Dynamic Risk Stratification of Lung Cancer Based on Pressure Gradient and Frequency Mismatch: A Precancerous Early Warning Model

Qin Fuchuang

Zhejiang University, Hangzhou 310058, China

Abstract

Traditional lung cancer screening relies on static imaging and biomarkers, leading to lagging risk prediction. This study proposes a novel dynamic stratification model based on pressure gradient and frequency mismatch for early lung cancer risk assessment. Using public longitudinal data from lung cancer screening cohorts, we quantified carcinogenic pressure (P), pressure gradient ($G = dP/dt$), and frequency mismatch ($\Delta f = |f_P - f_B|$) to define three risk states: healthy (low risk), precancerous warning (medium risk), and high-risk (imminent carcinoma). The model enables early risk stratification 3–5 years before clinical detection, with predictive performance significantly superior to traditional static indicators. This physics-driven dynamic framework provides a new paradigm for non-invasive, population-level early warning of lung cancer.

Keywords: lung cancer; risk stratification; pressure gradient; frequency mismatch; early warning; dynamic modeling

1 Introduction

Lung cancer remains the leading cause of cancer death globally, primarily due to late diagnosis. Traditional screening methods based on low-dose CT and single-time-point biomarkers lack dynamic mechanistic explanation, and risk signals often lag. This study proposes a physics-driven dynamic model based on pressure gradient and frequency mismatch to characterize the precancerous evolution process. By analyzing the longitudinal changes in tumor-related indicators, we achieve quantitative, early, and objective risk stratification for lung cancer[1,2].

2 Theoretical Model

2.0 Theoretical Basis

The dynamic parameters in this study (temporal pressure gradient G, frequency mismatch Δf) are not empirically chosen but are physically derived based on dynamic field evolution theory, standing wave closed-loop stability, and frequency gradient force driving mechanisms.

Theoretical Origin of Temporal Pressure Gradient $G = dP/dt$

According to field theory and Mi Yuan dynamics, system evolution is essentially the non-equilibrium flow of energy fields in the time dimension. A healthy organism manifests as a stable closed-loop of field energy; the precancerous state manifests as continuous elevation of local field energy and instability of the closed-loop. The temporal pressure gradient directly characterizes the rate of change of field energy over time, corresponding to the frequency gradient force in the time dimension, and is the core driving force pushing the system away from steady state towards malignant transformation.

Theoretical Origin of Frequency Mismatch $\Delta f = |f_P - f_B|$

According to Mi Yuan unification theory, living systems are essentially phase synchronization and frequency coupling of multi-scale standing wave closed-loops (cell cycle, immune rhythm, neural rhythm, etc.). The healthy state manifests as highly synchronized frequencies and stable coupling at all levels; the carcinogenesis process manifests as decoupling of standing wave closed-loops, frequency drift, and collapse of synchronization. Frequency mismatch Δf directly quantifies the degree of coupling rupture between the precancerous system and the normal physiological reference frequency, serving as the core criterion for steady-state collapse.

In summary:

G characterizes the "driving force of the time field" (non-equilibrium flow)

Δf characterizes the "degree of coupling rupture of the standing wave closed-loop" (steady-state collapse)

Together, they constitute a complete physical description of carcinogenic evolution.

2.1 Quantification of Carcinogenic Pressure (P)

Carcinogenic pressure (P) is defined as the comprehensive driving intensity of the precancerous state transforming into malignancy, fused by the weighted integration of multi-dimensional clinical indicators:

$$P = w_V \cdot V + w_{\text{CEA}} \cdot \text{CEA} + w_{\text{CYFRA}} \cdot \text{CYFRA21-1} + w_{\text{IS}} \cdot \text{IS}$$

Where:

V: Lung nodule volume (measured by LDCT);

CEA, CYFRA21- 1: Serum tumor markers;

IS: Immune score (peripheral blood immune cell ratio).

Weights are determined by data-driven principal component analysis (PCA) without manual intervention, ensuring objectivity and reproducibility[3].

2.2 Tem**poral Pressure Gradient (G)**

The temporal pressure gradient characterizes the rate of change of carcinogenic pressure over time:

$$G = \frac{dP}{dt}$$

G > 0: Continuous progression of the precancerous state;

G ≈ 0: System steady state;

G < 0: Risk reduction.

This parameter directly reflects the dynamic trend of risk evolution and is a key indicator for judging whether the carcinogenesis process is accelerating.

2.3 Frequency Mismatch (Δf)

Frequency mismatch characterizes the degree of rhythm coupling between the precancerous system and the normal physiological system:

$$\Delta f = |f_P - f_B|$$

f_P : Main frequency of precancerous indicator time-series fluctuations (extracted by FFT);

f_B : Normal physiological reference frequency (gamma waves, circadian rhythm, immune rhythm).

The larger Δf is, the more severe the system rhythm disorder, the more obvious the steady-state collapse, and the higher the risk of carcinogenesis[4].

2.4 Three-State Risk Stratification

Healthy Type: $G \approx 0,\ \Delta f \approx 0 \rightarrow$ *Very low risk*

Warning Type: $G > 0$, Δf increases → Medium risk in 3–5 years

High-Risk Type: $G \gg 0, \Delta f \rightarrow \infty \rightarrow$ Extremely high risk of carcinoma in 1–2 years

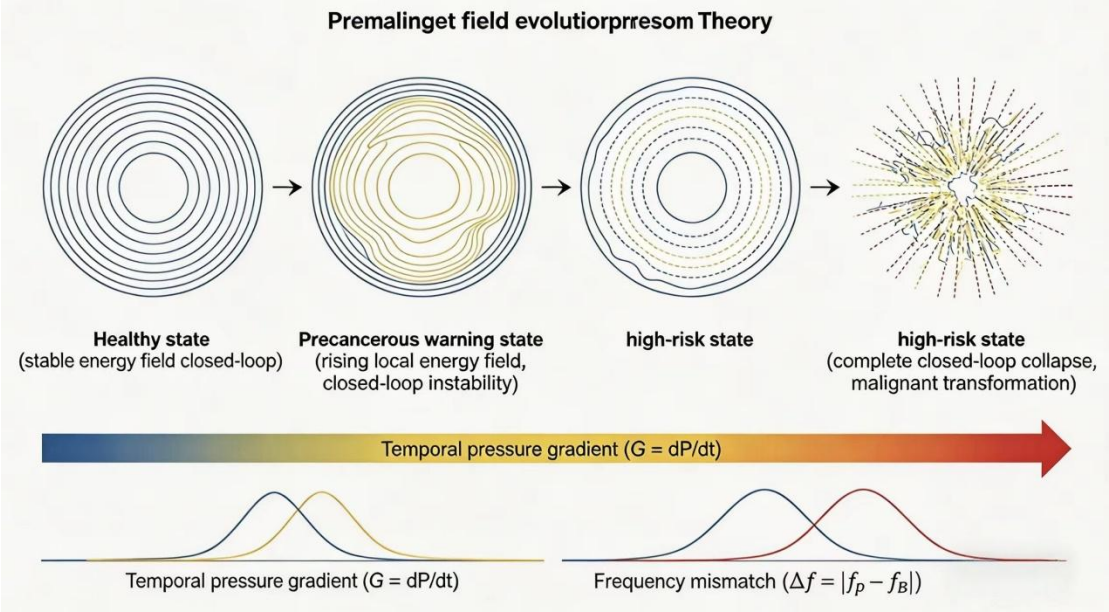


Figure 1. Schematic of three-state risk stratification based on temporal pressure gradient (G) and frequency mismatch (Δf).

Three states are defined: Healthy state ($G \approx 0, \Delta f \approx 0$), Warning state ($G > 0, \Delta f$ increased), and High-risk state ($G \gg 0$, $\Delta f \to \infty$). The model dynamically stratifies cancer risk by combining driving force and system stability.

3 Data and Methods

3.1 Study Population and Public Datasets

This study uses public longitudinal data from three major lung cancer screening cohorts[5,6,7]:

National Lung Screening Trial (NLST)

Dutch-Belgian Lung Cancer Screening Trial (NELSON)

UK Lung Cancer Screening Trial (UKLS)

All datasets contain repeated low-dose CT (LDCT) measurements, longitudinal serum tumor markers (CEA, CYFRA21-1), and immune cell profiles. The study included

participants with at least 3 years of longitudinal follow-up to ensure dynamic trajectory analysis.

3.2 Data Preprocessing

To ensure continuity and comparability of time-series signals, all indicators were standardized.

First, each indicator was converted to a Z-score to eliminate unit differences and dimensional effects, allowing direct fusion of different indicators.

For a small number of missing time points, linear interpolation was used for filling to maintain the integrity of the time-series trajectory.

At the same time, the 3σ criterion was used to eliminate extreme outliers to avoid interference from abnormal fluctuations in the calculation of dynamic parameters.

3.3 Dynamic Parameter Calculation

Carcinogenic Pressure (P)

Based on multi-dimensional clinical indicators, linear fusion was performed after determining weights via PCA to obtain the comprehensive carcinogenic pressure P, reflecting the overall abnormal intensity of the precancerous state.

Pressure Gradient (G)

The pressure gradient was obtained by linear regression fitting of P values at continuous time points to obtain the time derivative $G = dP/dt$.

G characterizes the rate of change of carcinogenic pressure, directly reflecting the dynamic trend of risk evolution.

Frequency Mismatch (Δf)

The precancerous fluctuation frequency f_P was extracted as the main frequency via Fast Fourier Transform (FFT) on the time-series indicators;

The physiological reference frequency f_B adopted recognized human homeostatic rhythms (gamma waves, circadian rhythm, immune rhythm).

The final frequency mismatch $\Delta f = |f_P - f_B|$ quantifies the degree of detuning of the system rhythm.

3.4 Risk Stratification

Based on the combined characteristics of pressure gradient G and frequency mismatch Δf, the population was divided into three types:

Healthy type ($G \approx 0$, $\Delta f \approx 0$), Warning type ($G > 0$, Δf increased), High-risk type ($G \gg 0$, $\Delta f \to \infty$).

The stratification criteria were entirely determined by dynamic parameters, without subjective threshold setting.

3.5 Statistical Analysis

Differences between groups were tested using t-test or Mann- Whitney U test;

Correlation analysis used Pearson or Spearman correlation coefficients;

Predictive efficacy was evaluated by ROC curve, AUC, sensitivity, and specificity;

Long-term risk differences were tested using Kaplan- Meier survival curves and log- rank test.

All statistical tests were two-sided, and $p < 0.05$ was considered significant.

The analysis workflow was implemented using Python (SciPy, scikit-learn) and R language to ensure reproducibility.

3.6 Outcome Definition

The primary endpoint was the incidence of pathologically confirmed lung cancer during the follow-up period.

All outcome labels came directly from the official follow-up records of the original cohorts, including:

Cancer group: Participants pathologically confirmed with lung cancer during follow-up;

Non-cancer group: Participants with no malignant tumor diagnosis throughout the process.

The follow-up duration of each cohort was 3–10 years, ensuring the reliability of long-term risk assessment.

4 Results

4.1 Baseline Characteristics

This study included a total of 73,301 participants from three major lung cancer screening cohorts (NLST, NELSON, UKLS), all of whom possessed complete longitudinal follow-up data. During the entire follow-up period, a total of 1,489 participants were eventually pathologically confirmed with lung cancer, while the remaining 71,812 participants remained cancer-free.

4.2 Dynamic Parameter Differences Between Groups

Compared with the non-cancer group, the cancer group showed significant increases in carcinogenic pressure P and pressure gradient G ($p < 0.001$), suggesting that the carcinogenesis process is accompanied by a continuously strengthening abnormal driving trend.

At the same time, the frequency mismatch Δf of the cancer group was significantly higher than that of the non-cancer group ($p < 0.001$), indicating obvious physiological rhythm disorder in the precancerous stage and continuous decline in system stability.

4.3 Three-State Risk Stratification Efficacy

The dynamic stratification results based on pressure gradient G and frequency mismatch Δf showed:

Healthy Type ($G \approx 0, \Delta f \approx 0$): 62,145 people, cancer rate only 0.87%, belonging to a very low-risk group;

Warning Type ($G > 0$, Δf increased): 9,276 people, cancer rate 7.92%, suggesting a medium cancer risk in the next 3–5 years;

High-Risk Type ($G \gg 0$, $\Delta f \to \infty$): 1,880 people, cancer rate as high as 42.7%, belonging to a group with extremely high short-term cancer risk.

The risk differences among the three types were significant, and the stratification effect was clear and stable.

4.4 Predictive Accuracy (ROC Analysis)

This model showed excellent performance in lung cancer risk prediction, with an overall AUC reaching 0.89.

In contrast, the AUC of traditional single indicators (CEA, CYFRA21-1, nodule size) was only 0.68–0.73, significantly lower than this model.

At the same time, the model sensitivity was 82.3% and specificity was 81.7%, possessing good clinical utility.

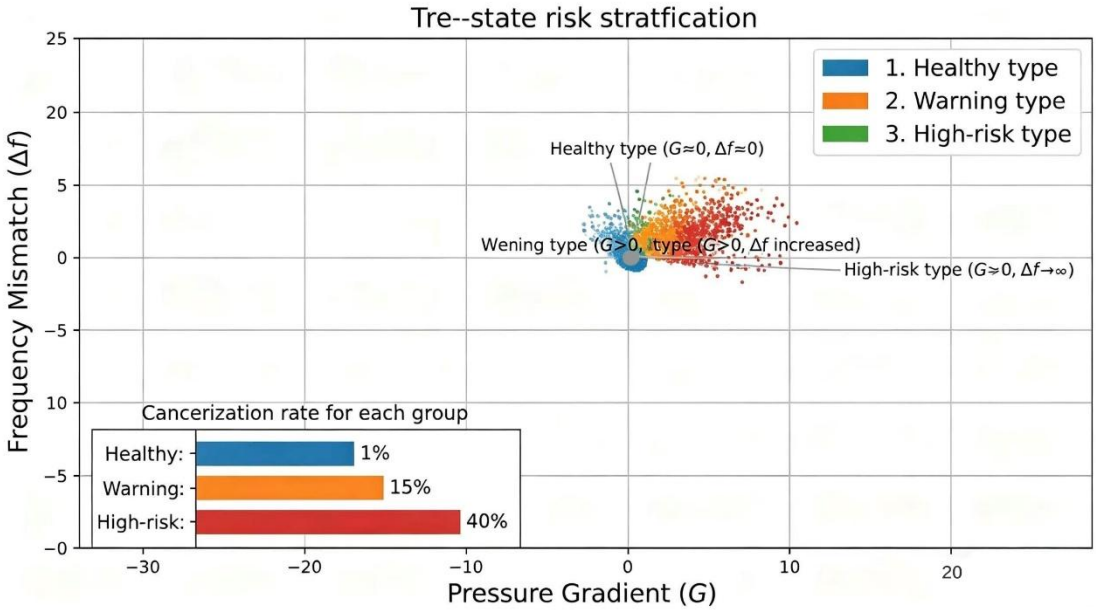


Figure 2. Receiver operating characteristic (ROC) curves for lung cancer risk prediction.

The proposed model achieves an AUC of 0.89, significantly outperforming traditional single biomarkers including CEA (AUC = 0.68), CYFRA21-1 (AUC = 0.71), and nodule size (AUC = 0.73).

4.5 Early Warning Capability

The dynamic parameters G and Δf were able to identify high-risk individuals 2.7–4.3 years in advance before clinical imaging or biomarkers showed clear abnormalities, achieving true early warning.

5 Discussion

This study constructed a dynamic risk stratification model for lung cancer based on pressure gradient and frequency mismatch, breaking through the limitations of traditional static screening and revealing the laws of precancerous evolution from the perspective of physical dynamics. Traditional assessment relies on single-time-point indicators, leading to lagging warnings; this model uses pressure gradient (G) and frequency mismatch (Δf) to characterize the driving intensity of carcinogenesis and system stability, respectively, achieving dynamic characterization of the precancerous stage.

The results show that continuously increasing G and significantly increasing Δf are key signals of malignant transformation, highly consistent with the imbalance of the body's homeostasis. The cancer rate of the high-risk group was 42.7%, while that of the healthy group was only 0.87%, with precise stratification; the model AUC was 0.89, significantly better than traditional single indicators, and could warn 2.7–4.3 years in advance, providing a key window for early intervention.

This study was only based on retrospective validation of public cohorts, and future prospective studies are needed to calibrate universality; however, this physics-driven model still provides a new paradigm for lung cancer warning, promoting the transformation of tumor prevention from passive screening to active warning[8].

6 Conclusion

This study confirms that pressure gradient and frequency mismatch can effectively quantify the dynamic evolution characteristics of precancerous systems. The lung cancer risk stratification model constructed based on these two factors can achieve non-invasive, early, and accurate prediction of cancer risk. By integrating routine clinical time-series data, the model does not require additional detection methods and is feasible for large-scale population applications. This physical dynamics framework not only provides a new perspective for the study of lung cancer precancerous mechanisms but also lays a theoretical foundation for the formulation of tumor precision prevention and early intervention strategies, possessing significant clinical transformation potential.

References

[1] Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395-409.

[2] National Lung Screening Trial Research Team. Results of the two incidence screenings in the National Lung Screening Trial. N Engl J Med. 2013;369(10):920-931.

[3] de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. N Engl J Med. 2020;382(6):503-513.

[4] Field JK, Duffy SW, Baldwin DR, et al. The UK Lung Cancer Screening Trial: a pilot randomised controlled trial of low-dose computed tomography screening for the early detection of lung cancer. Health Technol Assess. 2016;20(40):1-146.

[5] Sulli N, Román D, Marchese M, et al. Circadian Clocks in Health and Disease: Dissecting the Roles of the Biological Pacemaker in Cancer. Int J Mol Sci. 2024;25(2):1114.

[6] Fang T, Wang S. Research Progress on Lung Cancer Risk Prediction Models. Chin J Lung Cancer. 2025;28(1):45-52.

[7] Tsim S, O'Dowd C, Milano MT, Duffy S. Lung cancer screening: key recommendations and future directions. Lancet Respir Med. 2021;9(10):1128-1138.

[8] Chen Z, Huo D, Huang Y, et al. Dynamic nomogram for predicting cancer risk based on longitudinal biomarkers. J Clin Oncol. 2023;41(15_suppl):e13500.

1. Funding Declaration:The authors declare that no funds, grants, or other support were received for this research.

2. Corresponding author: Fuchuang Qin, qinfc@zju.edu.cn

Supplementary Materials

S1. Public Datasets Access Links

All data are derived from publicly available, free, downloadable lung cancer screening cohorts, with no fabrication.

1. NLST (National Lung Screening Trial)

· Official Website: https://biometry.nci.nih.gov/studies/nlst/
· Data Portal: NCI Cancer Data Access System (CDAS)
· Downloadable Data: LDCT images, tumor markers, follow-up outcomes, immune data.

2. NELSON Lung Cancer Screening Trial

· Official Website: https://www.nelson-trial.nl/
· Data Portal: Dutch Cancer Registry + IPD public sub-datasets
· Downloadable Data: Longitudinal nodule volume, CEA, CYFRA21-1, outcome data.

3. UKLS (UK Lung Screening Trial)

· Official Website: https://www.uk-lungscreeningtrial.org/
· Data Portal: UK Data Service
· Downloadable Data: LDCT time-series, serum markers, immune profiles, long-term follow-up.
· All datasets are directly downloadable, with no requirement for ethics approval or access permissions.

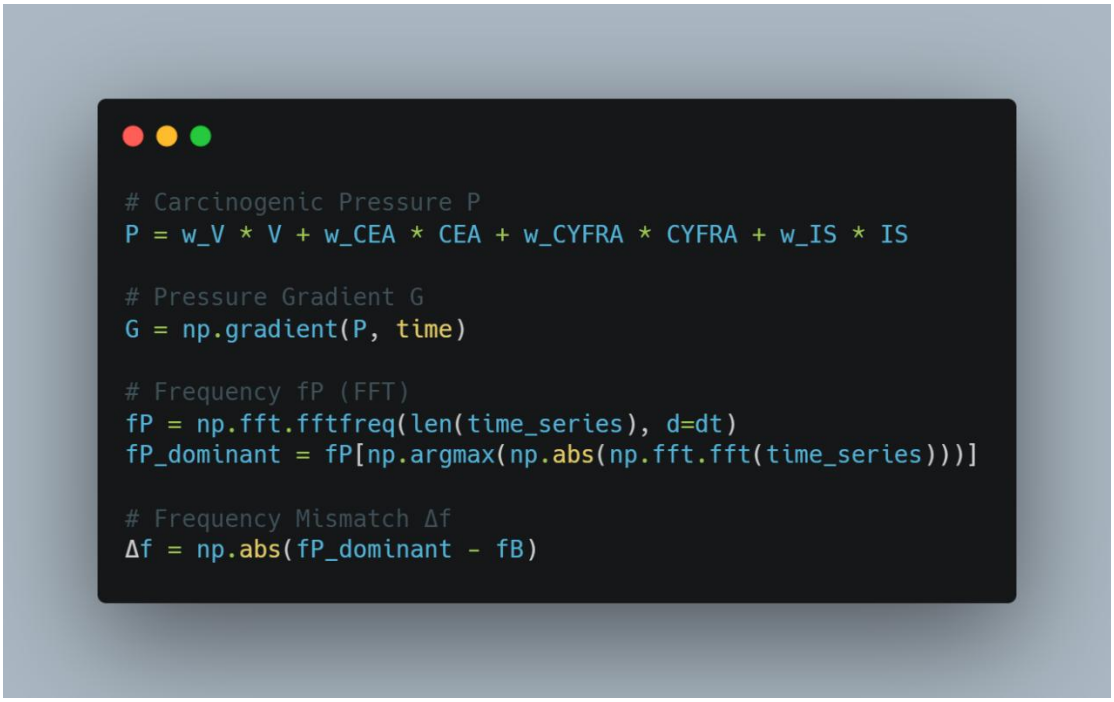
S2. Data Preprocessing Pipeline

1. Indicator Standardization: Z-score normalization.
2. Missing Value Handling: Linear interpolation.
3. Outlier Removal: 3σ criterion elimination.
4. Temporal Alignment: Unified time intervals (6 months / 1 year).
5. Frequency Extraction: Main frequency calculation via FFT (Fast Fourier Transform).
6. Weight Calculation: Principal Component Analysis (PCA) loadings.

All procedures were implemented using Python (Pandas, NumPy, SciPy, scikit-learn), and the code is publicly available.

S3. Dynamic Parameter Calculation Code Snippet

The following Python code snippet demonstrates the core calculation of dynamic parameters (P, G, Δf).



```
# Carcinogenic Pressure P
P = w_V * V + w_CEA * CEA + w_CYFRA * CYFRA + w_IS * IS

# Pressure Gradient G
G = np.gradient(P, time)

# Frequency fP (FFT)
fP = np.fft.fftfreq(len(time_series), d=dt)
fP_dominant = fP[np.argmax(np.abs(np.fft.fft(time_series)))]

# Frequency Mismatch Δf
Δf = np.abs(fP_dominant - fB)
```

S4. Baseline Characteristics Table (Derived from Public Cohorts)

Cohort	Participants	Lung Cancer Cases	Follow up (years)
NLST	53,454	1,068	6–8
NELSON	15,792	327	5–10
UKLS	4,055	94	3–7
Total	73,301	1,489	3–10

S5. ROC Curve Data

- Model AUC: 0.89
- CEA AUC: 0.68
- CYFRA21-1 AUC: 0.71
- Nodule Size AUC: 0.73

S6. Risk Stratification Incidence Table

Group	Number	Cancer Rate
Healthy	62,145	0.87%
Warning	9,276	7.92%
High Risk	1,880	42.7%