Asia-Pacific Journal of Oncology Nursing 12 (2025) 100680

Contents lists available at ScienceDirect

Asia-Pacific Journal of Oncology Nursing

journal homepage: www.apjon.org

Original Article

Artificial intelligence-assisted machine learning models for predicting lung cancer survival

Yue Yuan[a,#], Guolong Zhang[b,#], Yuqi Gu[c], Sicheng Hao[d], Chen Huang[e], Hongxia Xie[f], Wei Mi[a,*], Yingchun Zeng[g,*]

[a] School of Nursing, Hunan University of Medicine, Huaihua, China
[b] Respiratory Intervention Center, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
[c] School of Medicine, Hangzhou City University, Hangzhou, China
[d] Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[e] Nursing Department, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, China
[f] School of Computing Sciences, Hangzhou City University, Hangzhou, China
[g] Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

ARTICLE INFO

Keywords:
Lung cancer survival
Large language model
Predictive analytics
Nursing decision-making

ABSTRACT

Objective: This study aimed to evaluate the feasibility of large language model-Advanced Data Analysis (ADA) in developing and implementing machine learning models to predict survival outcomes for lung cancer patients, with a focus on its implications for nursing practice.
Methods: A retrospective study design was employed using a dataset of lung cancer patients. Data included sociodemographic, clinical, treatment-specific, and comorbidity variables. Large language model-ADA was used to build and evaluate three machine learning models. Model performance was validated, and results were presented using calibration plots.
Results: Of 737 patients, the survival rate of this cohort was 73.3%, with a mean age of 59.32 years. Calibration plots indicated robust model reliability across all models. The Random Forest model demonstrated the highest predictive accuracy among the models. Most critical features identified were preoperative white blood cells (2.2%), preoperative lung function of Forced Expiratory Volume in one second (2.1%), preoperative arterial oxygen saturation (1.9%), preoperative partial pressure of oxygen (1.7%), preoperative albumin (1.6%), preoperative preparation time (1.5%), age at admission (1.5%), preoperative partial pressure of carbon dioxide (1.5%), preoperative hospital stay days (1.5%), and postoperative total days of thoracic tube drainage (1.4%).
Conclusions: Large language model-ADA effectively facilitates the development of machine learning models for lung cancer survival prediction, enabling non-technical health care professionals to harness the power of advanced analytics. The findings underscore the importance of preoperative factors in predicting outcomes, while also highlighting the need for external validation across diverse settings.

Introduction

According to the International Agency for Research on Cancer of the World Health Organization, the number of new cancer cases worldwide reached 20 million in 2022, with lung cancer being the most common malignant tumor.[1] The number of new lung cancer cases reached 2.5 million, accounting for 12.4% of all newly diagnosed cancer cases globally.[1] In 2022, China recorded 4.82 million new cancer cases, with lung cancer accounting for 1.06 million cases, representing 21.9% of all newly diagnosed cancers, making it the most prevalent type.[2] With an increasingly aging population, nearly half of all lung cancer patients are over the age of 60. By 2050, the number of lung cancer cases in this age group is expected to double from current figures.[3] Lung cancer has become a major threat to the life and health of residents in China, causing significant harm to both physical and mental well-being and placing a heavy burden on health care resources.[1–3] With the introduction of

* Corresponding authors.
E-mail addresses: directorhuli@gmail.com (W. Mi), zengy25@nus.edu.sg (Y. Zeng).
[#] These authors contributed equally to this work.

https://doi.org/10.1016/j.apjon.2025.100680
Received 6 November 2024; Accepted 4 March 2025
2347-5625/© 2025 The Author(s). Published by Elsevier Inc. on behalf of Asian Oncology Nursing Society. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

advanced diagnostic techniques and new cancer treatment methods, the 5-year survival rate for early-stage lung cancer patients has reached as high as 67%.[4]

Predicting survival outcomes for lung cancer patients involves identifying various demographic, clinical, and treatment-related factors that influence prognosis. Machine learning models have shown superior predictive power by analyzing non-linear relationships between diverse patient data sets, including clinical features (e.g., disease stage, treatment types), and demographic factors (e.g., age, gender).[5–7] These algorithms can integrate large datasets, handle complex interactions between variables, and continuously improve with new data, offering more personalized survival predictions.[7,8] Moreover, machine learning models such as random forests, and support vector machines have demonstrated high predictive accuracy in lung cancer survival by identifying key factors influencing prognosis. For instance, machine learning models have shown superior discriminative ability over traditional methods, like logistic regression, in several studies, offering a more nuanced and personalized approach to treatment planning.[9,10] However, most of these studies have relied on the SEER database.[11] While clinical features such as comorbidities,[12] lung function,[13] and certain serum markers[14] have been shown to be associated with lung cancer prognosis, they have been included in fewer studies. This could lead to bias or less personalized predictions. Therefore, a more precise machine learning model that incorporates individual clinical features is still needed to provide a more reliable tool for predicting outcomes in lung cancer.

Despite the growing potential of machine learning models to enhance clinical decision-making, several barriers exist for nurses in effectively applying these technologies in practice. One key challenge is the lack of technical expertise among nurses, who often have limited training in data science, algorithms, and the interpretation of complex machine learning outputs.[7,10] Integrating machine learning into clinical workflows requires nurses to understand not only how these models function but also their limitations, such as bias in data and the need for continuous model validation.[15] Furthermore, the complexity of patient data used in machine learning models often goes beyond the scope of traditional nursing education, making it difficult for nurses to interpret results or make informed decisions based on these models.[9,16]

As nurses increasingly need to integrate advanced technologies into clinical practice, artificial intelligence (AI) assisted tools, like large language model-Advanced Data Analysis (ADA), provide a simple and efficient way to perform machine learning analysis without requiring extensive technical knowledge.[17] By allowing users to execute tasks with natural language prompts, such as building predictive models for patient outcomes, large language model-ADA democratizes access to complex machine learning techniques and significantly reduces technical barriers.[17] This tool automates tasks like data preprocessing, model selection, and outcome prediction,[17] minimizing the cognitive load on nurses and enabling them to focus more on patient care. With its ability to deliver results comparable to those developed by experienced data scientists, large language model-ADA allows nurses to confidently conduct machine learning model development and validation for outcome predictions, enhancing their decision-making capabilities in clinical environments.

Therefore, this study aimed to evaluate the feasibility of large language model-ADA in simplifying the development and implementation of machine learning models for predicting lung cancer survival outcomes, making advanced tools more accessible to non-technical health care professionals.

Methods

Study population

Eligible cases included all patients who were diagnosed with lung cancer at The First Affiliated Hospital of Guangzhou Medical University from January to December 2021. Patients were included in this study if they could be followed up by telephone, while those who could not be reached via phone were excluded. Detailed inclusion criteria are age ranging from 18 to 80 years, and pathologically confirmed lung cancer with disease stages I–IV. Exclusion criteria are expected survival of less than 6 months, and incomplete information for follow-up, or unexpected death during the study period.

Data collection

This retrospective study collected patients' demographic, clinical information and comorbidity information by retrieving from the studied hospital's medical recording system, and the survival outcomes of lung cancer patients were collected by the research nurse. Structured variables were categorized into four groups: sociodemographic information, clinical data, treatment specifics, and comorbid conditions, as shown in Table 1.

Data analysis

This study utilized large language model-ADA to assist in the machine learning model analysis. Based on the smaller sample size and the complexity of training deep learning models, we opted for more interpretable and computationally efficient models including Random Forest, Support Vector Machines, and Cat Boost. These models were able to manage complex, high-dimensional datasets and capable of capturing non-linear relationships and interactions between features and have been proved to be efficient in various predictive tasks.[18] The analysis procedure of this study is illustrated in Fig. 1.

The dataset was divided into a training set (80%) and a validation set (20%). The training set was used for model development, while the validation set was used for hyperparameter tuning and performance evaluation. Prior to model training, necessary preprocessing steps were carefully implemented. Missing values were handled using specific imputation techniques: mean imputation for continuous variables and mode imputation for categorical variables. For variables with a high proportion of missing data, removal methods were considered. Outliers were identified using the Z-score method, where data points with an absolute Z-score greater than 3 were considered as outliers. These outliers were then either transformed or removed, depending on their nature and impact on the overall analysis.

Predictor variables were then input into the three machine learning models using large language model-ADA to examine survival prediction and construct the prediction model. Model performance was evaluated using Accuracy (ACC) and Area Under the Curve (AUC). ACC refers to the ratio of correctly predicted samples to the total number of samples, reflecting overall classification accuracy. An AUC ≥ 0.7 indicates good predictive ability. The performance on the validation dataset was further assessed using Precision-Recall curves and calibration plots. Precision was used to evaluate the proportion of true positive predictions among all positive predictions made by the model, which is important for assessing the model's reliability in identifying poor survival outcomes. Recall measured the model's ability to correctly identify patients with poor

Table 1
Categories of predicting variables.

Category	Variables
Demographics	Age, gender, smoking history, alcohol history, BMI, etc.
Clinical information	Pathological classification, TNM stage, lung function, routine blood test, total protein, adenocarcinoma, duration of hospital stays, etc.
Surgery type	VATS, lobectomy, segmental resection, wedge resection, sleeve resection, etc.
Comorbidities	Hypertension, asthma, diabetes mellitus, cardiovascular disease, chronic obstructive pulmonary disease, etc.

BMI, Body Mass Index; VATS, Video-Assisted Thoracoscopic Surgery; TNM: tumor-node-metastasis.

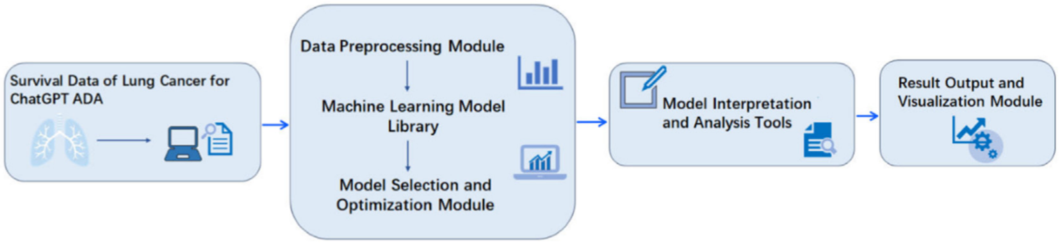

Fig. 1. Analytic schema of machine learning models. Large language model ADA, Generative Pre-Trained Transformer Advanced Data Analysis.

survival outcomes (true positives), which is critical for ensuring high-risk patients are accurately identified. The F1 Score, which represents the harmonic mean of precision and recall, was used to provide a balanced measure of model performance, particularly in the context of class imbalances.

This study employed SHapley Additive exPlanations (SHAP) to analyze the feature contributions to the model's predictions. Large language model ADA was tasked with autonomously performing a SHAP analysis, SHAP values were computed for the final trained Random Forest model to determine the impact of individual features on both the overall prediction and for specific instances. The top 10 contributing features was extracted to a SHAP summary plot to identify which features were the most impactful in the model's decision-making process.

To ensure accuracy, one computing data scientist reviewed the Python code generated by large language model ADA and independently re-implemented the procedure and confirmed these three models' output.

Results

Characteristics of lung cancer patients

This study included 737 lung cancer patients, with 197 patients died within 12-month follow-up, and the survival rate was 73.3%. The mean age at diagnosis was 59.32 (SD = 10.89). More than half were male patients (n = 414, 56.2%). All patients had surgery, most of them had Video-Assisted Thoracoscopic Surgery (n = 611, 82.9%). Adenocarcinoma was the most common tumor type (n = 497, 67.4%). The demographic and clinical characteristics of the study participants are presented in Table 2.

Prediction performance of machine learning models

Of the machine learning models analyzed, three models have good predictive performance: the Random Forest classifier exhibited relatively superior predictive performance than Support Vector Machines and CatBoost classifiers (Table 3 and Fig. 2), with accuracy of 0.71. The calibration plot demonstrated that all three classifiers closely approximated the ideal calibration line, indicating robust calibration reliability. The performances of the three machine learning models on the validation set are visually represented in Fig. 3.

Feature importance for predicting lung cancer survival

Feature importance for predicting survival outcomes of lung cancer was assessed using the feature importance metric provided by the Random Forest model. The 10 most critical features identified were preoperative white blood cells (2.2%), preoperative lung function of Forced Expiratory Volume in one second (2.1%), preoperative arterial oxygen saturation (SaO_2) (1.9%), preoperative partial pressure of oxygen (PaO_2) (1.7%), preoperative albumin (protein in your blood plasma) (1.6%), preoperative preparation time (1.5%), age at admission (1.5%), preoperative partial pressure of carbon dioxide (PCO_2) (1.48%), preoperative hospital stay days (1.5%), and postoperative total days of thoracic

Table 2
Characteristics of the study participants (N = 737).

Variables	Mean (SD)	n (%)
Age (years)	59.32 (10.89)	
BMI (kg/m^2)	22.87 (3.39)	
Sex		
Male		414 (56.2)
Female		323 (43.8)
Cigarette smoking		
Yes		118 (16.0)
Never		375 (50.9)
Cessation		244 (33.1)
Alcohol consumption		
Yes		263 (35.7)
None		474 (64.3)
Comorbidities		
Asthma		51 (6.9)
Hypertension		151 (20.5)
Diabetes mellitus		58 (7.9)
Cardiovascular disease		65 (8.8)
COPD		24 (3.3)
Preoperative lung function		
FEV1 ratio	88.92 (19.59)	
FEV1/FVC ratio	96.78 (13.78)	
Preoperative respiratory function		
SaO_2 (%)	96.92 (2.28)	
PaO_2 (mmHg)	109.69 (32.05)	
PCO_2 (mmHg)	43.8 (5.85)	
Preoperative blood test		
White blood cells (cells/μL)	13.69 (5.76)	
Red blood cells (million cells/μL)	4.23 (0.54)	
Platelets (thousand/μL)	211.57 (66.17)	
Hemoglobin (g/dL)	124.56 (14.83)	
Total protein (g/dL)	68.15 (23.04)	
Albumin (g/dL)	38.08 (4.79)	
Bilirubin (mg/dL)	12.20 (5.17)	
Blood glucose (mmol/L)	8.69 (1.98)	
Pathological classification		
Adenocarcinoma		567 (76.9)
Squamous carcinoma		118 (16.0)
Others		52 (7.1)
TNM grade		
I		399 (54.1)
II		132 (17.9)
III		153 (20.8)
IV		53 (7.2)
Surgery type		
VATS		611 (82.9)
Others (lobectomy, segmentectomy, wedge/sleeve resection, etc.)		126 (17.1)
Thoracic tube drainage (yes)		540 (73.3)
Total days of thoracic tube drainage	5.84 (3.35)	
Total days of hospital stays	20.40 (8.00)	
Preoperative hospital stays (days)	10.51 (6.04)	
Postoperative hospital stays (days)	8.89 (4.82)	

BMI, Body Mass Index; COPD, chronic obstructive pulmonary disease; FEV1: Forced Expiratory Volume in one second; FEV1/FVC Ratio: the ratio of FEV1 to Forced Vital Capacity (FVC); VATS, Video-Assisted Thoracoscopic Surgery; TNM, tumor-node-metastasis.
Note: TNM staging system for lung cancer: T (Tumor): Refers to the size and extent of the primary tumor. T1: Tumor $\leq$ 3 cm in greatest dimension; T2: Tumor $>$ 3 cm but $\leq$ 5 cm; T3: Tumor $>$ 5 cm but $\leq$ 7 cm or one that invades nearby tissues; T4: Tumor $>$ 7 cm or invades major structures such as the heart, large blood vessels, or esophagus.

Table 3
Predictive performance of three machine learning models ($N = 737$).

	Accuracy	AUC	Precision	Recall	F1
Random forest	0.71	0.62	0.74	0.91	0.81
Support vector machines	0.70	0.67	0.70	1.0	0.82
CatBoost model	0.70	0.60	0.72	0.93	0.81

AUC, area under the curve.

tube drainage (1.4%). Features importance highlights their significance in predicting survival outcomes of lung cancer (Fig. 4).

Discussion

The key findings of this study show that machine learning models, specifically those developed using large language model-ADA, provide strong predictive capabilities for lung cancer survival outcomes. This study found that Random Forest model demonstrated the highest accuracy in predicting survival, which is consistent with previous research.[9,15] Findings of this study support the growing role of predictive analytics in personalized cancer care, where tailored nursing interventions can lead to better patient outcomes in complex conditions such as lung cancer.

By incorporating both clinical and demographic factors, such as age, tumor size, and treatment modality, these models achieved significant accuracy and reliability in prognosis prediction. These results align with previous studies, such as Didier et al.,[9] which also identified Random Forest as a superior model for lung cancer survival prediction, highlighting its ability to handle complex, high-dimensional clinical data effectively.[9] The Random Forest model, which uses decision trees as base classifiers, mitigates the impact of individual trees by employing a voting or averaging mechanism, thereby reducing sensitivity to outliers. Additionally, Random

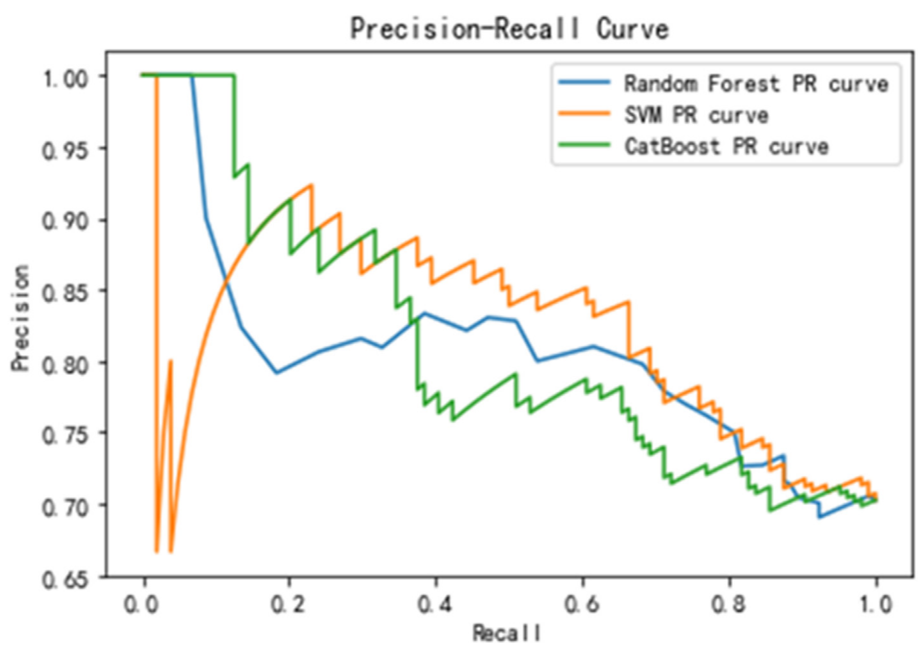

Fig. 2. Precision–recall curves of the three models in the validation set. PR, prevision-recall; SVM, Support Vector Machines.

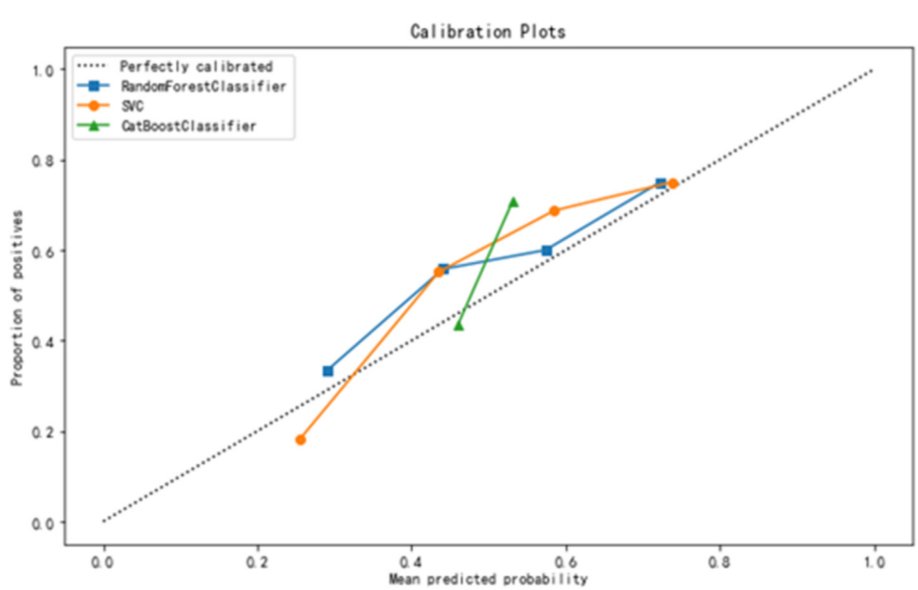

Fig. 3. Calibration plots of the three models in the validation set. SVM, Support Vector Machines.

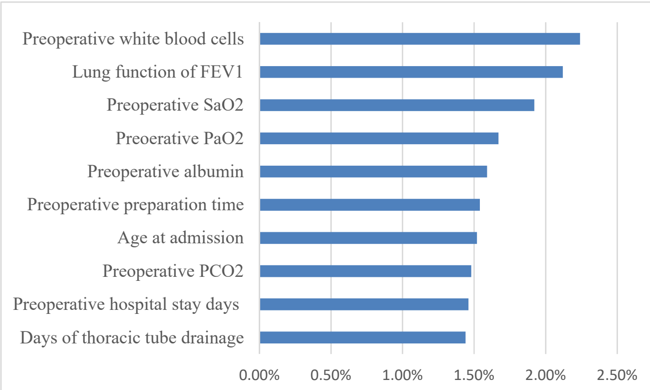

Fig. 4. Top 10 feature importance for predicting survival outcomes. FEV1: Forced Expiratory Volume in one second. SaO_2: Arterial Oxygen Saturation (the percentage of hemoglobin saturated with oxygen in arterial blood); PaO_2: Partial Pressure of Oxygen in Arterial Blood (the amount of oxygen gas in the blood); PCO_2: Partial Pressure of Carbon Dioxide (the amount of carbon dioxide gas in the blood).

Forest combats overfitting by randomly selecting subsets of features and samples during training, which enhances its predictive performance.[19] Consequently, applying machine learning, particularly Random Forest, to predict postoperative outcomes based on patient characteristics enables earlier prediction of survival prediction. This, in turn, facilitates the optimization of medical resource allocation and treatment protocols, ultimately improving treatment outcomes. Other research also reported that integrating clinical and imaging data into machine learning models further enhances prediction accuracy, supporting the findings of this study.[20] As increasing the predictive capacity of machine learning models is crucial in clinical settings, reliable prognostic tools can directly influence treatment decisions and patient management strategies.

Feature importance analysis revealed that the 10 most critical predictors of survival outcomes included preoperative clinical indicators like white blood cell count, Forced Expiratory Volume in one second (FEV1), arterial oxygen saturation (SaO_2), partial pressure of oxygen (PaO_2), and albumin levels. These features are vital as they reflect patients' physiological and functional status before surgery, which can significantly impact their recovery and overall prognosis. Total protein and albumin have been identified as significant predictive factors for lung cancer survival, consistent with previous research.[21] As key markers for assessing nutritional status, total protein levels are critical, with malnutrition being strongly associated with increased mortality.[22,23] Low preoperative albumin and total protein levels are markers of malnutrition, which has been widely associated with poor outcomes in cancer patients. Implementing preoperative nutritional support and monitoring may help to enhance patient resilience to surgical stress and postoperative recovery, as malnourished patients are more prone to infections, slower wound healing, and other complications.

This study also found that several preoperative factors, such as age at admission, preoperative respiratory and lung function, and the duration of hospital stays, were significant predictors of lung cancer survival. Age is an established risk factor for survival outcomes, with older patients often having comorbidities that increase surgical risks. Additionally, a longer preoperative hospital stay may reflect greater baseline severity or the need for extensive preoperative assessments, which can correlate with a poorer prognosis. Consequently, the importance of preoperative nutritional and functional management for lung cancer patients is increasingly recognized by health care providers, as it can enhance postoperative recovery and reduce complications.[21,24] By focusing on modifiable preoperative factors, health care providers, including nurses, can implement targeted strategies to optimize patients' health before surgery, potentially improving their postoperative recovery and survival.

Machine learning models offer several advantages in clinical decision-making, particularly their ability to analyze large datasets with complex, non-linear relationships.[15] These models can efficiently process vast amounts of data from various sources—such as clinical records, imaging, and biomarker information—surpassing the limitations of traditional statistical methods and enabling more personalized and accurate predictions.[15,21] However, a key limitation of machine learning models is their reliance on high-quality, well-structured data. Incomplete or biased data can greatly impact a model's accuracy and reliability.[14,20] Furthermore, while AI assisted tools like large language model-ADA help reduce technical barriers, the models still require extensive validation, particularly across diverse clinical environments, to ensure their generalizability.[17]

Implications for nursing practice and research

The integration of AI-assisted machine learning models, such as large language model-ADA, into lung cancer care offers transformative potential for nursing. By automating data preprocessing—which involves cleaning and standardizing complex clinical datasets—AI-assisted machine learning models streamline the integration of preoperative biomarkers (e.g., white blood cell counts, lung function) and postoperative variables (e.g., recovery time, complication rates) into predictive analytics for lung cancer survival.[17] Its built-in library of advanced algorithms, including Random Forests and Gradient Boosting Machines, enables nurses to generate survival predictions tailored to individual patient profiles without requiring coding expertise. Transparent AI tools like SHAP highlight critical decision drivers (e.g., preoperative oxygen levels, albumin status), empowering nurses to interpret results and prioritize high-risk patients proactively.[17] Dynamic calibration dashboards further allow real-time tracking of model accuracy, ensuring predictions evolve with new clinical data. For research, large language model-ADA accelerates hypothesis testing by automating model optimization and validation,[19] while clinically, it equips nurses to design precision care plans that address modifiable preoperative risks, ultimately bridging the gap between predictive analytics and actionable bedside interventions in lung cancer management.

Limitations

This study has several limitations. First, the dataset used was derived from a single medical center, the underrepresentation of patients enrolled in this study may affect the external validity of our predictions. Future studies incorporating more diverse and representative populations from other centers are necessary to better understand the models' broader applicability and to reduce the risk of bias. While the machine learning models demonstrated promising predictive accuracy, the lack of external validation across diverse health care settings limits their widespread clinical application. Future research is essential to confirm the utility of machine learning models in predicting lung cancer survival. Although SHAP is useful for analyzing feature importance, understanding the mechanisms of the models is crucial for clinical acceptance. We would like to utilize additional exploitability techniques or visual aids expanding on the interpretability of predicting models that break down the decision process in more accessible terms for clinicians in future studies. For nursing, integrating these advanced tools can greatly enhance decision-making, particularly in the development of personalized care plans. However, there is currently no application guideline available for the large language model-ADA, further research on how to efficiently apply this innovative technology in clinical nursing is still needed. In addition, the visualization of prediction models needs further improvement. It is crucial that nurses receive proper training in using and interpreting model outputs to avoid misinterpretation and ensure that the models complement, rather than replace, clinical judgment.[8,21]

Conclusions

The application of AI assisted machine learning models demonstrated a significant potential in predicting survival outcomes for lung cancer patients. The Random Forest classifier exhibited superior predictive performance, emphasizing the power of machine learning in analyzing complex clinical and demographic data. This study also highlighted the importance of preoperative factors such as nutritional status, lung function, and age, which were key predictors of survival outcomes. These findings align with previous research and underscore the utility of machine learning models in enhancing personalized medicine and improving clinical decision-making. Future research should focus on validating these findings across diverse populations and medical centers, as well as exploring the role of machine learning in other areas of cancer care. The continued integration of machine learning tools like large language model-ADA has the potential to revolutionize clinical workflows, offering more accurate prognostic tools and contributing to improved patient outcomes. Findings of this study support the growing role of predictive analytics in personalized cancer care, where tailored nursing interventions can lead to better patient outcomes.

CRediT authorship contribution statement

YY, GZ: Writing–Original draft, Conceptualization, Formal Analysis, Software. **YY, GZ, WM:** Writing – Original draft, Data curation, Validation. **YG, SH, CH, HX:** Investigation, Visualization, Methodology, Validation, Data curation. **GZ, WM, YZ:** Writing- Reviewing and Editing, Supervision. All authors have read and approved the final manuscript.

Ethics statement

The study was approved by the Medical Ethics Committee of the First Affiliated Hospital of Guangzhou Medical University (Approval No. ES202307203) and was conducted in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was waived by our Institutional Review Board because of the retrospective nature of our study.

Data availability statement

The data that support the findings of this study are available from the corresponding author, YZ, upon reasonable request.

Declaration of generative AI and AI-assisted technologies in the writing process

No AI tools/services were used during the preparation of this work.

Funding

This study received no external funding.

Declaration of competing interest

The authors declare no conflict of interest. The corresponding author, Dr. Yingchun Zeng, is an editorial board member of *Asia–Pacific Journal of Oncology Nursing*. The article was subject to the journal's standard procedures, with peer review handled independently of Dr. Zeng and their research groups.

References

1. International Agency for Research on Cancer (IARC), World Health Organization. Global cancer burden growing, amidst mounting need for services. https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing–amidst-mounting-need-for-services/2024-2-1; 2024. Accessed September 28, 2024.
2. Han B, Zheng R, Zeng H, et al. Cancer incidence and mortality in China, 2022. *J Natl Cancer Cent*. 2024;4(1):47–53. https://doi.org/10.1016/j.jncc.2024.01.006.
3. Huang D, Li L, Li X, et al. Chinese expert consensus on medical treatment for elderly patients with advanced lung cancer (2022 edition). *Chin J Lung Cancer*. 2022;25(6): 363–384.
4. Lee HW, Lee CH, Park YS. Location of stage I-III non-small cell lung cancer and survival rate: systematic review and meta-analysis. *Thorac Cancer*. 2018;9(12): 1614–1622. https://doi.org/10.1111/1759-7714.12869.
5. Jin L, Zhao Q, Fu S, Cao F, Hou B, Ma J. Development and validation of machine learning models to predict survival of patients with resected stage-III NSCLC. *Front Oncol*. 2023;13:1092478. https://doi.org/10.3389/fonc.2023.1092478.
6. Kinoshita F, Takenaka T, Yamashita T, et al. Development of artificial intelligence prognostic model for surgically resected non-small cell lung cancer. *Sci Rep*. 2023; 13(1):15683. https://doi.org/10.1038/s41598-023-42964-8.
7. Yang Y, Xu L, Sun L, Zhang P, Farid SS. Machine learning application in personalised lung cancer recurrence and survivability prediction. *Comput Struct Biotechnol J*. 2022; 20:1811–1820. https://doi.org/10.1016/j.csbj.2022.03.035.
8. Owusuaa C, van der Padt-Pruijsten A, Drooger JC, et al. Development of a clinical prediction model for 1-year mortality in patients with advanced cancer. *JAMA Netw Open*. 2022;5(11):e2244350. https://doi.org/10.1001/jamanetworkopen.2022.44350.
9. Didier AJ, Nigro A, Noori Z, Omballi MA, Pappada SM, Hamouda DM. Application of machine learning for lung cancer survival prognostication-A systematic review and meta-analysis. *Front Artif Intell*. 2024;7:1365777. https://doi.org/10.3389/frai.2024.1365777.
10. Li Y, Wu X, Yang P, Jiang G, Luo Y. Machine learning for lung cancer diagnosis, treatment, and prognosis. *Genom Proteom Bioinform*. 2022;20(5):850–866. https://doi.org/10.1016/j.gpb.2022.11.003.
11. Altuhaifa FA, Win KT, Su G. Predicting lung cancer survival based on clinical data using machine learning: a review. *Comput Biol Med*. 2023;165:107338. https://doi.org/10.1016/j.compbiomed.2023.107338.
12. Pluchart H, Chanoine S, Moro-Sibilot D, et al. Lung cancer, comorbidities, and medication: the infernal trio. *Front Pharmacol*. 2024;14:1016976. https://doi.org/10.3389/fphar.2023.1016976.
13. Pu CY, Batarseh H, Zafron ML, et al. Effects of preoperative breathing exercise on postoperative outcomes for patients with lung cancer undergoing curative intent lung resection: a meta-analysis. *Arch Phys Med Rehabil*. 2021;102(12):2416–2427. https://doi.org/10.1016/j.apmr.2021.03.028.
14. Rhee J, Graubard BI, Purdue MP. Blood lead levels and lung cancer mortality: an updated analysis of NHANES II and III. *Cancer Med*. 2021;10(12):4066–4074. https://doi.org/10.1002/cam4.3943.
15. Rony MKK, Parvin MR, Ferdousi S. Advancing nursing practice with artificial intelligence: enhancing preparedness for the future. *Nurs Open*. 2024;11(1). https://doi.org/10.1002/nop2.2070, 10.1002/nop2.2070.
16. Maurya SP, Sisodia PS, Mishra R, Singh DP. Performance of machine learning algorithms for lung cancer prediction: a comparative approach. *Sci Rep*. 2024 Aug 9; 14(1):18562. https://doi.org/10.1038/s41598-024-58345-8.
17. Tayebi Arasteh S, Han T, Lotfinia M, et al. Large language models streamline automated machine learning for clinical studies. *Nat Commun*. 2024;15(1):1603. https://doi.org/10.1038/s41467-024-45879-8.
18. Huang S, Yang J, Shen N, Xu Q, Zhao Q. Artificial intelligence in lung cancer diagnosis and prognosis: current application and future perspective. *Semin Cancer Biol*. 2023;89:30–37.
19. Lin K, Hu Y, Kong G. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. *Int J Med Inf*. 2019;125:55–61. https://doi.org/10.1016/j.ijmedinf.2019.02.002.
20. Zhang S, Gong C, Wu L, Liu X, Zhou M. AutoML-GPT: automatic machine learning with GPT. Preprint at http://arxiv.org/abs/2305.02499; 2023.
21. Zhang G, Liu X, Hu Y, et al. Development and comparison of machine-learning models for predicting prolonged postoperative length of stay in lung cancer patients following video-assisted thoracoscopic surgery. *Asia Pac J Oncol Nurs*. 2024;11(6): 100493. https://doi.org/10.1016/j.apjon.2024.100493.
22. De Rosa S, Greco M, Rauseo M, Annetta MG. The good, the bad, and the serum creatinine: exploring the effect of muscle Mass and nutrition. *Blood Purif*. 2023;52(9-10):775–785. https://doi.org/10.1159/000533173.
23. Schuetz P, Fehr R, Baechli V, et al. Individualised nutritional support in medical inpatients at nutritional risk: a randomised clinical trial. *Lancet*. 2019;393(10188): 2312–2321. https://doi.org/10.1016/S0140-6736(18)32776-4.
24. Bargetzi L, Brack C, Herrmann J, et al. Nutritional support during the hospital stay reduces mortality in patients with different types of cancers: secondary analysis of a prospective randomized trial. *Ann Oncol*. 2021;32(8):1025–1033. https://doi.org/10.1016/j.annonc.2021.05.793.