 times as high as the non-lesion area.

200

# 201 Stratified analysis of different malignant subgroups

The results of the subgroup analysis for age, sex and lesion size were shown in the Table S3 andTable 2.

In the subgroup with age of <= 60, RGM-RX 4 and CTM 4 achieved the similar highest model performance, with AUC of 0.837 (0.746-0.924) and 0.831 (0.749-0.904), respectively; In

206 the subgroup with age of > 60. RGM-RX 4 achieved the highest model performance with an AUC 207 of 0.845 (0.713-0.949), which outperformed the best CTM (CTM4 with an AUC of 0.790). In the 208 male subgroup, CTM4 and RGM-RX 4 performed best, with a similar AUC of 0.804 (0.706-0.882) 209 and 0.810 (0.707-0.897), respectively; In the female subgroup, RGM-RX 3 achieved the highest 210 performance with an AUC of 0.885 (0.818-0.945), far exceeding the best CTM (CTM4 with an 211 AUC of 0.823 (0.720-0.920)). In the subgroup with lesion size <=23mm, RGM-RX 3 achieved the 212 highest model performance with an AUC of 0.847 (0.781-0.916), far exceeding the best CTM 213 (CTM4 with an AUC of 0.806); In the subgroup of lesion size >23mm, CTM 4 and RGM-RE 4 214 showed the similar highest model performance, with AUC of 0.819 (0.719-0.906) and 0.833 215 (0.703-0.925), respectively. As for the lesion location subgroups, in addition to the similar performance of CTM 4 and RGM-RX 4 in the subgroup of superior lobe of left lung, RGM 216 217 outperformed the CTM, with AUC of 0.840 vs. 0.812 in the subgroup of inferior lobe of left lung, 218 0.849 vs. 0.807 in the subgroup of superior lobe of right lung, and 0.872 vs. 0.843 in the subgroup 219 of inferior lobe of right lung.

220

# 221 Discussion

222 In this prospective study, for the first time, we validated the potential value of rawdata in real 223 clinical practice. Interestingly, the rawdata analysis showed comparable performance with CT 224 images, which indicates that leveraging non-image information holds promise as an alternative to 225 image-based methods. Moreover, the add value of rawdata to CT images was also confirmed in 226 this study, which means that the combination of non-image and image data will further promote 227 the advance of disease diagnosis. This study proposed and validated a feasible method for 228 diagnosis without image reconstruction, and it has the potential to change existing imaging-based 229 diagnosis and treatment strategies.

The classification of benign and malignant pulmonary nodules is a matter of great clinical concern[24][25][26]. This study explores the feasibility of rawdata analysis in classifying indeterminate lung nodules greater than 2 cm in size. The results indicated that rawdata can well discriminate malignant nodules from benign nodules. Meanwhile, the AUCs of the rawdata in the training cohort, validation cohort and test cohort are 0.768 (95%CI: 0.681~0.851), 0.760 (95%CI: 0.558~0.922) and 0.782 (95%CI: 0.592~0.924), respectively (**Extended Data Fig. 1**), and there is

no statistical difference between the performances of rawdata and CT, which means that the
classification of lung nodules may not need image reconstruction and clinician participation.
Think further, the rawdata model could be applied to the majority of grassroots hospitals, who
have mainstream CT systems but lack technical personnel and clinicians.

240 Our study showed that the introduction of rawdata to CT had an overall improvement over 241 different CTMs, no matter which backbone network was used. This indicates that rawdata has 242 unique information which may be lost during the reconstruction processing. Moreover, the 243 compared with CTM, RGMs showed better stability on the training cohort, validation cohort, and test cohort. The combination of both non-image and image data could make the model robust. In 244 245 addition, we also performed intra-CT and inter-CT analyses. For intra-CTM, fused rawdata 246 prediction has a higher optimization rate than the error rate which shows a similar gain trend, and 247 about 80% of the optimized patients appear in at least 2 feature extraction networks. For inter-248 CTMs, eighty-five percent of the patients that all CTMs predicted incorrectly have optimizable 249 RGMs within each CTM. The results also proved that the gain of rawdata is stable across different 250 convolutional networks and different CTM approaches. Therefore, exploring the drawbacks of 251 post-reconstructed CT image analysis and developing models for direct diagnosis from rawdata 252 are the keys to future research. The results of the subgroup analysis showed that the RGMs 253 performed better than the CTM in most subgroups, especially in the subgroup of older, female, 254 and smaller lesion size, indicating that the rawdata could provide more valuable information that 255 brings model gains in subgroups, while this information may have been lost in the process of CT 256 reconstruction.

Our study has some limitations. First, this study involved a small number of patients and the proportion of positive and negative samples is unbalanced. Further study on large-scale multicenter datasets should be performed. Second, only patients with single nodule are included in this study, further validation of our method on patients with multiple nodules should be further studied. Third, although the rawdata had a comparable performance with CT, it still had a certain gap with the best CT diagnosis. There is an urgent need to develop novel AI methods specifically for rawdata.

264 Meanwhile, strong computing power is a problem that cannot be ignored when calculating 265 rawdata. It is not realistic to read the complete high-frequency scanning data directly to the

computing device. Designing appropriate pre -processing algorithms and building deep networks
in combination with characteristics of rawdata are the potential breakthrough points in the future.
Finally, the CT scan scheme is designed for image reconstruction and it may be not suitable for
rawdata analysis. Therefore, novel scan strategies, e.g., scanning for specific diagnostic purposes,
should be developed to maximize the gain of rawdata.

271 In conclusion, for the first time, we validated the potential value of rawdata in real clinical 272 practice. The rawdata analysis showed comparable performance with CT images, which indicates 273 that leveraging non-image information holds promise as an alternative to image-based methods. 274 Moreover, the added value of rawdata to CT images was also confirmed in this study, which 275 means that the combination of non-image and image will further promote the advance of disease diagnosis. This study proposed and validated a new feasible direction for diagnosis without image 276 277 reconstruction, and it may facilitate the development of fully automated scanning and diagnostic 278 processes.

279

```
280 Methods
```

#### 281 **Patients**

282 In this prospective study, 626 consecutive patients who had a chest CT scan in the First Hospital of Jilin University from November 2019 to May 2021 were recruited. Eligible patients were 283 284 included according to the following inclusion criteria: (i) patients who had a pulmonary lesion 285 more than 2 cm with contrast enhanced chest CT scan, (ii) rawdata obtained from CT machine 286 after the imaging examination, (iii)pathological diagnosis of pulmonary lesion with two weeks 287 interval from CT scan. Patients were excluded on the basis of the following: (i) previous systemic 288 antineoplastic treatments, (ii) CT images with poor image quality or unreadable scan. After 289 exclusion, a total of 276 patients were included for modeling experiments.

290 The methods were performed in accordance with Standards for Reporting Diagnostic 291 accuracy studies (STARD) and approved by the Ethics Committee of the First Hospital of Jilin 292 University (AF-IRB-032-05).

293

#### 294 Collection of CT image and rawdata

295 Both the CT images and rawdata were consecutively collected from the First Hospital of Ji Lin

296 University and were acquired with a NeuViz Prime CT system (Neusoft Medical Systems Co., 297 Ltd., Shenyang, China). The system parameters of CT scanner included source-to-isocenter 298 distance of 570 mm, source-to-detector distance of 1040mm, and scanning FOV of 500 mm. The 299 imaging protocol included contrast-enhanced CT of the chest with variable imaging parameters. 300 Contrast-enhanced CT scans were performed at a spiral scan mode using 324 mA tube current, 301 100 kVp tube voltage, 0.5 s ration time, and 0.9 spiral pitch. CT rawdata were reconstructed using 302 a kernel F20 at slice thickness of 1.0 mm with image pixel range from 0.59 mm to 0.98 mm and 303 image matrix of 512 by 512. In addition, we also acquired the initial height and initial view angle 304 of the CT detector each time the patient underwent a scan. Finally, CT images and rawdata from 305 each scanner were randomly stratified into one of three cohorts in a 6:2:2 ratio: a training cohort, a 306 validation cohort and a test cohort. All in all, Table S4 describes CT scanner information, system 307 parameters and imaging parameters.

308

### 309 lesion segmentations in CT images

The segmentations of primary lesion were manually delineated across all the sections in the axial view using annotation tool in ISD (IntelliSpace Discovery, Philips, German). The regions of interest were annotated and reviewed by four radiologists with 8 to 25 years' chest CT experience. All radiologists were blinded to any clinical or histopathologic information. The annotation was labeled as five common categories according lesions' pulmonary lobe.

315

#### 316 Realization of typical CT models

There are many studies on benign-malignant lung nodule classification on chest CT. We selected four typical papers from the major journals, including *IEEE Transactions on medical imaging*, *Medical image analysis*, and *Nature medicine*, which refer to multi-scale ensemble method (CTM1) [27], global and local information fusion method (CTM2) [26], loss function-based method (CTM3) [28], and multi-view fusion (CTM4) method [29]. We further performed experiments on four typical models with our dataset, and all the realization details are described in **Supplementary 1**.

324

#### 325 Extraction of lesion region from raw data

326 After acquiring 4 CTMs, we proceeded to perform rawdata gain experiments. The first step of the 327 experiment is to select projection surface containing lesions in the rawdata. The rawdata of CT 328 scans contains three dimensions: 1) the index dimension representing the acquisition order; 2) the 329 projection surface which is the detector receives the x-ray attenuation, where the channel and row 330 directions are defined as x and y, respectively. All lesion segmentation regions of rawdata were 331 derived from the binarized segmentation of CT image after being represented in a unified 332 coordinate system. The complete derivation can be condensed into three steps: orientation, 333 querying and mapping.

334 1) Orientation.

On the derivation of localization, we took the segmented regions in the CT image as the research object, and the localization calculation mainly includes cross-sectional localization and height localization. For cross-sectional positioning, we set the center point between the CT source and detector as the coordinate origin (which is also the rotation center of the CT gantry), parallel to the cross-section of the CT image. Next, the motion trajectory can be characterized by the scan index t, the rotation radius r and the angle  $\theta$ . In order to obtain the above parameters, we first read the origin coordinates ( $x_{origin}$ ,  $y_{origin}$  and  $z_{origin}$ ), and the offset values ( $x_{offset}$  and  $y_{offset}$ )

342 can be calculated through voxel spacing and image size ( $x_{size}$  and  $y_{size}$ ).

343 
$$x_{\text{offset}} = x_{\text{origin}} + x_{\text{spacing}} \times \frac{1}{2} x_{\text{size}}$$

344 
$$y_{\text{offset}} = y_{\text{origin}} + y_{\text{spacing}} \times \frac{1}{2} y_{\text{size}}$$

Next, with the help of the offset values and the voxel coordinates  $x_{CT}$  and  $y_{CT}$  in the CT image, the distance from the coordinate origin ( $x_{length}$  and  $y_{length}$ ) can be calculated as:

347 
$$x_{\text{length}} = x_{\text{spacing}} \times \left( x_{\text{CT}} - \frac{1}{2} x_{\text{size}} \right) - x_{\text{offset}}$$

348 
$$y_{\text{length}} = y_{\text{spacing}} \times \left( y_{\text{CT}} - \frac{1}{2} y_{\text{size}} \right) - y_{\text{offset}}$$

349 Otherwise, the radius r, and the starting angle  $\theta_0$  can be obtained.

$$r = \sqrt{x^2 + y^2}$$

351 
$$\theta_0 = \arctan \frac{y_{\text{length}}}{x_{\text{length}}}$$

352 Meanwhile, by introducing the scanning period of the machine, we got the angle change  $\Delta \theta$ 353 with the following relationship:

 $y_{\text{length}} = r \times \cos(\Delta \theta \cdot t + \theta_0)$ 354

355 
$$x_{\text{length}} = r \times \sin\left(\Delta \theta \cdot t + \theta_0\right)$$

356 For height positioning, we directly obtained the initial height  $h_{\text{start}}$  of the voxel through the 357 coordinate z, slice thickness and origin of the voxel point in CT images.

358 2) Querying.

359 Our purpose in this step is to determine the interval of index dimension t in which tumor 360 voxel appears in the raw data. Since there is a cone beam in the projection, we first calculated the 361 change function  $h_{area}$  of the voxel, where d is the distance from the voxel to the X-ray focal spot on the x-axis; l is the distance from the focal spot to the detector. The number of detector rows is 362 363  $n_{y}$ ; the channel spacing along y-axis is  $\Delta n_{y}$ .

$$d = \frac{1}{2}l + x_{\text{length}}$$

365 
$$h_{\text{area}} = \frac{1}{l} \left( \frac{1}{2} n_y \times \Delta n_y \times d \right)$$

Next, we determined the index (t) range of the voxel in the rawdata by the following 366 367 inequality, where  $h_0$  is the initial height at which the detector start to scan;  $\Delta h$  is the height 368 change in a scan.

 $h_0 - \Delta h \cdot t \le h_{\text{start}} + h_{\text{area}}$ 369

$$h_0 - \Delta h \cdot t \ge h_{\text{start}} - h_{\text{area}}$$

371 To reduce computational complexity, we first extracted the highest and lowest masks in the 372 segmentation images, and calculated the start and end indices of two voxels. Then, we initially 373 located the range of index dimensions. Within this interval, we computed the mapping result of

374 voxels within the layer.

375 3) Mapping.

Through the above calculation, we have obtained the index interval corresponding to the voxel, then the voxel appearing in index is obtained by calculating the projection data of the index layer by layer. The coordinates of each voxel on the projection surface are defined as  $x_{raw}$  and  $y_{raw}$ , respectively.  $y_{raw}$  is related to the height  $h_t$ ,  $h_{area}$  at the *t* index and the number of detector rows  $n_y$  in the detector, so we determined its height difference relative to the detector by the following formula, and then calculated its coordinates in the projection.

382 
$$y_{\text{voxel}} = \frac{h_{\text{t}}}{h_{\text{area}}} \cdot \frac{(n_{\text{y}} - 1)}{2}$$

Since the x-axis of the projection plane is equiangularly sampled,  $x_{raw}$  can be acquired through the angle at the *t* index  $\theta_t$ , the view angle  $\theta_d$  of the detector, and the number of channels in the detector *n*.

$$x_{\text{voxel}} = \frac{2\theta_{\text{t}}}{\theta_{\text{d}}} \cdot \frac{(n_{\text{x}} - 1)}{2}$$

After obtaining the segmentation files of lesions in the raw data, we saved the raw data segment through the initial index interval, and used this as the training data for this gain experiment. It should be added that there are different directions in the actual retrieval of raw data (From head to foot or foot to head). We used the same spatial relationship to modify the inequality for different directions and then located the lesion.

392

### 393 Construction of RGM based on CT images

To explore whether the rawdata contained unique information, we built residual fusion models through the rawdata and fused it with CTMs' output to determine whether the rawdata could bring benefits. First, we built three feature extraction networks using the rawdata. Based on the memory need of calculation, we sampled the index dimension of rawdata fragments containing lesions to one-eighth, and the same size was resampling based on the average value by equal interval sampling. For the Channel dimension, we directly removed the data outside the reconstruction

area from both sides, and resampling with the row dimension into half of the size. For model building, we did not modify Densenet121 [21], Resnet18 [22] and Resnext18 [23] in 3D with the purpose of directing the direct gain of nude data as much as possible. The training settings and parameters are detailed in **Supplementary 2**.

The core of the residual fusion model is to obtain the correction of the CT model output, and the origin of the idea is that the learning residual is easier which is mentioned in Resnet. The probabilities of predicting the patients as positive by the CTMs were fused with the predicted probabilities of the rawdata models. This fusion is performed during the training process. Specifically, the probability of predicting one patient as positive was calculated as:

409 
$$Output_{\text{positive}} = Raw_{\text{positive}} + CT_{\text{positive}}$$

410 The probability of predicting one patient as negative was calculated as:

411

 $Output_{negative} = Raw_{negative} + (1 - CT_{positive})$ 

After the output fusion of the CTM and rawdata model, the loss function was used to calculate the loss and optimize the model. The three feature extraction networks built with rawdata were fused with the four representative CTMs described above to obtain four raw gain models respectively, which were: raw gain model-Densenet121 (RGM-DN 1/2/3/4), raw gain model-Resnet18 (RGM-RE 1/2/3/4), raw gain model-Resnext18 (RGM-RX 1/2/3/4). Therefore, we obtain 12 raw gain models. Then the RGMs were compared with the CTMs to evaluate the benefits of the rawdata.

419

### 420 The calculation of the average attention score

For calculating the average attention score of each voxel, we first used the segmentation data of lesion in rawdata to obtain non-lesion area by unary complement. Next, we dotted and summed the segmentation data of lesion and non-lesion areas with the attention matrix. Finally, the average attention score was obtained by dividing the total amount of attention in the two areas by the number of voxels in the segmented regions, respectively. It should be noted that we also normalized the average attention score of the lesion area and the non-lesion area in each rawdata, so as to obtain a more intuitive comparison result.

# 429 **Data availability**

- 430 The image features and output values associated with the CTMs and the RGMs are stored on
- 431 GitHub (https://github.com/CASIAMI/rawdata gain). The original data that support the findings
- 432 of this study are available from the corresponding author upon reasonable request.
- 433

#### 434 Code availability

- 435 Source code for related CT and raw data methods can be found from GitHub
  436 (https://github.com/CASIAMI/rawdata gain).
- 437

# 438 **References**

- 439 [1] Ciccarelli, E., Jacobs, A. & Berman, P. Looking back on the millennium in medicine. *N Engl*440 *J Med* 342, 42-49 (2000).
- Lauwerends, L.J., *et al.* Real-time fluorescence imaging in intraoperative decision making for
  cancer surgery. *The Lancet Oncology* 22, e186-e195 (2021).
- 443 [3] Lehman, C.D., *et al.* Diagnostic accuracy of digital screening mammography with and 444 without computer-aided detection. *JAMA internal medicine* **175**, 1828-1837 (2015).
- [4] Bi, W.L., *et al.* Artificial intelligence in cancer imaging: clinical challenges and applications. *CA: a cancer journal for clinicians* 69, 127-157 (2019).
- 447 [5] Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H. & Aerts, H.J. Artificial intelligence
  448 in radiology. *Nature Reviews Cancer* 18, 500-510 (2018).
- 449 [6] Litjens, G., *et al.* A survey on deep learning in medical image analysis. *Medical image*450 *analysis* 42, 60-88 (2017).
- [7] Lambin, P., *et al.* Radiomics: the bridge between medical imaging and personalized medicine. *Nature reviews Clinical oncology* 14, 749-762 (2017).
- [8] Liu, X., *et al.* A comparison of deep learning performance against health-care professionals in
  detecting diseases from medical imaging: a systematic review and meta-analysis. *The lancet digital health* 1, e271-e297 (2019).
- 456 [9] Killock, D. AI outperforms radiologists in mammographic screening. *Nature Reviews* 457 *Clinical Oncology* 17, 134-134 (2020).
- [10] Cruz Rivera, S., Liu, X., Chan, A.-W., Denniston, A.K. & Calvert, M.J. Guidelines for
  clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI
  extension. *Nature medicine* 26, 1351-1363 (2020).
- [11] Dong, D., *et al.* Deep learning radiomic nomogram can predict the number of lymph node
   metastasis in locally advanced gastric cancer: an international multicenter study. *Annals of Oncology* 31, 912-920 (2020).
- 464 [12] Huang, Y.-q., *et al.* Development and validation of a radiomics nomogram for preoperative
  465 prediction of lymph node metastasis in colorectal cancer. *Journal of clinical oncology* 34,
  466 2157-2164 (2016).
- 467 [13] Mu, W., Schabath, M.B. & Gillies, R.J. Images Are Data: Challenges and Opportunities in

It is made available under a CC-BY-NC-ND 4.0 International license .

- the Clinical Translation of Radiomics. *Cancer Research* **82**, 2066-2068 (2022).
- 469 [14] Gillies, R.J., Kinahan, P.E. & Hricak, H. Radiomics: images are more than pictures, they are
  470 data. *Radiology* 278, 563 (2016).
- [15] Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R. & Rosen, M.S. Image reconstruction by domaintransform manifold learning. *Nature* 555, 487-492 (2018).
- 473 [16] Wang, G., Ye, J.C. & De Man, B. Deep learning for tomographic image reconstruction.
   474 *Nature Machine Intelligence* 2, 737-748 (2020).
- [17] Kalra, M., Wang, G. & Orton, C.G. Radiomics in lung cancer: Its time is here. *Medical physics* 45, 997-1000 (2018).
- [18] Wang, G., Ye, J.C., Mueller, K. & Fessler, J.A. Image reconstruction is a new frontier of
  machine learning. *IEEE transactions on medical imaging* 37, 1289-1296 (2018).
- 479 [19] De Man, Q., et al. A two-dimensional feasibility study of deep learning-based feature
  480 detection and characterization directly from CT sinograms. *Medical physics* 46, e790-e800
  481 (2019).
- [20] Dong, D., *et al.* Abstract CT274: Diagnosis based on signal: The first time break the routinely
  used circle of signal-to-image-to-diagnose. *Cancer Research* 80, CT274-CT274 (2020).
- 484 [21] Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K.Q. Densely connected
  485 convolutional networks. in *Proceedings of the IEEE conference on computer vision and*486 *pattern recognition* 4700-4708 (2017).
- 487 [22] He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in
  488 *Proceedings of the IEEE conference on computer vision and pattern recognition* 770-778
  489 (2016).
- 490 [23] Xie, S., Girshick, R., Dollár, P., Tu, Z. & He, K. Aggregated residual transformations for deep
   491 neural networks. in *Proceedings of the IEEE conference on computer vision and pattern* 492 recognition 1492-1500 (2017).
- 493 [24] Shen, W., *et al.* Multi-crop convolutional neural networks for lung nodule malignancy
  494 suspiciousness classification. *Pattern Recognition* 61, 663-673 (2017).
- 495 [25] Mukherjee, P., et al. A shallow convolutional neural network predicts prognosis of lung
  496 cancer patients in multi-institutional computed tomography image datasets. Nature machine
  497 intelligence 2, 274-282 (2020).
- 498 [26] Ardila, D., *et al.* End-to-end lung cancer screening with three-dimensional deep learning on
  499 low-dose chest computed tomography. *Nature medicine* 25, 954-961 (2019).
- [27] Xu, X., *et al.* MSCS-DeepLN: Evaluating lung nodule malignancy using multi-scale cost sensitive neural networks. *Medical Image Analysis* 65, 101772 (2020).
- [28] Liu, L., Dou, Q., Chen, H., Qin, J. & Heng, P.-A. Multi-task deep model with margin ranking
  loss for lung nodule analysis. *IEEE transactions on medical imaging* 39, 718-728 (2019).
- 504 [29] Xie, Y., et al. Knowledge-based collaborative deep learning for benign-malignant lung
   505 nodule classification on chest CT. *IEEE transactions on medical imaging* 38, 991-1004
   506 (2018).
- 507 [30] Chen, S., Ma, K. & Zheng, Y. Med3d: Transfer learning for 3d medical image analysis. *arXiv* 508 *preprint arXiv:1904.00625* (2019).
- 509

# 510 Acknowledgements

511 This work was supported by the National Key R&D Program of China (2017YFA0205200), National Natural Science Foundation of China (82022036, 91959130, 81971776, 62027901, 512 81930053, 81771924), the Beijing Natural Science Foundation (Z20J00105), Strategic Priority 513 514 Research Program of Chinese Academy of Sciences (XDB38040200), Chinese Academy of Sciences under Grant No. GJJSTD20170004 and QYZDJ-SSW-JSC005, the Project of High-515 Level Talents Team Introduction in Zhuhai City (Zhuhai HLHPTP201703), the Youth Innovation 516 517 Promotion Association CAS (Y2021049) and the China Postdoctoral Science Foundation 518 (2021M700341). The authors would like to acknowledge the instrumental and technical support of 519 Multi-modal biomedical imaging experimental platform, Institute of Automation, Chinese 520 Academy of Sciences.

521

# 522 Author contributions

B.X.H., Y.B.Z., C.X.S., T.T., K.S. and H.L.L. developed the network architecture and 523 524 data/modeling infrastructure, training and testing setup. B.X.H., Y.B.Z., C.X.S., T.T., K.S. and H.L.L. wrote the methods. B.X.H., C.X.S. and T.T. created the figures. B.X.H. and M.J.F. 525 performed statistical analysis. J.T., D.D., Z.Y.L., K.W., Z.Q.W., W.M., S.W., Z.C.T., S.T.Z., 526 J.W.W. and L.Z.S. advised on the modeling techniques. L.X.T., L.W.W., S.P.Z. and J.T.L. 527 provided raw data structure information. D.D., B.X.H., Y.B.Z., C.X.S., T.T., K.S. H.L.L., L.W.W. 528 529 and Y.G. wrote the manuscript. H.M.Z., Y.G., M.W., F.Y.M., L.D., L.L.Z. and S.W. provided clinical expertise and guidance on the study design. H.M.Z., Y.G., M.W., F.Y.M. and L.D. created 530 the clinical datasets, interpreted the data and defined the clinical labels, L.X.T., W.L.W., and F.H. 531 532 created the rawdata sets. J.T., D.D., F.H. and H.M.Z. initiated the project and provided guidance 533 on the concept and design. J.T., F.H. and H.M.Z. supervised the project. 534

### 535 **Competing interests**

536 The authors declare that they have no competing interests.

# 538 Figures and Tables



539

540

Figure 1. Flow chart of rawdata gain experiment.





residual fusion model based on different backbone feature extraction networks; (B) Bar graph of

- reaction AUC gain. DM, Densenet121; RE, Resnet18, RX, Renext18.



- and prediction categories. (B) shows the prediction results of all real categories, CTMs and RGMs.





552

553

554

Figure 4. Lesion trajectory in rawdata and the Grad-Cam graphs of the RGMs.



558

# Table 1 Detailed result statistics of raw data for CT models

Model	True positive proportion	False negative proportion	False positive proportion	True negative proportion	True positive rate	True negative rate
CTM1	0.616	0.199	0.069	0.116	0.756	0.627
RGM-DN1	0.699	0.116	0.091	0.094	0.858	0.508
RGM-RE1	0.543	0.272	0.043	0.141	0.666	0.766
RGM-RX1	0.761	0.054	0.120	0.065	0.934	0.351
CTM2	0.678	0.138	0.101	0.083	0.831	0.451
RGM-DN2	0.656	0.159	0.076	0.109	0.805	0.589
RGM-RE2	0.594	0.221	0.051	0.134	0.729	0.724
RGM-RX2	0.605	0.210	0.058	0.127	0.742	0.686
CTM3	0.460	0.355	0.022	0.163	0.564	0.881
RGM-DN3	0.529	0.286	0.025	0.159	0.649	0.864
RGM-RE3	0.656	0.159	0.072	0.112	0.805	0.609
RGM-RX3	0.685	0.130	0.080	0.105	0.840	0.568
CTM4	0.634	0.181	0.043	0.141	0.778	0.766
RGM-DN4	0.652	0.163	0.047	0.138	0.800	0.746
RGM-RE4	0.681	0.134	0.051	0.134	0.836	0.724
RGM-RX4	0.696	0.120	0.043	0.141	0.853	0.766

	Table 2. The performance of 12 faw gain models in subgroup analysis. S												
A 95	AUC %CI	RGM- DN 1	RGM- RE 1	RGM- RX 1	RGM- DN 2	RGM- RE 2	RGM- RX 2	RGM- DN 3	RGM- RE 3	RGM- RX 3	RGM- DN 4	RGM- RE 4	RGM- RX 4
Age	<=60	0.731 (0.647- 0.825)	0.760 (0.667- 0.844)	0.756 (0.667- 0.837)	0.764 (0.668- 0.845)	0.790 (0.709- 0.874)	0.786 (0.705- 0.867)	0.800 (0.719- 0.870)	0.825 (0.743- 0.895)	0.824 (0.750- 0.894)	0.829 (0.749- 0.903)	0.836 (0.746- 0.915)	0.837 (0.746- 0.924)
Age	>60	0.801 (0.680- 0.908)	0.781 (0.664- 0.894)	0.804 (0.700- 0.900)	0.740 (0.610- 0.859)	0.792 (0.678- 0.901)	0.815 (0.699- 0.932)	0.744 (0.636- 0.833)	0.767 (0.648- 0.863)	0.799 (0.673- 0.914)	0.808 (0.661- 0.926)	0.835 (0.690- 0.945)	0.845 (0.713- 0.949)
0	Male	0.693 (0.584- 0.794)	0.698 (0.591- 0.793)	0.697 (0.601- 0.792)	0.737 (0.626- 0.826)	0.725 (0.608- 0.837)	0.714 (0.600- 0.824)	0.702 (0.605- 0.791)	0.736 (0.634- 0.834)	0.717 (0.617- 0.822)	0.793 (0.695- 0.877)	0.791 (0.686- 0.889)	0.810 (0.707- 0.897)
Sex	Female	0.808 (0.716- 0.898)	0.825 (0.728- 0.901)	0.835 (0.749- 0.901)	0.786 (0.695- 0.871)	0.842 (0.749- 0.916)	0.859 (0.777- 0.927)	0.852 (0.780- 0.921)	0.863 (0.784- 0.920)	0.885 (0.818- 0.945)	0.849 (0.741- 0.944)	0.873 (0.764- 0.958)	0.863 (0.739- 0.954)
Tumor	<=23mm	0.771 (0.677- 0.849)	0.786 (0.689- 0.869)	0.815 (0.732- 0.893)	0.775 (0.685- 0.863)	0.821 (0.743- 0.894)	0.828 (0.752- 0.901)	0.800 (0.722- 0.869)	0.823 (0.736- 0.888)	0.847 (0.781- 0.916)	0.821 (0.729- 0.906)	0.836 (0.744- 0.925)	0.839 (0.731- 0.926)
Size	>23mm	0.737 (0.603- 0.847)	0.746 (0.627- 0.854)	0.709 (0.582- 0.815)	0.748 (0.639- 0.847)	0.753 (0.634- 0.864)	0.754 (0.641- 0.859)	0.759 (0.657- 0.850)	0.782 (0.684- 0.874)	0.775 (0.666- 0.868)	0.822 (0.712- 0.924)	0.833 (0.703- 0.925)	0.831 (0.729- 0.928)

Table 2. The performance of 12 raw gain models in subgroup analysis. S

	superior left	0.679 (0.520- 0.836)	0.667 (0.464- 0.836)	0.743 (0.616- 0.881)	0.638 (0.469- 0.812)	0.683 (0.480- 0.847)	0.737 (0.517- 0.899)	0.824 (0.702- 0.925)	0.812 (0.701- 0.916)	0.810 (0.680- 0.914)	0.861 (0.754- 0.953)	0.885 (0.773- 0.964)	0.909 (0.829- 0.979)
Tumor	inferior left	0.796 (0.638- 0.915)	0.789 (0.652- 0.921)	0.769 (0.600- 0.896)	0.741 (0.584- 0.881)	0.759 (0.607- 0.905)	0.771 (0.599- 0.911)	0.775 (0.633- 0.894)	0.793 (0.656- 0.907)	0.812 (0.650- 0.934)	0.813 (0.651- 0.953)	0.840 (0.688- 0.969)	0.833 (0.661- 0.961)
Location	superior right	0.739 (0.618- 0.855)	0.814 (0.698- 0.912)	0.785 (0.673- 0.876)	0.810 (0.704- 0.900)	0.843 (0.740- 0.930)	0.811 (0.685- 0.901)	0.756 (0.633- 0.864)	0.802 (0.675- 0.909)	0.815 (0.675- 0.899)	0.843 (0.730- 0.936)	0.848 (0.741- 0.934)	0.849 (0.729- 0.939)
	Inferior right	0.821 (0.638- 0.957)	0.809 (0.655- 0.957)	0.795 (0.617- 0.937)	0.872 (0.744- 0.970)	0.849 (0.701- 0.970)	0.849 (0.700- 0.970)	0.786 (0.617- 0.923)	0.835 (0.705- 0.961)	0.812 (0.616- 0.959)	0.721 (0.430- 0.942)	0.764 (0.470- 0.981)	0.775 (0.524- 0.962)

Characteristics	Training	Validation	Р	Test	Р
	cohort (N=166)	cohort (N=55)	value *	cohort (N=55)	value **
Age	$58.80 \pm 10.16$	$58.48 \pm 10.07$	0.06	$59.38\pm8.25$	0.06
Sex			0.17		0.84
Female	80 (0.48)	33 (0.60)		25 (0.45)	
Male	86 (0.52)	22 (0.40)		30 (0.55)	
Tumor height	$28.29 \pm 17.22$	$27.76 \pm 16.91$	0.20	$27.35\pm14.50$	0.20
Pathology subtype			0.90		0.26
Adenocarcinoma	84 (0.71)	26 (0.67)		39 (0.83)	
Squamous carcinoma	22 (0.18)	8 (0.21)		5 (0.11)	
Small cell carcinoma	13 (0.11)	5 (0.13)		3 (0.06)	
Left upper Lobe			0.15		0.25
No	121 (0.73)	46 (0.84)		45 (0.82)	
Yes	45 (0.27)	9 (0.16)		10 (0.18)	
Left lower Lobe			0.88		0.54
No	124 (0.75)	41 (0.75)		44 (0.80)	
Yes	42 (0.25)	14 (0.25)		11 (0.20)	
Right upper Lobe			0.88		0.04
No	118 (0.71)	39 (0.71)		30 (0.55)	
Yes	48 (0.29)	16 (0.29)		25 (0.45)	
Right middle lobe			0.56		0.53
No	154 (0.93)	49 (0.89)		53 (0.96)	
Yes	12 (0.07)	6 (0.11)		2 (0.04)	
Right lower Lobe			0.83		0.77
No	137 (0.83)	44 (0.80)		47 (0.85)	
Yes	29 (0.17)	11 (0.20)		8 (0.15)	
Category			0.91		0.91
Malignancy	135 (0.81)	45 (0.82)		45 (0.82)	
Benign	31 (0.19)	10 (0.18)		10 (0.18)	

Noted that other categories of malignant tumors were not included in the statistics as sparse categories. \* The p Value is the test result of the training cohort and the validation cohort; \*\* the p Value is the test result of the training cohort and the test cohort.

563

564

# Table S2 Detailed gain statistics of raw data for CT models

		-	
ID of CT model	Optimization rate of error sample	Error rate of correct sample	Ratio of at least 2 gain model optimizations
CTM1	0.771 (81/105)	0.329 (57/173)	0.827 (67/81)
CTM2	0.676 (46/68)	0.286 (60/210)	0.848 (39/46)
CTM3	0.724 (76/105)	0.208 (36/173)	0.789 (60/76)
CTM4	0.406 (26/64)	0.079 (17/214)	0.846 (22/26)

566	Table S3. The performance of four CT models in subgroup analysis.											
	AU 95%	JC 6CI	CTM 1	CTM 2	CTM 3	CTM 4						
		<=60	0.790 (0.707-0.861)	0.739 (0.626-0.831)	0.801 (0.729-0.866)	0.831 (0.749-0.904)						
	Age	>60	0.705 (0.546-0.823)	0.725 (0.592-0.849)	0.701 (0.599-0.799)	0.790 (0.672-0.899)						
-	Sor	Male	0.765 (0.680-0.848)	0.748 (0.641-0.866)	0.721 (0.626-0.803)	0.804 (0.706-0.882)						
	Sex	Female	0.752 (0.638-0.858)	0.729 (0.626-0.827)	0.803 (0.703-0.888)	0.823 (0.720-0.920)						
-	Tumor	<=23mm	0.760 (0.647-0.843)	0.696 (0.569-0.809)	0.753 (0.664-0.833)	0.806 (0.699-0.888)						
	Size	>23mm	0.764 (0.670-0.852)	0.798 (0.697-0.890)	0.774 (0.684-0.862)	0.819 (0.719-0.906)						
		superior left	0.802 (0.671-0.901)	0.671 (0.463-0.880)	0.873 (0.749-0.961)	0.921 (0.843-0.980)						
	Tumor	inferior left	0.698 (0.526-0.852)	0.732 (0.557-0.868)	0.698 (0.562-0.840)	0.812 (0.675-0.934)						
	Location	superior right	0.759 (0.615-0.878)	0.732 (0.578-0.852)	0.731 (0.603-0.844)	0.807 (0.683-0.912)						
		Inferior right	0.781 (0.630-0.906)	0.843 (0.712-0.950)	0.769 (0.617-0.901)	0.695 (0.443-0.928)						

# 568 **Table S4.** Scanner information, system parameters and imaging parameters for the Chest CT

569

examinations

Manufacturer	Scanner	System parameters	Contrast-enhanced CT imaging parameters
		Source-to-detector	Tube current, 324mA;
		distance, 1040 mm;	Tube voltage, 100kV;
		Source-to-isocenter	Rotation time, 0.5 s;
Nousoft Modical		distance, 570 mm;	Spiral pitch, 0.9;
Sustama Co	NeuViz Prime	Scanning FOV, 500 mm;	Image matrix, 512x512;
Systems Co.,		Scanning frequency,	Pixel spacing, 0.59~0.94 mm
China		2320 times/round;	Slice thickness, 1mm;
China		Detector channel, 672 pcs;	Kernel, F20;
		Detector row, 64 pcs;	Image time relative to onset of
		Detector channel spacing,	contrast material injection, pre-
		0.625 mm.	contrast 60s~70s.

#### 571 Supplement 1. Implementation and optimization details of typical CT models

572

573 Considering the differences in experimental design between our study and the typical models, we 574 need to make appropriate modifications to make the typical models have the best performances on 575 our dataset. The difference is mainly reflected in data volume and data imbalance. The amount of 576 data in previous articles ranged from 1018 to more than 20,000, much larger than that in this 577 experiment (276). Therefore, we applied small data learning related techniques (including pre-578 trained model and sharing network weight) for four typical CT models. Hence, in our study, a pre-579 trained 3D-resnet18 was used for 3D input, which was pre-trained using eight medical datasets 580 [30], a pre-trained 2D-resnet50 was used for 2D input. To train the subnets effectively using our 581 small dataset, only the last layer (layer 4) and full connected network were trained in this study. In 582 addition, for data imbalance, we used resampling strategy during training.

583 Specific to each of these models, we describe them in detail below. CTM1 was a multi-scale 584 ensemble model, which ensembled three subnets whose networks were all the same, except for the 585 input size. The input data were cropped from CT images using three different sizes,  $32 \times 32 \times 32$ , 586  $48 \times 48 \times 48$ , and  $64 \times 64 \times 64$  pixels. The outputs of three subnets were weighted to obtain the final 587 ensemble results, and a grid search was used to tune the weight values. In addition, AUC loss 588 proposed in original paper was used for dealing with data imbalance. CTM2 combined the 589 features of both the entire CT volume and the region of interest cropped from CT volume, so that 590 all predictions relied on both nodule-level local information and global context from the entire CT 591 volume. In addition, CTM2 was trained with focal loss to mitigate the data imbalance. CTM3 592 designed a deep network with a margin ranking loss to enhance the discrimination capability on 593 ambiguous nodule cases. In our study, 3D input was used to preserve the spatial information of 594 pulmonary nodule. CTM4 was a multi-view deep network, which ensembled the outputs of nine 595 2D-view subnets to characterize the 3D nodule. Each subnet combined three types of image 596 patches, and each patch was input into a pre-trained ResNet-50 network. In this study, all networks 597 shared weights. Finally, nine subnets were used jointly to classify nodules with an average 598 weighting scheme. For model training, the batch size was set as 32. The cross-entropy loss was 599 used as the loss function, and the SGD optimizer was applied. The start learning rates were set as 600 0.001, and the models were trained for 100 epochs.

#### Supplement 2. Model training and statistical analysis methods 601

602

603 For model building, the batch size was set as 64. The Cross-Entropy loss was used as the loss function, and the Adam optimizer was applied. The start learning rates were set as  $1 \times 10^{-5}$ ,  $1 \times$ 604  $10^{-6}$  separately, and the learning rate decay was set as 0.001. Meanwhile, we also added weight 605 606 attenuation. The weight decay was set as 0.01, 0.05, and 0.1 separately. The model was trained for 607 50 epoches and select the best parameters with the lowest loss of validation cohort. The three 608 RGMs built based on the three structures were all trained with the above setting.

609 For statistical analysis, discrete variables and continuous variables are calculated in the chi-

squared test and Man-Whitney U test, respectively. All models are implemented in Python 3.7.3 610

611 (https://www.python.org/) with Numpy (version  $\geq 1.16.4$ ), SciPy (version  $\geq 1.3.0$ ), Matplotlib

612 (version  $\geq$  3.1.1), Scikit-Learn (version  $\geq$  1.10.1), Statsmodels (version  $\geq$  0.12.2) and Pandas

(version  $\geq 0.25.0$ ). All models were trained in python package named Pytorch (version  $\geq 1.10.1$ ; 613

614 https://pytorch.org/) with 4 Graphics Processing Units of NVIDIA TITAN RTX (24G).