Risk-stratified classification of pulmonary nodule malignancy via a machine learning model integrating imaging and cell-free DNA: a model development and validation study (DECIPHER-NODL)

Huiting Wang,[a,i] Hairong Huang,[b,i] Feng Li,[a,i] Ying Deng,[a,i] Changyong Wang,[b] Wei Wei,[c] Song Wang,[d] Dongqin Zhu,[d] Hao Xu,[d] Hua Bao,[d] Zheng Li,[a] Wenjun Ye,[a] Yuan Zhang,[a] Caichen Li,[a] Bo Cheng,[a] Xiwen Liu,[a] Liping Liu,[a] Zheng Li,[e] Jing Yang,[e] Wei Chen,[f] Peng He,[d] Fufeng Wang,[d] Wen Zhong,[e] Weisheng Guo,[g] Yang Shao,[d,h,****] Yi Shen,[b,***] Jianxing He,[a,**] and Wenhua Liang[a,*]

[a]Department of Thoracic Surgery and Oncology, The First Affiliated Hospital of Guangzhou Medical University, China State Key Laboratory of Respiratory Disease & National Clinical Research Center for Respiratory Disease, Guangzhou, 510165, China
[b]Department of Cardiothoracic Surgery, Jinling Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, Jiangsu, 210002, China
[c]Departments of Esophageal Surgery and Thoracic Surgery, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, 210008, Jiangsu, China
[d]Geneseeq Research Institute, Nanjing Geneseeq Technology Inc., Nanjing, 210032, China
[e]Department of Thoracic Surgery and Oncology, Guangzhou National Laboratory, Guangzhou, China
[f]Department of Pulmonology, Taizhou Affiliated Hospital of Nanjing University of Chinese Medicine, Taizhou, 225300, China
[g]Department of Minimally Invasive Interventional Radiology, The Second Affiliated Hospital, School of Biomedical Engineering Guangzhou Medical University, Guangzhou, China
[h]School of Public Health, Nanjing Medical University, Nanjing, 211166, China

The Lancet Regional Health - Western Pacific 2025;64: 101730

Published Online 12 November 2025
https://doi.org/10.1016/j.lanwpc.2025.101730

Summary

Background Accurate risk stratification of pulmonary nodules is critical for early lung cancer detection. This study aimed to improve malignancy classification and invasiveness prediction using machine learning models integrating low-dose computed tomography (LDCT) radiomics and plasma cell-free DNA (cfDNA) fragmentomics.

Methods This multicenter study enrolled 1356 participants across discovery (n = 1147) and external validation (n = 209) cohorts. A deep learning-based imaging model processed LDCT scans for automated lung nodule detection and malignancy classification. A parallel cfDNA model analyzed four whole-genome fragmentation features: copy number variation, fragment size ratio, fragment-based methylation, and mutation context and signature. The two models were integrated via a stacked ensemble algorithm. An invasion prediction model evaluated tumor aggressiveness.

Findings The integrated imaging-cfDNA model outperformed individual models, with an AUC of 0.950 (95% CI: 0.926–0.975) in the internal test set and 0.966 (95% CI: 0.940–0.991) in the external validation. The combined model's specificity increased to 0.60 (95% CI: 0.49–0.71) while maintaining 95% sensitivity, compared to specificities of 0.50 (95% CI: 0.41–0.59) and 0.33 (95% CI: 0.23–0.44) at equivalent sensitivity levels for the imaging and cfDNA models, respectively. The combined model consistently outperformed the other two models across nodule characteristics, with particular improvement for 10–20 mm and pure solid nodules. The invasion prediction model stratified lung cancers with an AUC of 0.884 (internal) and 0.880 (external). Prediction scores increased stepwise with tumor aggressiveness, from adenocarcinoma in situ to minimally invasive adenocarcinoma, and were highest for invasive adenocarcinoma.

Interpretation This multimodal approach enhances pulmonary nodule risk stratification by integrating radiomic and molecular biomarkers. The model significantly improves diagnostic accuracy, potentially reducing unnecessary procedures while minimizing missed diagnoses, supporting its clinical utility in lung cancer screening.

*Corresponding author.
**Corresponding author.
***Corresponding author.
****Corresponding author. Geneseeq Research Institute, Nanjing Geneseeq Technology Inc., Nanjing, 210032, China.
E-mail addresses: liangwh1987@163.com (W. Liang), drjianxing.he@gmail.com (J. He), dr.yishen@163.com (Y. Shen), yang.shao@geneseeq.com (Y. Shao).
[i]These authors contributed equally to this work.
Translation: For the Chinese translation of the abstract see the Supplementary Materials section.

Funding Noncommunicable Chronic Diseases-National Science and Technology Major Project, National Key Research & Development Programme, China National Science Foundation, the Science and Technology Planning Project of Guangzhou, and Guangzhou National Laboratory.

Keywords: Liquid biopsy; Cell-free DNA; Pulmonary nodule; Radiomics; Machine learning

Research in context

Evidence before this study

We conducted a comprehensive PubMed search using combinations of the following keywords: ("lung cancer" OR "pulmonary nodule*" OR "lung nodule*") AND ("computed tomography" OR "low-dose CT" OR "imaging" OR "radiomic*") AND ("cell-free DNA" OR "cfDNA" OR "liquid biopsy" OR "DNA methylation" OR "fragmentation") AND ("machine learning" OR "deep learning" OR "artificial intelligence" OR "predictive model"). Numerous studies have investigated the utility of low-dose CT (LDCT) radiomics and plasma cell-free DNA (cfDNA) fragmentomics for lung nodule classification. However, most multimodal approaches to date have primarily focused on integrating LDCT with cfDNA methylation profiles. Models that combine LDCT with whole-genome fragmentomic features, such as copy number variation (CNV), fragment size ratio (FSR), fragment-based methylation (FBM), and mutation context and signature (MCS), remain limited and insufficiently explored.

Added value of this study

To our knowledge, the DECIPHER-NODL study is the first to integrate deep learning-derived LDCT radiomics with a cfDNA model encompassing four genome-wide fragmentation features (CNV, FSR, FBM, and MCS) via a stacked ensemble algorithm. This multimodal model demonstrated superior accuracy compared to individual imaging or cfDNA models in both internal and external validation cohorts. Additionally, we developed a novel invasion prediction model that effectively stratifies tumor aggressiveness, differentiating adenocarcinoma in situ, minimally invasive adenocarcinoma, and invasive adenocarcinoma. Our findings highlight the complementary value of combining structural imaging features with molecular fragmentation signals to overcome key challenges in lung nodule classification, particularly for small or radiologically ambiguous lesions.

Implications of all the available evidence

Overall, current evidence supports the use of multimodal diagnostic strategies to improve early lung cancer detection and risk stratification. Our study contributes a robust, non-invasive, interpretable, and externally validated model that enhances both malignancy classification and tumor invasiveness assessment. Clinical integration of such an approach has the potential to reduce unnecessary invasive procedures, minimize false positives, and facilitate earlier, more personalized treatment decisions. The model holds particular promise for high-risk populations and clinical scenarios where radiologic assessments alone are inconclusive.

Introduction

Lung cancer remains the leading cause of cancer-related mortality worldwide, with poor survival rates primarily due to late-stage diagnosis.[1] Early cancer diagnosis provides a unique opportunity for curative treatment, and low-dose computed tomography (LDCT) has been widely adopted as a screening tool for high-risk individuals, significantly reducing lung cancer mortality.[2–4] While LDCT effectively detects pulmonary nodules, differentiating between benign and malignant nodules remains a major clinical challenge. According to the National Lung Cancer Screening Trial (NLST), 24.2% of participants had at least one nodule, yet only 3.6% of these LDCT-detected nodules were malignant.[2] Other studies have also reported high false positive rates, with up to 20% of surgically resected nodules and 38% of biopsied nodules ultimately found to be benign.[5,6] Additionally, a substantial proportion (~50–76%) of detected nodules are indeterminate, leading to unnecessary invasive procedures, patient anxiety, and increased healthcare burden. Moreover, the interpretation of CT scans is time-consuming, labor-intensive, and susceptible to inter- and intra-rater variability. Therefore, an automated and accurate classification method that supplements existing diagnostic workflows by evaluating pulmonary nodule malignancy risk would be beneficial for risk stratification and clinical decision-making.

In this context, artificial intelligence (AI)-based approaches for automated lung nodule detection and classification are highly desirable. Most algorithms employ a two-step strategy, with a first step to detect candidate nodules with a high specificity, followed by a second step to reduce false positives.[7] Once a nodule is detected, measured, and classified, management decisions can be determined based on current clinical guidelines.[8–11] Patients with highly suspicious nodules typically undergo histological sampling,

fluorodeoxyglucose positron emission tomography (FDG-PET), nonsurgical biopsy, or surgical resection, whereas those with indeterminate nodules are often monitored through serial CT surveillance. Meanwhile, cell-free DNA (cfDNA) fragmentomic analysis has emerged as a promising liquid biopsy approach, offering insights into tumor-derived genomic alterations for early cancer detection. Features such as fragment size ratio (FSR),[12] copy number variations (CNV),[13] methylation patterns,[14] and nucleosome positioning[15] provide valuable cancer signals. However, individual cfDNA features often lack sufficient diagnostic accuracy, particularly for early-stage cancers. Integrating multiple cfDNA fragmentomic features with LDCT-based image models holds significant potential to enhance early lung cancer detection and risk stratification, addressing current diagnostic limitations and improving clinical outcomes.

In the DECIPHER-NODL study, we propose a machine learning (ML)-based approach to classify lung nodule malignancy risk by integrating radiomic features from LDCT scans with cfDNA fragmentomic characteristics. Our model aims to address the limitations of current diagnostic methods by improving accuracy, reducing unnecessary interventions, and providing a robust, non-invasive risk stratification tool. By leveraging multi-modal data and advanced computational techniques, our study contributes to the growing body of research on AI-assisted lung cancer diagnostics and personalized risk assessment.

Methods

Study population analyzed

The discovery cohort consisted of 1147 retrospectively collected participants (831 with lung cancer and 316 with benign nodule) who presented at the First Affiliated Hospital of Guangzhou Medical University between May 16, 2014, and December 10, 2021, with positive LDCT findings and surgically confirmed lung nodules. The external test cohort (n = 209; 158 with lung cancer and 51 with benign nodule) was prospectively enrolled at the General Hospital of Eastern Theater Command between October 11, 2021, and July 11, 2023. Blood samples were collected at each patient's first clinic visit before any systemic treatment. All Digital Imaging and Communications in Medicine (DICOM) images were recorded onto CDs and sent to the central laboratory for image feature extraction. Written informed consent was obtained from all participants. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the First Affiliated Hospital of Guangzhou Medical University (No. 2021-95).

Eligible patients in both cohorts were adults ≥18 years old. Enrollment was not restricted by smoking history or underlying lung disease to capture the full spectrum of surgically treated nodules. Standardized radiologic criteria based on Lung-RADS (version 1.0) were applied to determine whether a definitive diagnosis could be made based on LDCT.[16] Patients were included if they had pulmonary nodules primarily 5–30 mm in diameter on LDCT and surgical intervention was clinically anticipated. This included both malignant nodules and high-risk benign nodules that are difficult to distinguish from malignancies. All patients were required to provide written informed consent and have sufficient qualified blood samples available for analysis.

Exclusion criteria included pregnancy or lactation, confirmed diagnosis of any other cancer, prior cancer treatment (surgery, radiotherapy, chemotherapy, targeted therapy, or immunotherapy) before blood sampling, current febrile illness or inflammatory disease requiring treatment within 14 days before blood sampling, history of organ transplantation or non-autologous (allogeneic) bone marrow or stem cell transplantation, and any condition deemed unsuitable for study enrollment by the investigator. Representative examples include uncontrolled cardiovascular disease (e.g., NYHA [New York Heart Association] class III-IV heart failure, recent myocardial infarction), active infections (e.g., uncontrolled HIV [human immunodeficiency virus], replicating HBV [hepatitis B virus]/HCV [hepatitis C virus]), severe autoimmune disorders, long-term high-dose immunosuppressive therapy, psychiatric or cognitive disorders impairing consent or compliance, poor adherence to prior clinical protocols, imaging contraindications, or laboratory abnormalities beyond protocol thresholds. For patients with multiple nodules, only the most suspicious nodule was included in the analysis, and all performance metrics were computed at the patient level.

Blood sample processing and low-coverage whole-genome sequencing

Plasma sample collection, cfDNA extraction, library preparation, and WGS were conducted sequentially at a clinical testing laboratory certified by the Clinical Laboratory Improvement Amendments (CLIA) and accredited by the College of American Pathologists (CAP) (Nanjing Geneseeq Technology Inc., Nanjing, China). Blood samples from the discovery and external test cohorts were processed in batches: cfDNA extraction for the discovery cohort was completed in 2021, with sequencing finalized in 2024, while all sample processing and sequencing for the external test cohort were completed in 2023. Importantly, blood samples were processed within 48 h, and cfDNA was extracted promptly after plasma separation and frozen directly at −80 °C until use, without any intermediate freeze–thaw cycles. In brief, 10 mL of peripheral blood was collected from each eligible patient in Streck Cell-free DNA BCT tubes (Streck, #218962). Plasma was separated using a two-step centrifugation protocol: an initial

centrifugation at 1600 g for 10 min at 4 °C, followed by centrifugation of the supernatant at 16,000 g for 10 min at 4 °C. cfDNA was extracted immediately after plasma separation using the QIAamp Circulating Nucleic Acid Kit (Qiagen Cat. No. 55114) and then frozen directly at −80 °C until use. cfDNA concentration was assessed using the dsDNA HS assay kit on a Qubit 3.0 fluorometer (Life Technology, US), and quality was determined on the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, USA). For each sample, 5–10 ng of cfDNA was used to construct a PCR-free WGS library using the KAPA Hyper Prep kit (KAPA Biosystems). Libraries were sequenced on the DNBSEQ-T7 platform at 5× coverage.

Sequencing data analysis

Raw sequencing data underwent initial quality control using Trimmomatic,[17] where sequencing adapters and 3' low-quality bases were removed. Duplicated reads were filtered using the Picard toolkit (http://broadinstitute.github.io/picard/). Clean reads were then mapped to the human reference genome (Hg19) using the Burrows-Wheeler Aligner (BWA-mem, v0.7.12; https://github.com/lh3/bwa/tree/master/bwakit) with default parameters. Samples with coverage less than 3× were excluded from downstream analysis due to insufficient ability for feature calling.

Imaging model development

To mitigate inter-scanner variability resulting from different LDCT devices across centers, all DICOM images were resampled to a uniform voxel spacing (1.25 mm × 1.0 mm × 1.0 mm) using trilinear interpolation for image data and nearest-neighbor interpolation for mask data, implemented via the MONAI framework (https://docs.monai.io/en/stable/transforms.html#spacingd). This spatial normalization ensures that each voxel represents the same physical dimension, reducing scanner-dependent bias while preserving anatomical fidelity.

Based on these preprocessed and standardized images, we performed a two-stage deep learning pipeline comprising lung nodule detection and benign-malignant classification. For nodule detection, the process begins with segmentation of the lung region from raw CT images using the U-Net (R231) model implemented in the Lungmask tool (https://github.com/JoHof/lungmask).[18] This step isolates the lung parenchyma and eliminates interference from extra-thoracic tissues. Image intensity values are truncated to the range [−1024, 300] Hounsfield Units (HU) and normalized to the [0,1] interval to preserve relevant features of the lung parenchyma and nodules.[19]

For our study, nodule detection was performed using a pre-trained 3D RetinaNet model, implemented through the MONAI framework (https://github.com/Project-MONAI/MONAI),[20] using the MONAI lung nodule CT detection model available on NVIDIA NGC (https://catalog.ngc.nvidia.com/orgs/nvidia/teams/monaitoolkit/models/monai_lung_nodule_ct_detection). This model generated an initial set of nodule candidates, which were subsequently manually reviewed and confirmed by an experienced radiologist, establishing the ground truth for our predictive model. To achieve this, the model was trained on the LUNA16 dataset and utilizes a 3D ResNet backbone integrated with a feature pyramid network (FPN) structure. The anchor generator was configured with three base shapes — [[6,8,4], [8,6,5], [10,10,6]] — to accommodate nodules of varying sizes. Multiple data augmentation strategies were applied during training to enhance generalizability, including random rotations (±30°), flipping (50% probability along each axis), scaling (0.7–1.4×), and Gaussian noise addition ($\sigma = 0.1$). Focal Loss ($\alpha = 0.25$, $\gamma = 2$) was adopted as the classification loss function, and parameters were updated using the stochastic gradient descent (SGD) optimizer (learning rate = 1e-2, momentum = 0.9). A learning rate schedule with a 10-epoch warm-up followed by stepwise decay every 150 epochs ($\gamma = 0.1$) was employed. Each training batch comprises four image patches of size 192 × 192 × 80, and training was accelerated via mixed-precision computing on an NVIDIA 4090 GPU. The training process spans 300 epochs, and the model with the best performance on the validation set is retained. During model validation, a sliding window inference strategy is applied with a window size of 512 × 512 × 208 and 25% overlap. Non-maximum suppression (NMS) with an IoU threshold of 0.22 is used to retain the top 100 predicted bounding boxes with the highest confidence scores. Model performance was primarily evaluated using the COCO evaluation metrics, with mean average precision (mAP) at IoU = 0.1 as the main criterion (Supplementary Material 1).

All detected nodules undergo standardization by cropping 3D image patches centered on the nodules' centroid, with a uniform size of 32 × 32 × 32 voxels to ensure input consistency. For malignancy classification, a 3D convolutional neural network (3D CNN) model is constructed using the AutoKeras platform (http://jmlr.org/papers/v24/20-1355.html), trained on an internal dataset. The model employs the Binary Focal Cross entropy loss function with a focal factor ($\alpha = 0.75$, $\gamma = 2.0$), which increases sensitivity to minority classes and emphasizes hard-to-classify samples. Model construction is automated via neural architecture search (NAS) within the AutoKeras framework. The search is configured with a maximum of 200 trials, using the area under the curve (AUC) of the validation set as the primary selection metric, while also monitoring the F1 score. The batch size is fixed at 16 to accommodate GPU memory constraints. The final classifier is selected as the model achieving the highest validation AUC among the 200 candidate architectures.

Whole-genome cfDNA fragmentation features

Four cfDNA features related to fragmentation patterns were extracted from WGS data: copy number variation (CNV), fragment size ratio (FSR), fragment-based methylation (FBM), and mutation context and signature (MCS).

Copy number variation (CNV)

The copy number variation (CNV) feature extraction was performed using ichorCNA, as previously described by Wan et al.[13] The genome was segmented into 1 Mb non-overlapping bins (2475 total) based on the reference genome. Coverage depth for each bin was corrected for GC content and compared against the software baseline, after which ichorCNA computed the log2 ratio for each bin.

Fragment size ratio (FSR)

The extraction of fragment size ratio (FSR) features followed the method introduced by Mathios et al.[12] The genome was partitioned into 5 Mb non-overlapping bins, resulting in 541 bins with an average GC content $\geq$0.3 and average mappability $\geq$0.9. Reads shorter than 100 bp or longer than 220 bp were filtered out, and the remaining short (100–150 bp) and long (151–220 bp) fragments were standardized and transformed into z-scores with a mean of zero and standard deviation of 1. The ratio of short-to-long fragments was computed for each bin.

Mutation context and signature (MCS)

Raw FASTQ files were trimmed and aligned to the hg19 human reference genome, followed by duplicate read removal. Reads were then sorted and indexed into BAM files using SAMtools (v1.9), and the average sequencing depth was calculated. Unpaired and low-quality reads (base quality <30 or mapping quality <40), as well as those with alternative alignments or template lengths exceeding 300 bp were excluded. Reads mapping to repeats or low-complexity regions were masked before downstream analyses. Paired reads were retained only if they contained at least one mismatch to the reference genome at the center of a 3-nt context, where the substitution was concordant between R1 and R2. Insertions and deletions were excluded.

Before aggregating single-base substitution (SBS) features, rigorous quality filters were applied to remove germline mutations and low-quality noise from cfDNA, thereby minimizing false positives. Germine mutations were filtered using an in-house panel of normals and gnomAD (v2.0). Next, a machine learning ensemble model,[21] consisting of a convolutional neural network (CNN) and a multilayer perceptron (MLP), was employed to further eliminate low-quality noise and generate a high-confidence ctDNA SNV mutation list. True labels were assigned to high-confidence ctDNA SNV calls in lung cancer patients by comparison to high-purity tumor samples and white blood cells, while false labels were assigned to cfDNA variant-containing fragments from healthy controls processed in the same batch and sequenced under the same settings. Samples used for ctDNA SNV filtering gold standards were excluded from downstream analyses, such as cancer vs. non-cancer classification. For modeling, the CNN model integrated various fragment-level features, including reference genome sequences, SBS patterns, fragment length, and quality metrics (e.g., read edit distance and position within the read). The MLP model classified regional genomic features overlapping with mutation-laden fragments, such as replication timing, which correlates with mutation frequency. Outputs from both the fragment-level CNN and regional MLP models were combined in an ensemble framework, leveraging spatial genomic context and nucleotide-level sequence information to distinguish ctDNA SNVs from background cfDNA artifacts.

The final mutation list was used to generate the mutation context and signature (MCS) features. By incorporating the adjacent upstream and downstream nucleotides of each SBS event, classified into six types (C > A, C > G, C > T, T > A, T > C, and T > G), a total of 96 distinct SBS sequence patterns (6 × 4 × 4) were defined. The count of each pattern was normalized by the mean sequencing depth. The final MCS feature comprised a 96-context mutational profile, with contributions fitted to COSMIC SBS signatures (ver. 3.3.1, GRCh37) using the R package "MutationalPatterns".

Fragment-based methylation (FBM)

The extraction of fragment-based methylation (FBM) features was adapted from Zhou et al.,[14] examining cleavage patterns around cytosine-phosphate-guanine (CpG) sites to reflect the methylation status of cfDNA. Theoretically, increased cfDNA cleavage at adjacent cytosines indicates methylated CpG sites, whereas unmethylated sites exhibit reduced cleavage. Within *Alu* regions, fragment ratios were computed for 8 patterns where the 5' end breakpoint is either CGN or NCG, along with the CGN/NCG ratio, resulting in 9 features. Beyond *Alu* regions, FBM extends to genome-wide CGCG sites, where fragment ratios were calculated for 10 patterns based on two categories: CGC and NCG (C1) or CGN and GCG (C2), alongside the CGC/NCG and CGN/GCG ratios, contributing an additional 12 features. In total, 21 FBM features are extracted and processed using a Fully Connected Neural Network (FCN) to capture intricate relationships within the methylation landscape, leveraging the interplay between methylation patterns and cfDNA fragmentation to enhance cancer detection, classification, and monitoring with high precision.

cfDNA model development

Each cfDNA feature type (CNV, FSP, FBM, and MCS) was used as input to construct base models using the

H2O AutoML package,[22] which integrates four algorithms: gradient boosting machine (GBM), generalized linear model (GLM), XGBoost, and neural networks. Hyperparameter tuning was conducted using a random grid search approach over a list of candidate values. Model performance was evaluated via 5-fold cross-validation, where the dataset was split into five subsets, with four used for training and one for validation in each iteration. Additionally, runtime constraints, ensemble techniques, and the use of an independent temporal dataset were implemented to enhance model generalizability. For each feature type, the four models with the best AUC were preserved, and the mean prediction score from these models was used as the final classification output.

Stacked ensemble model development

The stacked ensemble model was developed by fitting a GLM using scores from both the imaging and cfDNA models as two individual covariates. A complete case analysis was employed, training the GLM only on participants with data available for both modalities. To optimize performance and reduce overfitting, we used 5-fold cross-validation in the training set. In each iteration of the 5-fold cross-validation, the entire training cohort was divided into five subsets, with stratification based on malignancy classification. Each subset was iteratively used as the test set, while the remaining four subsets served as the training set. For each train-test pair, a logistic regression model was independently trained on the training set and evaluated on the corresponding test set, and the average performance across the five test sets was used as the cross-validation result. Once the cross-validation was completed, the final model was locked and evaluated in the test sets with all parameters fixed.

The prediction score of the combined model for benign-malignant classification was derived using the following formula:

$$\text{Prediction Score} = \frac{1}{1 + e^{-\left(-8.94769160 + 5.49825152 \times \text{Image score} + 9.99599241 \times \text{cfDNA score}\right)}}$$

where the image score is the risk score generated by the image model, and the cfDNA score is the risk score generated by the cfDNA model.

For the invasiveness prediction task, only patients with pathologically confirmed lung cancer were included in both training and testing, as the goal was to differentiate tumor aggressiveness rather than to distinguish benign from malignant lesions. The prediction score of the invasion-combined model was derived as follows:

$$\text{Prediction Score (Invasion)} = \frac{1}{1 + e^{-\left(-12.22995711 + 14.79213838 \times \text{Image score} + 4.39406150 \times \text{cfDNA score}\right)}}$$

Statistical analysis

All statistical analyses were performed using R software (v4.3.2). Continuous variables were reported as means and ranges, while categorical variables were presented as counts. Student's t-tests or Wilcoxon rank-sum tests were used for comparing continuous variables, and Chi-square tests or Fisher's Exact tests were used for categorical variables, as appropriate. Receiver operating characteristic (ROC) analysis was conducted using the "pROC" R package, with AUC values assessing discriminatory ability and DeLong's test used for comparisons between models. Multiple testing corrections within each subgroup were applied using the Benjamini-Hochberg method. For proportion-based performance metrics, 95% confidence intervals (CIs) were derived from the exact binomial distribution. Model calibration was assessed using calibration curves, calibration intercept, calibration slope, and Brier score to evaluate the agreement between predicted probabilities and observed outcomes. Decision-curve analysis (DCA) was performed to determine the potential clinical usefulness of the models across a range of threshold probabilities. Net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were calculated to quantify the incremental predictive value of the integrated model compared with the imaging-only and cfDNA-only models. Bootstrap resampling was used to estimate 95% CIs for calibration, DCA, NRI, and IDI. The stacked model, which integrates both imaging and cfDNA features, was designed to identify malignant lung nodules for further clinicopathological examination, including PET-CT and tissue biopsy, while directing low-risk patients to active LDCT surveillance. Therefore, a sensitivity threshold of 95% was set, allowing performance comparisons across models. Unless otherwise specified, all reported *P*-values were two-tailed, with statistical significance defined as $P < 0.05$.

Ethics approval

The study was approved by the Ethics Committee of the First Affiliated Hospital of Guangzhou Medical University (Approval number: No. 2021-95; Approval date: September 30, 2021). Written informed consent was obtained from each patient before sample collection.

Role of the funding source
The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Dataset characteristics
We recruited a total of 1356 participants for the development and evaluation of three models, including 1147 in the discovery cohort and 209 in the external test cohort (Fig. 1; Figures S1 and S2). The discovery cohort consisted of 831 lung cancer patients and 316 participants with surgically confirmed benign nodules. Of these, 417 participants were used to train the imaging-cfDNA combined model (hereafter referred to as the "combined model"), while 288 were allocated to the internal test set for model performance evaluation (Table S1). The combined model was independently validated on an external cohort comprising 136 malignant and 24 benign nodules. Among the 705 participants in the discovery cohort, 78.1% (550/705) were above 50 years old, and approximately half were male. About 45% of nodules were subsolid, including 18.4% (1330/705) pure GGOs and 25.7% (181/705) part-solid nodules. In terms of nodule size, 43.5% (307/705) were classified as intermediate (10–20 mm), while large (≥20 mm) and small (<10 mm) nodules accounted for 38.9% and 17.4%, respectively. Overall, 74.1% (522/705) of cases were malignant, whereas 25.9% (183/705) were benign. Two-thirds of participants with malignant nodules had stage I/II cancer. The external test set comprised 160 participants, with 59.4% females and 40.6% males. The participant demographic and clinical characteristics are shown in Table S1.

Performance evaluation of the imaging model
The imaging model, developed using deep learning (DL) algorithms for automated nodule detection, lung window filtering, and benign-malignant classification, demonstrated strong diagnostic accuracy across datasets (Fig. 1; Fig. 2a; Figures S1 and S3). The baseline characteristics of participants used for imaging model construction and validation are detailed in Table S2. The model achieved an AUC of 0.934 (95% CI: 0.915–0.951) in the training set, 0.883 (95% CI: 0.851–0.915) in the internal test set, and 0.906 (95% CI: 0.849–0.963) in the external test set (Fig. 2b). Imaging scores were significantly higher in malignant cases compared to benign nodules across all datasets ($P < 0.001$, Wilcoxon rank-sum test; Fig. 2c). Given that the lung nodules were suspicious and surgery was anticipated based on LDCT scans, our study prioritized high sensitivity to minimize the risk of missed diagnoses. Using a cutoff value of 0.292 calculated based on 95% sensitivity on the training set, the imaging model achieved a specificity of 0.50 (56/112) in the internal test set and 0.29 (7/24) in the external test set. The lower specificity in the external test set was likely influenced by its patient demographics and the limited number of benign nodules, which may amplify variance. These findings underscore the need for further

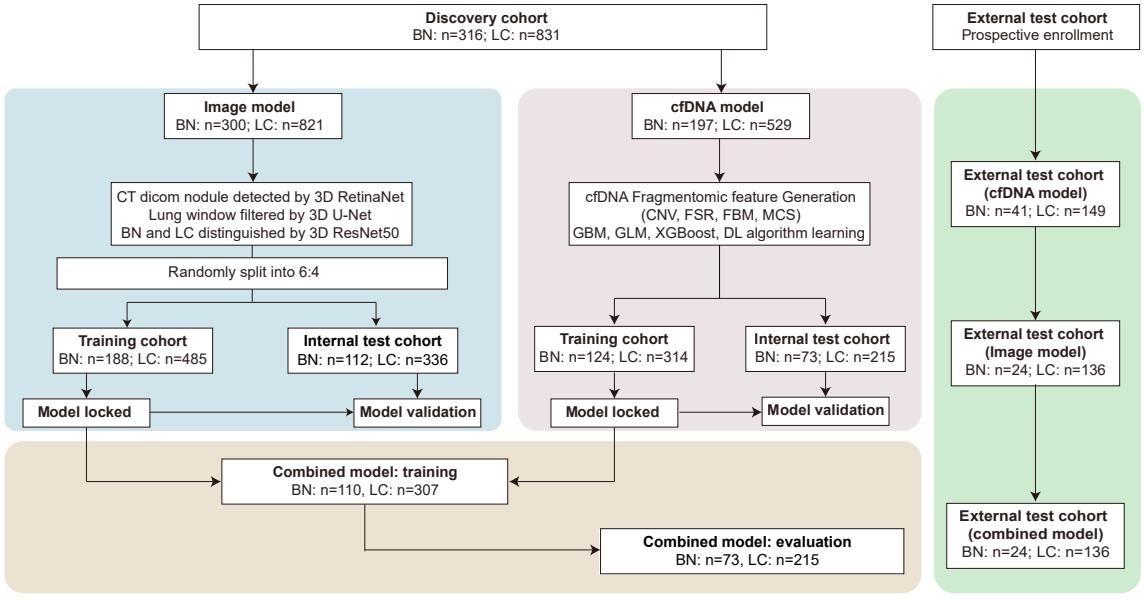

***Fig. 1:* Flowchart depicting the workflow for developing malignancy prediction models**. Abbreviations: BN, benign nodule; LC, lung cancer; CT, computed tomography; CNV, copy number variation; FSR, fragment size ratio; FBM, fragment-based methylation; MCS, mutation context and signature; GBM, gradient boosting machine; GLM, generalized linear model.

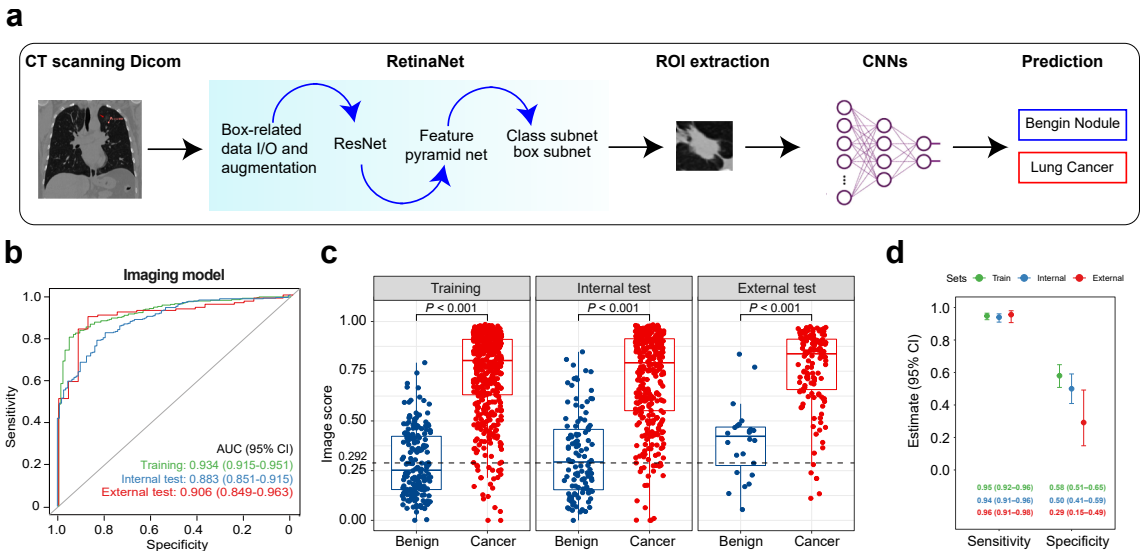

Fig. 2: **Construction and evaluation of the imaging model for lung cancer prediction. (a) Workflow for constructing the imaging model**. (**b**) Receiver operating characteristic analyses assessing the imaging model in malignancy prediction across datasets. (**c**) Imaging score distribution between benign and cancer samples. (**d**) Performance evaluation of the imaging model using a 0.292 cutoff in the training set (95% sensitivity) across datasets.

evaluation in larger cohorts to improve specificity and generalizability.

Performance evaluation of the cfDNA model

The cfDNA model was developed based on four whole-genome fragmentation features derived from plasma ctDNA, including copy number variation (CNV), fragment size ratio (FSR), mutation context and signature (MCS), and fragment-based methylation (FBM) (Fig. 1; Fig. 3a; Figures S1 and S4). These features were selected for their ability to provide comprehensive insights into tumor biology by analyzing ctDNA characteristics. The baseline characteristics of participants used for cfDNA model construction and validation are presented in Table S3. The cfDNA model achieved an AUC of 0.860 (95% CI: 0.821–0.900) in the training set, 0.855 (95% CI: 0.807–0.904) in the test set, and 0.858 (95% CI: 0.791–0.924) in the external set (Fig. 3b). cfDNA scores were markedly higher in malignant cases compared to benign nodules ($P < 0.001$; Fig. 3c). Using a cutoff value of 0.582 for 95% sensitivity on the training set, the model accurately identified 94.0% (202/215) and 98.7% (147/149) of malignant nodules, while maintaining reasonable specificity across the test sets (internal test set: 0.33; external test set: 0.44) (Fig. 3d).

Improved malignancy risk assessment via combining the imaging and cfDNA models

Theoretically, the imaging model may excel at capturing structural and morphological characteristics, while the cfDNA model offers molecular insights that can reveal genetic changes associated with malignancy. Building on previous evaluations of single models, we developed an ensemble stacked model that integrates the strengths of both models to further improve diagnostic accuracy (Fig. 1; Figure S1; Table S4). Our results demonstrated that the combined model significantly outperformed single-feature models, achieving an AUC of 0.952 (95% CI: 0.933–0.971) in the training set, 0.950 (95% CI: 0.926–0.975) in the internal test set, and 0.966 (95% CI: 0.940–0.991) in the external test set (Fig. 4a). Additionally, the combined model generated prediction scores that effectively distinguished between malignant and benign nodules ($P < 0.001$; Fig. 4b). No significant differences were observed among benign nodule subtypes (Figure S5). Notably, the specificity improved from 0.50 (95% CI: 0.41–0.59) for the imaging model and 0.33 (95% CI: 0.23–0.44) for the cfDNA model to 0.60 (95% CI: 0.49–0.71) for the combined model in the internal test set (Fig. 4c). In independent validation, the specificity further increased from 0.44 (95% CI: 0.30–0.59) in the cfDNA model to 0.46 (95% CI: 0.28–0.65) in the combined model. Calibration analysis showed good agreement between predicted and observed malignancy probabilities (Figure S6a, b). In the internal test set, the intercept was –0.192 (95% CI: –0.646 to 0.262), the slope 1.272 (95% CI: 0.968–1.577), and the Brier score 0.077 (95% CI: 0.060–0.097). In the external test set, the intercept was –0.750 (95% CI: –1.748 to 0.248), the slope 1.429 (95% CI: 0.876–1.982), and the Brier score 0.056 (95% CI: 0.032–0.081). Hosmer–Lemeshow tests were non-significant in both cohorts (internal: $P = 0.555$;

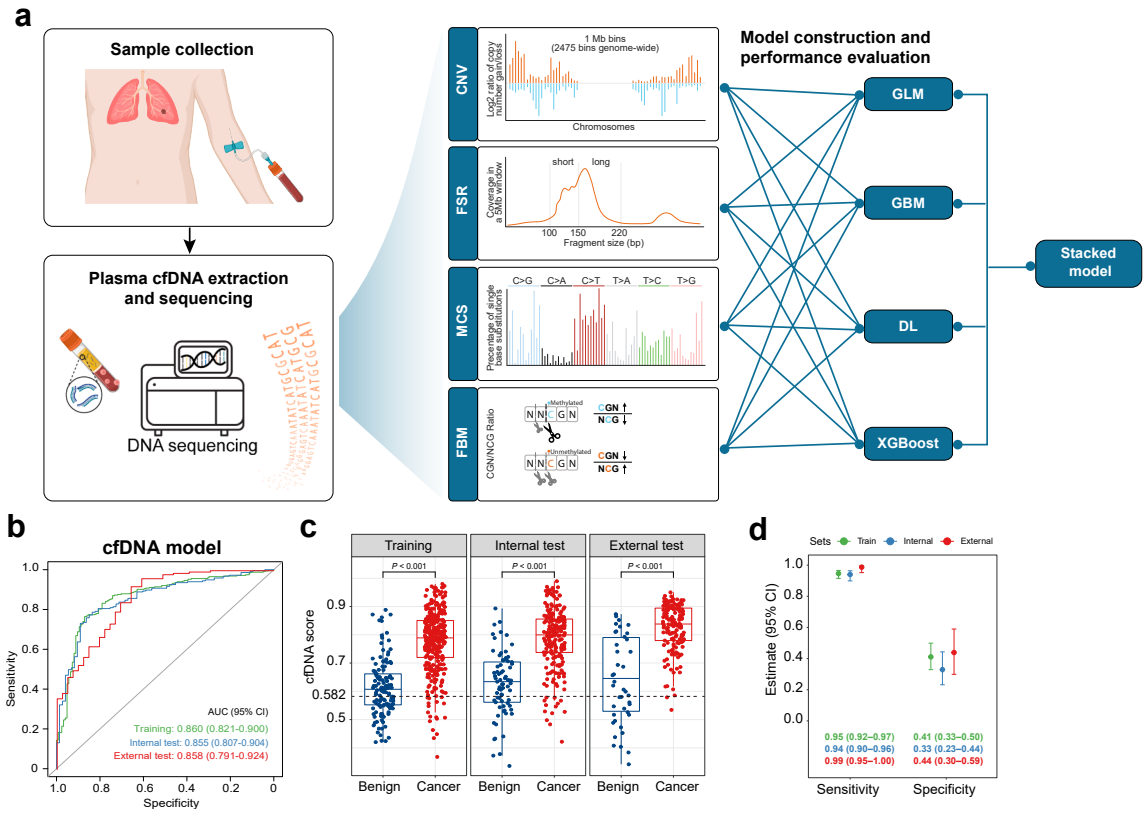

***Fig. 3:* Utilizing plasma-derived fragmentation features for lung cancer prediction**. (**a**) Workflow for constructing the cfDNA model integrating four cfDNA features, including copy number variation (CNV), fragment size ratio (FSR), mutation context and signature (MCS), and fragment-based methylation (FBM). (**b**) Receiver operating characteristic analyses assessing the cfDNA model in malignancy prediction across datasets. (**c**) Distribution of cfDNA scores between benign and cancer samples. (**d**) Performance evaluation of the cfDNA model using a 0.582 cutoff in the training set (95% sensitivity) across datasets.

external: P = 0.594), indicating adequate calibration. DCA suggested higher net clinical benefit of the combined model compared with single-feature models, and the incremental predictive value was supported by NRI and IDI analyses (Figure S6c, d; Table S5). To ensure that performance comparisons between the imaging-only, cfDNA-only, and combined models were not affected by cohort differences, we re-evaluated all models on a unified subset of participants with complete imaging and cfDNA data. The results were highly consistent with those obtained from the full cohorts, with no meaningful differences in AUC, sensitivity, or specificity (Table S6). This confirms that our performance comparisons are robust and not influenced by variations in cohort composition.

The diagnostic performance of the combined model was further evaluated across various clinical scenarios. The cfDNA model generally showed lower performance than the imaging model, except in pure solid nodules, where it achieved comparable or slightly higher AUCs in test sets (Fig. 4d). Notably, incorporating cfDNA features significantly improved the performance of the combined model over imaging alone for nodules measuring 10–20 mm (adjusted P = 0.014, DeLong's test) and for pure solid nodules (adjusted P = 0.011, DeLong's test), yielding substantially higher AUCs (Fig. 4e). In addition, the combined model also significantly outperformed the cfDNA model for lesion measuring 10–20 mm in both the training (AUC: 0.97 vs. 0.87; adjusted P = 0.009) and internal test sets (AUC: 0.99 vs. 0.89; adjusted P = 0.008), as well as for pure solid lesions in the training (AUC: 0.93 vs. 0.85; adjusted P < 0.001) and internal tests (AUC: 0.93 vs. 0.85; adjusted P = 0.010). In multivariate logistic regression adjusting for clinical and radiological characteristics, older age, subsolid nodules, and higher combined model score emerged as independent predictors of lung nodule malignancy in the combined test sets (n = 448; 351 cancer and 97 benign) (Table 1). Collectively, the combined model provides the best diagnostic accuracy across various clinical scenarios, as indicated by the consistently high AUCs. The performance of individual methods may vary depending on the nodule size and radiological type.

Furthermore, our combined model demonstrated a superior ability for malignancy risk stratification

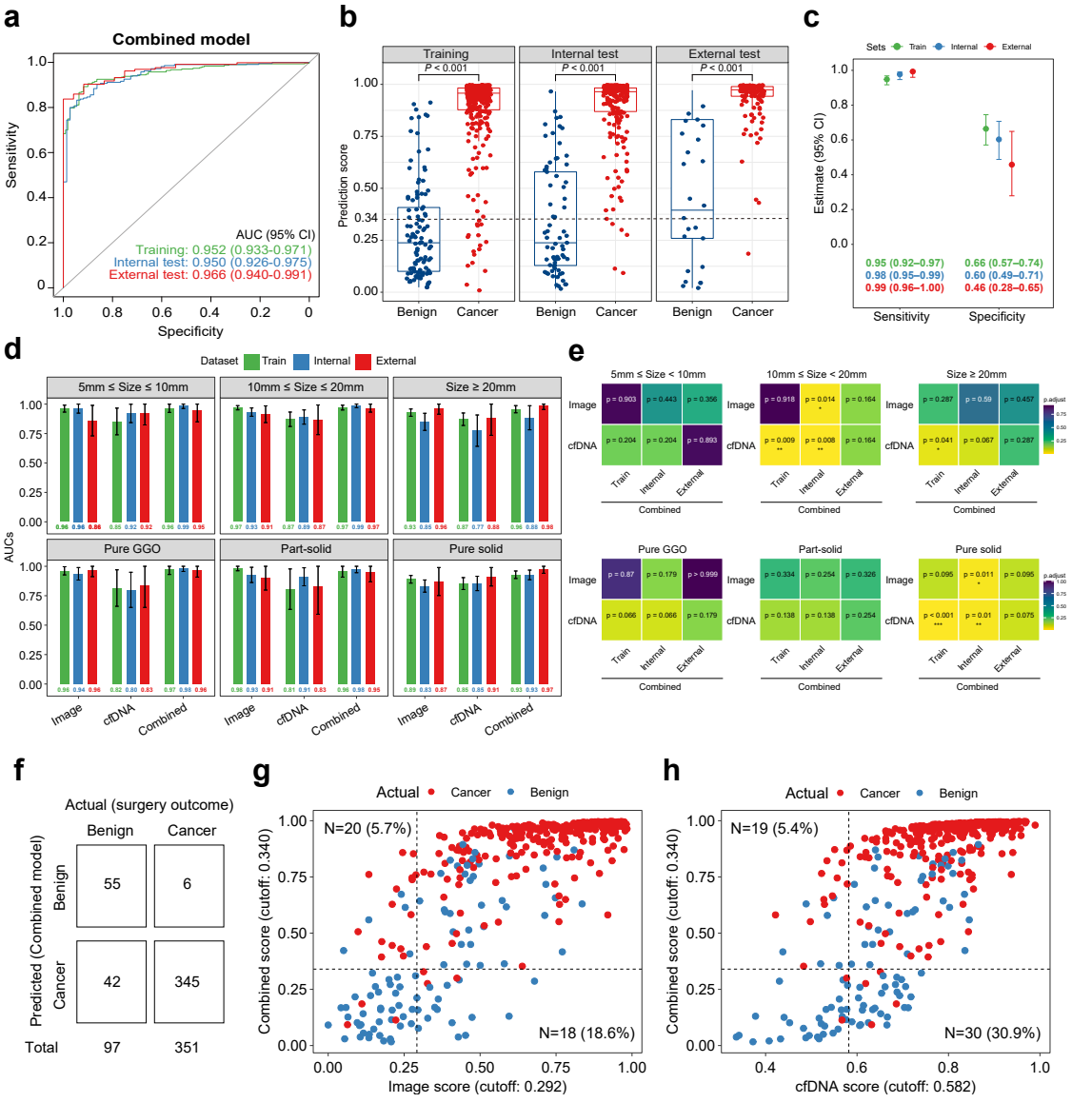

Fig. 4: **Performance analyses of the imaging-cfDNA combined model**. (a) Receiver operating characteristic analyses assessing the combined model in malignancy prediction across datasets. (**b**) Distribution of cancer prediction scores between benign and cancer samples. (**c**) Performance evaluation of the combined model using a 0.340 cutoff in the training set (95% sensitivity) across datasets. (**d**) Bar plot showing the AUCs of single-feature models (imaging model and cfDNA model) and the combined model across different subgroups. Each bar represents the mean AUC, and error bars indicate the 95% CI. (**e**) Heatmap showing the comparison of AUCs between the combined model and single-feature models across subgroups and datasets, with *P*-values adjusted for multiple testing using the Benjamini-Hochberg method. (**f**) Confusion matrices illustrating the predicted outcomes generated by the combined model in comparison with the actual outcomes determined by surgery on the combined test sets. (**g, h**) Scatter plots showing the reclassification performance achieved by the combined model in comparison to the imaging model (**g**) or the cfDNA model (**h**).

compared to single-feature models on the combined test sets (Fig. 4f–h). Using the cutoff value of 0.340 derived from the training set, the combined model correctly classified 345 malignant and 55 benign cases (Fig. 4f). Notably, compared to the imaging model, 18.6% (18/97) of benign cases were reclassified as low-risk by the combined model, potentially reducing unnecessary invasive procedures for definitive cancer diagnosis. In contrast, 5.7% (20/351) of malignant cases were reclassified as high-risk, aiding in treatment planning and preventing delayed cancer treatment (Fig. 4g). Similarly, compared to the cfDNA model, 30.9% (30/97) of benign cases and 5.4% (19/351) of malignant cases could benefit from the combined

Characteristics	Univariate analysis		Multivariate analysis	
	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
Age at diagnosis, year	1.02 (1.00–1.04)	0.026	1.03 (1.00–1.06)	0.026
Sex (Male vs. Female)	0.91 (0.50–1.75)	0.763		
Size (>10 mm vs. ≤10 mm)	1.49 (0.91–2.43)	0.111		
Radiological type (Subsolid[a] vs. Pure solid)	4.46 (2.71–7.60)	<0.001	3.95 (2.07–7.88)	<0.001
Combined model score (>0.34 vs. ≤0.34)	75.30 (32.87–205.15)	<0.001	68.73 (28.8–194.79)	<0.001

Abbreviations: OR, odds ratio; CI, confidence interval. [a]Subsolid includes pure ground-glass opacity and part-solid subtypes.

Table 1: **Logistic regression analyses for cancer prediction in the combined test set.**

model, either by reducing unnecessary invasive procedures or enabling timely cancer intervention (Fig. 4h). In direct comparisons between single-feature models, 34 participants and 47 participants were reclassified as high-risk by either the imaging or cfDNA model (Figure S7). Notably, cancer patients correctly identified by the imaging model but missed by the cfDNA model were exclusively adenocarcinomas ($P = 0.07$) and predominantly presented with subsolid nodules ($P = 0.07$) (Table S7). The statistical significance became more pronounced when increasing the sample size by including participants from the training set for the histology subgroup ($P = 0.03$). Additionally, 30 participants were correctly classified as low-risk by the imaging model but misclassified by the cfDNA model, while 20 benign nodules were correctly classified by the cfDNA model but not by the imaging model. No significant trends in clinical features were observed between the groups.

Taken together, by integrating radiomic and cfDNA features, the combined model effectively captures both the morphological characteristics and molecular insights of lung nodules, significantly enhancing malignancy risk stratification. This approach provides a robust and generalizable framework for lung nodule classification.

Invasive prediction for further risk stratification and personalized treatment

While the malignancy model effectively differentiates between benign and malignant nodules, it does not distinguish invasive from non-invasive malignancies, which is crucial for treatment decision-making (Figure S8). To bridge this gap, we developed an invasive prediction model that integrates both imaging and cfDNA features to enhance risk assessment. The model construction and evaluation followed a similar approach to the malignancy model but included lung cancer patients with nodules <15 mm and invasive characteristics (Fig. 5a; Tables S8–S10). This combined model outperformed single imaging or cfDNA models, achieving the highest diagnostic accuracy, with an AUC of 0.884 in the test set and 0.880 in the external validation set (Fig. 5b–d). Notably, the integrated prediction score effectively classified patients based on tumor aggressiveness, showing a stepwise increase from less-invasive to invasive subtypes ($P < 0.001$; Jonckheere–Terpstra test; Fig. 5e–g).

In real-world clinical practice, our machine learning model provides a precise and automated solution for lung nodule evaluation and risk stratification with minimal human intervention (Fig. 6). For individuals with ≥ 5 mm LDCT-detected nodules, plasma cfDNA and DICOM images can be analyzed using our malignancy classification model. Those testing negative can proceed with routine annual screening, while those testing positive are recommended for further assessments through invasive diagnostic methods, such as PET-CT and tissue biopsy. Furthermore, with the support of the invasion prediction model, clinicians can assess tumor aggressiveness to optimize treatment planning. Overall, this multimodal approach provides a stepwise risk stratification framework that helps reduce unnecessary procedures while minimizing missed diagnoses.

Discussion

Accurate risk stratification of lung nodules is essential for the early detection of lung cancer and subsequent treatment planning. However, conventional screening methods such as LDCT often lead to unnecessary invasive procedures and potential delays in treatment due to their limited specificity and reliance on morphological features alone.[2] In the DECIPHER-NODL study, we developed and evaluated a combined model that integrates radiomic features with cfDNA fragmentomic characteristics, demonstrating superior performance over single-feature models in both malignancy classification and tumor invasiveness prediction. Our findings highlight the potential of this integrated approach to address key limitations in current diagnostic practices.

Radiomic features provide structural and morphological insights from LDCT scans, while cfDNA fragmentomics reflects underlying tumor biology at the molecular level. The combined model achieved significantly higher AUCs compared to individual models

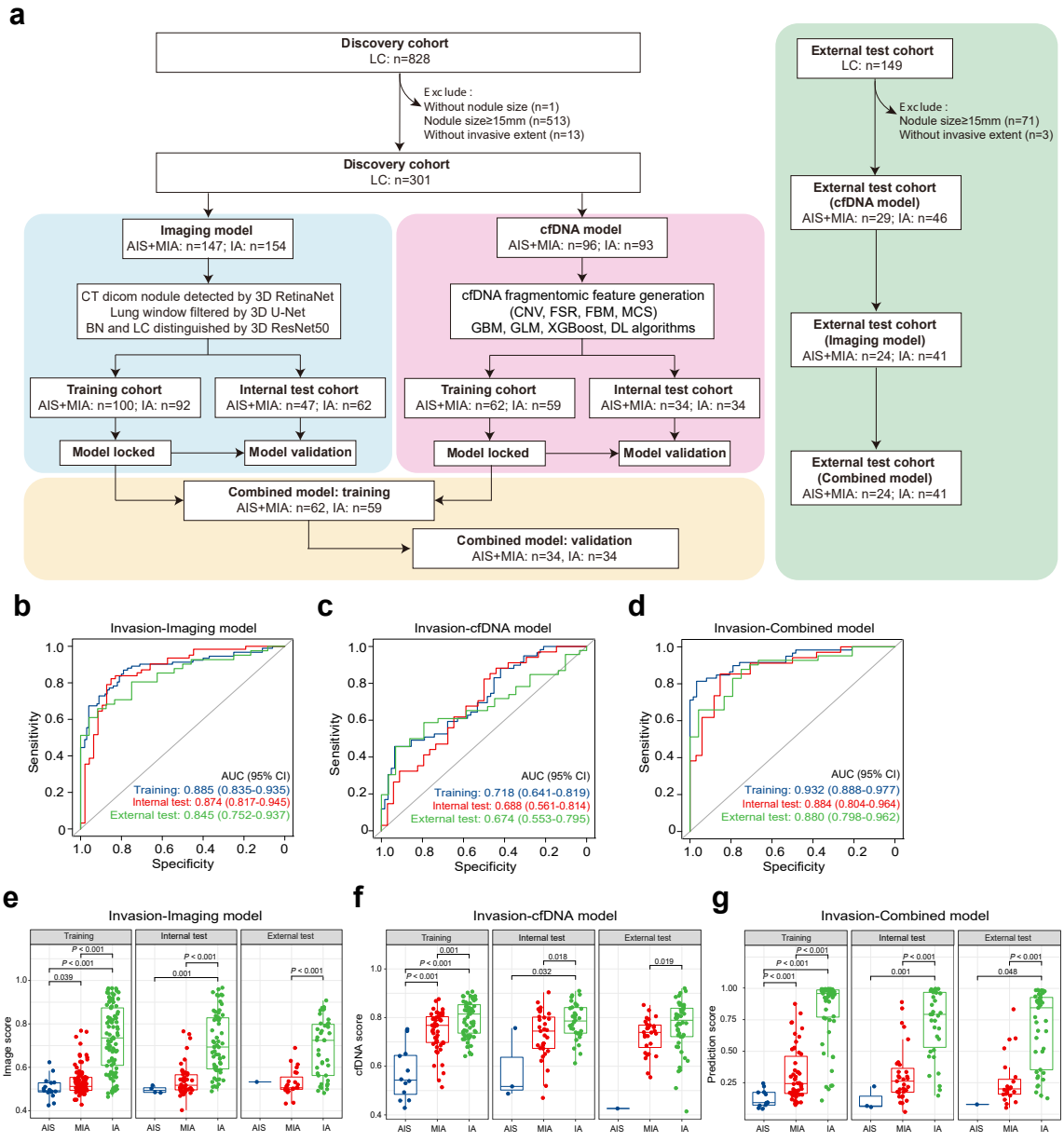

Fig. 5: **Development of the invasion model for assessing tumor aggressiveness and treatment planning**. (**a**) Flowchart illustrating the workflow for constructing invasion prediction models. (**b-d**) Receiver operating characteristic analyses evaluating the performance of the invasion-imaging model (b), invasion-cfDNA model (c), and the invasion-combined model (d) across datasets. (**e-g**) Distribution of invasiveness prediction scores across subgroups stratified by tumor invasiveness for each model. Abbreviations: AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IA, invasive adenocarcinoma.

(internal and external test: 0.950 and 0.966, respectively, vs. single models, all below 0.9), reinforcing the value of multi-modal integration in optimizing diagnostic accuracy. This aligns with recent studies that have demonstrated the complementary nature of radiomics and liquid biopsy-based approaches in aiding decision-making in clinical practice.[23,24] Notably, our model prioritizes high sensitivity to minimize the risk of misclassifying high-risk patients as benign. With a sensitivity of 0.95, the combined model significantly improved specificity in both test sets compared to the individual models, achieving 0.60 in the internal test and 0.46 in the external test set, respectively. The model's performance also varied based on nodule size and radiological type, which is consistent with previous findings.[25,26] Our model performed particularly well in

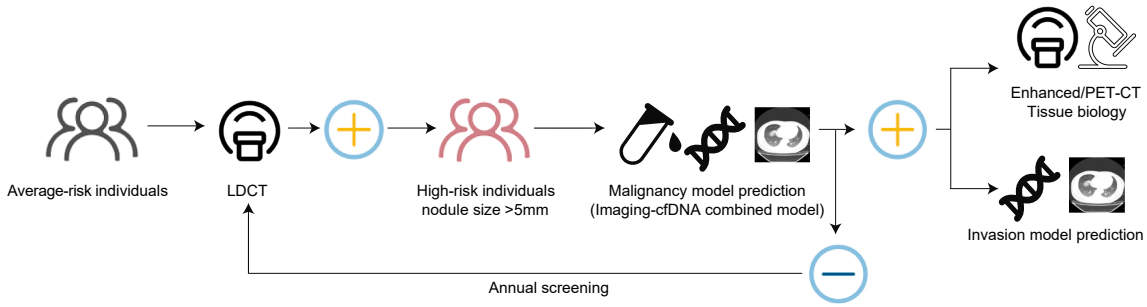

***Fig. 6:* Clinical application of models for malignancy risk evaluation and informing personalized treatment plans**. Application of the combined models for patients with both computed tomography (CT) scan results and plasma cfDNA data.

identifying malignancy in nodules <20 mm, a cohort often presenting a significant clinical dilemma due to their size and indeterminate appearance on LDCT. This strong performance likely stems from the study's enrollment of patients with radiologically indeterminate nodules but a higher pre-test probability of malignancy. While a slight decrease was observed in nodules >20 mm, the model maintained a robust AUC above 0.88 in validation sets, underscoring the robustness of the approach even in larger nodules. Notably, the observed complementary performance patterns between the cfDNA and imaging models across different radiological subtypes provide key biological and clinical insights. The imaging model excelled in classifying subsolid nodules (e.g., pure GGO), which are often visually distinct but biologically indolent and consequently associated with reduced ctDNA release. Conversely, the cfDNA model showed superior performance for pure solid nodules, which are typically more aggressive and characterized by increased vascularity and necrosis, leading to higher rates of ctDNA shedding. This dichotomy not only explains the differential performance of single-modality models but also provides a strong rationale for integration. In this regard, the combined model synergistically leverages these complementary strengths to deliver robust and consistently high performance across all subtypes. Furthermore, these findings highlight a potential clinical role for cfDNA analysis. In pure GGO nodules, which are easily detected by LDCT yet create uncertainty regarding the need for invasive surgical resection vs. ongoing surveillance, a positive result from our cfDNA-integrated model serves as a non-invasive readout of the potential presence of occult invasive components and invasiveness, thereby providing a molecular rationale for more aggressive intervention. Thus, the value of cfDNA integration extends beyond improved detection to enabling non-invasive molecular risk stratification for already visible nodules, thereby enabling more individualized management strategies based on the underlying biology of each lesion.

A key strength of the combined model is its ability to refine malignancy risk classification, enabling more precise clinical decision-making. Compared to the imaging model, it correctly identified 18.6% of benign nodules as low risk, potentially reducing unnecessary invasive procedures and overtreatment. When compared to the cfDNA model, this proportion increased to 30.9%. Additionally, the model classified 5.7% of malignant nodules as high-risk relative to the imaging model (5.4% relative to the cfDNA model), ensuring that these patients receive timely and appropriate interventions. Our findings align with those reported by Zhao et al., who integrated cfDNA methylation, clinical features, and imaging characteristics to improve lung nodule diagnosis.[26] However, a key distinction is that their model applied the Youden index for cutoff selection, whereas we prioritized sensitivity to reduce the risk of misclassifying high-risk patients as benign. This sensitivity-driven approach demonstrates greater clinical significance, particularly in balancing early detection with minimizing overdiagnosis.

Beyond malignancy classification, assessing tumor invasiveness is critical for prognostication and personalized treatment planning. In general, AIS and MIS have excellent prognoses, with nearly 100% disease-specific survival when surgically resected, whereas IA carries a higher risk of recurrence and metastasis, necessitating more aggressive treatment.[27,28] Accurate differentiation of these subtypes is crucial, as patients with less-invasive tumors may be candidates for sublobar resection without adjuvant therapy, while IA often requires lobectomy, lymph node dissection, and, in some cases, adjuvant targeted therapy or immunotherapy.[29] Our invasion-combined model exhibited strong predictive accuracy, achieving an AUC of 0.884 in the internal test and 0.880 in the external test cohort. Given its robust performance, integrating both malignancy classification and invasiveness prediction models could support clinical decision-making by guiding whether patients require further diagnostic procedures or treatment, thereby potentially reducing unnecessary

invasive interventions. Although this approach differs from typical screening strategies based on age and smoking history, it addresses a clinically relevant problem of minimizing overtreatment of benign nodules without compromising lung cancer detection. This is particularly relevant given the increasing incidence of lung cancer among younger, non-smoking populations, especially in Asia.[30,31] In the current study, a cutoff maintaining 95% sensitivity was applied to a high-risk cohort, ensuring that most malignant nodules were correctly identified. Even with a modest gain in specificity, this approach can meaningfully reduce unnecessary follow-up procedures in high-risk patients. While cfDNA WGS incurs additional costs, its use is justified in this high-risk population, as the clinical benefit of avoiding unnecessary interventions outweighs the resource burden. Furthermore, while the current study focused on surgically resected nodules, the model could be adapted as a risk stratification tool in screening settings for populations at elevated risk, such as smokers or individuals with a family history of lung cancer, where CT-based assessments may be inconclusive. This highlights the model's dual purpose of improving clinical management and supporting early detection efforts.

Despite these promising results, several limitations should be acknowledged. First, the study population consisted exclusively of Chinese participants, which may limit the generalizability of our findings to populations with different genetic, environmental, or demographic backgrounds. Second, the external test cohort had a relatively small sample size, highlighting the need for future studies with larger, independent cohorts to confirm the robustness and generalizability of the model's performance. Third, the study predominantly focused on adenocarcinoma, limiting its applicability in differentiating other lung cancer subtypes. Fourth, while our model effectively stratifies malignancy risk and predicts invasiveness, its ability to forecast long-term outcomes such as recurrence and overall survival remains to be evaluated. In addition, the study was limited by its cross-sectional design and the lack of longitudinal follow-up for both imaging and cfDNA. Future studies should expand the cohort to include a broader range of tumor subtypes, conduct multi-center validation, and investigate the model's prognostic value in long-term follow-up, including repeated imaging and blood sampling. Finally, patient-level predictions in our study were based on the most suspicious lesion, typically the largest nodule, combined with cfDNA results. Future studies could explore fully automated strategies that integrate all nodules with cfDNA using different approaches to generate robust patient-level predictions.

In conclusion, the DECIPHER-NODL study highlights the significant potential of integrating imaging and cfDNA models to enhance malignancy risk stratification in lung cancer. By combining structural and molecular data, the model outperforms single-feature approaches, providing a robust, non-invasive tool for lung nodule classification. Incorporating the invasiveness prediction model further enhances its utility in assessing tumor aggressiveness and guiding personalized treatment strategies. These findings support the clinical utility of machine learning-based risk assessment tools in lung cancer screening, diagnosis, and treatment planning, highlighting their role in the future of precision oncology.

Contributors

H. Wang, H. Huang, F. Li, and Y. Deng contributed to the study design. H. Wang, H. Huang, F. Li, Y. Deng, S. Wang, D. Zhu, H. Xu, and P. He performed data analysis. H. Wang, H. Huang, F. Li, Y. Deng, and S. Wang edited the manuscript. Y. Shao, Y. Shen, J. He, and W. Liang conceived and supervised the study. All authors had access to raw data. Data verification was undertaken by H. Wang, H. Huang, F. Li, and Y. Deng. All authors reviewed and approved the final version of the manuscript, and the corresponding authors had final responsibility for the decision to submit it for publication.

Data sharing statement

The datasets generated and/or analyzed during this current study are available from the corresponding author upon reasonable request.

Declaration of interests

S. Wang, D. Zhu, H. Xu, H. Bao, P. He, F. Wang, and Y. Shao are employees of Nanjing Geneseeq Technology Inc. The remaining authors declare no competing interests.

Acknowledgements

This study was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2024ZD0520004, 2024ZD0520000, 2024ZD0529501, 2024ZD0529500), the National Key Research & Development Programme (2022YFC2505105), China National Science Foundation (No. 82373121), the Science and Technology Planning Project of Guangzhou (202206080013), and Guangzhou National Laboratory (SRPG22-017). The funder had no role in the study design, data collection, analysis, or decision to publish the manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lanwpc.2025.101730.

References

1 Siegel RL, Kratzer TB, Giaquinto AN, Sung H, Jemal A. Cancer statistics, 2025. *CA Cancer J Clin*. 2025;75(1):10–45.
2 Team TNLSTR. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395–409.
3 Vachani A, Tanner NT, Aggarwal J, et al. Factors that influence physician decision making for indeterminate pulmonary nodules. *Ann Am Thorac Soc*. 2014;11(10):1586–1591.
4 Pastorino U, Silva M, Sestini S, et al. Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy. *Ann Oncol*. 2019;30(7):1162–1169.
5 Madariaga ML, Lennes IT, Best T, et al. Multidisciplinary selection of pulmonary nodules for surgical resection: diagnostic results and long-term outcomes. *J Thorac Cardiovasc Surg*. 2020;159(4):1558–15566.e3.
6 Zhang W, Duan X, Zhang Z, et al. Combination of CT and telomerase+ circulating tumor cells improves diagnosis of small pulmonary nodules. *JCI Insight*. 2021;6(11).
7 de Margerie-Mellon C, Chassagnon G. Artificial intelligence: a critical review of applications for lung nodule and lung cancer. *Diagn Interv Imaging*. 2023;104(1):11–17.

8 Callister ME, Baldwin DR, Akram AR, et al. British thoracic society guidelines for the investigation and management of pulmonary nodules. *Thorax*. 2015;70(Suppl 2):ii1–ii54.
9 Clark TJ, Flood TF, Maximin ST, Sachs PB. Lung CT screening reporting and data system speed and accuracy are increased with the use of a semiautomated computer application. *J Am Coll Radiol*. 2015;12(12 Pt A):1301–1306.
10 Bai C, Choi CM, Chu CM, et al. Evaluation of pulmonary nodules: clinical practice consensus guidelines for Asia. *Chest*. 2016; 150(4):877–893.
11 MacMahon H, Naidich DP, Goo JM, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the fleischner society 2017. *Radiology*. 2017;284(1):228–243.
12 Mathios D, Johansen JS, Cristiano S, et al. Detection and characterization of lung cancer using cell-free DNA fragmentomes. *Nat Commun*. 2021;12(1):5060.
13 Wan N, Weinberg D, Liu TY, et al. Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA. *BMC Cancer*. 2019;19(1):832.
14 Zhou Q, Kang G, Jiang P, et al. Epigenetic analysis of cell-free DNA by fragmentomic profiling. *Proc Natl Acad Sci U S A*. 2022;119(44):e2209852119.
15 Doebley AL, Ko M, Liao H, et al. A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. *Nat Commun*. 2022;13(1):7475.
16 American College of Radiology Committee on Lung-RADS®. Lung-RADS assessment categories, version 1.0. Available from: https://edgesitecorecloudio/americancoldf5f-acrorgf92a-productioncb02-3650/media/ACR/Files/RADS/Lung-RADS/LungRADS-v1-Assessment-Categoriespdf; 2014.
17 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120.
18 Hofmanninger J, Prayer F, Pan J, Rohrich S, Prosch H, Langs G. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *Eur Radiol Exp*. 2020;4(1):50.
19 Zeng R, Badal A, Ezzati AO, et al. *Deep virtual CT workflow for evaluating AI in low-dose CT lung cancer screening: a 2D demonstration. Medical imaging 2025: Physics of medical Imaging*. 2025.
20 Cardoso MJ, Li W, Brown R, et al. MONAI: an open-source framework for deep learning in healthcare. *ArXiv*. 2022;abs/2211.02701.
21 Widman AJ, Shah M, Frydendahl A, et al. Ultrasensitive plasma-based monitoring of tumor burden using machine-learning-guided signal enrichment. *Nat Med*. 2024;30(6):1655–1666.
22 H2O AutoML. Automatic machine learning — H2O 3.44.0.3 documentation. Available from: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html.
23 Neri E, Del Re M, Paiar F, et al. Radiomics and liquid biopsy in oncology: the holons of systems medicine. *Insights Imaging*. 2018; 9(6):915–924.
24 Cucchiara F, Del Re M, Valleggi S, et al. Integrating liquid biopsy and radiomics to monitor clonal heterogeneity of EGFR-positive non-small cell lung cancer. *Front Oncol*. 2020;10:593831.
25 He J, Wang B, Tao J, et al. Accurate classification of pulmonary nodules by a combined model of clinical, imaging, and cell-free DNA methylation biomarkers: a model development and external validation study. *Lancet Digit Health*. 2023;5(10):e647–e656.
26 Zhao M, Xue G, He B, et al. Integrated multiomics signatures to optimize the accurate diagnosis of lung cancer. *Nat Commun*. 2025;16(1):84.
27 Boland JM, Froemming AT, Wampfler JA, et al. Adenocarcinoma in situ, minimally invasive adenocarcinoma, and invasive pulmonary adenocarcinoma–analysis of interobserver agreement, survival, radiographic characteristics, and gross pathology in 296 nodules. *Hum Pathol*. 2016;51:41–50.
28 Yotsukura M, Asamura H, Motoi N, et al. Long-term prognosis of patients with resected adenocarcinoma in situ and minimally invasive adenocarcinoma of the lung. *J Thorac Oncol*. 2021;16(8):1312–1320.
29 Zhang Y, Ma X, Shen X, et al. Surgery for pre- and minimally invasive lung adenocarcinoma. *J Thorac Cardiovasc Surg*. 2022; 163(2):456–464.
30 Yang W, Qian F, Teng J, et al. Community-based lung cancer screening with low-dose CT in China: results of the baseline screening. *Lung Cancer*. 2018;117:20–26.
31 Chang G-C, Chiu C-H, Yu C-J, et al. Low-dose CT screening among never-smokers with or without a family history of lung cancer in Taiwan: a prospective cohort study. *Lancet Respir Med*. 2024; 12(2):141–152.