**Supplementary methods**

**1. Study design**

This study is a retrospective cohort analysis involving patients who were first evaluated in 2023 at the Department of Thoracic Surgery, The First Affiliated Hospital, Zhejiang University School of Medicine. The exclusion criteria were as follows: 1) non-pulmonary lesions, 2) pulmonary lesions that were not indeterminate nodules, such as pulmonary bullae or cases requiring foreign body removal, 3) patients with a preoperative diagnosis of lung cancer, 4) patients lacking high-resolution CT imaging of the lungs, and 5) patients with pure ground-glass nodules. The patient selection process is illustrated in Supplementary Figure 1.After enrollment, patients were categorized into two groups based on the pathological nature of the nodules: the benign group (human judgment difficulty group) and the malignant group (human judgment accuracy group). CT images from both groups were standardized and provided to the GPT-4o model for analysis.This study was approved by the Ethics Committee of the First Affiliated Hospital, Zhejiang University School of Medicine.

**2. Patient information**

All patient data were extracted from the hospital's electronic medical record database. High-resolution CT images were used for pulmonary evaluation. At our institution, CT reports are typically reviewed by two radiologists. For this study, the extracted imaging features were further reviewed by a third radiologist, who was blinded to the pathological nature (benign or malignant) of the pulmonary lesions. Definition of Comorbidities: Hypertension: Defined as an office-measured systolic blood pressure (SBP) exceeding 140 mm Hg or a diastolic blood pressure (DBP) above 90 mm Hg1.

Diabetes mellitus: Diagnosed according to standardized diagnostic criteria, including: (1) fasting plasma glucose levels ≥7.0 mmol/L; (2) 2-hour postprandial plasma glucose levels ≥11.1 mmol/L following a 75-g oral glucose tolerance test; or (3) glycated hemoglobin levels ≥6.5%. Additionally, in patients presenting with classic symptoms of hyperglycemia (such as polyuria, polydipsia, and unexplained weight loss) or hyperglycemic crisis, a single random plasma glucose measurement ≥11.1 mmol/L was considered diagnostic2. Dyslipidemia: Defined as meeting any of the following criteria: total cholesterol ≥200 mg/dL, triglycerides ≥150 mg/dL, LDL-cholesterol ≥130 mg/dL, or HDL-cholesterol <40 mg/dL (men) or <50 mg/dL (women) 3.

**3. CT image processing and GPT-4o prompt**

To simulate the clinical scenario where physicians review lung CT images to assess nodule characteristics, we selected the thinnest continuous high-resolution CT slices corresponding to the nodules for image extraction. Each image captured the entire viewing interface of the imaging software (Window width: 400; window level: 40), and the number of images represented the layers from the initial appearance to the disappearance of the nodule. The nodule regions were annotated with red square boxes, and all images were processed using a standardized protocol. This process was jointly performed by a clinical physician and a radiologist. We did not manually crop the nodule regions because GPT-4o might require information from the entire image, including comparisons with non-nodule areas, to make accurate judgments. This approach aimed to provide information similar to that used in human judgment and to explore the feasibility of simplifying the application scenario. The images were sequentially annotated with red boxes in Microsoft PowerPoint (v16.92, 2024 Microsoft), exported as PDFs, opened using the Preview app (v11.0) on Mac OS (macOS Sequoia, v15.2), and uniformly exported as sequential PNG images at 100 pixels per inch. These images were then uploaded to the GPT-4o dialog for analysis. The GPT-4o dialog was customized using the "Create a GPT" function for the specific task of differentiating between benign and malignant pulmonary nodules. The following prompt was provided: I will upload multiple layers of lung CT images where specific regions containing lung nodules are marked with red boxes. Please note: The red boxes indicate the general location of the nodules but do not strictly follow the exact edges of the nodules; they may include surrounding normal lung tissue. Analyze the provided images to determine: 1. Whether the nodule is classified as lung cancer (yes or no). 2. The probability of the nodule being benign (non-cancerous). 3. The probability of the nodule being malignant (cancerous). 4. Assign a weight score to the following modules based on their contribution to your judgment process. Use integer scores ranging from -5 to 5, where: Negative scores (closer to -5) indicate features more likely associated with benign nodules. Positive scores (closer to 5) indicate features more likely associated with malignant nodules. Scores closer to 0 indicate minimal or no influence. If a feature is not present, assign it a score of 0. The modules to be scored are: Nodule Shape (e.g., round, lobulated, irregular); Nodule Margins (e.g., smooth, spiculated, indistinct); Bubble Lucency (if present); Walled Cavities (e.g., thin-walled or thick-walled cavities); Pleural Indentation (if present); Other Factors (please specify if applicable). After completing the judgment, I will provide you with the pathological diagnosis (gold standard) of whether the nodule is benign or malignant. To avoid confusion, I will use the following format: "The above images (the latest uploaded) are benign/malignant." Use this feedback to continuously learn the characteristics of benign and malignant nodules and improve your accuracy in subsequent evaluations. Do not include any additional explanations or unrelated information. The output is strictly for educational purposes.

**4. Statistical analyses**

All statistical analyses were performed using R software (v4.2.2). Continuous variables were expressed as median [interquartile range (IQR)] for non-normally distributed data and as mean ± standard deviation (SD) for normally distributed data. Categorical variables were expressed as counts [percentages]. For baseline comparisons, categorical variables were analyzed using the chi-square test or Fisher’s exact test, as appropriate. Continuous variables were compared using the Mann-Whitney U test for non-normally distributed data and the t-test for normally distributed data. To compare the malignancy rate determined by GPT-4o with the assumed 100% malignancy rate by human judgment, a binomial test was used. For comparisons of GPT-4o’s malignancy rates between different pathological groups and random judgment rates, analysis of variance (ANOVA) followed by Tukey’s HSD test was applied for multiple comparisons. The scoring results of GPT-4o’s judgment criteria were visualized using bubble plots, where the size of the bubbles represents the magnitude of the scores, and the specific scores are annotated below the bubbles. For group comparisons of each feature, the Shapiro-Wilk test was first used to assess normality. If all groups met the normality assumption, an independent samples t-test was applied; otherwise, the Wilcoxon rank-sum test was used. The p-values were adjusted for multiple comparisons using the Bonferroni correction, and the adjusted p-values were annotated below the corresponding features. The proportions of actual imaging features across different groups were also visualized using bubble plots, where the size of the bubbles represents the proportion, and the specific proportions are annotated below the bubbles. For group comparisons of each feature, the chi-square test or Fisher’s exact test (if any cell in the contingency table had an expected frequency <5) was used. The p-values were adjusted for multiple comparisons using the Bonferroni correction, and the adjusted p-values were annotated below the corresponding features.

**References**

1. McEvoy, J.W.*, et al.* 2024 ESC Guidelines for the management of elevated blood pressure and hypertension. *Eur Heart J* **45**, 3912-4018 (2024).

2. American Diabetes, A. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. *Diabetes Care* **43**, S14-S31 (2020).

3. Mach, F.*, et al.* 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J* **41**, 111-188 (2020).