Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Composite Coagulation–Inflammation–Nutrition–Hematologic (DANH) Axis for Clinical Risk Stratification in Lung Cancer

Tung Dao Van
Hai Phong Medical College
Hung Tran Quang
Viet Tiep Friendship Hospital
Hung Nguyen Duy
Hai phong University Of Medicine and Pharmacy
Anh Do Nguyen Mai
Hai phong University Of Medicine and Pharmacy
Giang Bui Thi Huong
Hai phong University Of Medicine and Pharmacy
Thuc Pham Van
Hai phong University Of Medicine and Pharmacy
Dung Luu Vu
Hai Phong Hospital of Obstetrics and Gynecology
Huy Chu Quang
Independent Researcher
Loc Pham Thi
ptloc@hpmu.edu.vn

Hai phong University Of Medicine and Pharmacy

Short Report

Keywords: Lung cancer, D-dimer, Neutrophil-to-lymphocyte ratio, Albumin, Hemoglobin, DANH score, Risk stratification

Posted Date: April 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9480809/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Lung cancer progression involves systemic alterations in coagulation, inflammation, nutritional decline, and hematologic impairment. Although individual biomarkers show prognostic relevance, their integrated clinical value remains insufficiently explored. This study aimed to evaluate the prognostic performance of a novel DANH axis comprising D-dimer, albumin, neutrophil to lymphocyte ratio (NLR), hemoglobin for risk stratification in lung cancer.

Methods

This retrospective observational study included 184 patients with histologically confirmed lung cancer. Pretreatment laboratory parameters, including D-dimer, NLR, serum albumin, and hemoglobin, were analyzed. Composite biomarker indices were constructed by integrating three-factor combinations (DAN, DAH, DNH, ANH) and a four-factor model (DANH). Predictive performance for advanced disease, metastasis, and cancer associated thrombosis was analysis. Bootstrap resampling (1,000 iterations) was applied to evaluate model stability.

Results

Individual biomarkers showed moderate discrimination for advanced disease (AUC 0.687–0.747), while composite models improved accuracy, with DANH performing best (AUC = 0.784). For thrombosis, D-dimer was the top single marker (AUC = 0.711), and composite models improved performance, led by DANH (AUC = 0.750). In contrast, both single and composite biomarkers had limited ability to predict metastasis. Composite scores based on optimal cut offs (D-dimer ≥ 986 ng/mL, NLR ≥ 4.07, hemoglobin ≤ 120 g/L, albumin ≤ 36.6 g/L) showed significant stage dependent stratification, with higher risk categories associated with advanced disease (all $p \leq 0.01$; strongest for DANH: $p < 0.0001$). Bootstrap analysis confirmed robustness, with the DANH model achieving the highest performance and stability (F1 = 0.9595; 95% CI: 0.9323–0.9819; win rate 76.9%), while other models showed greater degeneration and lower discrimination.

Conclusions

Integration of coagulation, inflammatory, nutritional, and hematologic biomarkers into a DANH axis improves discrimination of advanced disease and cancer-associated thrombosis in lung cancer. Although not predictive of metastasis, the DANH score provides a simple, cost-effective framework for systemic risk stratification in routine clinical practice.

INTRODUCTION

Lung cancer remains the leading cause of cancer-related mortality worldwide, largely due to late-stage diagnosis and the systemic biological disturbances that accompany tumor progression. Recent global cancer statistics estimate that lung cancer accounts for approximately 1.8 million deaths annually, representing the most lethal malignancy worldwide [1]. Although advances in histopathological classification and molecular profiling have improved therapeutic stratification, these tumor-centered approaches only partially capture the

complex host–tumor interactions that influence disease progression and clinical outcomes [2], [3]. Increasing evidence indicates that lung cancer progression is not solely determined by tumor-intrinsic characteristics but also by systemic host responses. Chronic inflammation, immune disregulation, activation of coagulation pathways, metabolic and nutritional impairment, and hematologic dysfunction collectively contribute to a permissive microenvironment that supports tumor growth, immune evasion, and vascular complications [4], [5], [6].

Routine laboratory biomarkers reflecting these processes—such as D-dimer, neutrophil-to-lymphocyte ratio (NLR), serum albumin, and hemoglobin—have individually demonstrated prognostic relevance in lung cancer [7], [8]. Elevated D-dimer reflects cancer-associated hypercoagulability and fibrin turnover, processes linked to tumor burden, angiogenesis, and thrombotic complications. Cancer-associated thrombosis is increasingly recognized as both a complication and a biological hallmark of tumor progression [9], [10], [11]. Similarly, the neutrophil-to-lymphocyte ratio integrates pro-tumor inflammatory activity with impaired antitumor immune surveillance and has been widely validated as a prognostic biomarker across multiple malignancies including lung cancer [12], [13], [14]. Hypoalbuminemia indicates systemic inflammation and nutritional decline, reflecting both tumor-driven catabolism and host metabolic dysfunction. Low serum albumin levels have been consistently associated with poor prognosis and treatment tolerance in cancer patients [15]. Anemia, frequently observed in patients with advanced malignancies, reflects impaired erythropoiesis, chronic inflammation, and increased tumor metabolic demand, and has been associated with reduced survival in lung cancer cohorts [16], [17]. Despite the growing recognition of these systemic alterations, most studies have evaluated these biomarkers in isolation. However, coagulation activation, inflammation, nutritional deterioration, and hematologic impairment are biologically interconnected processes that collectively shape the host response to malignancy and influence tumor evolution [4], [5], [6]. Integrating these complementary biological domains may therefore provide a more comprehensive representation of systemic tumor–host interactions than single biomarkers alone. Composite biomarker indices have increasingly emerged as practical tools for clinical risk stratification in oncology [18]. Indices such as the systemic immune-inflammation index (SII) and the prognostic nutritional index (PNI) illustrate how integrating multiple routine laboratory parameters can enhance prognostic discrimination and improve risk stratification in cancer populations [19], [20], [21]. Several studies have demonstrated the predictive value of composite inflammatory or nutritional scores for survival, treatment response, and postoperative outcomes in lung cancer [22], [23], [24]. Nevertheless, an integrated framework simultaneously capturing coagulation, inflammatory, nutritional, and hematologic axes has not been systematically explored in lung cancer. A multidimensional biomarker model incorporating these domains may therefore provide improved insight into systemic disease burden and thrombotic vulnerability [25].

In this study, we propose a unified systemic biomarker framework - the DANH axis, comprising D-dimer, albumin, neutrophil-to-lymphocyte ratio, and hemoglobin - to capture the multidimensional host response to lung cancer. We hypothesized that integrating these routinely available laboratory markers would improve clinical risk stratification compared with individual parameters alone. Accordingly, we evaluated the prognostic performance of individual biomarkers and composite indices derived from the DANH axis for predicting disease stage, metastasis, and cancer-associated thrombosis in patients with lung cancer.

METHODS

1. Study Design and Participants

This cross-sectional study included adults (≥ 18 years) with histologically or cytologically confirmed lung cancer treated at Viet Tiep friendship hospital from 2024 to feb 2026. Eligible patients had baseline D-dimer, complete blood count, and serum albumin measured prior to treatment.

Patients with acute infection or inflammatory disease, active anticoagulation for venous thromboembolism, severe liver disease or nephrotic syndrome, or incomplete data were excluded.

2. Variables

Dependent Variables

Outcomes included advanced TNM stage, metastasis, and cancer-associated thrombosis.

Independent Variables

Clinical variables comprised age, sex, treatment modality, and imaging findings.

Laboratory variables included D-dimer; red blood cell count, hemoglobin, hematocrit, platelet count, total and differential white blood cell counts; serum albumin. The neutrophil-to-lymphocyte ratio (NLR) was calculated as the absolute neutrophil-to-lymphocyte count ratio.

3. Laboratory Measurements

Baseline venous blood was collected before treatment initiation. Citrated plasma, EDTA blood, and serum samples were analyzed in the central laboratory under standard quality control procedures. D-dimer was measured using an immunoturbidimetric assay (ng/mL, FEU). Serum albumin was determined by the bromocresol green method (g/L) with 147 of the 184 enrolled patients. Coagulation parameters D-dimer were measured using an automated coagulation analyzer (ACL TOP 550 CTS). Complete blood counts were performed using an automated hematology analyzer (DxH 600). Serum albumin levels were determined using an automated biochemistry analyzer (DxC 700).

4. Composite Index

Patients were stratified by D-dimer, NLR, and albumin using median or ROC-derived cut-offs. One point was assigned for each adverse biomarker (elevated D-dimer, elevated NLR, low albumin, and low hemoglobin), generating a composite score ranging from 0 to 4 with higher scores indicating greater coagulation–inflammatory–nutritional dysregulation.

Formula: NLR = Neutrophil/Lymphocyte, SII = (Neutrophil x platelet) /Lymphocyte, PLR = Platelet /Lymphocyte, PNI = Albumin + 5x Lymphocyte.

5. Statistical Analysis

Analyses were conducted using stata. Continuous variables were expressed as mean ± SD or median (IQR); categorical variables as frequencies (%).

Comparisons used t-test or Mann–Whitney U test and χ^2 or Fisher's exact test, as appropriate. Spearman's correlation assessed biomarker associations. Logistic regression identified independent predictors. Diagnostic performance was evaluated by ROC curve analysis with area under the curve (AUC). A two-sided $p < 0.05$ was considered statistically significant. To evaluate the stability of the ROC estimates under conditions of unequal sample sizes, bootstrap resampling was performed with 1,000 iterations. For each iteration, AUC values were recalculated and summarized to obtain bootstrap-based estimates and confidence intervals.

RESULTS

1. Baseline characteristics of the study population

1.1 Clinical characteristics

A total of 184 patients with lung cancer were enrolled. Albumin data were available for a subset of 147 patients. The cohort was predominantly male and older, consistent with the typical demographic profile of lung cancer in real-world settings. The majority presented at an advanced stage, with a high proportion showing metastatic disease at initial evaluation. Common metastatic sites included bone, brain, and liver, reflecting established patterns of disease dissemination.

Table 1. Baseline Clinical and Laboratory Characteristics of the Study Population (n = 184; ¹ albumin subset n = 147)

No.	Characteristic	Subgroup	n (%)	Mean ± SD
1. Demographic characteristics				
1.1	Age (years)		184 (100%)	65.5 ± 9.26
1.2	Sex	Male	142 (77.2%)	
		Female	42 (22.8%)	
2. Disease characteristics				
2.1	Stage at diagnosis	Early stage	16 (8.7%)	
		Advanced stage	168 (91.3%)	
2.2	Thrombosis at diagnosis	No thrombosis	172 (93.5%)	
		Thrombosis present	12 (6.5%)	
2.3	Metastasis at diagnosis	No metastasis	51 (27.7%)	
		Metastasis present	133 (72.3%)	
2.4	Sites of metastasis (n = 133)	Brain	31 (23.3%)	
		Liver	16 (12.0%)	
		Bone	34 (25.6%)	
		Multi-organ	34 (25.6%)	
3. Laboratory parameters				
3.1	D-dimer (ng/mL)			4559 ± 7636
3.2	NLR			7.9 ± 7.7
3.3	Albumin (g/L) [1]			34.01 ± 5.66
3.4	Hemoglobin (g/L)			111.28 ± 20.84
3.5	HGB ≤ 120 g/L		124 (67.4%)	
3.6	HGB ≤ 80 g/L		15 (8.15%)	

Abbreviations: NLR, neutrophil-to-lymphocyte ratio; HGB, hemoglobin; SD, standard deviation.

[1] Albumin data available for n = 147 patients only. p values < 0.05 are considered statistically significant.

Table 1 summarises the baseline characteristics of all enrolled patients. The mean age was 65.5 ± 9.26 years, with a marked male predominance (77.2%). The majority were diagnosed at an advanced stage (91.3%), and 72.3% had metastatic disease at presentation. The most frequent metastatic sites were bone (25.6%), brain (23.3%), and liver (12.0%). Mean D-dimer and NLR were markedly elevated (4559 ± 7636 ng/mL and 7.9 ± 7.7, respectively), indicative of heightened coagulation activity and systemic inflammation. Mean albumin was

34.01 ± 5.66 g/L and mean haemoglobin was 111.28 ± 20.84 g/L, consistent with impaired nutritional and haematologic status.

1.2 General clinical laboratory characteristics of the study population according to disease stage.

To delineate stage-specific biological alterations, laboratory parameters were compared between early- and advanced-stage patients, focusing on coagulation, inflammatory, nutritional, and haematologic indices (Table 2).

Table 2. Comparison of Laboratory Parameters Between Early- and Advanced-Stage Patients

Parameter	Early stage (n = 16)	Advanced stage (n = 168)	p value
Coagulation marker			
D-dimer (ng/mL)	2112 ± 3071	4792 ± 7901	**0.005**
Inflammatory markers			
NLR	4.54 ± 3.41	8.22 ± 7.91	**0.02**
PLR	208.67 ± 82.22	323.75 ± 271.89	0.11
SII	1467.50 ± 1432.94	2602.17 ± 3415.34	0.08
Nutritional / immune status			
Albumin (g/L) [1]	37.37 ± 5.43	33.66 ± 5.59	**0.02**
PNI	44.16 ± 6.80	39.52 ± 6.54	**0.017**
Hematologic parameters			
Platelet count (×10⁹/L)	279.75 ± 146.70	302.37 ± 160.21	0.45
Hemoglobin (g/L)	123.5 ± 15.41	110.12 ± 20.95	**0.01**
White blood cell (×10⁹/L)	8.65 ± 5.49	9.85 ± 5.95	0.32
Neutrophil count (×10⁹/L)	5.98 ± 4.45	7.55 ± 5.30	0.17
Lymphocyte count (×10⁹/L)	1.35 ± 0.52	1.20 ± 0.60	0.24
Monocyte count (×10⁹/L)	0.90 ± 0.65	0.80 ± 0.46	0.80
Other biochemical parameters			
Thrombosis, n (%)	0 (0%)	12 (7.1%)	0.269
HGB ≤ 120 g/L, n (%)	6 (37.5%)	118 (70.2%)	**0.008**

Abbreviations: NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; SII, systemic immune-inflammation index; PNI, prognostic nutritional index; HGB, hemoglobin. [1] Early stage n = 14; Advanced stage n = 133. Bold p values indicate statistical significance (p < 0.05).

Advanced-stage patients exhibited significantly higher D-dimer (p = 0.005) and NLR (p = 0.02), indicating enhanced hypercoagulability and systemic inflammation. PLR and SII were also elevated but did not reach statistical significance. Albumin, haemoglobin, and PNI were significantly lower in advanced disease (p = 0.02, p = 0.01, and p = 0.017, respectively), reflecting progressive nutritional deterioration and anaemia. Other haematologic indices were comparable between groups. Thrombosis and severe anaemia were more frequent in advanced-stage patients, though these differences were not statistically significant.

2. Risk Stratification Performance of the 3 factors score: DAN, DAH, DNH, ANH and DANH - 4 factors Composite Score

To evaluate the predictive utility of routinely available laboratory biomarkers, five composite scores were constructed by integrating markers of coagulation (D-dimer), inflammation (NLR), nutritional status (albumin), and haematologic function (haemoglobin). Four three-factor models were derived: DAN (D-dimer + albumin + NLR), DAH (D-dimer + albumin + haemoglobin), DNH (D-dimer + NLR + haemoglobin), and ANH (albumin + NLR + haemoglobin). A four-factor model—DANH—incorporating all four biomarkers was also assessed. These composite indices were evaluated for their ability to stratify patient risk across three clinical outcomes: disease stage, thrombosis, and metastasis.

ROC analysis revealed heterogeneous predictive performance across biomarkers and clinical endpoints. For disease stage, D-dimer demonstrated the highest discriminative ability among single markers (AUC ≈ 0.75), outperforming NLR, haemoglobin, and albumin. For thrombosis, D-dimer and albumin achieved moderate performance (AUC ≈ 0.71 and 0.68, respectively), whereas NLR and haemoglobin showed lower accuracy. For metastasis, all individual biomarkers exhibited limited discriminative capacity (AUC < 0.60), indicating poor standalone predictive value.

Composite models—particularly the four-factor DANH model—demonstrated a marked improvement over individual biomarkers. The optimal haemoglobin threshold was ≤120 g/L for disease stage and metastasis prediction; for thrombosis, the threshold was ≤80 g/L. The ≤120 g/L cut-off did not reach statistical significance for thrombosis prediction, and the corresponding composite AUC is reported in Figure S1. For disease stage, the DANH model achieved AUC ≈ 0.78, comparable to or exceeding the three-factor models. For thrombosis, DANH maintained good performance (AUC ≈ 0.69), outperforming individual biomarkers and most partial combinations. For metastasis, AUCs remained moderate for both four- and three-factor models (≈0.55). Overall, the DANH model demonstrated consistent performance across all clinical outcomes, reflecting its capacity to integrate complementary pathophysiological information.

In the full four-factor matrix (DANH), D-dimer showed a weak positive correlation with NLR (r = 0.220) and moderate negative correlations with albumin (r = −0.433) and haemoglobin (r = −0.350), consistent with an interplay between hypercoagulability, systemic inflammation, and impaired nutritional–haematologic status. NLR was inversely correlated with both albumin (r = −0.426) and haemoglobin (r = −0.197), supporting the link between inflammatory activation and host deterioration. Albumin and haemoglobin exhibited a moderate positive correlation (r = 0.511), suggesting convergence between nutritional reserve and erythropoietic capacity.

Correlation coefficients remained directionally consistent across three-factor subsets, indicating structural stability regardless of model configuration. Overall, the matrix revealed modest-to-moderate inter-marker associations without evidence of strong collinearity, supporting the rationale for integrating these parameters into composite prognostic indices while preserving complementary biological information.

3. Risk stratification and predictive performance of biomarker indices for disease progression based on calculated cut-off values.

3.1 Determination of optimal cut-off values and score construction

Optimal cut-off values for D-dimer, neutrophil-to-lymphocyte ratio (NLR), and albumin were determined using receiver operating characteristic (ROC) curve analysis against the primary clinical outcomes. For each biomarker, the threshold corresponding to the maximum Youden index (sensitivity + specificity − 1) was selected to maximise discrimination between outcome groups. The haemoglobin cut-off of ≤120 g/L (anaemia vs. no anaemia) was derived from clinical guidelines for predicting disease stage and metastasis, whereas the threshold of ≤80 g/L (severe anaemia) was applied specifically for thrombosis prediction.

The derived optimal cut-off values were: D-dimer ≥ 986 ng/mL, NLR ≥ 4.07, haemoglobin ≤ 120 g/L (or ≤ 80 g/L for thrombosis), and albumin ≤ 36.6 g/L. Each biomarker was dichotomised accordingly, with an adverse value assigned 1 point and a non-adverse value assigned 0 points. This approach enabled objective threshold selection, minimised arbitrary categorisation, and facilitated clinically interpretable risk stratification.

Composite scores were constructed by summing individual biomarker points within each combination: DAN (D-dimer + albumin + NLR), DAH (D-dimer + albumin + haemoglobin), DNH (D-dimer + NLR + haemoglobin), ANH (albumin + NLR + haemoglobin), and DANH (all four biomarkers). For three-factor models (DAN, DAH, DNH, ANH), scores ranged from 0 to 3 and were stratified as low risk (0–1), intermediate risk (2), and high risk (3). For the four-factor DANH model, scores ranged from 0 to 4 and were categorised as low risk (0–1), intermediate risk (2–3), and high risk (4). This framework enabled standardised comparison of predictive performance across models and facilitated evaluation of disease stage, metastasis, and thrombosis risk across biologically integrated biomarker profiles.

3.2 Association of Composite Risk Scores with Disease Stage, Thrombosis, and Metastasis

Statistical significance of trends across risk categories was assessed using the chi-square test for trend.

Among patients with haemoglobin ≤120 g/L, the distribution of disease stage differed significantly across composite risk categories (Figure 3). Across all models, the proportion of patients in the high-risk category increased progressively from early- to advanced-stage disease, with a corresponding decrease in the low-risk category. Trend tests confirmed significant associations between higher risk scores and advanced disease stage for all indices: DAN ($p = 0.005$), DAH ($p = 0.008$), DNH ($p = 0.007$), ANH ($p = 0.01$), and DANH ($p < 0.0001$). The DANH model demonstrated the strongest discriminatory ability for disease stage among all evaluated indices.

For the DAN score, the proportion of high-risk patients was markedly greater in the thrombosis group than in the non-thrombosis group (77.78% vs. 37.68%), although the trend did not reach statistical significance (p =

0.059). The DAH score demonstrated a significant association with thrombosis (p = 0.002), with a notably higher proportion of high-risk patients among those with thrombosis (33.33% vs. 4.35%). Stronger associations were observed for DNH, ANH, and the four-factor DANH model (all p < 0.0001), in which the high-risk category increased from approximately 3% in non-thrombosis patients to 33.33% in those with thrombosis, consistent with a clear dose–response relationship across composite risk strata.

3.3 Robustness of ROC Analysis Assessed by Bootstrap Resampling Under Class Imbalance

To assess the robustness of ROC-based estimates under conditions of group imbalance (early stage n = 16 vs. advanced stage n = 168; non-thrombosis n = 172 vs. thrombosis n = 12), bootstrap resampling (1000 iterations) was performed. Bootstrapped AUC values were consistent with original ROC results across all models, indicating stable discriminatory performance (Table 3; supplementary Table S3). Given the limited number of thrombotic events, findings should be interpreted with appropriate caution.

Table 3. Bootstrap-Based Performance, Stability, and Degeneration Assessment of Composite Biomarker Models for Predicting Advanced-Stage Lung Cancer

Model	Optimal threshold	Mean F1 score	95% CI	Win rate (%)	Threshold ≥ 0 Win rate	Score 0 probability (mean)	Interpretation
DANH	≥ 1	**0.9595**	[0.9323 – 0.9819]	**76.9**	23.1%	**0.38**	Best overall performance, stable, non-degenerate
DNH	≥ 1	0.9541	[0.9268 – 0.9781]	16.6	35.8%	0.44	Moderate performance, partially stable
DAH	≥ 0	0.9504	[0.9231 – 0.9720]	83.4	83.4%	0.58	Degenerate (predicts most as advanced)
DAN	≥ 0	0.9502	[0.9231 – 0.9720]	67.9	67.9%	0.53	Degenerate, poor discrimination
ANH	≥ 0	0.9491	[0.9191 – 0.9756]	82.0	82%	0.58	Degenerate, limited clinical utility

Abbreviations: CI, confidence interval; DANH, D-dimer + albumin + NLR + haemoglobin; DAN, D-dimer + albumin + NLR; DAH, D-dimer + albumin + haemoglobin; DNH, D-dimer + NLR + haemoglobin; ANH, albumin + NLR + haemoglobin. Win rate reflects the proportion of bootstrap resamples in which the model outperformed all alternatives. Threshold ≥ 0 win rate indicates degenerate behaviour (predicting most samples as advanced stage). Bootstrap resamples: n = 1000.

Bootstrap analysis demonstrated that the DANH model achieved the highest predictive performance (mean F1 = 0.9595, 95% CI: 0.9323–0.9819), the greatest stability (win rate 76.9%), and the lowest degeneration rate

(threshold ≥0 win rate 23.1%; score-0 probability 0.38). In contrast, the remaining models—despite achieving comparable F1 scores—exhibited substantially higher degeneration rates (threshold ≥0 win rate up to 83.4%) and reduced discriminatory capacity, indicating limited clinical utility.

Collectively, the DANH composite score demonstrated the most robust association with cancer-associated thrombosis across all analytical approaches. The stepwise risk gradient observed across score categories supports the biological plausibility of the composite model in capturing cancer-associated hypercoagulability through the integrated assessment of inflammatory, coagulation, nutritional, and haematologic parameters.

DISCUSSION

4.1. Biological significance of coagulation, inflammation, nutritional, and hematologic alterations in lung cancer progression

In this study of 184 patients with lung cancer, Table 1 reflects a typical real-world lung cancer cohort, with older age (65.5 ± 9.26 years), male predominance (77.2%), and a high proportion of advanced-stage (91.3%) and metastatic disease (72.3%), consistent with late diagnosis and poor prognosis [1], [26]. Elevated D-dimer and NLR, together with low albumin and hemoglobin, indicate concurrent hypercoagulability, systemic inflammation, and nutritional–hematologic impairment [7], [9], [15], [27]. Notably, hemoglobin showed context-dependent roles: anemia at 120 g/L was associated with advanced disease, whereas severe anemia (≤80 g/L) better stratified thrombotic risk [16], [17]. These dual thresholds capture distinct biological dimensions—tumor progression and thrombosis—thereby enhancing the clinical utility of the model.

Table 2 demonstrates clear biological differences between early- and advanced-stage patients, reflecting the systemic nature of lung cancer progression. Advanced-stage patients exhibited significantly higher D-dimer and NLR levels ($p = 0.005$ and $p = 0.02$), indicating concurrent activation of coagulation and inflammatory pathways [9], [13], [24]. In parallel, albumin and hemoglobin levels were significantly lower ($p = 0.02$ and $p = 0.01$), reflecting worsening nutritional status and anemia associated with increased tumor burden [15], [16], [17]. Although other markers such as PLR, SII, and platelet count showed an increasing trend, these differences did not reach statistical significance, suggesting that not all individual biomarkers are sufficiently sensitive to discriminate disease stage [21], [22]. Notably, the prevalence of anemia (HGB ≤ 120 g/L) was significantly higher in advanced-stage patients ($p = 0.008$), further supporting the role of hematologic dysfunction in disease progression [28]. Overall, these findings highlight that the interplay between hypercoagulability, inflammation, and nutritional–hematologic impairment constitutes a biologically coherent axis associated with disease advancement and provides a rationale for integrated biomarker models.

4.2. Predictive value of individual biomarkers for disease stage, thrombosis, and metastasis

Figure 1A (A.1, A.2, A.3) highlights the variable performance of single biomarkers across outcomes. D-dimer showed the best discrimination for disease stage (AUC ≈ 0.75), consistent with prior evidence linking hypercoagulability to tumor burden, angiogenesis, and advanced disease [9].

For thrombosis, D-dimer and albumin had moderate accuracy (AUC ≈ 0.71 and 0.68), in line with their established roles in cancer-associated thrombosis and inflammation-related endothelial dysfunction, while

NLR and hemoglobin performed less well—possibly reflecting the limited number of thrombotic events [9], [10], [11], [29].

For metastasis, all markers showed poor performance (AUCs < 0.60), consistent with the multifactorial nature of metastatic spread. Overall, these findings support previous studies, emphasizing the relative strength of D-dimer but also the limited utility of single biomarkers, underscoring the need for integrated models [3], [6].

4.3. Superiority of composite biomarker models over single parameters

As shown in **Figure 1B (B.1, B.2, B.3),** composite biomarker models consistently outperformed individual parameters across clinical outcomes. For disease stage (B.1), all combination models achieved higher AUCs (~0.76–0.78) compared with single markers, with the four-factor model (DANH) demonstrating the best overall performance. This supports the concept that integrating coagulation, inflammation, nutritional, and hematologic parameters provides a more comprehensive reflection of tumor biology than any single marker alone [13], [22].

In thrombosis prediction (B.2), composite models also showed improved and more balanced performance (AUC ~0.65–0.70), with DAN and DNH performing slightly better than other combinations. This is consistent with previous studies suggesting that cancer-associated thrombosis arises from the interaction between hypercoagulability, systemic inflammation, and host-related factors, which cannot be fully captured by isolated biomarkers [10], [11], [25], [27].

For metastasis (B.3), although the overall discriminative ability remained modest (AUCs ~0.52–0.58), composite models still showed slight improvement over single markers, particularly DAH. This finding aligns with prior evidence that metastatic progression is highly multifactorial, and even combined indices may have limited predictive power without incorporating molecular or imaging data [2], [6], [30].

4.4. Association between composite risk scores and clinical outcomes

In the figure 2, composite risk stratification analyses underscore the clinical relevance of integrated biomarker models. All scores exhibited a consistent monotonic gradient, with higher risk categories strongly associated with advanced-stage disease, most prominently in the DANH model ($p < 0.0001$), suggesting that cumulative dysregulation across coagulation, inflammatory, nutritional, and hematologic pathways parallels tumor progression. This relationship remained evident in anemic patients (HGB ≤ 120 g/L), supporting the robustness of the model in clinically vulnerable subgroups [27], [28].

For cancer-associated thrombosis (figure 3), the association was more pronounced, with higher-risk categories markedly enriched among affected patients, particularly in the DNH, ANH, and DANH models (all $p < 0.0001$), demonstrating a clear dose–response pattern with cut off of HGB ≤ 80 g/L. however, with cut off of HGB ≤ 120 g/L, no significant trend was observed across any of the models (all p for trend > 0.05), indicating limited discriminatory capacity under this cutoff (Figure S2). These findings are biologically plausible, reflecting the interplay between hypercoagulability, inflammation, and host deterioration [27], [28].

The observed associations support the concept that tumor progression is strongly driven by systemic host responses. Cancer-related inflammation promotes neutrophil expansion, cytokine release, and endothelial

activation, facilitating angiogenesis and immune evasion [4], [5], [6], [18] while tumor-derived procoagulant factors activate the coagulation cascade, reflected by elevated D-dimer levels [9], [10], [11], [25]. These pathways are closely interconnected, as inflammation enhances coagulation and hypercoagulability further promotes tumor progression and metastasis [25]. The positive correlation between D-dimer and NLR in our cohort likely reflects this shared inflammatory–coagulatory axis. Additionally, systemic inflammation contributes to host deterioration, with hypoalbuminemia indicating impaired protein synthesis and chronic catabolism [15], and anemia arising from inflammation, nutritional deficiency, and bone marrow dysfunction [16], [17]. Together, these findings support an integrated coagulation–inflammation–nutrition–hematologic axis in lung cancer progression.

4.5. Robustness and stability of predictive models under class imbalance

Bootstrap analysis underscores the importance of model robustness beyond conventional performance metrics (Table 3, Table S3) [31]. For advanced-stage prediction, the DANH model achieved the most balanced profile, combining high F1 score, strong stability, and minimal degeneration, whereas other models despite similar F1 values - showed high degeneration rates, indicating reduced discriminatory capacity and limited clinical applicability (Table 3) [29]. In thrombosis prediction, overall performance was lower and more variable. However, models such as DNH and DANH demonstrated improved discrimination at higher thresholds, suggesting that risk stratification for thrombotic events may benefit from focusing on high-risk subgroups (Table S3). Collectively, these findings highlight the superiority of the DANH model while emphasizing the need to consider threshold effects and stability when applying composite biomarker models in clinical practice [30].

4.6. Clinical implications and potential applications

Our findings align with a growing body of evidence demonstrating the prognostic value of systemic biomarkers in oncology. Previous studies have consistently shown that elevated D-dimer levels are associated with advanced disease stage, increased thrombotic risk, and worse survival in cancer patients [9], [10], [11]. Similarly, numerous meta-analyses have demonstrated that NLR is a robust marker of systemic inflammation and is associated with poor prognosis across multiple malignancies, including lung cancer [7], [12], [13], [14]. Markers reflecting nutritional and metabolic status have also been widely studied. Serum albumin has long been recognized as an indicator of both nutritional reserve and systemic inflammation, with hypoalbuminemia consistently associated with adverse clinical outcomes [15]. Likewise, cancer-related anemia has been linked to impaired survival and tumor hypoxia-driven disease progression [16], [17]. More recently, composite indices integrating multiple systemic biomarkers have been proposed to improve prognostic discrimination. Examples include the systemic immune-inflammation index and the prognostic nutritional index, which combine hematologic and nutritional parameters to better capture the multidimensional nature of cancer-associated systemic responses [19], [20], [21], [32]. Our DANH framework extends this concept by incorporating coagulation activation, a key but often underrepresented component of systemic cancer biology. From a clinical perspective, the DANH axis provides a simple and readily accessible framework for systemic risk stratification in lung cancer. All four biomarkers are routinely measured in standard clinical practice, allowing immediate applicability without additional cost or specialized testing. Although such composite indices cannot replace imaging-based staging or molecular profiling, they may provide rapid insight into systemic disease burden and thrombotic vulnerability [1], [2], [3], [33]. This may be particularly relevant in clinical settings with

limited access to advanced molecular diagnostics. Identification of patients with elevated composite risk scores could facilitate closer monitoring for thrombotic complications and inform supportive care strategies aimed at mitigating cancer-associated morbidity [25], [27].

In summary, our findings support the concept that coagulation activation, systemic inflammation, nutritional decline, and hematologic impairment form an integrated host-response axis in lung cancer. Composite modeling of these domains through the DANH framework improves discrimination of advanced disease and cancer-associated thrombosis while highlighting the limitations of systemic biomarkers for predicting metastatic biology. These results provide a foundation for future prospective validation and suggest that integrated host–tumor biomarker models may represent a promising direction for improving clinical risk stratification in oncology.

Limitation

Several limitations merit consideration. This single-center study included a modest sample size with a predominance of advanced-stage disease, potentially limiting generalizability. The cross-sectional design precluded longitudinal assessment, and biomarker dichotomization may have reduced predictive resolution. In addition, the limited performance for metastasis highlights the need to incorporate tumor-intrinsic molecular features into future models. Finally, the relatively small number of thrombotic events may have constrained statistical power.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee in Biomedical Research of Hai Phong University of Medicine and Pharmacy. (Approval number: 13/ IRB_HPMU, dated September 15, 2023). All procedures performed in this study involving human participants were conducted in accordance with the ethical standards of the institutional research committee and with the Declaration of Helsinki. Written informed consent was obtained from all participants.

Ethical Approval

The study was approved by the Institutional Review Board of Hai Phong University of Medicine and Pharmacy (Approval No.: 13/IRB_HPMU). All patient data were anonymized and handled confidentially. Given the retrospective design of the study, the requirement for informed consent was waived in accordance with institutional ethical regulations.

Consent for publication

Written informed consent for participation waived by the Institutional Review Board due to the retrospective study design.

Availability of data and materials

The raw data supporting the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.19676154

Competing interests

The authors declare that they have no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions

Loc Thi Pham contributed equally to the conception, design, data acquisition, interpretation, and QPL drafting of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Viet Tiep friendship Hospital and Hai Phong University of Medicine and Pharmacy for their valuable support and collaboration throughout this study. Their contribution provided essential clinical and academic resources that made this work possible.

References

1. Sung H, Global Cancer Statistics. 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021. 10.3322/caac.21660
2. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. Jan. 2018;553:446–54. 10.1038/nature25183.
3. Skoulidis F, Heymach JV. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. Nat Rev Cancer. Sep. 2019;19(9):495–509. 10.1038/s41568-019-0179-8.
4. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation, *Nature*, vol. 454, no. 7203, pp. 436–444, Jul. 2008, 10.1038/nature07205
5. Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? Lancet. Feb. 2001;357(9255):539–45. 10.1016/S0140-6736(00)04046-0.
6. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. Jan. 2022;12(1):31–46. 10.1158/2159-8290.CD-21-1059.
7. Templeton AJ, et al. Prognostic role of neutrophil-to-lymphocyte ratio in solid tumors: a systematic review and meta-analysis. J Natl Cancer Inst. Jun. 2014;106(6):dju124. 10.1093/jnci/dju124.
8. Gu X-B, Tian T, Tian X-J, Zhang X-J. Prognostic significance of neutrophil-to-lymphocyte ratio in non-small cell lung cancer: a meta-analysis. Sci Rep. Jul. 2015;5:12493. 10.1038/srep12493.
9. Ay C, et al. High D-dimer levels are associated with poor prognosis in cancer patients. Haematologica. Aug. 2012;97(8):1158–64. 10.3324/haematol.2011.054718.

10. Falanga A, Marchetti M, Russo L. The mechanisms of cancer-associated thrombosis. Thromb Res. Feb. 2015;135(1):S8–11. 10.1016/S0049-3848(15)50432-5.
11. Khorana AA, Dalal M, Lin J, Connolly GC. Incidence and predictors of venous thromboembolism (VTE) among ambulatory high-risk cancer patients undergoing chemotherapy in the United States, *Cancer*, vol. 119, no. 3, pp. 648–655, Feb. 2013, 10.1002/cncr.27772
12. Ozyurek BA, Ozdemirel TS, Ozden SB, Erdogan Y, Kaplan B, Kaplan T. Prognostic Value of the Neutrophil to Lymphocyte Ratio (NLR) in Lung Cancer Cases. Asian Pac J Cancer Prev. 2017;18(5):1417–21. 10.22034/APJCP.2017.18.5.1417.
13. Guthrie GJK, Charles KA, Roxburgh CSD, Horgan PG, McMillan DC, Clarke SJ. The systemic inflammation-based neutrophil-lymphocyte ratio: experience in patients with cancer, *Crit Rev Oncol Hematol*, vol. 88, no. 1, pp. 218–230, Oct. 2013, 10.1016/j.critrevonc.2013.03.010
14. Ethier J-L, Desautels D, Templeton A, Shah PS, Amir E. Prognostic role of neutrophil-to-lymphocyte ratio in breast cancer: a systematic review and meta-analysis, *Breast Cancer Res*, vol. 19, no. 1, p. 2, Jan. 2017, 10.1186/s13058-016-0794-1
15. Gupta D, Lis CG. Pretreatment serum albumin as a predictor of cancer survival: a systematic review of the epidemiological literature. Nutr J. Dec. 2010;9:69. 10.1186/1475-2891-9-69.
16. Caro JJ, Salas M, Ward A, Goss G. Anemia as an independent prognostic factor for survival in patients with cancer: a systemic, quantitative review, *Cancer*, vol. 91, no. 12, pp. 2214–2221, Jun. 2001.
17. Knight K, Wade S, Balducci L. Prevalence and outcomes of anemia in cancer: a systematic review of the literature. Am J Med. Apr. 2004;116. 10.1016/j.amjmed.2003.12.008. Suppl 7A, pp. 11S-26S.
18. Dolan RD, McSorley ST, Horgan PG, Laird B, McMillan DC. The role of the systemic inflammatory response in predicting outcomes in patients with advanced inoperable cancer: Systematic review and meta-analysis, *Crit Rev Oncol Hematol*, vol. 116, pp. 134–146, Aug. 2017, 10.1016/j.critrevonc.2017.06.002
19. Sun K, Chen S, Xu J, Li G, He Y. The prognostic significance of the prognostic nutritional index in cancer: a systematic review and meta-analysis, *J Cancer Res Clin Oncol*, vol. 140, no. 9, pp. 1537–1549, Sep. 2014, 10.1007/s00432-014-1714-3
20. Onodera T, Goseki N, Kosaki G. Prognostic nutritional index in gastrointestinal surgery of malnourished cancer patients, *Nihon Geka Gakkai Zasshi*, vol. 85, no. 9, pp. 1001–1005, Sep. 1984.
21. Hu B, et al. Systemic immune-inflammation index predicts prognosis of patients after curative resection for hepatocellular carcinoma. Clin Cancer Res. Dec. 2014;20(23):6212–22. 10.1158/1078-0432.CCR-14-0442.
22. Liu J, et al. Prognostic value of combined nutritional and inflammatory markers in NSCLC patients receiving ICIs. Discov Oncol. Apr. 2025;16:571. 10.1007/s12672-025-02391-9.
23. Zaitsu J et al. Jul., Systemic Inflammatory Score Predicts Response and Prognosis in Patients With Lung Cancer Treated With Immunotherapy, *Anticancer Res*, vol. 41, no. 7, pp. 3673–3682, 2021, 10.21873/anticanres.15158
24. Tomita M, Shimizu T, Ayabe T, Yonei A, Onitsuka T. Preoperative neutrophil to lymphocyte ratio as a prognostic predictor after curative resection for non-small cell lung cancer, *Anticancer Res*, vol. 31, no. 9, pp. 2995–2998, Sep. 2011.

25. Khorana AA. Cancer-associated thrombosis: updates and controversies, *Hematology Am Soc Hematol Educ Program*, vol. 2012, pp. 626–630, 2012. 10.1182/asheducation-2012.1.626
26. Nguyen KT et al. Nov., Lung Cancer Outcomes in Vietnam: A 6-Year Retrospective Study, *J Immunother Precis Oncol*, vol. 8, no. 4, pp. 254–262, 2025, 10.36401/JIPO-25-12
27. Khorana AA, Kuderer NM, Culakova E, Lyman GH, Francis CW. Development and validation of a predictive model for chemotherapy-associated thrombosis. Blood. May 2008;111(10):4902–7. 10.1182/blood-2007-10-116327.
28. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Accessed: Mar. 26. 2026. [Online]. Available: https://www.who.int/publications/i/item/WHO-NMH-NHD-MNM-11.1
29. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis, *J Clin Epidemiol*, vol. 54, no. 8, pp. 774–781, Aug. 2001, 10.1016/s0895-4356(01)00341-9
30. Riley RD, Collins GS. Stability of clinical prediction models developed using statistical or machine learning methods, *Biom J*, vol. 65, no. 8, p. e2200302, Dec. 2023, 10.1002/bimj.202200302
31. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman and Hall/CRC; 1994. 10.1201/9780429246593.
32. Yang R, Chang Q, Meng X, Gao N, Wang W. Prognostic value of Systemic immune-inflammation index in cancer: A meta-analysis. J Cancer. 2018;9(18):3295–302. 10.7150/jca.25691.
33. Proctor MJ, McMillan DC, Morrison DS, Fletcher CD, Horgan PG, Clarke SJ. A derived neutrophil to lymphocyte ratio predicts survival in patients with cancer, *Br J Cancer*, vol. 107, no. 4, pp. 695–699, Aug. 2012, 10.1038/bjc.2012.292

Figures

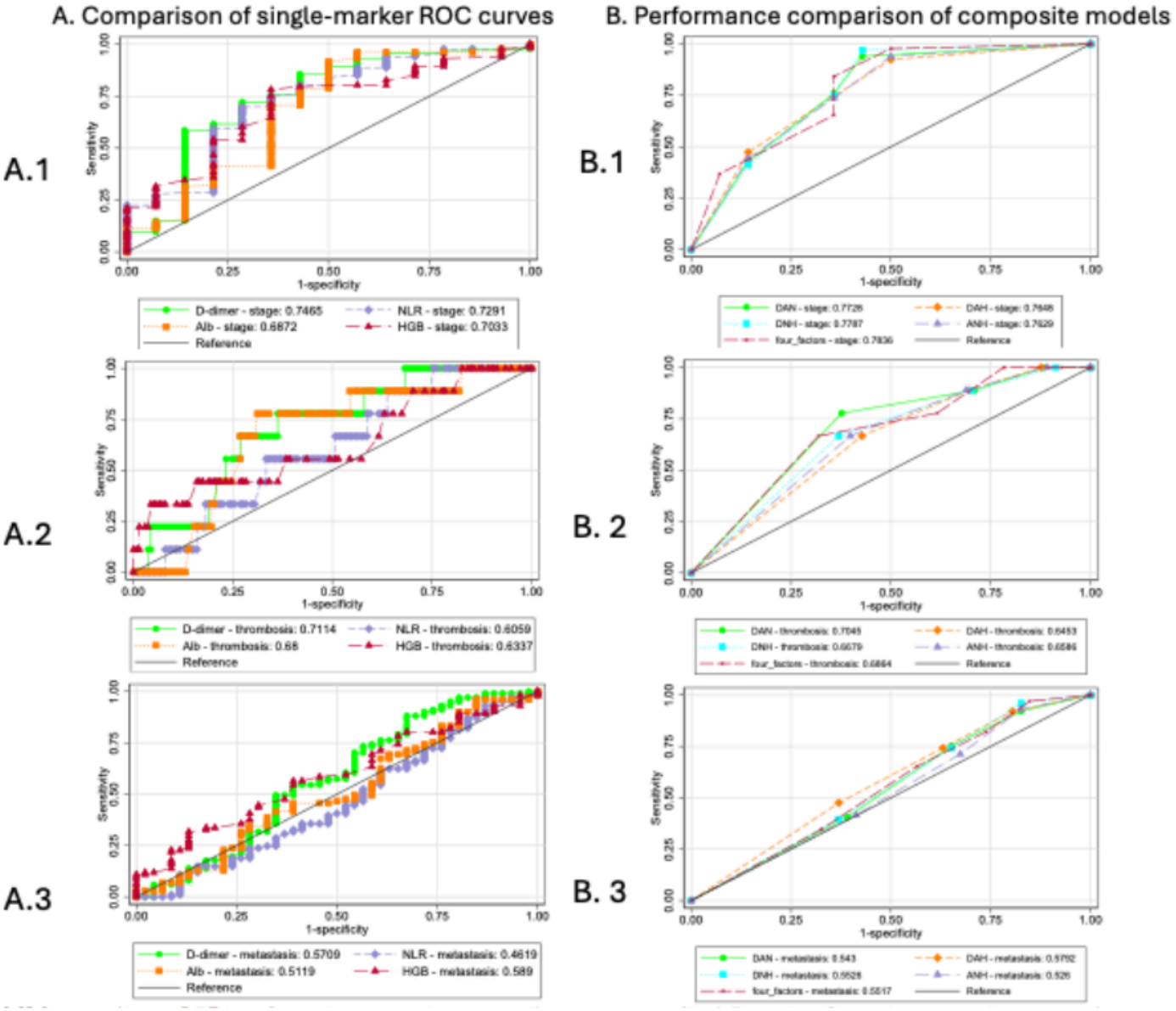


Figure 1

Comparison of diagnostic performance of individual biomarkers and composite models. *(A) ROC curves of single biomarkers for prediction of (A.1) disease stage, (A.2) thrombosis, and (A.3) metastasis. Evaluated biomarkers include D-dimer, albumin (Alb), neutrophil-to-lymphocyte ratio (NLR), and haemoglobin (HGB). (B) Performance of composite models for (B.1) disease stage, (B.2) thrombosis, and (B.3) metastasis. Models include DAN, DAH, DNH, ANH, and the four-factor DANH model. AUC, area under the ROC curve; diagonal line, reference (random prediction). n = 147.*

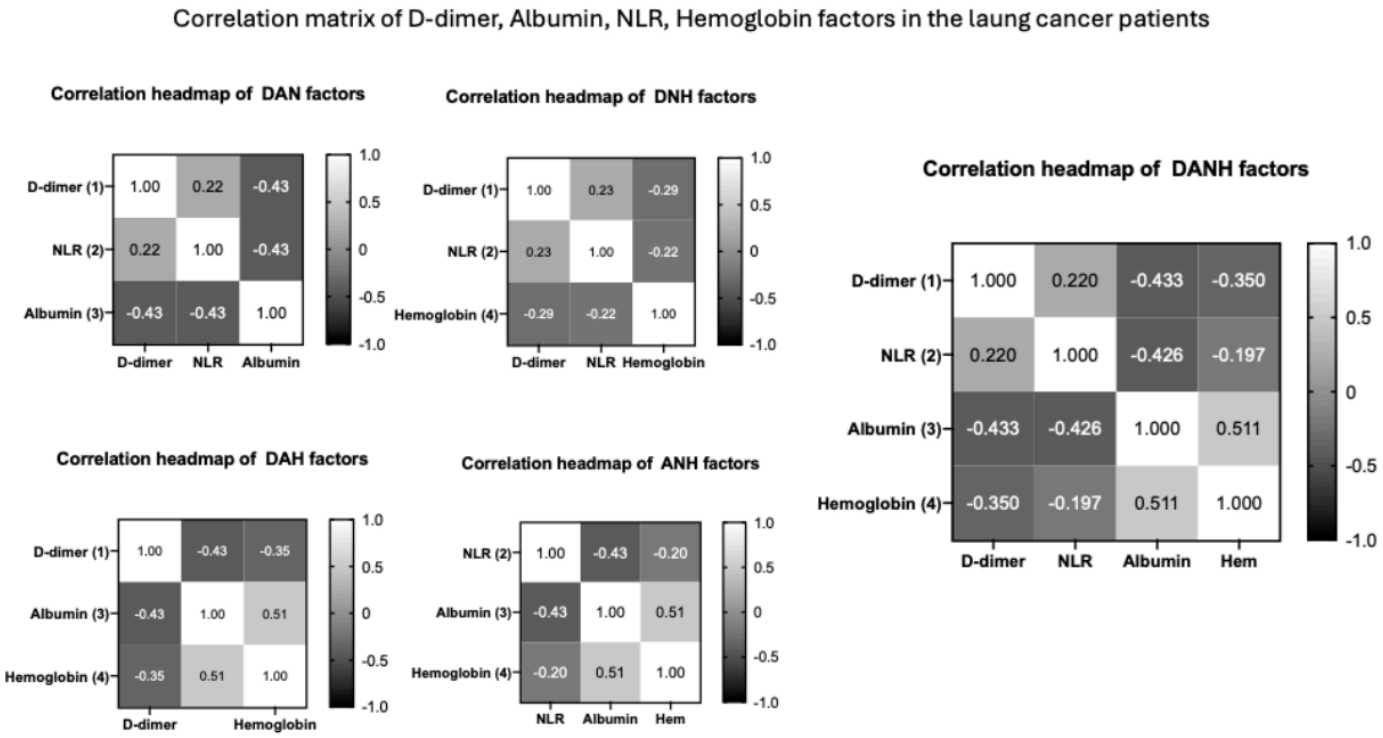


Figure 2

Correlation structure among coagulation, inflammatory, nutritional, and haematologic biomarkers. Heatmaps illustrate pairwise Spearman correlations among D-dimer, NLR, albumin, and haemoglobin across composite groupings (DAN, DNH, DAH, ANH, and DANH).

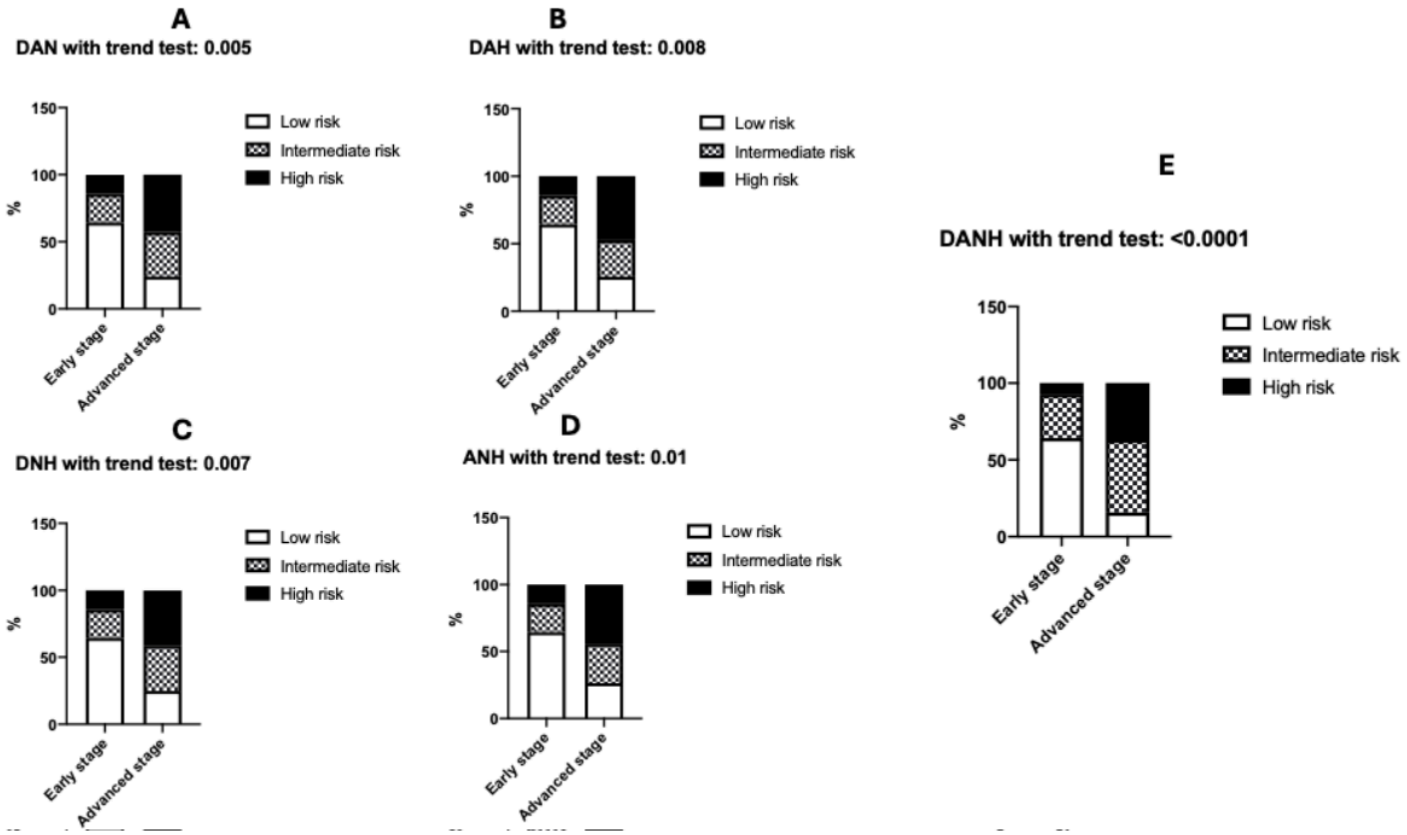


Figure 3

Distribution of disease stage across composite risk strata. (A–E) Stacked bar charts illustrating the proportion (%) of patients classified into low-, intermediate-, and high-risk groups according to five predictive models—DAN (A), DAH (B), DNH (C), ANH (D), and DANH (E)—stratified by early- and advanced-stage disease. *P value calculated using chi-square test for trend.*

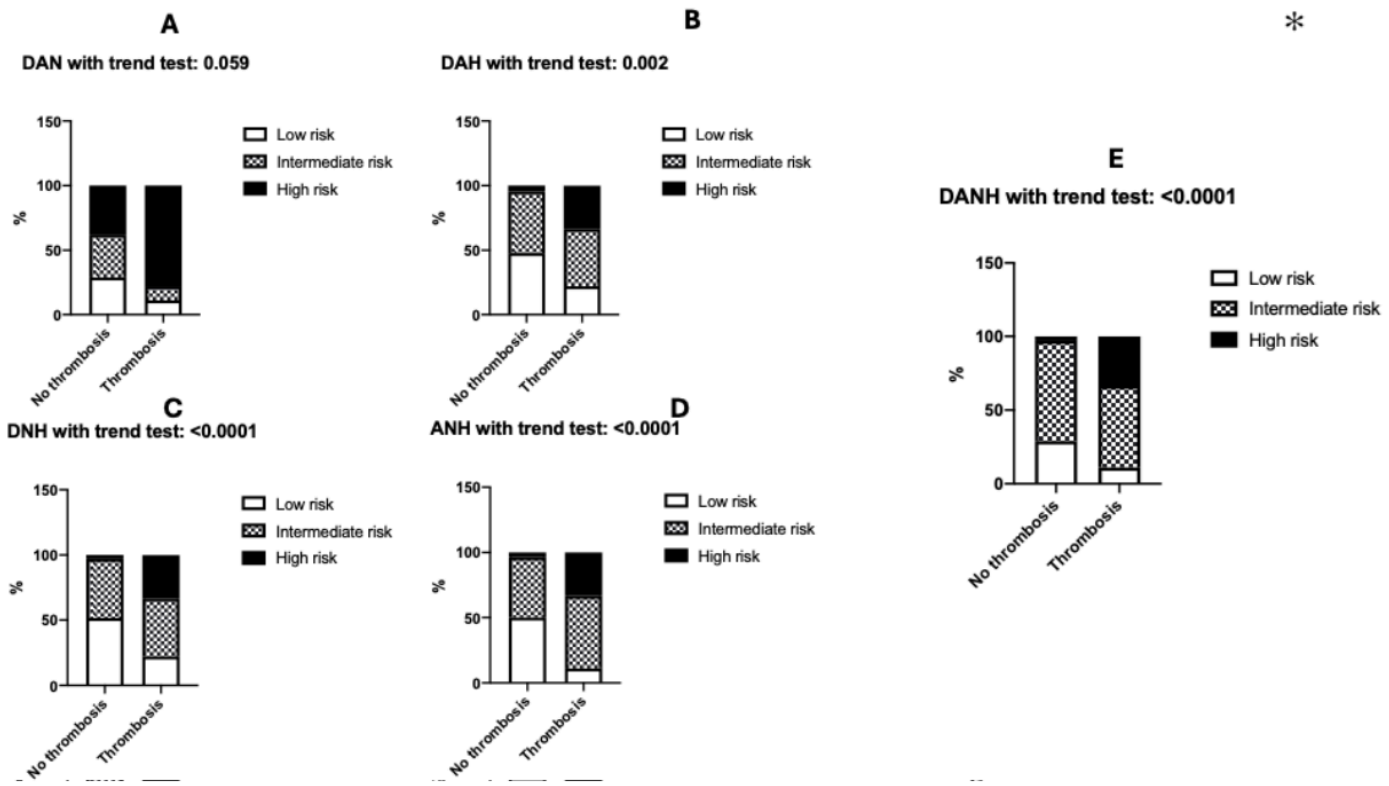


Figure 4

Distribution of cancer-associated thrombosis across composite risk score categories (haemoglobin threshold ≤ 80 g/L). (A–E) Stacked bar charts showing the proportion (%) of patients in low-, intermediate-, and high-risk groups according to DAN (A), DAH (B), DNH (C), ANH (D), and DANH (E), stratified by thrombosis status (no thrombosis vs. thrombosis present).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supp.Info.docx
- Abstract.png