Heliyon 10 (2024) e33888

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Development and validation of a TAAbs and TAAs based non-invasive model for diagnosing lung cancer

Yan Jiang [a], Gong Zhang [a], Jiayi Zhu [b], Xuchu Wang [a], Zhihua Tao [a,**], Pan Yu [a,*]

[a] Department of Laboratory Medicine, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang, 310009, China
[b] Department of Pathology, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang, 310009, China

ARTICLE INFO

Keywords:
Diagnostic model
Lung cancer
Tumor-associated autoantibody
Tumor-associated antigen
Lasso regression analysis

ABSTRACT

Background: Single Tumor-associated autoantibodies (TAAbs) and tumor-associated antigens (TAAs) have been found to have lower diagnostic efficacy in lung cancer. Our objective is to develop and validate a lung cancer prediction model that utilizes TAAbs and TAAs and to enhance the accuracy of lung cancer detection.

Methods: 1830 subjects were randomly divided into training and validation sets at a 7:3 ratio for this study. Lasso regression analysis was used to remove collinear variables, whereas univariate logistic regression analysis was employed to identify potential independent risk factors for lung cancer. A diagnostic model was constructed using multivariate logistic analysis. The results were presented as a nomogram and assessed for various performance measures, including area under the curve, calibration curve, and decision curve analysis.

Results: The diagnostic model was developed using gender, age, GAGE7, MAGE-A1, CA125, and CEA as variables. The training set had an AUC of 0.787, while the validation set had an AUC of 0.750. The calibration curves of the training and validation sets showed a strong agreement between anticipated and observed values. The nomogram performed better than any individual variable in both the training and validation sets in terms of net benefits for lung cancer detection, according to DCA analysis.

Conclusions: This study proposes a diagnostic model for lung cancer that uses TAAbs and TAAs and incorporates individual characteristics. This model can be easily applied to personalized diagnosis.

1. Introduction

Lung cancer (LC) has consistently held the highest incidence rate among malignancies and remains the primary contributor to cancer-related mortality on a global scale [1]. As anticipated, the Chinese population exhibits the highest incidence of new cases and mortality rates associated with lung cancer [2]. The 5-year survival rate of individuals diagnosed with lung cancer exhibits a range of 4 %–17 %, contingent upon the stage of the tumor upon initial detection [3]. Nevertheless, the implementation of premise-based lung cancer screening and timely initiation of early treatments have the potential to reduce lung cancer mortality and increase the 5-year survival rate to a range of 70–80 % [3,4]. Early detection of lung cancer remains the most efficacious approach for mitigating lung

* Corresponding author.
** Corresponding author.
E-mail addresses: zrtzh@zju.edu.cn (Z. Tao), yupan107@zju.edu.cn (P. Yu).

https://doi.org/10.1016/j.heliyon.2024.e33888
Received 16 November 2023; Received in revised form 26 June 2024; Accepted 28 June 2024
Available online 29 June 2024

cancer mortality, despite the fact that advancements in the treatment of advanced-stage lung tumors have only yielded marginal enhancements in survival rates [5,6]. The efficacy of sputum exfoliative cytologic examination in reducing lung cancer mortality has been demonstrated to be limited [7]. On the other hand, using low-dose spiral computed tomography (LDCT) has shown a significant 20 % decrease in lung cancer (LC) deaths and a high level of sensitivity in finding small lung nodules at an early stage [8]. However, the prevalence of false-positive results and the associated expenses may impede its widespread adoption as a standard screening technique [7,9]. Consequently, there is a need for enhanced and more precise methods of minimally invasive detection in both general cases and specifically for early lung cancer screening.

Currently, the tumor-associated antigen (TAA) is widely utilized as a non-invasive biomarker for lung cancer screening, as evidenced by its high frequency of usage [10]. However, due to the fact that TAAs are not exclusively found in tumor tissues or cells and can also be present in normal tissues or cells, the majority of TAAs lack tissue specificity and are not suitable as standalone biomarkers for the detection of asymptomatic patients with lung cancer [11]. Multiple investigations conducted in recent decades have demonstrated that autologous cells possessing the ability to undergo tumorigenesis also bear tumor-associated antigens. Tumor-associated autoantibodies (TAAbs), which are autoantibodies targeting specific cellular antigens, can be induced by abnormal exposure to or presentation of these antigens within the human immune system. The conclusion has been reached based on the evidence that TAAbs serve as immunological "sentinels" associated with the molecular mechanisms that form the foundation of cancer [12–14]. During the initial phases of carcinogenesis, there is an observed increase in the levels of TAAbs [15]. Despite low levels of related antigens or their elimination, TAAbs remain persistent in high quantities in the sera of patients [16–18]. The aforementioned advantages have contributed to the increasing prevalence of TAAbs in the realm of cancer early detection in recent times.

TAAbs have garnered interest as potential biomarkers owing to their convenient accessibility in serum samples. Nevertheless, their utility as biomarkers has been hindered by inconsistent findings across various studies, thereby limiting their practical applicability. There exists a plausible notion that cancer immunosurveillance encompasses distinct phases, namely elimination, escapes, equilibrium, and subversions. Throughout these stages, a diverse array of immune cells, including natural killer cells, various subsets of T-cells, B-cells, dendritic cells, and other relevant cell types, function in a coordinated manner within both the localized tumor microenvironment and the systemic immune response. The utilization of a single targeted anticancer agent is hindered by the presence of tumor heterogeneity. Hence, it is suggested that a more favorable approach for diagnosis could involve the development of a panel of TAAbs through the utilization of high-throughput, impartial screening in extensive populations.

The combination of various molecular indicators is hypothesized to be more efficacious in discerning cancer than relying solely on individual types of markers, owing to the distinct characteristics exhibited by each indicator. This study aimed to develop and validate risk predictive models for lung cancer screening through the examination of TAAs, TAAbs, and individual patient parameters. The evaluation of model performance involved the utilization of analytical methods such as Area Under the Curve (AUC), calibration curve, and Decision Curve Analysis (DCA). These analyses were conducted to ascertain the suitability of the model for clinical lung cancer prediction.

2. Materials and methods

2.1. Study population and data collection

The retrospective collection of clinical data was conducted at The Second Affiliated Hospital Zhejiang University School of Medicine, encompassing 828 hospital patients and 1286 healthy volunteers who underwent LDCT. The data collection period spanned from January 2018 to March 2023. The exclusion of 284 patients and volunteers was influenced by several factors, as follows: (i) Patients with certain diagnoses of additional tumors; (ii) Patients without fine-needle aspiration biopsy; (iii) Incomplete clinical and absence of serum biomarker testing data. Finally, a total of 646 patients and 1184 healthy volunteers were ultimately enrolled in this study. Prior to the commencement of the trial, the protocol was approved by the ethics committee of The Second Affiliated Hospital of Zhejiang University School of Medicine, thereby validating the protocol. The TAA and TAAb data we collected from healthy volunteers and patients were detected by chemiluminescence immunoassay.

2.2. Statistical analysis

To facilitate the construction and validation of the nomogram, the participants were randomly allocated into two distinct sets: a training set and a validation set. The allocation ratio between these sets was 7:3, respectively. An assessment of comparability between the two sets was subsequently conducted. The continuous variables, which followed a normal distribution, were characterized using the means ± standard deviation; the differences between the two sets were then examined using the Student's t-test. Continuous variables that exhibited a skewed distribution were examined for differences between the two sets by using the Mann-Whitney *U* test; these variables were described using the median, along with the 25th and 75th percentiles. The categorical data were expressed as numerical values in percentages and were subjected to statistical analysis using either the chi-square test or the Fisher's exact test for the purpose of making comparisons.

2.3. Prediction model

The researchers employed Lasso regression analysis to determine the most effective predictive features within the training dataset. Additionally, univariate logistic regression analysis was employed to eliminate risk factors that were not independent. Subsequently,

multivariate logistic regression analysis was utilized to ascertain the independent factors. Following the completion of the multivariate logistic regression analysis, factors that exhibited a two-sided p value of less than 0.05 were deemed to possess statistical significance. The evaluation of nomogram prediction accuracy in both the training and validation sets involved the utilization of the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve. The calibration curve evaluated the concordance between observed results and expected probabilities. The clinical application value of the nomograms was evaluated for a population size of 1000 using DCA and clinical impact curves. The statistical analyses were conducted using R Studio software version 4.2.2 and SPSS version 25.0.

3. Results

3.1. Characteristics of subjects

A total of 1830 patients were included in the study for the purpose of developing and validating the model. The overall prevalence of lung cancer among these patients was found to be 35.3 %. A total of 1281 participants were allocated to the training set, while 549 participants were assigned to the validation set, following a ratio of 7:3. The prevalence of lung cancer in both the training set and the validation set was 35.44 % and 34.97 %, respectively. There was no significant difference in the prevalence of lung cancer observed between the training set and the validation set. There were also no statistically significant differences observed in the other characteristics of the subjects between the training set and the validation set, including gender, smoking status, age, P53, PGP9.5, SOX2, GAGE7, GBU4-5, MAGE-A1, GAGE, CA125, CA211, CEA, NSE, and SCC. Table 1 presents the characteristics of the participants.

3.2. Features selection

The Lasso regression analysis, utilizing a 10-fold cross-validation approach, identified ten variables with nonzero coefficients that were selected from the training set (Fig. 1, Table 2). Next, a univariate logistic regression analysis was conducted on the ten coefficients in order to identify independent factors. The candidate predictors for model building were identified as gender, age, GAGE7, MAGE-A1, CA125, and CEA (Table 3).

3.3. Model development and validation

A multivariate logistic regression analysis was conducted on the training set, incorporating gender, age, GAGE7, MAGE-A1, CA125, and CEA as predictor variables. The prediction model was then constructed using the selected candidate predictors. The performance of the receiver operating characteristic curve (ROC) was evaluated by calculating the area under the curve (AUC). The AUC of the

Table 1
Baseline characteristics of individuals in training set and validation set.

Variables	Training set (N = 1281)	Validation set (N = 549)	P values
Lung cancer diagnosis status, n (%)			
No	827 (64.56 %)	357 (65.03 %)	0.873
Yes	454 (35.44 %)	192 (34.97 %)	
Lung cancer type			
Adenocarcinoma	426 (93.83 %)	181 (94.27 %)	0.831
Squamous carcinoma	28 (6.17 %)	11 (5.73 %)	
Gender			
Female	580 (45.28 %)	250 (45.54 %)	0.918
Male	701 (54.72 %)	299 (54.46 %)	
Smoking status			
Yes	979 (76.42 %)	410 (74.68 %)	0.438
No	302 (23.58 %)	139 (25.32 %)	
Age, year	50.00 (41.00, 57.00)	52.00 (41.00, 60.00)	0.3225
TAAb biomarkers, median (the 25 % percentile, the 75 % percentile)			
P53, U/mL	0.10 (0.10, 0.50)	0.20 (0.10, 0.60)	0.2791
PGP9.5, U/mL	0.10 (0.10, 0.20)	0.10 (0.10, 0.20)	0.6926
SOX2, U/mL	0.30 (0.10, 1.15)	0.40 (0.10, 1.40)	0.0942
GAGE7, U/mL	0.30 (0.10, 0.90)	0.40 (0.10, 0.95)	0.1894
GBU4-5, U/mL	0.30 (0.10, 1.20)	0.30 (0.10, 1.10)	0.7191
MAGE-A1, U/mL	0.10 (0.10, 0.10)	0.10 (0.10, 0.10)	0.4079
GAGE, U/mL	0.10 (0.10, 0.10)	0.10 (0.10, 0.10)	0.6242
TAA biomarkers, median (the 25 % percentile, the 75 % percentile)			
CA125, U/mL	10.5 (7.70, 14.20)	10.2 (7.70, 13.85)	0.4932
CA211, ng/mL	2.30 (1.70, 3.00)	2.30 (1.70, 3.10)	0.6161
CEA, ng/mL	1.70 (1.10, 2.50)	1.70 (1.20, 2.60)	0.3581
NSE, ng/mL	14.20 (12.30, 16.60)	14.20 (12.15, 16.40)	0.6457
SCC, ng/mL	0.80 (0.60, 1.00)	0.80 (0.60, 1.00)	0.2195

Data were presented as median (the 25 % percentile, the 75 % percentile) for continuous variables and count (percentage) for categorical variables.

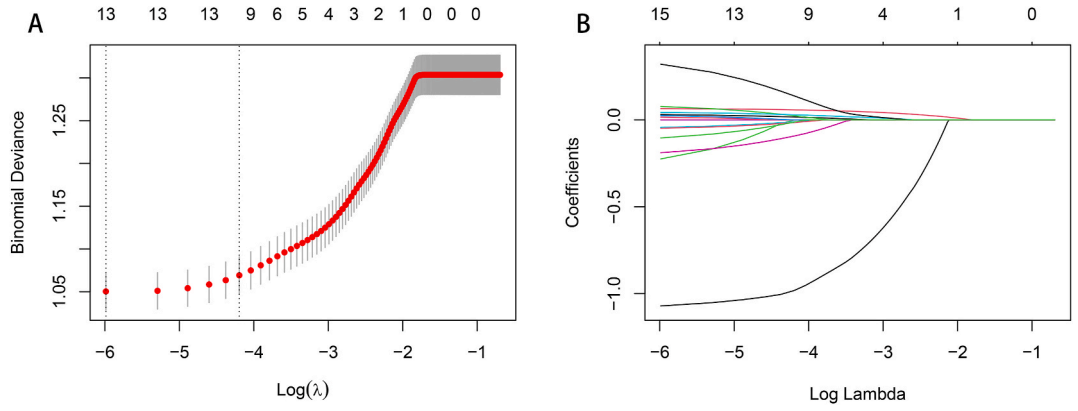

Fig. 1. Feature selection using the least absolute shrinkage and selection operator (Lasso) binary logistic regression model in the training set. (A) Identification of the optimal penalization coefficient lambda (λ) in the Lasso model with 10-fold cross-validation. (B) Lasso coefficient profiles of 15 features. The trajectory of each lung cancer-related feature's coefficient was observed in the Lasso coefficient profiles with the change of the lambda in the Lasso algorithm.

Table 2
Nonzero coefficient variables selected by Lasso regression model.

nonzero coefficient variables	Coefficients
Gender	−0.98212337
Age	0.05815464
PGP9.5	−0.01095153
GAGE7	0.01683627
MAGE.A1	0.02404285
CA125	0.02965106
CA211	−0.09270514
CEA	0.13860046
NSE	−0.01507374
SCC	−0.01282176

Table 3
Univariate logistic regression analysis of lung cancer risk factors.

Risk factors	HR (95 % CI)	P values
Gender	0.339 [0.267, 0.429]	<0.001
Age	1.067 [1.056, 1.079]	<0.001
PGP9.5	0.964 [0.917, 1.004]	0.106
GAGE7	1.054 [1.028, 1.088]	<0.001
MAGE.A1	1.099 [1.034, 1.190]	0.008
CA125	1.045 [1.027, 1.064]	<0.001
CA211	0.993 [0.916, 1.071]	0.859
CEA	1.508 [1.367, 1.673]	<0.001
NSE	0.989 [0.955, 1.023]	0.512
SCC	0.941 [0.821, 1.078]	0.382

model (Nomogram) in the training set was found to be 0.787, with a 95 % confidence interval (CI) ranging from 0.761 to 0.813. The AUC values for age, gender, GAGE7, MAGE-A1, CA125, and CEA were 0.713 (95 % CI: 0.682–0.745), 0.632 (95 % CI: 0.604–0.660), 0.586 (95 % CI: 0.554–0.619), 0.502 (95 % CI: 0.483–0.522), 0.560 (95 % CI: 0.527–0.593), and 0.648 (95 % CI: 0.617–0.679), respectively (Fig. 2A). Additionally, we assess the effectiveness of the nomogram and individual predictors in the validation dataset. The researchers calculated the AUC for the model (Nomogram) to be 0.750, accompanied by a 95 % CI spanning from 0.706 to 0.794. The AUC values for age, gender, GAGE7, MAGE-A1, CA125, and CEA were determined to be 0.662 (95 % CI: 0.611–0.713), 0.626 (95 % CI: 0.584–0.669), 0.581 (95 % CI: 0.531–0.631), 0.496 (95 % CI: 0.468–0.525), 0.567 (95 % CI: 0.517–0.616), and 0.620 (95 % CI: 0.571–0.670), respectively (Fig. 2B). The cut-off values of each indicator and the comparison between them are presented in Table 4. In addition, the calibration curve exhibited a high level of concordance between the anticipated values and the actual values in both the training set (Slope = 1.000) and the validation set (Slope = 0.818) (Fig. 2C and D).

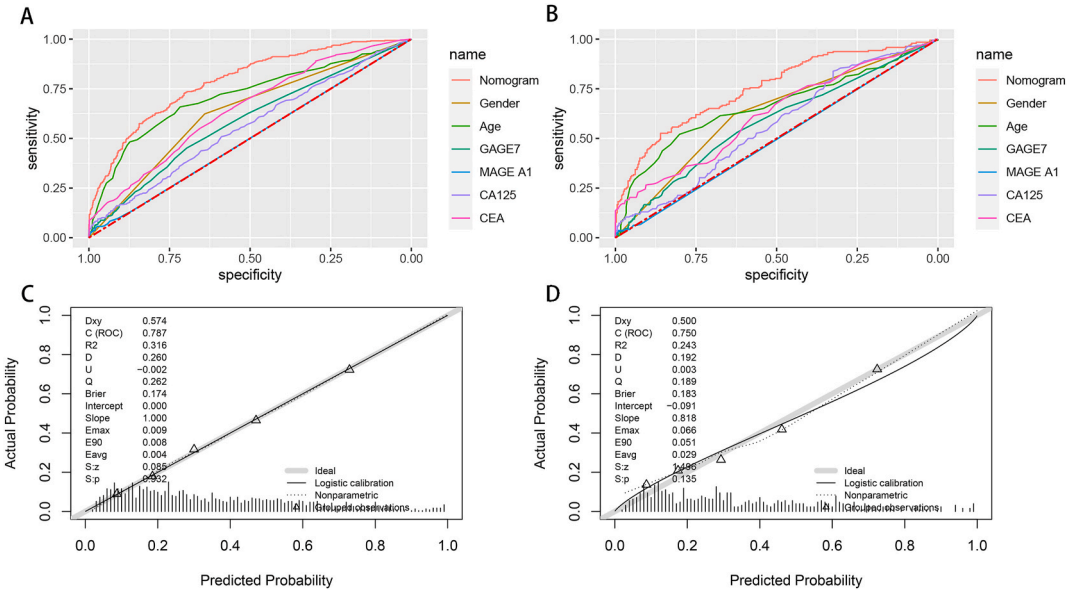

Fig. 2. Model performance assessment by ROC and calibration curve. (A) ROC curves of the factors and nomogram in the training set. (B) ROC curves of the factors and nomogram in the validation set. (C) Calibration curves of nomogram prediction in the validation set. (D) Calibration curves of nomogram prediction in the validation set.

3.4. Clinical utility

A clinical nomogram was developed to facilitate model visualization and enhance operational convenience (Fig. 3). A score was assigned to the serum biomarker and individual characteristics of patients, and a total score was calculated to determine the probabilities of lung cancer. The findings from the decision curve analysis (DCA) demonstrated that the nomogram exhibited superior net benefits in the detection of lung cancer compared to any individual factor in both the training and validation sets (Fig. 4A). The validation sets yielded comparable outcomes (Fig. 4B). Furthermore, utilizing the outcomes of DCA, we proceeded to construct clinical impact curves in order to assess the clinical efficacy of the nomograms. The nomogram's clinical impact curves demonstrated a strong alignment between the predicted probability and the actual probability within the training sets (Fig. 4C). The validation sets yielded comparable outcomes (Fig. 4D).

4. Discussion

The majority of individuals diagnosed with lung cancer are found to be in advanced stages, resulting in unfavorable prognoses and elevated mortality rates [19]. Hence, it is imperative to enhance the rate of early detection of lung cancer in order to enhance the prognosis and mitigate mortality. There is a consensus in the medical community that low-dose computed tomography (LDCT) has emerged as a widely accepted screening tool for the detection of lung cancer and has demonstrated significant efficacy in reducing mortality rates associated with this disease [20,21]. Nevertheless, there is still uncertainty regarding the specificity of LDCT, which has limited its potential application in lung cancer screening [8,22]. In recent years, there has been a significant focus on the utilization of serum biomarkers in the diagnosis of lung cancer, primarily due to their non-invasive nature and ease of accessibility [10]. The

Table 4
Performance of non-invasive models for diagnosing lung cancer.

Variables	AUC	95%CI	Sensitivity	Specificity	Cut-off	P values
Training Cohort (n = 1281)						
Nomogram	0.787	[0.761–0.813]	0.727	0.705	0.338	NA
CEA	0.648	[0.617–0.679]	0.621	0.589	1.75	<0.001
CA125	0.560	[0.527–0.593]	0.509	0.589	11.15	<0.001
GAGE7	0.586	[0.554–0.619]	0.452	0.696	0.55	<0.001
MAGE-A1	0.502	[0.483–0.522]	0.055	0.976	1.05	<0.001
Validation Cohort (n = 549)						
Nomogram	0.750	[0.706–0.794]	0.661	0.731	–	NA
CEA	0.620	[0.571–0.670]	0.615	0.571	–	<0.001
CA125	0.567	[0.517–0.616]	0.839	0.325	–	<0.001
GAGE7	0.581	[0.531–0.631]	0.531	0.622	–	<0.001
MAGE-A1	0.496	[0.468–0.525]	0.036	0.992	–	<0.001

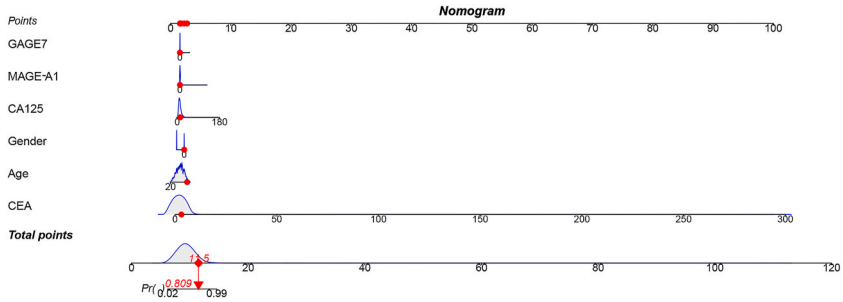

Fig. 3. Nomogram for predicting the probability of lung cancer. Nomogram used by summing the scores identified at top scale for each variable. The totaled point score then located on total score scale to determine the lung cancer probability.

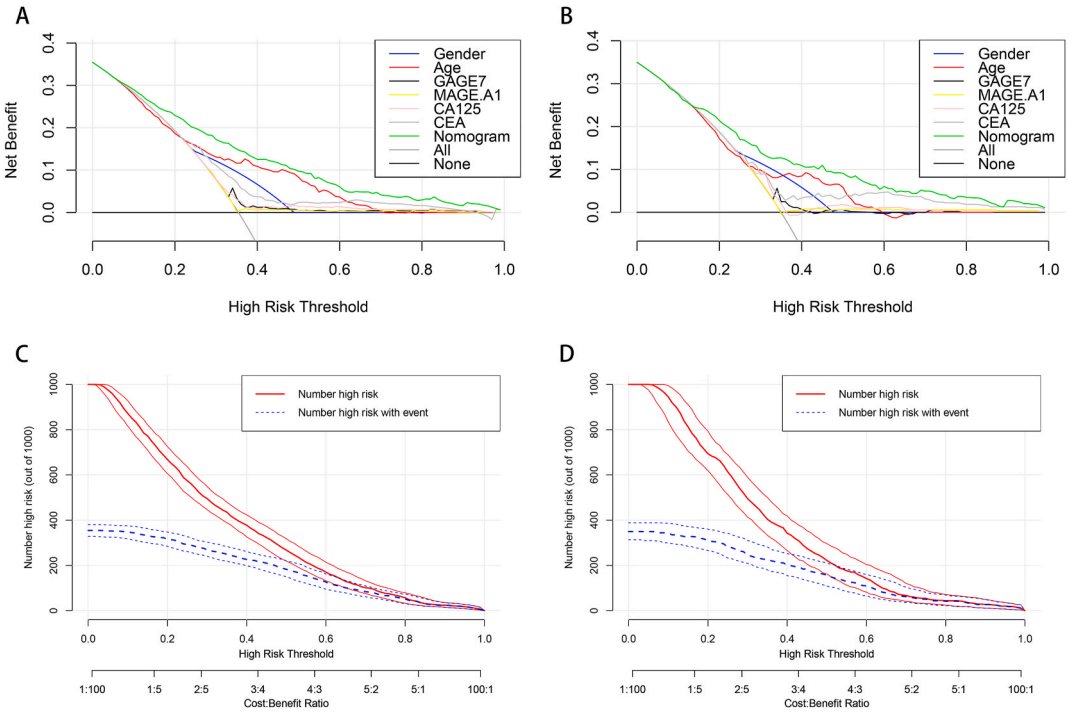

Fig. 4. Clinical utility evaluation of the nomogram by DCA and clinical impact curves. (A) DCA of nomogram prediction in the training set. (B) DCA of nomogram prediction in the validation set. (C) Clinical impact curves of nomogram prediction in the training set. (D) Clinical impact curves of nomogram prediction in the validation set.

utilization of a combined panel of multiple biomarkers has been recognized as an effective strategy to enhance the efficacy of individual biomarkers in the context of lung cancer [23].

With regards to the durability and consistency of tumor-associated autoantibodies (TAAbs), an expanding body of research has identified TAAbs as potentially valuable biomarkers for the purpose of cancer detection [24]. Li et al. developed a lung cancer prediction model using TAAbs, which were identified by analyzing differential TAAbs in the sputum of individuals diagnosed with lung cancer. The predictive model demonstrated strong performance, achieving a sensitivity of 81 % and a specificity of 83 %. However, the standardization of TAAbs levels in sputum remains uncertain due to the variability in sputum sample volume and the omission of individuals' characteristics as predictive factors in constructing the model [25]. Furthermore, certain studies employed TAAbs as a diagnostic tool for lung cancer without conducting a thorough examination of potential confounding variables or conducting internal or external validation of the model [26,27]. Considering the aforementioned factors, the current approach of detecting biomarker levels in serum remains in use. This study involves deliberately integrating tumor-associated autoantibodies (TAAbs) with tumor-associated antigens (TAAs) and individual characteristics to construct a diagnostic model for lung cancer that is suitable for clinical implementation. Lasso regression analysis screens for non-collinear variables; univariate logistic regression analysis determines independent risk factors for lung cancer; and available variables are used to build a diagnostic model whose performance is evaluated by ROC analysis. Furthermore, an internal validation was conducted to assess the reliability of the lung cancer diagnostic

model.

In this study, the initial step involves conducting Lasso regression in order to eliminate variables that exhibit collinearity. Remarkably, the inclusion of smoking, a widely recognized risk factor for lung cancer, was conspicuously absent. In general, non-small cell lung cancers (NSCLCs) are categorized into adenocarcinomas and squamous cell carcinoma based on their histopathological features. Lung adenocarcinoma (LUAD) constituted approximately 55%–60 % of cases among Chinese patients, as reported in a study [28]. The findings of this study revealed that the proportion of LUAD patients constituted approximately 90 % of the overall population of cancer patients. There is a widely held belief that smoking is closely linked to the development of lung squamous cell carcinoma [29]. Hence, it is reasonable to exclude smoking as a variable in this study. Additionally, the variables p53, SOX2, GBU4-5, and GAGE were excluded from the analysis due to their collinearity, as determined by Lasso regression. Subsequently, a univariate logistic regression analysis was conducted to eliminate non-independent risk factors. Ultimately, a nomogram was developed to predict lung cancer, utilizing six variables: age, gender, CEA, CA125, MAGE-A1, and GAGE7. All parameters are easily accessible during routine health examinations. Hence, the utilization of nomograms can prove valuable in conducting comprehensive evaluations independently, without the need for medical professionals.

The analysis of the receiver operating characteristic (ROC) revealed that the nomogram exhibited exceptional predictive ability for lung cancer. Additionally, the DCA and clinical impact curves in both the training and validation datasets indicate that a significant proportion of the threshold probabilities of the nomogram yielded favorable net benefits. However, despite the favorable performance demonstrated by the nomogram, the predictive ability of TAAbs in lung cancer was not as compelling as anticipated for a standalone predictor, which contradicts the findings of the majority of existing studies. Upon investigating the underlying reason for the observed discrepancy, it was discovered that a particular TAAb indicator exhibited satisfactory performance in these studies. Additionally, a statistically significant difference in TAAb levels was observed between individuals with lung cancer and those who were healthy volunteers [30,31]. Nevertheless, our study revealed minimal disparity in the levels of TAAb between the two cohorts, potentially attributed to variations in the methodologies and kits employed for TAAb detection across diverse investigations. Given the wide range of TAAb kits available on the market, there are variations in the methods employed to measure serum concentrations of TAAbs. Autoantibodies found in human blood are a diverse collection of polyclonal antibodies, with significant anticipated variations among individuals. Hence, the selection of a suitable calibrator poses an additional challenge in the process of commercializing autoantibody detection kits. The inclusion of gender as a significant factor in the prediction of lung cancer is an unexpected finding in our nomogram. In the present study, it is plausible that the high prevalence of adenocarcinoma among the population may have contributed to the observed phenomenon. Additionally, it is worth noting that a substantial proportion of patients diagnosed with adenocarcinoma were female [32]. In view of the variable composition of the diagnostic model, four variables GAGE7, MAGE-A1, CA125, and CEA need to be detected when using this model to diagnose lung cancer. The cost of completing the diagnostic process of this model is worthy of attention. Coincidentally, the total cost of testing for these four biomarkers in our area is almost the same as the cost of CT testing. Therefore, regardless of the detection performance or cost, this model cannot yet replace CT examination for lung cancer screening. It can be used as a supplementary approach to lung cancer screening.

5. Limitations

First of all, there is no external multicenter validation was conducted to enhance the reliability of the nomogram for predicting lung cancer. Another constraint that should be acknowledged is the absence of the inclusion of family history of lung cancer as a predictive factor in the development of a nomogram for lung cancer. This omission has the potential to impact the diagnostic accuracy of the nomograms used in the assessment of lung cancer. Furthermore, the assessment of the utility of nomograms in the grading of lung cancer necessitates the availability of comprehensive clinical staging data pertaining to patients with lung cancer, and our lack of this part of the data limits the scope of application of the nomogram. Finally, we are short of several key information such as specific lung cancer tissue types and molecular characteristics, which also limits the further exploration of the application value of the model in lung cancers with different tissue types and molecular characteristics. In summary, we have successfully incorporated tumor-associated antigens (TAAs), tumor-associated antibodies (TAAbs), and individual characteristics to develop a nomogram for the prediction of lung cancer. Our findings demonstrate favorable diagnostic performance. However, it is important to acknowledge that there are certain limitations that require further refinement. By addressing these shortcomings, we can enhance the suitability of the constructed nomogram for clinical implementation.

Data availability statement

Data associated with the study has not been deposited into a publicly available repository; all data relevant to the study is included in the article or uploaded as supplementary information.

Ethical statement

This study (no. 2023LSYD0955) was approved by the Ethics Committee of the Second Affiliated Hospital Zhejiang University School of Medicine on September 25, 2023. The study adhered to the tenants of Declaration of Helsinki. The study adhered to the tenants of Declaration of Helsinki. In view of the following reasons, implied consent was obtained when submitting the questionnaire for this study: The medical records or clinical data used in this study were obtained from the subjects' previous normal medical procedures for disease diagnosis and clinical diagnosis and treatment. There is no need to collect samples again. There are no risks to

subjects in this study. There will be no adverse impact on the rights and health of the subjects. Subjects' privacy and personally identifiable information will be strictly protected in this study.

Funding source

The financial support of this work was provided by Zhejiang Provincial Natural Science Foundation of China (no. LQ21H160016).

CRediT authorship contribution statement

Yan Jiang: Writing – original draft, Methodology. **Gong Zhang:** Visualization, Software, Formal analysis. **Jiayi Zhu:** Investigation, Data curation, Conceptualization. **Xuchu Wang:** Validation, Resources, Project administration. **Zhihua Tao:** Writing – review & editing, Supervision. **Pan Yu:** Writing – review & editing, Writing – original draft, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the support from Department of Pathology of The Second Affiliated Hospital of Zhejiang University School of Medicine, Zhejiang, China for providing information on the pathological examination results of all patients.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e33888.

References

[1] R.L. Siegel, K.D. Miller, N.S. Wagle, A. Jemal, Cancer statistics, 2023, CA A Cancer J. Clin. 73 (1) (2023) 17–48.
[2] S. Gao, N. Li, S. Wang, F. Zhang, W. Wei, N. Li, et al., Lung cancer in people's Republic of China, J. Thorac. Oncol. : official publication of the International Association for the Study of Lung Cancer 15 (10) (2020) 1567–1576.
[3] F.R. Hirsch, G.V. Scagliotti, J.L. Mulshine, R. Kwon, W.J. Curran Jr., Y.L. Wu, et al., Lung cancer: current therapies and new targeted treatments, Lancet (London, England) 389 (10066) (2017) 299–311.
[4] S.J. Adams, E. Stone, D.R. Baldwin, R. Vliegenthart, P. Lee, F.J. Fintelmann, Lung cancer screening, Lancet (London, England) 401 (10374) (2023) 390–408.
[5] H. Borghaei, L. Paz-Ares, L. Horn, D.R. Spigel, M. Steins, N.E. Ready, et al., Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer, N. Engl. J. Med. 373 (17) (2015) 1627–1639.
[6] N. Howlader, G. Forjaz, M.J. Mooradian, R. Meza, C.Y. Kong, K.A. Cronin, et al., The effect of advances in lung-cancer treatment on population mortality, N. Engl. J. Med. 383 (7) (2020) 640–649.
[7] R. Manser, A. Lethaby, L.B. Irving, C. Stone, G. Byrnes, M.J. Abramson, et al., Screening for lung cancer, Cochrane Database Syst. Rev. 2013 (6) (2013) Cd001991.
[8] D.R. Aberle, A.M. Adams, C.D. Berg, W.C. Black, J.D. Clapp, R.M. Fagerstrom, et al., Reduced lung-cancer mortality with low-dose computed tomographic screening, N. Engl. J. Med. 365 (5) (2011) 395–409.
[9] J.M. Croswell, S.G. Baker, P.M. Marcus, J.D. Clapp, B.S. Kramer, Cumulative incidence of false-positive test results in lung cancer screening: a randomized trial, Ann. Intern. Med. 152 (8) (2010) 505–512, w176-80.
[10] X. Zhang, M. Liu, X. Zhang, Y. Wang, L. Dai, Autoantibodies to tumor-associated antigens in lung cancer diagnosis, Adv. Clin. Chem. 103 (2021) 1–45.
[11] M.J. Duffy, K. O'Byrne, Tissue and blood biomarkers in lung cancer: a review, Adv. Clin. Chem. 86 (2018) 1–21.
[12] E.M. Tan, Autoantibodies as reporters identifying aberrant cellular mechanisms in tumorigenesis, J. Clin. Invest. 108 (10) (2001) 1411–1415.
[13] M.A. Cheever, J.P. Allison, A.S. Ferris, O.J. Finn, B.M. Hastings, T.T. Hecht, et al., The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research, Clin. Cancer Res. : an official journal of the American Association for Cancer Research 15 (17) (2009) 5323–5337.
[14] N.P. Restifo, M.E. Dudley, S.A. Rosenberg, Adoptive immunotherapy for cancer: harnessing the T cell response, Nat. Rev. Immunol. 12 (4) (2012) 269–281.
[15] P. Zayakin, G. Ancāns, K. Siliņa, I. Meistere, Z. Kalniņa, D. Andrejeva, et al., Tumor-associated autoantibody signature for the early detection of gastric cancer, Int. J. Cancer 132 (1) (2013) 137–147.
[16] H. Lu, V. Goodell, M.L. Disis, Humoral immunity directed against tumor-associated antigens as potential biomarkers for the early diagnosis of cancer, J. Proteome Res. 7 (4) (2008) 1388–1394.
[17] J.W. Pedersen, H.H. Wandall, Autoantibodies as biomarkers in cancer, Lab. Med. 42 (10) (2011) 623–628.
[18] K.S. Anderson, D.W. Cramer, S. Sibani, G. Wallstrom, J. Wong, J. Park, et al., Autoantibody signature for the serologic detection of ovarian cancer, J. Proteome Res. 14 (1) (2015) 578–586.
[19] H. Huang, Y. Yang, Y. Zhu, H. Chen, Y. Yang, L. Zhang, et al., Blood protein biomarkers in lung cancer, Cancer Lett. 551 (2022) 215886.
[20] F. Wang, F. Tan, S. Shen, Z. Wu, W. Cao, Y. Yu, et al., Risk-stratified approach for never- and ever-smokers in lung cancer screening: a prospective cohort study in China, Am. J. Respir. Crit. Care Med. 207 (1) (2023) 77–88.
[21] A.L. Potter, A.L. Rosenstein, M.V. Kiang, S.A. Shah, H.A. Gaissert, D.C. Chang, et al., Association of computed tomography screening with lung cancer stage shift and survival in the United States: quasi-experimental study, BMJ (Clinical research ed) 376 (2022) e069008.
[22] S. Sone, F. Li, Z.G. Yang, T. Honda, Y. Maruyama, S. Takashima, et al., Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner, Br. J. Cancer 84 (1) (2001) 25–32.
[23] D. Crosby, S. Bhatia, K.M. Brindle, L.M. Coussens, C. Dive, M. Emberton, et al., Early detection of cancer, Science 375 (6586) (2022) eaay9040.
[24] C. Desmetz, A. Mange, T. Maudelonde, J. Solassol, Autoantibody signatures: progress and perspectives for early cancer detection, J. Cell Mol. Med. 15 (10) (2011) 2013–2024.

[25] N. Li, V.K. Holden, J. Deepak, N.W. Todd, F. Jiang, Autoantibodies against tumor-associated antigens in sputum as biomarkers for lung cancer, Transl Oncol 14 (2) (2021) 100991.
[26] D. Jiang, X. Zhang, M. Liu, Y. Wang, T. Wang, L. Pei, et al., Discovering panel of autoantibodies for early detection of lung cancer based on focused protein array, Front. Immunol. 12 (2021) 658922.
[27] Y. Mu, J. Li, F. Xie, L. Xu, G. Xu, Efficacy of autoantibodies combined with tumor markers in the detection of lung cancer, J. Clin. Lab. Anal. 36 (8) (2022) e24504.
[28] P. Chen, Y. Liu, Y. Wen, C. Zhou, Non-small cell lung cancer in China, Cancer Commun. 42 (10) (2022) 937–970.
[29] S. Saito, F. Espinoza-Mercado, H. Liu, N. Sata, X. Cui, H.J. Soukiasian, Current status of research and treatment for non-small cell lung cancer in never-smoking females, Cancer Biol. Ther. 18 (6) (2017) 359–368.
[30] J.R. Jett, L.J. Peek, L. Fredericks, W. Jewell, W.W. Pingleton, J.F. Robertson, Audit of the autoantibody test, EarlyCDT®-lung, in 1600 patients: an evaluation of its performance in routine clinical practice, Lung cancer (Amsterdam, Netherlands) 83 (1) (2014) 51–55.
[31] S. Ren, S. Zhang, T. Jiang, Y. He, Z. Ma, H. Cai, et al., Early detection of lung cancer by using an autoantibody panel in Chinese population, OncoImmunology 7 (2) (2018) e1384108.
[32] X.N. Zou, D. Lin, A. Chao, X. Wan, Q. Feng, J. Li, et al., Histological subtypes of lung cancer in Chinese women from 2000 to 2012, Thoracic cancer 5 (5) (2014) 447–454.