Elevate Series

ORIGINAL RESEARCH

Machine Learning for Prediction of Non-Small Cell Lung Cancer Based on Inflammatory and Nutritional Indicators in Adults: A Cross-Sectional Study

Qiaoli Wang [1],*, Tao Liang [2],*, Yuexi Li [1],*, Xiaoqin Liu[1]

[1]Department of Health Screening Center, Deyang Peoples' Hospital, Deyang, Sichuan, 618000, People's Republic of China; [2]Department of Gastroenterology, Deyang Peoples' Hospital, Deyang, Sichuan, 618000, People's Republic of China

*These authors contributed equally to this work

Correspondence: Xiaoqin Liu, Email 66823572@qq.com

Purpose: The aim of this study was to evaluate the potential benefit of blood inflammation in the diagnosis of non-small cell lung cancer (NSCLC) and propose a machine-learning-based method to predict NSCLC in asymptomatic adults.

Patients and Methods: A cross-sectional study was evaluated using medical records of 139 patients with non-small cell lung cancer and physical examination data from May 2022 to May 2023 of 198 healthy controls. The NSCLC cohort comprised 128 cases of adenocarcinoma, 3 cases of squamous cell carcinoma, and 8 cases of other NSCLC subtypes. The correlation between inflammatory and nutritional markers, such as monocytes, neutrophils, LMR, NLR, PLR, PHR and non-small cell lung cancer was examined. Features were selected using Python's feature selection library and analyzed by five algorithms. The predictive ability of the model for non-small cell lung cancer diagnosis was assessed by precision, accuracy, recall, F1 score, and area under the curve (AUC).

Results: The results showed that the top 14 important factors were PDW, age, TP, RBC, HGB, LYM, LYM%, RDW, PLR, LMR, PHR, MONO, MONO%, gender. Additionally, the naive Bayes (NB) algorithm demonstrated the highest overall performance in predicting adult NSCLC among the five machine learning algorithms, achieving an accuracy of 0.87, a macro average F1 score of 0.85, a weighted average F1 score of 0.87, and an AUC of 0.84.

Conclusion: In feature ranking, platelet distribution width was the most important feature, and the NB algorithm performed best in predicting adult NSCLC diagnosis.

Keywords: machine learning, non-small cell lung cancer, inflammatory indicators, nutritional indicators, ratio, diagnosis

Introduction

According to the International Agency for Research on Cancer (IARC) estimates of new cancer cases and deaths worldwide in 2020, lung cancer continues to be a significant burden of global morbidity and mortality, has become the most common malignancy in men and the second most common malignancy in women, and remains the leading cause of cancer deaths (18.0% of all cancer types)[1.] Among lung cancers, non-small cell lung cancer (NSCLC) is the most common histologic type, accounting for nearly 85% of all diagnosed lung cancers. The clinical prognosis of lung cancer is directly related to the stage of the cancer at the time of diagnosis. The 5-year survival rate for patients with stage I lung cancer is 68.4%, whereas the 5-year survival rate for patients with stage IV lung cancer is only 5.8%.[2,3] Unfortunately, most patients with NSCLC are not diagnosed until advanced stage III, which is associated with a poorer prognosis.[4] Unfortunately, most patients with NSCLC are not diagnosed until advanced stage III, which is associated with a poorer prognosis. Recently, there has been increasing evidence that inflammation may contribute to the onset, tumor staging, and progression of malignant tumors.[5,6] In addition, studies have analyzed the impact of hematological indicators such as monocytes, lymphocytes, neutrophils, albumin, and hemoglobin on cancer detection and prognosis. These indicators can be detected in peripheral blood and have the potential to assist in the diagnosis of lung cancer. It was also found that

527

combinations of ratios derived from these measurements help to better diagnose cancer and predict prognosis compared to single indicators.[7] For example, monocyte-to-albumin ratio (MAR), neutrophil percentage-to-hemoglobin ratio (NPHR), neutrophil-to-lymphocyte ratio (NLR), and lymphocyte-to-monocyte ratio (LMR) were reported by several published studies to play an extremely important role in cancer prediction and diagnosis.[8] However, the clinical significance of PHR and MHR in NSCLC has not been fully investigated.

With the development of artificial intelligence (AI), machine learning (ML) has made technological breakthroughs in analyzing clinical data in the biomedical field and has been widely used to predict the diagnosis and prognosis of diseases.[9] Therefore, we collected inflammation-related data from preoperative peripheral blood indices, including leukocytes, monocytes, neutrophils, monocytes, and platelet count, and calculated their ratios, including lymphocyte-to-monocyte ratio (LMR), neutrophil-to-lymphocyte ratio (NLR); neutrophil-to-monocyte ratio (NMR), and platelet-to- Lymphocyte Ratio (PLR), Monocyte-to -Albumin Ratio (MAR), Monocyte-to-Hemoglobin Ratio (MHR), and Platelet-to-Hemoglobin Ratio (PHR) to determine a machine learning method for predicting the risk of adult non-small cell lung cancer. On this basis, five machine learning algorithm models, including logistic regression (LR), decision tree (DT), adaboost (AB) random forest (RF) and naive Bayes (NB) algorithm were used to build a prediction model to explore the clinical significance of inflammatory markers for NSCLC diagnosis. The diagnostic ability of the model was assessed using the area under the ROC curve, sensitivity, and specificity. The aim was to explore the added value of single inflammatory markers and ratio combinations of these measurements in the diagnosis of NSCLC. This will provide a new theoretical basis for future screening and diagnostic studies of NSCLC.

Material and Methods
Study Design and Participants
This retrospective analysis included medical record data from 139 patients diagnosed with NSCLC by surgery and 198 healthy subjects diagnosed with NSCLC by physical examination at Deyang People's Hospital (Sichuan Province, China) between May 2022 to May 2023. Patients with NSCLC were diagnosed by cytology or histopathology. Healthy subjects were those who had no suspicious lesions on chest CT examination. The study was approved by the Ethics Committee of Deyang People's Hospital. As the paper was a retrospective study, the requirement of informed consent was waived. The data was anonymized.

Inclusion and Exclusion Criteria
Inclusion criteria (1) The case group consisted of patients with surgical and pathological diagnosis of NSCLC and the control group consisted of healthy subjects who showed no evidence of suspicious lesions detected by chest CT; (2) Aged 18 years or above; (3) No previous history of cancer; (4) No antitumor therapy prior to examination.

Exclusion criteria (1) patients with acute and chronic inflammatory diseases; (2) cardiovascular and cerebrovascular diseases; (3) chronic obstructive pulmonary disease; (4) diabetes mellitus; (5) hematological diseases; (6) history of other malignant tumors; and (7) incomplete records of clinical and/or auxiliary examinations.

Collecting Clinical Data
Clinical data included participants' identification numbers, age, sex, smoking history, blood cell counts, histopathology, and diagnostic information. Then, the values of other hematological indices were calculated. NLR is the ratio of absolute neutrophil count to absolute lymphocyte count per liter of whole blood; PLR is the ratio of platelet count to lymphocyte count; MAR is the ratio of monocytes to albumin; TAR is the ratio of total bilirubin to albumin; MHR is the ratio of monocytes to hemoglobin; and PHR is the ratio of platelets to hemoglobin. LMR is the lymphocyte-to-monocyte ratio; NMR is the neutrophil-to-monocyte ratio; and PLR is the platelet-to-lymphocyte ratio.

Statistical Analysis
Statistical analyses were performed using SPSS version 23.0 (SPSS Inc., Chicago, USA), and diagnostic predictive models were constructed using Python 3.7.6. Variables that conformed to normal distribution were expressed as mean ± standard deviation and subjected to Student-t test. Variables that did not fit the normal distribution were expressed as median and interquartile range and analyzed with the independent samples Kruskal–Wallis U-test. Categorical variables

were expressed as frequencies and proportions and analyzed with the chi-square test. Risk factors for NSCLC were determined by binary logistic regression. p <0.05 was considered statistically significant.

Machine Learning-Based Diagnostic Model

In this study, five ML algorithms were used for model selection, namely logistic regression (LR), decision tree (DT), adaboost (AB) random forest (RF) and naive Bayes (NB). All machine learning processes were done in Python environment (version 3.9.0) using sklearn, numpy, pandas, matplotlib, seaborn and scipy packages. The first step is to import the data set into a Pandas data frame and then split it into the training set (80%) and test set (20%) using the train_test_split function for building the model and ensuring the accuracy of the results. The independent variables are first standardized based on their feature range using the StandardScaler package from the sklearn library. To prevent overfitting, we employed 5-fold cross-validation during training. The aim of feature selection is to rank and prioritize the most significant predictors within the dataset. The impact of these features was evaluated by calculating permutation importance. Additionally, we manually fine-tuned the parameters in each model. After the completion of model training, an evaluation is conducted to compare and assess the performance of each individual model. Accuracy, precision, recall and F1 score of the prediction model were the evaluation metrics[10.] True Positive (TP) refers to a positive sample predicted as positive, whereas False Positive (FP) refers to negative samples that are falsely predicted as positive. TN (True Negative) refers to a negative sample predicted as negative, while False Negative (FN) indicates positive samples that are falsely classified as negative by the model. For performance comparison, four parameters, accuracy, precision, recall, and F1-score are used.[10] Codes related to this research can be downloaded from GitHub website (https://github.com/ganbingliangyi/ganbingliangyi).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2PR}{P + R} = \frac{2TP}{2TP + FP + FN}$$

Results

Participant Characteristics

The baseline data of the two groups are shown in Table 1. A total of 337 volunteers were recruited in this study, including 139 patients with NSCLC and 198 subjects with no detectable disease (Table 1). The patients with NSCLC consisted of 128 patients with adenocarcinoma, 3 patients with squamous cell carcinoma, and 8 patients with other NSCLC. The analysis of single factor revealed that the NSCLC group exhibited a higher proportion of females compared to males, with a statistically significant gender disparity between the two groups. Compared with healthy volunteers, monocytes and monocyte percentage were significantly higher in NSCLC patients (P < 0.05), while hemoglobin, albumin, total bilirubin, total protein, RBC, and PDW were decreased (P < 0.05). NLR, PLR, MAR, MHR, and PHR levels were significantly higher in NSCLC patients compared to controls (P < 0.05). However, LMR, TAR and NMR were much higher in healthy volunteers than in NSCLC patients (shown in Table 1).

Analysis of Multifactor Logistic Regression

The variables that showed statistical significance in Table 1 were included in the multivariate logistic regression analysis. The following factors were identified as independent associations for non-small cell lung cancer (NSCLC): gender, age, red blood cell count (RBC), platelet distribution width (PDW), total protein (TP), and neutrophil-to-lymphocyte ratio (NLR) (shown in Table 2).

Table 1 Participant Characteristics

Variable	NSCLC Patients	Healthy Participants	P values
Age (years)	53.11±12.08	41.34±8.66	<0.001
Gender			<0.001
Male	54	118	
Female	85	80	
Smoking history			0.869
No	108	152	
Yes	31	46	
Laboratory parameters			
RBC	4.712±0.570	4.844±0.512	0.027
NEUT (×109/L) 1	3.658±1.348	3.629±1.123	0.828
Mono (×109/L)	0.396±0.122	0.351±0.109	0.001
LYM	1.928±0.604	2.056±0.578	0.051
PDW	15.514±2.714	16.319±0.387	0.001
PLT	214.690±56.251	212.040±61.937	0.688
HGB	140.060±20.918	148.540±17.866	<0.001
RDW	13.565±1.759	13.480±1.190	0.623
MONO%	6.545±1.591	5.702±1.329	<0.001
LYM%	32.275±9.027	33.372±6.759	0.226
NEUT%	58.757±9.535	57.861±7.251	0.351
TP	72.248±4.481	75.431±3.765	<0.001
ALB (g/L)	45.668±2.286	46.721±2.299	<0.001
TBIL	15.809±6.398	18.116±7.851	0.003
LMR	5.175±1.897	6.181±1.962	<0.001
NLR	2.092±1.098	1.860±0.660	0.027
NMR	9.569±3.064	10.769±3.121	0.001
PLR	120.99±48.87	109.90±45.61	0.034
TAR	0.346±0.138	0.387±0.163	0.018
MAR	0.009±0.003	0.008±0.002	<0.001
MHR	0.003±0.002	0.002±0.001	0.002
PHR	1.646±1.050	1.463±0.541	0.037

Abbreviations: MONO, monocyte; NEUT, neutrophil; LMY, lymphocyte; RBC, red blood cell; HGB, hemoglobin; PDW, the platelet distribution width; RDW, the Red Cell Distribution Width; TP, total protein; ALB, albumin; TBIL, total bilirubin; LMR, lymphocyte to monocyte ratio; NLR, neutrophil to lymphocyte ratio; NMR, neutrophil to monocyte ratio; PLR, platelet to lymphocyte ratio; TAR, total bilirubin to albumin ratio; MAR, monocyte to albumin ratio; MHR, monocyte to hemoglobin ratio; PHR, platelet to hemoglobin ratio.

Table 2 Multivariate Regression Analysis of Independent Factors for NSCLC

Variables	Odds ratio	P value	95% Confidence interval
Gender	5.713	<0.001	2.318–14.080
Age	1.102	<0.001	1.068–1.137
RBC	2.754	0.010	1.276–5.943
PDW	0.728	0.001	0.600–0.884
TP	0.791	<0.001	0.719–0.871
NLR	3.131	0.048	1.010–9.712

Ranking of Influencing NSCLC Feature

Non-small cell lung cancer was set as the target variable. Importance scores for all risk factors were calculated based on patient characteristics using the "feature selector" method. The importance ranking of all variables was done using the Gini scale. Factors with higher risk scores indicated a greater impact on NSCLC occurrence. We selected first 14 indicators as the characterization factors to build the risk prediction model. The order of importance of the 14 variables is shown in Figure 1.
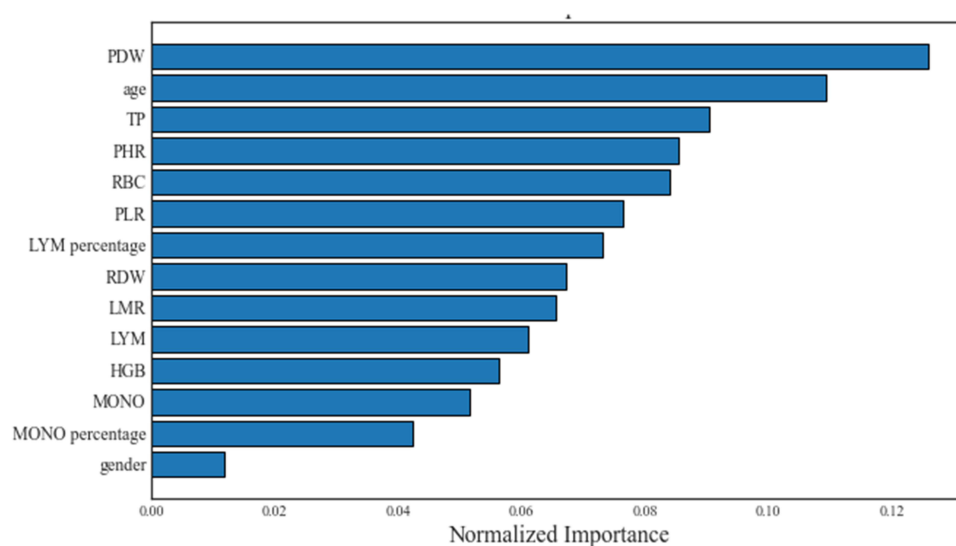
Figure 1 The rank of the importance of indicators in NSCLC.

The cumulative importance score of these 14 features was subsequently computed and presented in Table 3. The cumulative importance of PDW, age, TP, RBC, HGB,LYM, LYM%, RDW, PLR, LMR, and PHR approached 0.9 (Figure 2).

Prediction Model of ML

To further validate the potential of these indicators as diagnostic indicators for NSCLC, we selected top 14 indicators (PDW, age, TP, RBC, HGB, LYM, LYM%, RDW, PLR, LMR, PHR, MONO, MONO%, gender) as the characterization factors to build the risk prediction model, including RF, LR, DT, AB and NB algorithms. As shown in Figure 3, the ROC curves were plotted according to each of the five algorithms. The ROC-AUC plots for each algorithm (RF (AUC = 0.77), LR (AUC = 0.74), DT (AUC = 0.67), AB (AUC=0.87), and NB (AUC=0.84)) are shown in Figure 3, confirming the reliability of the models in predicting risk factors. The AUC values of all five models were greater than 0.6, indicating that these models have moderate potential in the diagnosis of non-small cell lung cancer patients. In addition, Table 4 shows the accuracy, precision, recall, and F1 score of the ML models, with the five algorithmic models predicting non-small cell lung cancer with an accuracy of 79%, 76%, 69%, 78% and 87%, respectively. This indicates that the naive

Table 3 The Cumulative Importance Score of Features

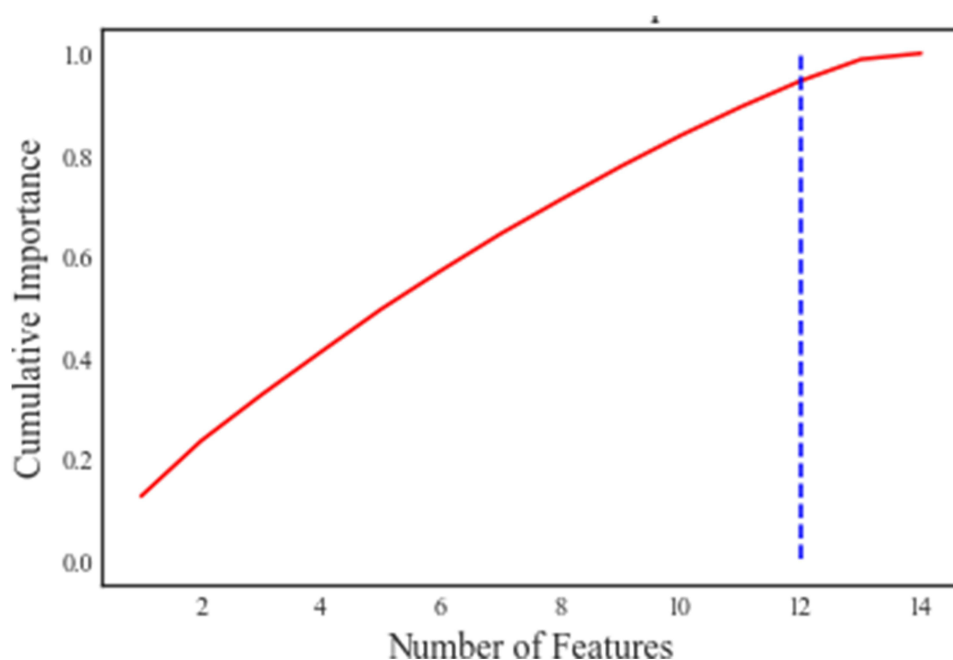| Feature	Importance	Normalized-importance	Cumulative-importance
PDW	1184	0.126	0.126
Age	1028	0.109	0.235
TP	850	0.090	0.325
PHR	804	0.085	0.411
RBC	791	0.084	0.495
PLR	719	0.076	0.571
LYM percentage	687	0.073	0.644
RDW	633	0.067	0.711
LMR	616	0.065	0.777
LYM	575	0.061	0.838
HGB	530	0.056	0.895
MONO	484	0.051	0.946
MONO percentage	398	0.042	0.988
gender	110	0.012	1.000

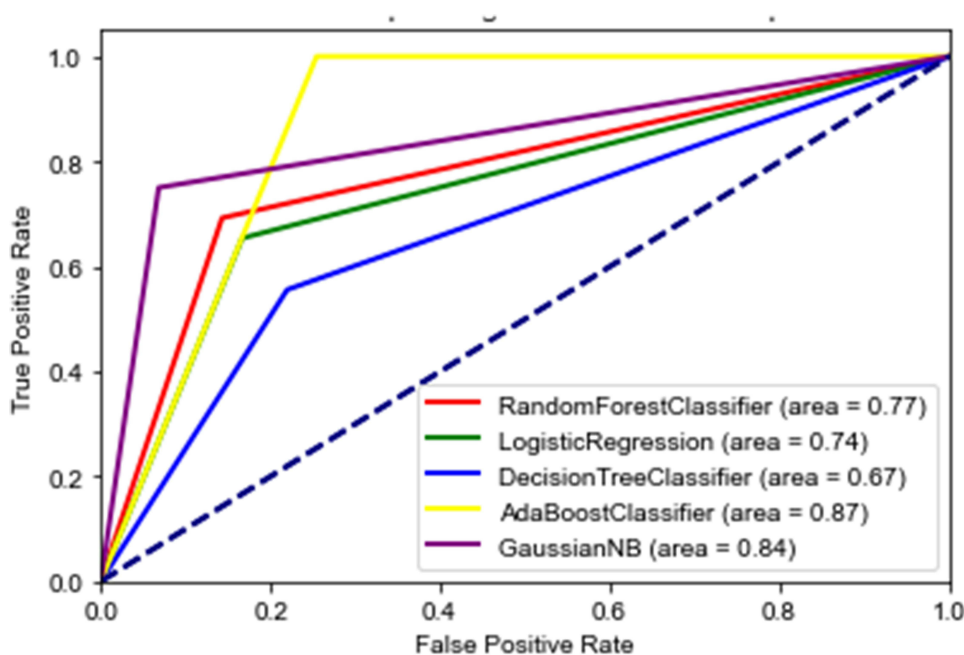**Figure 2** The cumulative importance of factors.

Figure 3 The ROC-AUC graph.

Bayes (NB) algorithm demonstrated the highest overall performance in predicting adult NSCLC among the five machine learning algorithms, achieving an accuracy of 0.87, a macro average F1 score of 0.85, a weighted average F1 score of 0.87, and an AUC of 0.84.

Discussion

In this study, records of 139 NSCLC patients and 198 healthy individuals were retrospectively analyzed. It showed that elevated levels of monocytes, monocyte percentage, NLR, PLR, MAR, MHR, PHR and decrease levels of hemoglobin, albumin, total bilirubin, total protein, RBC, PDW, LMR, TAR, and NMR were risk factors for NSCLC. And the results of

Table 4 Compared in Terms of Precision, Recall, F1 Value and AUC

	Accuracy	AUC	Precision		Recall		F1-score	
			Macro avg	Weighted avg	Macro avg	Weighted avg	Macro avg	Weighted avg
Random Forest	0.79	0.77	0.77	0.80	0.78	0.79	0.78	0.80
Logistic Regression	0.76	0.74	0.74	0.77	0.75	0.76	0.75	0.77
Decision Tree	0.69	0.67	0.67	0.70	0.68	0.69	0.67	0.69
AdaBoost	0.78	0.87	0.87	0.84	0.69	0.78	0.70	0.75
Naive Bayes	0.87	0.84	0.86	0.87	0.84	0.87	0.85	0.87

further multivariate analysis revealed that the following factors were identified as independent associations for non-small cell lung cancer (NSCLC): gender, age, RBC, PDW, TP, and NLR, which is consistent with previous studies.[5] Studies have suggested that inflammation may play an important role in the development and progression of cancer.[11–14] For example, monocytes play an important role in tumor development and metastasis by creating a link between innate and adaptive immune responses by promoting immunosuppression, remodeling of the extracellular matrix, angiogenesis, and tumor cell conductance.[15] A study observed that elevated NLR, and PLR were associated with increased lung cancer risk, while LMR decreased.[5] In addition, in this study, hematological parameters such as MHR and PHR were used for the first time to predict NSCLC. Albumin and hemoglobin can reflect nutritional status. Hypoalbuminemia is usually the result of insufficient nutrient absorption and excessive tumor consumption, which affects metabolism and immune function. Tumor cells in a state of hypoxia promote tumor growth while hemoglobin concentration decreases.[16] Our data showed that the ratio of inflammatory indicators to nutritional indicators (including PHR, MHR, and MAR) was higher in NSCLC patients than in healthy subjects, which may be a new independent factor for predicting NSCLC. Further analysis revealed that MAR, MHR, and PHR may have potential value in identifying NSCLC patients from the healthy population. Previous studies have shown that MAR is an independent risk factor for NSCLC, which is consistent with our results.[8] In addition, a combination of inflammatory and nutritional indicators may be better than a single inflammatory indicator in diagnosing NSCLC patients. In addition, our study showed the ranking of the characteristics affecting NSCLC, and PDW, age, total protein, PHR, and RBC were the top five ranked characteristics.

Recently, machine learning has been more widely applied in predicting diseases in the medical field,[17] particularly in medical imaging and decision support systems.[18,19] Machine learning can effectively learn the features of huge amounts of data, which provides new research ideas and methods for precise prediction.[20] A lot of ML algorithms have been used for clinical applications, including support vector machines (SVMs), random forests (RFs), neural networks (NNs), Decision Trees (DTs), and other algorithms.[21] ML techniques allowed the development of forecasting models for predicting cancer diagnosis and clinical outcomes.[22,23] Research indicates that the adoption of ML for other cancers substantially improves prediction accuracy.[24] In addition, the previous study developed ML models applied to Lung Cancer Classification and Prediction based on a decision tree, RF, logistic regression, SVM, naive Bayes, and neural network, and the results showed that models of the neural network, random forest, and naive Bayes performed well for the data of baseline. The classification accuracies of the models were 0.767, 0.718, and 0.688, respectively, and the AUCs were 0.793, 0.779, and 0.771.[25]

In this study, five machine learning algorithm models consisting of LR, DT, AB, RF and NB were designed to predict the diagnosis value in patients with NSCLC. Owing to the imbalanced data set, we utilize comprehensive scoring indicators, including accuracy, precision, recall rate, F1-score, and AUC value to improve model performance. In our research, the naive Bayes algorithm was determined to have the best prediction ability with 0.87 accuracy, 0.86 macro-avg precision, 0.87 weighted-avg precision,0.84 macro-avg recall rate, 0.87 weighted-avg recall rate, 0.85 macro-avg F1 score, 0.87 weighted-avg F1 score, and 0.84 AUC. In addition, the clinical importance of machine learning lies in the detection of risk factors that are closely associated with NSCLC. According to the permutation importance of feature variables, PDW, age, TP, RDW, MONO percentage, RBC, LYM percentage, PLR, MONO, LYM, HGB, gender, PHR and LMR were critical to NSCLC. According to this study, this model can help to improve the rate of diagnosis of NSCLC. With successful machine learning models,

the outcome of cancer can be improved, and hospitalization and healthcare expenses can be reduced due to early diagnosis. Therefore, a machine learning model with the ability to predict the lung cancer is required.

However, there are some limitations in our research. Firstly, it is a cross-sectional study, our data is based on medical records of single center clinical data from a hospital, which has differences in treatment strategies, ethnicity, and other factors. For example, the majority of the non-small cell lung cancers included in the sample were adenocarcinoma, accounting for 92.09%, which may not adequately represent the comprehensive characteristics of all non-small cell lung cancers.Secondly, potential factors that are not included may also have a certain impact on the results, and the algorithm model is skewed because some input features could affect the accuracy of the model. Thirdly, although the risk of bias was considered and calibrated in this study by two internal validation methods (split-sample validation and n-fold cross-validation), a larger sample size and prospective research are needed to further verify the practice of the model in future studies.

Conclusion

In recent years, artificial intelligence has been widely used in the medical field. Timely and accurate diagnosis is crucial in order to select the most appropriate treatment and reduce misdiagnosis of NSCLC patients. Machine learning models can be used to predict diagnosis and survival, and it is now used in many other fields to predict outcomes. In this study, the evaluation of NSCLC prediction was done using five machine-learning models. The NB model has better predictive ability with the higher level of precision, recall, F1, and AUC and shows the best performance in predicting NSCL in adults. Our results lay the foundation for an early warning system that can provide clinicians with relevant information for clinical decision-making.

Institutional Review Board Statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Deyang People's Hospital (protocol code:2022-04-097-K01 and date of approval: December 28, 2022).

Informed Consent Statement

As the paper was a retrospective study, the requirement of informed consent was waived.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca a Canc J Clinic*. 2021;71(3):209–249. doi:10.3322/caac.21660
2. Goldstraw P, Chansky K, Crowley J, et al. The iaslc lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2016;11(1):39–51. doi:10.1016/j.jtho.2015.09.009
3. Mithoowani H, Febbraro M. Non-small-cell lung cancer in 2022: a review for general practitioners in oncology. *Curr Oncol*. 2022;29(3):1828–1839. doi:10.3390/curroncol29030150
4. Liu G, Pei F, Yang F, et al. Role of autophagy and apoptosis in non-small-cell lung cancer. *Int J Mol Sci*. 2017;18(2).
5. Nøst TH, Alcala K, Urbarova I, et al. Systemic inflammation markers and cancer incidence in the UK biobank. *Eur J Epidemiol*. 2021;36 (8):841–848.

6. Qian S, Golubnitschaja O, Zhan X. Chronic inflammation: key player and biomarker-set to predict and prevent cancer development and progression based on individualized patient profiles. *Epma Journal*. 2019;10(4):365–381. doi:10.1007/s13167-019-00194-x

7. Xu F, Xu P, Cui W, et al. Neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios may aid in identifying patients with non-small cell lung cancer and predicting tumor-node-metastasis stages. *Oncol Lett*. 2018;16(1):483–490. doi:10.3892/ol.2018.8644

8. Zhao ST, Chen XX, Yang XM, He SC, Qian FH. Application of monocyte-to-albumin ratio and neutrophil percentage-to-hemoglobin ratio on distinguishing non-small cell lung cancer patients from healthy subjects. *Int J Gen Med*. 2023;16:2175–2185. doi:10.2147/IJGM.S409869

9. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375(13):1216–1219. doi:10.1056/NEJMp1606181

10. Ngabo D, Dong W, Ibeke E, Iwendi C, Masabo E. Tackling pandemics in smart cities using machine learning architecture. *Math Biosci Eng*. 2021;18(6):8444–8461. doi:10.3934/mbe.2021418

11. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. *Nature*. 2008;454(7203):436–444. doi:10.1038/nature07205

12. Diakos CI, Charles KA, Mcmillan DC, Clarke SJ. Cancer-related inflammation and treatment effectiveness. *Lancet Oncol*. 2014;15(11):e493–e503. doi:10.1016/S1470-2045(14)70263-3

13. Ding HP, Ling YQ, Chen W, et al. Effects of nutritional indices and inflammatory parameters on patients received immunotherapy for non-small cell lung cancer. *Curr Probl Cancer*. 2023;48:101035. doi:10.1016/j.currproblcancer.2023.101035

14. Berkman SJ, Roscoe EM, Bourret JC. Comparing self-directed methods for training staff to create graphs using GraphPad prism. *J Appl Behav Anal*. 2019;52(1):188–204. doi:10.1002/jaba.522

15. Olingy CE, Dinh HQ, Hedrick CC. Monocyte heterogeneity and functions in cancer. *J Leukoc Biol*. 2019;106(2):309–322. doi:10.1002/JLB.4RI0818-311R

16. Huang Y, Wei S, Jiang N, et al. The prognostic impact of decreased pretreatment haemoglobin level on the survival of patients with lung cancer: a systematic review and meta-analysis. *BMC Cancer*. 2018;18(1):1235. doi:10.1186/s12885-018-5136-5

17. Rajpurkar P, Chen E, Banerjee O, Topol EJ. Ai in health and medicine. *Nat Med*. 2022;28(1):31–38. doi:10.1038/s41591-021-01614-0

18. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine learning and deep learning in medical imaging: intelligent imaging. *J Med Imaging Radiat Sci*. 2019;50(4):477–487. doi:10.1016/j.jmir.2019.09.005

19. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of health care. *Am J Med*. 2019;132(7):795–801. doi:10.1016/j.amjmed.2019.01.017

20. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res*. 2016;18(12):e323. doi:10.2196/jmir.5870

21. Nwanosike EM, Conway BR, Merchant HA, Hasan SS. Potential applications and performance of machine learning techniques and algorithms in clinical practice: a systematic review. *Int J Med Inform*. 2022;159:104679. doi:10.1016/j.ijmedinf.2021.104679

22. Nageswaran S, Arunkumar G, Bisht AK, et al. Lung cancer classification and prediction using machine learning and image processing. *Biomed Res Int*. 2022;2022:1755460. doi:10.1155/2022/1755460

23. 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. doi:10.1016/j.compbiomed.2023.107338

24. Xie Y, Meng WY, Li RZ, et al. Early lung cancer diagnostic biomarker discovery by machine learning methods. *Transl Oncol*. 2021;14(1):100907. doi:10.1016/j.tranon.2020.100907

25. Shi Y, Wang H, Yao X, et al. Machine learning prediction models for different stages of non-small cell lung cancer based on tongue and tumor marker: a pilot study. *BMC Med Inform Decis Mak*. 2023;23(1):197. doi:10.1186/s12911-023-02266-5