Supportive Care in Cancer (2025) 33:573
https://doi.org/10.1007/s00520-025-09619-y

RESEARCH

Exploring supportive care needs of lung cancer patients in China and predicting with machine learning models

Zimeng Chen[1] · Sifan Qiu[1] · Chunmei Luo[1] · Jixiang Zeng[1] · Bifang Huang[1] · Juanxia Zheng[1]

Received: 26 August 2024 / Accepted: 1 June 2025 / Published online: 13 June 2025

Abstract

Purpose This study aims to explore the level of supportive care needs among hospitalized lung cancer patients in China, explore the key influencing factors and use machine learning (ML) to develop predictive models for the level of supportive care needs among hospitalized lung cancer patients.

Methods This cross-sectional study collected data on the supportive care needs, demographics, and clinical information of 486 hospitalized lung cancer patients. Univariate and multivariate analyses identified factors associated with these needs. Predictive models were developed using six machine learning methods—logistic regression, linear regression, k-nearest neighbors, support vector machine, random forest, and adaptive boosting—to assess their performance, followed by a visualization of feature importance. The code used for model development and analysis is publicly available at https://github.com/zimengcc/predict_cancer_scn.

Results Among the factors influencing the supportive care needs of hospitalized lung cancer patients, age, education level, occupation, tumor stage, and household per capita monthly income have a significant impact on supportive care needs scores. Multiple linear regression analysis revealed that education level and household per capita monthly income were statistically significant predictors of supportive care needs scores. In the predictive tasks, the random forest model performed the best, with a mean absolute error (MAE) of 4.45 for predicting the total supportive care needs score. Furthermore, to predict the dimension with the highest level of supportive care needs, the model achieved an accuracy of 88. 42%, an F1 score of 87. 49%, and an ROC-AUC of 0.9061.

Conclusion Our study explored the factors influencing the level of supportive care needs among hospitalized lung cancer patients. While the machine learning models demonstrate promising predictive performance, it is important to note that all results were derived solely through cross-validation. Therefore, potential overfitting and overestimation of model performance should be considered when interpreting these findings. Nevertheless, these models may serve as a foundation for developing tools to support personalized care planning in clinical settings.

Keywords Supportive care needs · Lung cancer · Machine learning

Introduction

According to a report released by the World Health Organization (WHO) in 2022, cancer is currently the second leading cause of death in the world after cardiovascular disease [1]. The latest global cancer burden data for 2021, published by the International Agency for Research on Cancer (IARC), indicate that there were 19.29 million new cancer cases worldwide [2]. Among these, lung cancer accounted for 2.2 million new cases, making it the second most common cancer globally and its mortality rate ranked first in the world. Lung cancer is the most frequently diagnosed cancer in both males and females, representing 11.6% of all cancer cases and accounting for 18.4% of cancer deaths [3]. Current projections suggest that by 2030, the number of lung cancer deaths in China will increase by 42.7% [4]. As the most prevalent and deadliest malignant tumor, lung cancer imposes significant health and economic burdens worldwide and has emerged as a major global public health issue [5]. Cancer treatment does not guarantee complete recovery. As a result, many survivors experience a reduced overall quality of life and health-related quality of life (QoL), which may persist over the long term [6].

Extended author information available on the last page of the article

Among patients with lung cancer, some research groups have demonstrated that unmet physical and psychological supportive care needs are independently associated with lower QoL [7]. Other groups have shown that unmet psychological supportive care needs are negatively correlated with global QoL [8]. For lung cancer patients, prolonged and repeated treatments impose significant physical and psychological stress, severely impacting their QoL. Compared to patients with other types of cancer, lung cancer patients experience more substantial physical and psychological distress and have higher needs in psychological, physiological, and daily living aspects [9]. A previous study reported that up to 90% of newly diagnosed cancer patients have unmet needs, and lung cancer patients, in particular, have more unmet needs than those with other types of cancer [10]. The dual pressures of psychological and physiological stress, making it difficult to accurately identify and meet their needs [11]. Therefore, it is crucial for medical staff to systematically understand the needs of lung cancer patients and provide appropriate supportive care to enhance their QoL, facilitate recovery, and alleviate the burden on patients and their families. Accurately identifying and meeting the specific needs of patients has given rise to the important concept of supportive care needs.

The concept of supportive care needs generally refers to the requirements that arise during the course of illness and treatment, including symptom and side effect management, facilitating patient adaptation and coping, enhancing understanding and informed decision-making, and minimizing functional decline [12]. However, the specifics of these needs can vary significantly. Based on the concepts of human needs, cognitive assessment, coping, and adaptation, Fitch et al. [13] define supportive care needs as essential services aimed at meeting the physiological, informational, practical, emotional, psychological, social, and spiritual needs of cancer patients or survivors at any stage—before diagnosis, during treatment, and after treatment ends. These needs revolve around issues of cancer survival, palliation, and bereavement. Hui et al. [14] emphasize that the core of supportive care is person-centered, transforming the healthcare model from a disease-centered approach to a patient-centered one. This approach ensures that cancer patients receive necessary care services that address their informational, emotional, social, and physical needs.

Lung cancer patients generally have various unmet needs, which are affected by many factors. A study by Sanders et al. [15] from Southern California showed that the most critical supportive need for lung cancer patients is information. Studies on the supportive care needs of Western lung cancer patients primarily focus on psychological, physical, and daily living dimensions. Emanuel's study [16] showed that some cancer patients have unmet supportive care needs, and research on the level and type of needs of lung cancer patients being particularly scarce. Clinically, obtaining information about the supportive care needs of hospitalized patients is extremely cumbersome. In order to better understand the supportive care needs of patients, medical staff often need to conduct a detailed assessment, which includes communicating with the patients and his or her families, collecting personal information, medical history, allergy history, and medication use [6]. Additionally, in order to systematically assess patients' needs, medical staff may use questionnaires or assessment tools to perform quantitative evaluations, thus determining the most appropriate care plans. Cancer significantly impacts patients' psychological well-being, including pain, anxiety, and fear of cancer recurrence [17]. This aspect of work can add extra stress to cancer patients.

With the expanding application of information technology in the field of nursing, the integration of Artificial Intelligence (AI) and big data has become a new development trend [18]. Therefore, to better assess the supportive care needs of patients individually, this study introduces a machine learning (ML) model for prediction. ML is a core technology in the field of AI. Its principle is to use different algorithms to learn patterns from large datasets in order to achieve predictions for unknown samples [19]. ML methods, as a branch of computer science increasingly utilized in medical research, are rapidly expanding in both application scope and clinical scenarios. Chaturvedi et al. [20] employed ML techniques to predict and classify lung cancer. Huang et al. [21] used ML algorithms to predict in-hospital mortality rates among critical lung cancer patients. Lynch et al. [22] applied ML to predict the survival rates of lung cancer patients. Nguyen et al. [23] utilized ML models to predict the level of care required for emergency patients upon hospital admission. Hercus et al. [24] utilized an ML-based logistic regression model to predict cases of Consultation-Liaison Psychiatry (CLP) that are misdiagnosed at the time of referral. Recent studies have also employed both non-ML techniques and ML techniques to predict the psychosocial needs of cancer patients. Raineri et al. [25] proposed a non-ML psycho-oncological assessment model for female cancer patients, aiming to identify and predict their psychosocial needs. Nunez et al. [26] utilized natural language processing (NLP) techniques to predict whether cancer patients require consultation with a psychiatrist or counsellor to address their psychosocial needs. Chien et al. [27] developed both unified and category-specific machine learning models to predict long-term care (LTC) service demands among cancer patients.

However, there is a lack of research on predicting the levels of supportive care needs using ML technology. Therefore, this study aims to develop an ML model that incorporates demographic and disease information to predict the supportive care needs and the most urgent requirements of hospitalized lung cancer patients. Predicting these needs with ML models can not only alleviate the workload of medical staff improve efficiency but also reduce the influence of subjective factors in the assessment of patient care needs, enhancing the accuracy and reliability of predictions [28]. This approach can better meet the personalized needs of patients and improve the quality and efficiency of healthcare services.

Methods

This cross-sectional study utilized data from an online survey. Ethical approval was obtained from the Shenzhen Baoan Chinese Medicine Hospital at Guangzhou University of Chinese Medicine (No. KY-2024-021-01). Before data collection, participants were provided with both oral and written explanations of the survey and the research objectives, and informed consent was obtained from all participants.

Study design and participants

This cross-sectional study was conducted from May 2024 to August 2024. The study population consisted of hospitalized lung cancer patients from a tertiary hospital in Guangdong Province, China. Inclusion criteria: participants with a clinical diagnosis of lung cancer; aged $\geq$ 18 years; subjects were conscious; informed consent and voluntary participation in this survey. Exclusion criteria: participants who were unaware of their condition; participants with mental or hearing impairment; incomplete questionnaire content; unreliable answers to questionnaire content.

Participants were selected using a convenience sampling method. Researchers screened subjects based on the inclusion and exclusion criteria, explaining the study's purpose, methods, and confidentiality of information. According to the sample estimation method proposed by Kendall, the sample size should be 5 to 10 times the number of variables [29]. This survey involved a total of 19 variables, comprising 14 items from a socio-demographic and disease questionnaire and 5 items from a supportive care needs questionnaire. Considering a 10% questionnaire invalidation rate, the required sample size for this study ranged from 105 to 209 participants [30]. In light of the actual circumstances, the study ultimately included 486 hospitalized patients with lung cancer. In addition, they completed a socio-demographic, disease questionnaire and a brief supportive care needs questionnaire (SCNS-SF34) in an appropriate setting. If participants were fatigued or physically unfit to complete the questionnaires at the time, they were allowed to complete the questionnaire at other times according to their wishes and convenience.

Data collection and measures

General information questionnaire

The General Information Questionnaire comprises two sections: socio-demographic information and disease characteristics. The socio-demographic section collects data on the patient's gender, age, marital status, occupation, long-term residence, education level, religious affiliation, and household per capita monthly income. The disease characteristics section includes pathological type, tumor stage, duration of illness, treatment methods, comorbidities, and the patient's level of understanding of their disease.

Short-form supportive care needs survey (SCNS-SF34)

The supportive care needs of patients were assessed using the Supportive Care Needs Survey-Short Form 34 (SCNS-SF34). This questionnaire, simplified by Boyes et al. [31] from the original SCNS, maintains high reliability and validity, with Cronbach's alpha coefficients for each dimension not less than 0.86. The SCNS-SF34 consists of 34 items across five dimensions: health systems and information (11 items), sexuality (3 items), patient care and support (5 items), physical and daily living (5 items), and psychological (10 items). It uses a Likert 5-point scale ranging from "no need" to "high need" with higher scores indicating greater unmet needs in that dimension (1 = no need; 2 = need satisfied; 3 = low need; 4 = moderate need; 5 = high need), with higher scores reflecting greater unmet needs for the patient and indicating a higher level of required support. Au et al. [32] translated it into Chinese and reported that the Cronbach's α coefficients for each of the five dimensions exceeded 0.75. The SCNS-SF34 has been widely used in China to measure the supportive care needs of cancer patients. Zhang et al. [33] used it to investigate unmet supportive care needs among Chinese prostate cancer survivors. Chan et al. [7] employed it to explore the supportive care needs and health-related quality of life issues among Chinese lung cancer survivors. Zhu et al. [8] utilized it to examine patterns of unmet supportive needs and their relationship to quality of life in Chinese cancer patients.

Research process

Based on the questionnaire results, we obtained the general information and supportive care needs results of each subject. Supportive care needs scores were considered the dependent variable. We conducted both univariate and multivariate analyses to explore the relationship between the demographic and clinical disease characteristics of hospitalized lung cancer patients and their supportive care needs scores. In the univariate analysis, chi-square tests were used to identify statistically significant variables. Subsequently, these significant variables were then included as independent variables in a multiple linear regression analysis to further investigate their impact on supportive care needs.

Furthermore, we developed six predictive models based on ML techniques using the demographic characteristics, disease characteristics, and supportive care needs scores of hospitalized lung cancer patients. The prediction process framework is shown in Fig. 1. This study aims to explore two predictive tasks:

(1) Predict the total score of supportive care needs for hospitalized lung cancer patients.
(2) Predict the dimensions (health systems and information, patient care and support, physical and daily living, psychological, and sexuality) with the highest level of supportive care needs among hospitalized lung cancer patients.

These two tasks were of great significance to clinical medical staff in providing comprehensive care and support. After the model training was completed, it required only the input of patient demographic and some clinical information to generate the two predictive results.

Statistical analysis

Data process

The data were analyzed using SPSS Statistics version 25.0 to explore the factors influencing supportive care needs.

Additionally, we initially used Chi-square tests, independent *t*-tests, and one-way ANOVA to analyze descriptive and statistical data, including frequency distributions and means (standard deviations). Variables that were statistically significant in the univariate analysis were used as independent variables in the multiple linear regression analysis, with $P < 0.05$ considered statistically significant. As appropriate, we used *t*-tests and ANOVA to compare statistical data. Chi-square tests were employed to determine the relationships between demographic characteristics, disease characteristics, and supportive care needs. In order to perform ML modeling, we first preprocessed the data. The demographic and clinical characteristics of hospitalized lung cancer patients, as presented in Table 1, were used as input features. For each categorical variable (e.g., sex, age group), one-hot encoding was applied to all subcategories (e.g., Male, Female for sex; 18–40, 41–60, 61–90 for age group). These one-hot encoded variables were then used as model inputs. In addition, the total score of supportive care needs, measured by the SCNS-SF34 scale, was calculated for each patient. The dimension with the highest need was identified and one-hot encoded to serve as the model's target label.

Model selection

In this study, six ML models were used for modeling: linear regression, logistic regression, support vector machine (SVM), K-nearest neighbors (KNN), random forest (RF), and Adaptive Boosting (AdaBoost).

Linear and logistic regression models are collectively known as generalized linear models (GLMs). Linear regression is used to solve regression problems and predict continuous variables, while logistic regression is used to solve classification problems and predict binary or multi-classification results [34]. GLMs link linear predictors and response variable distributions through link functions, allowing them to handle various types of response variables, such as normal distribution (for linear regression) and binomial distribution (for logistic regression) [35]. The SVM algorithm maps data into a high-dimensional feature space to identify the optimal boundary for class separation. Through kernel functions, SVM effectively classifies linearly inseparable data and constructs decision boundaries using support vectors, enhancing robustness to outliers [36]. In addition, SVM can also be used for regression problems, through support vector regression (SVR). SVR uses an ϵ-insensitive loss function to find a regression function that maximally includes training data, balancing model complexity, and training error [37]. The application of kernel functions enables SVR to handle nonlinear regression problems, enhancing its flexibility and applicability [38]. KNN is a nonparametric classification and regression method that classifies instances in feature space based on similarity measures such as Euclidean distance [39, 40]. In regression problems, a continuous output value is predicted by computing the average of the K-nearest neighbors

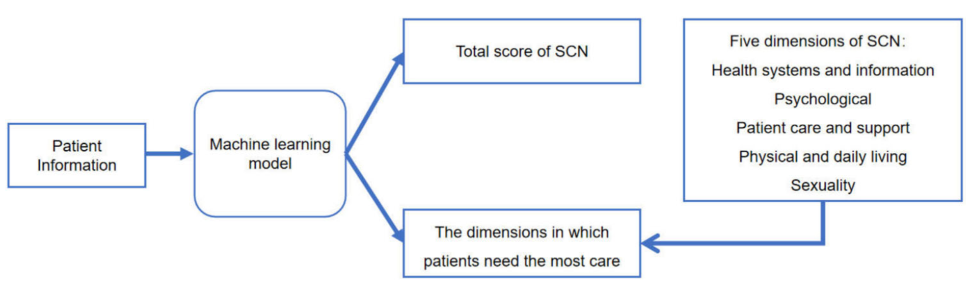


Fig. 1 Prediction framework for supportive care needs (SCN)

Table 1 Univariate analysis of supportive care needs in lung cancer patients (n = 486)

Variables	n	n (%)	SCNs ($\bar{x} \pm s$)	t/F	P
Gender				0.392[a]	0.695
Male	300	61.73	125.41 ± 7.21		
Female	186	38.27	124.80 ± 5.75		
Age (years)				3.814[b]	0.025
18–40	11	2.26	117.00 ± 6.60		
41–60	99	20.37	123.30 ± 6.77		
61–90	376	77.37	126.06 ± 6.77		
Marital status				0.345[a]	0.602
Married	398	81.89	125.30 ± 6.96		
Divorced or separated	88	18.11	124.23 ± 6.62		
Single	0	0.00	–		
Occupation				4.494[b]	0.004
Retired person	48	9.88	121.92 ± 6.02		
Employed	11	2.26	121.33 ± 12.66		
Farmer	125	25.72	128.62 ± 6.56		
Other	302	62.14	124.60 ± 6.42		
Long-term residence				1.798[b]	0.151
Downtown	169	34.77	124.50 ± 6.43		
Suburb	81	16.67	123.64 ± 7.35		
County	30	6.17	123.38 ± 8.12		
Town	206	42.39	126.88 ± 6.72		
Education level				2.335[b]	0.007
Primary school and below	350	72.02	126.17 ± 7.23		
Junior high school	107	22.02	123.79 ± 5.38		
High school (including technical secondary school)	18	3.70	120.60 ± 4.50		
University (including college, undergraduate and above)	11	2.26	120.00 ± 7.48		
Religious affiliation				−1.097[a]	0.275
Yes	59	12.14	123.50 ± 4.66		
No	427	87.86	125.54 ± 7.19		
Household per capita monthly income				40.639[b]	< 0.001
≤ 2000	40	8.23	138.18 ± 7.03		
2000<income≤ 4000	279	57.41	127.47 ± 4.90		
> 4000	167	34.36	126.06 ± 6.42		
Pathological type					
Squamous cell carcinoma	119	24.48	126.56 ± 7.06		
Adenocarcinoma	253	52.06	124.58 ± 7.39		
Large cell carcinoma	11	2.26	125.00 ± 1.41		
Small cell carcinoma	52	10.70	125.64 ± 5.34		
Unknown	51	10.49	125.71 ± 6.24		
Tumor stage				4.258[b]	0.017
I−II	35	7.20	123.20 ± 6.43		
III	105	21.60	122.18 ± 7.09		
IV	266	54.73	126.53 ± 6.95		
Unknown stage	80	16.46	126.18 ± 5.48		
Other	0	–	–		

Table 1 continued

Variables	*n*	*n* (%)	SCNs ($\bar{x} \pm s$)	*t/F*	*P*
Duration of illness				0.689^b	0.504
1-6 months	228	46.91	126.02 ± 6.78		
6-12 months	93	19.13	125.08 ± 7.65		
> 12 months	165	33.95	124.42 ± 6.70		
Treatment method				1.107^b	0.363
Surgery	15	3.09	128.50 ± 3.50		
Chemotherapy	316	65.02	126.14 ± 7.04		
Radiation therapy	11	2.26	122.67 ± 1.25		
Targeted therapy	11	2.26	123.33 ± 6.24		
Surgery + chemotherapy	55	11.32	122.40 ± 6.27		
Surgery + radiotherapy	4	0.82	126 ± 2.24		
Surgery + targeted therapy	4	0.82	114 ± 3.25		
Chemotherapy + radiotherapy	7	1.44	121 ± 2.00		
Chemotherapy + targeted therapy	63	12.96	124.76 ± 7.46		
Comorbidities				-1.323^a	0.188
Yes	443	91.15	125.07 ± 6.84		
No	43	8.85	128.10 ± 7.76		
Understanding of the disease				0.905^b	0.407
Familiar	327	67.28	125.31 ± 7.09		
Understand	48	9.88	123.08 ± 5.14		
Don't understand	111	22.84	126.20 ± 7.06		

(a) Independent *t*-test; (b) One-way ANOVA

[41]. RF is an ensemble method of decision trees notable for its ability to detect feature interactions and its interpretability [42, 43]. RF uses bootstrap sampling and random feature selection to avoid overfitting, leading to strong generalization. By aggregating predictions from multiple decision trees, the algorithm is highly robust to outliers and missing data. AdaBoost is an ensemble learning method that improves classification performance by combining multiple weak classifiers. It is simple to implement, requires no adjustment of the weak classifiers, and exhibits robustness to noise and outliers. AdaBoost can elevate a weak learning algorithm with slightly better accuracy than random guessing to an arbitrarily accurate strong learning algorithm [44].

The optimal parameters for each model are identified through grid search-based tuning. We employed cross-validation techniques to conduct the grid search on the training dataset. Through experiments, we assessed the performance of various parameter combinations across different folds and ultimately chose the model parameters that yielded the highest validation performance. For the linear regression model, we primarily adjusted the regularization parameter. For the logistic regression model, the regularization parameter and the type of penalty term were considered. For the SVM model, we tuned the regularization parameter, kernel type, polynomial degree, and kernel coefficient. For the KNN model, different K values and distance metrics were selected. For the RF model, we adjusted the number of estimators, maximum tree depth, number of features, minimum samples required to split a node, and minimum samples required for a leaf node. For the AdaBoost model, we adjusted the type and number of basic weak classifiers, as well as the learning rate of each classifier. Through these parameter adjustments, we ensured the best performance of each model on the dataset. The entire prediction process is shown in Fig. 2.

Assessing model performance

In order to better evaluate the performance of each model, we used different indicators to compare the models. For the first prediction task (predicting the total supportive care needs score), we used mean absolute error (MAE), mean squared error (MSE), and the coefficient of determination (R^2). Then, for the second prediction task (predicting the dimension with the highest level of supportive care needs), we used accuracy, precision, recall, Macro-F1 score, and the area under the receiver operating characteristic curve (ROC-AUC). Furthermore, through these evaluation indicators, we can fully understand the performance of each model in predicting supportive care needs and select the best model. We will use the ten-fold cross-validation method to ensure the accuracy and

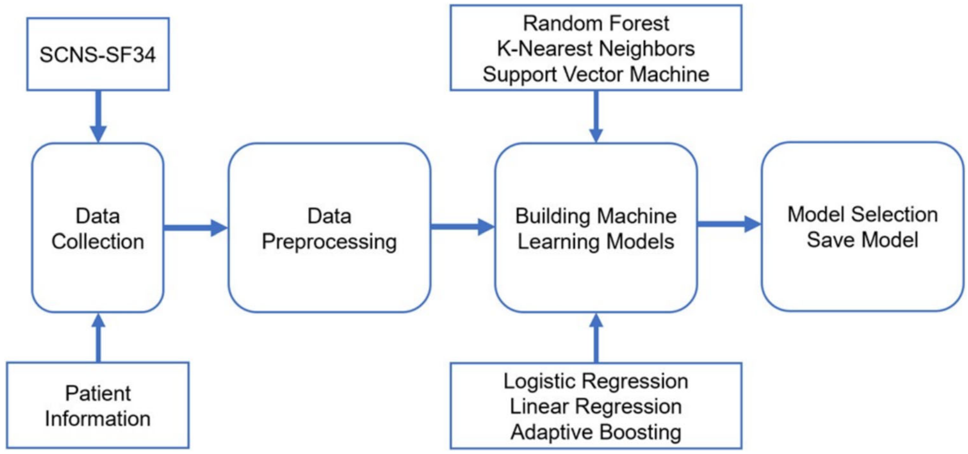


Fig. 2 Machine learning modeling framework

stability of the results, and the final result will be the average of each model in the ten-fold cross-validation.

Feature importance

To elucidate the predictive generation process of ML models, we opted to visualize the feature importance of the RF model. Feature importance is assessed using Gini importance. This technique determines the importance of each feature by calculating the average decrease in Gini impurity across the decision tree nodes [45]. Specifically, Gini importance measures each feature's contribution to reducing the impurity of the dataset. When a feature splits the data in the decision tree, it impacts the Gini impurity. If a feature frequently appears across multiple nodes and effectively partitions the data, its Gini importance score will be high. Therefore, a higher importance score indicates a greater impact of that feature on the model's predictions [46].

Results

Analysis of factors influencing supportive care needs

A total of 486 patients hospitalized with lung cancer were recruited. The patients included 300 males (61.73%) and 186 females (38.27%). The majority of patients were aged between 61 and 90 years, with 398 patients (81.89%) being married. The predominant occupations were other professions (62.14%), including examples such as self-employed individuals and freelancers. Most patients resided in towns (42.39%). The education level was primarily elementary or below, making up 72.02%. Most patients (87.86%) had no religious affiliation, and the majority of families had a household per capita monthly income between 2000 and 4000 Chinese yuan (57.41%). The most common cancer types were adenocarcinoma (52.06%) and squamous cell carcinoma (24.48%), with chemotherapy (65.02%) being the primary treatment modality. Among the patients, 91.15% had comorbidities, and 67.28% were familiar with their disease.

As shown in Table 1, univariate analysis revealed that differences in age, occupation, education level, tumor stage, and household per capita monthly income were significantly associated with supportive care needs scores ($P < 0.05$). However, gender, marital status, long-term residence, religious affiliation, pathological type, duration of illness, treatment method, comorbidities, and level of understanding of the disease did not reach statistical significance ($P < 0.05$). Other detailed demographic characteristics of hospitalized cancer patients and their relationship with supportive care needs are presented in Table 1.

Multiple linear regression analysis was performed with the supportive care needs score as the dependent variable and variables with statistical differences in the univariate analysis as independent variables. The variables were coded as follows: Household per capita monthly income ($\leq 2000 = 1$; $2000 - 4000 = 2$; $> 4000 = 3$); Education level (primary school and below = 1; junior high school = 2; high school (including technical secondary school) = 3; university (including junior college, undergraduate, and above) = 4); Age (18–40 years = 1; 41–60 years = 2; 61–90 years = 3); Tumor stage ($\text{I} - \text{II} = 1$; $\text{III} = 2$; $\text{IV} = 3$; unknown stage = 4); and Occupation (other = 1; farmer = 2; employed = 3; retired person = 4). The results of multiple linear regression analysis showed that education level and household per capita monthly income ($P < 0.05$) entered the regression equation. The results are shown in Table 2.

According to the results of the SCNS-SF34 questionnaire, the average total score of supportive care needs of all hospitalized lung cancer patients was 125.26 ± 6.84 points (out of 150 points), and the average score of all items was 3.68 ± 1.21 points (out of 5 points). The scores of each dimension are shown in Table 3. The needs of the five dimensions were ranked from highest to lowest: health systems

Table 2 Multivariate analysis of supportive care needs in hospitalized patients with lung cancer

Variables	B	SE	β	t	P	95%CI
Constant term	137.911	4.780	–	28.851	<0.001	128.451–147.370
Age	−2.196	1.184	−0.157	−1.855	0.066	−4.458–0.147
Education level	−3.166	0.879	−0.313	−3.602	<0.001	−4.905–1.426
Household per capita monthly income	−2.722	0.953	−0.242	−2.856	0.005	−4.609–0.836
Tumor stage	1.106	0.692	0.132	1.598	0.113	−0.264–2.476
Occupation	−0.017	0.602	−0.002	−0.027	0.978	−1.207–1.174

(B, regression coefficient; SE, standard error; β, standardized regression coefficient; CI, confidence interval; $R^2 = 0.697$, adjusted $R^2 = 0.665$; $F = 6.179$; $P < 0.001$)

and information, patient care and support, physical and daily living, psychological, and sexuality. We also calculated the relative desirability of each dimension by dividing the average score of each dimension by the total score of each dimension, which reflects the relative importance or score proportion of each dimension in the overall score. Thus, this method helps to understand the contribution of each dimension to the overall score and its relative weight in the overall score. A higher proportion indicates a greater influence of that dimension in the total score, highlighting the aspects that occupy a larger portion of the overall evaluation. In general, patients have the highest demand for health systems and information, reaching 84.80%. Next are patient care and support (75.20%), physical and daily living (74.00%), and psychological (72.20%), however, these differences are not obvious. Among them, the lowest demand is in sexuality, with a demand level of only 34.80%.

Model performance

The performance of the models was evaluated using the assessment metrics from the test results of each model. Based on five different machine learning models, the evaluation for Prediction Task 1 is illustrated in Fig. 3. The linear regression model performed the worst in terms of MSE and MAE, with values of 61.11 and 6.30, respectively, and an R^2 of only 0.04, indicating poor fit to the data. The KNN and Adaboost models exhibited moderate performance, while the SVM and RF models performed the best in terms of MAE and MSE. However, the RF model had a higher R^2 value, indicating it can better fit the data. Overall, the RF model exhibited the best performance in predicting the total supportive care needs score for lung cancer inpatients, with higher prediction accuracy and a better fit to the data.

The results of Prediction Task 2 are presented in Table 4. Based on the performance of each model, all five ML models achieved accuracy rates above 84%, with the RF model demonstrating the highest accuracy at 88.42%. Furthermore, the RF model also exhibited the best overall performance in terms of precision, recall, and F1 score, with values of 86.53%, 89.43%, and 87.49%, respectively. The ROC curves are shown in Fig. 4, where all models achieved an ROC-AUC value above 0.87, with the RF model having the highest ROC-AUC at 0.9061, indicating the best performance. Analyzing these results, it is evident that the RF model displayed outstanding comprehensive performance in handling Prediction Task 2. It not only led in accuracy but also performed excellently in precision, recall, and F1 score, suggesting a balanced predictive capability across different classes. The SVM model, while performing well in terms of accuracy (86.29%) and ROC-AUC (0.8752), had relatively lower precision (82.94%) and F1 score (83.52%). The KNN and logistic regression models exhibited balanced performance across various metrics, but without a distinct advantage. The Adaboost model's performance is also noteworthy, with a high accuracy of 87.12% and a ROC-AUC value of 0.8919, indicating its potential in enhancing classifier performance. However, the RF model's comprehensive performance across

Table 3 Scores of each dimension of supportive care needs and needs assessment results

Dimensions	Total score	Average score ($\bar{x} \pm s$)	Items average score ($\bar{x} \pm s$)	Relative demand (%)
Health systems and information	55	46.64±8.01	4.24±0.71	84.80
Psychological	50	36.10±12.10	3.61±1.21	72.20
Patient care and support	25	18.80±4.35	3.76±0.87	75.20
Physical and daily living	25	18.50±5.50	3.70±1.10	74.00
Sexuality	15	5.22±1.83	1.74±0.61	34.80

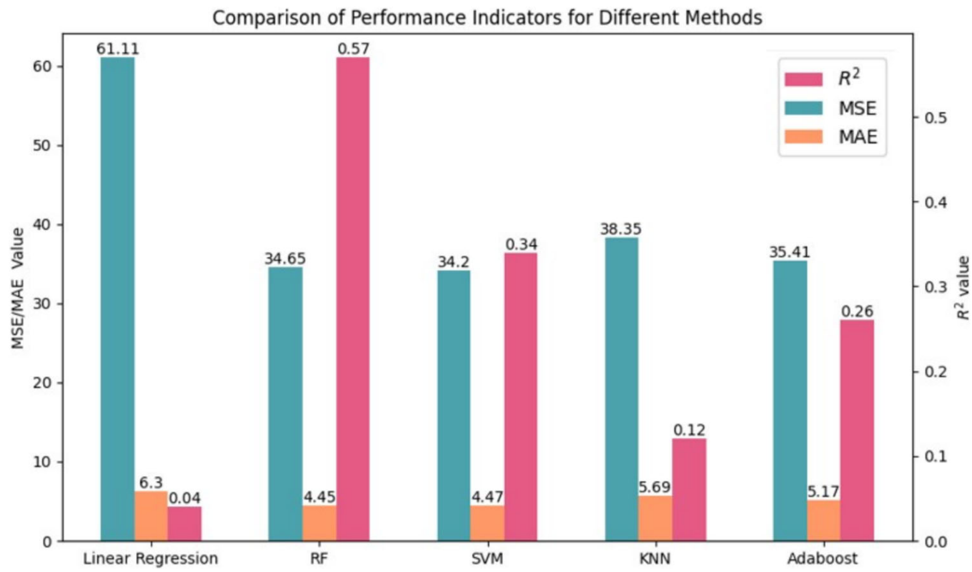


Fig. 3 Performance of different machine learning methods for Prediction Task 1

all evaluation metrics made it the preferred model for Prediction Task 2.

In summary, the RF model demonstrated the most outstanding predictive ability and stability in this study. To gain a deeper understanding of the model's decision-making mechanism, we will normalize the feature importance within the RF model and visualize it. Normalization allows for a clearer comparison of the relative contributions of different features to the prediction results. Additionally, visualizing feature importance aids in identifying the most critical features, thereby enabling model optimization and performance enhancement. The visualizations of feature importance for Prediction Task 1 and Task 2 are shown in Fig. 5.

As shown in Fig. 5a, the RF model visualized feature importance for Prediction Task 1 (i.e., predicting the total supportive care needs score for lung cancer inpatients). The most important features identified in the model's predictions include patients' education level, tumor stage, pathological type, household per capita monthly income, and occupation. Figure 5b illustrates a similar visualization of feature importance for Prediction Task 2 (i.e., predicting the most needed dimension within supportive care needs for lung cancer inpatients). In this case, the key predictive features include age, pathological type, tumor stage, education level, and household per capita monthly income. Further validation through univariate analysis confirmed significant differences in education level, occupation, tumor stage, and household per capita monthly income within the supportive care needs assessment. These findings not only highlight the importance of these factors in the supportive care needs assessment but also reinforce their reliability and effectiveness in predicting supportive care needs through the Random Forest model.

Table 4 Performance of different machine learning methods for Prediction Task 2

Model	Accuracy	Precision	Recall	F1 score
SVM	86.29	82.94	85.52	83.52
KNN	85.36	84.56	85.16	84.79
Logistic regression	84.82	86.37	81.41	83.70
Adaboost	87.12	85.21	87.21	86.58
RF	**88.42**	**86.53**	**89.43**	**87.49**

Discussion

Analysis of supportive care needs and demographic factors

In this study, we found that differences in age, education level, occupation, tumor stage, and household per capita monthly income showed significant differences in supportive care needs scores, consistent with prior research by Fitch et al. [47]. Zhang et al. [48] examined the supportive care needs of lung cancer patients in mainland China and concluded that gender, age, education, and income level significantly impacted supportive care needs, though their study did not consider occupation and tumor stage. Consistent with our findings, Sanders et al. [15] and Liao et al. [9] also reported no significant gender differences in supportive care needs. Our results indicated that 61.67% of the surveyed patients were male. However, Sarna et al. [49] found that females had more needs than males, may due to their roles within the family. Adenocarcinoma was the most common type of lung cancer in our study (52.12%), aligning with the

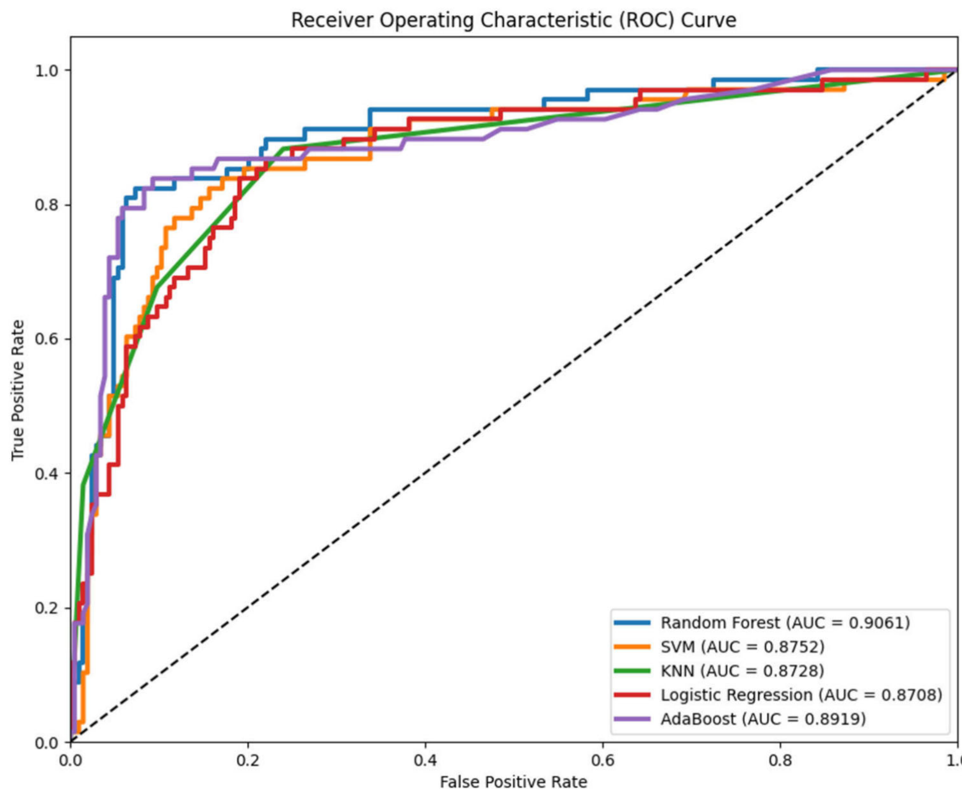


Fig. 4 ROC curves of different machine learning methods

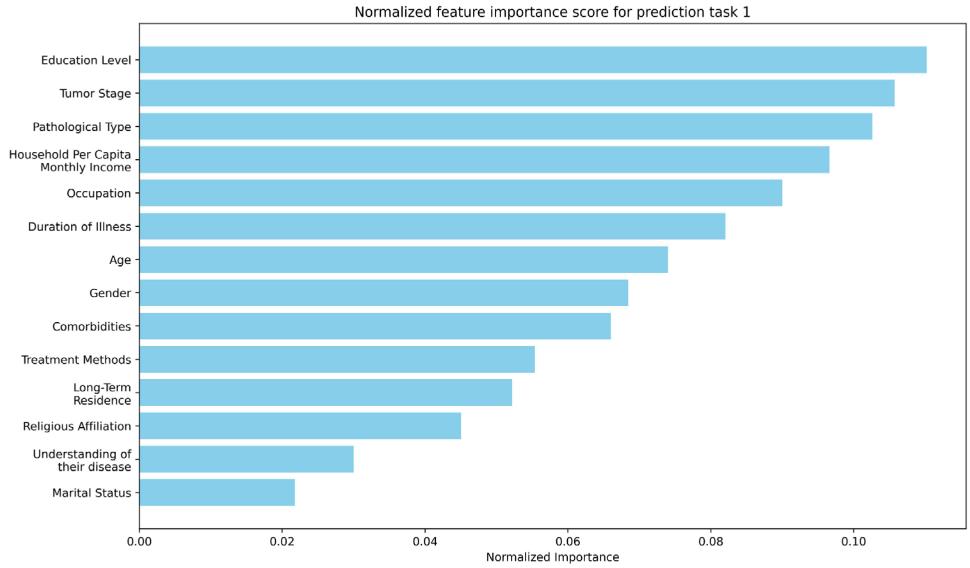
epidemiological characteristics of the disease [50]. Patients in the early treatment stage (1-6 months) accounted for 42.9%, indicating a higher demand for information and psychological support [51]. Additionally, having an education level of elementary school or below was a significant factor contributing to higher supportive care needs in lung cancer patients. At the same time, having an education level of primary school or below is a factor that affects the high demand for supportive care among lung cancer patients. The analysis found that lung cancer patients with low education levels have poor comprehension ability, find it difficult to effectively master relevant disease knowledge, and do not actively collect relevant information, so they have a higher demand for supportive care [52]. Yun et al. [53] found that lung cancer patients with a monthly income of less than 2,000 Chinese yuan faced a higher economic burden. Similarly, this study found that patients with lower incomes had higher demand levels, and differences in occupational and family monthly income directly affected the economic level of patients. Since the clinical symptoms of lung cancer appear late and lack specificity, a series of primary and auxiliary treatments are required, resulting in a large burden on patients and their families [54, 55]. Furthermore, Harrison et al. [56] explored the differences in unmet needs of different cancer patients at different stages of the disease and found that tumor stage affects the unmet needs of cancer patients, which is similar to our study.

Analysis of the overall level of supportive care needs

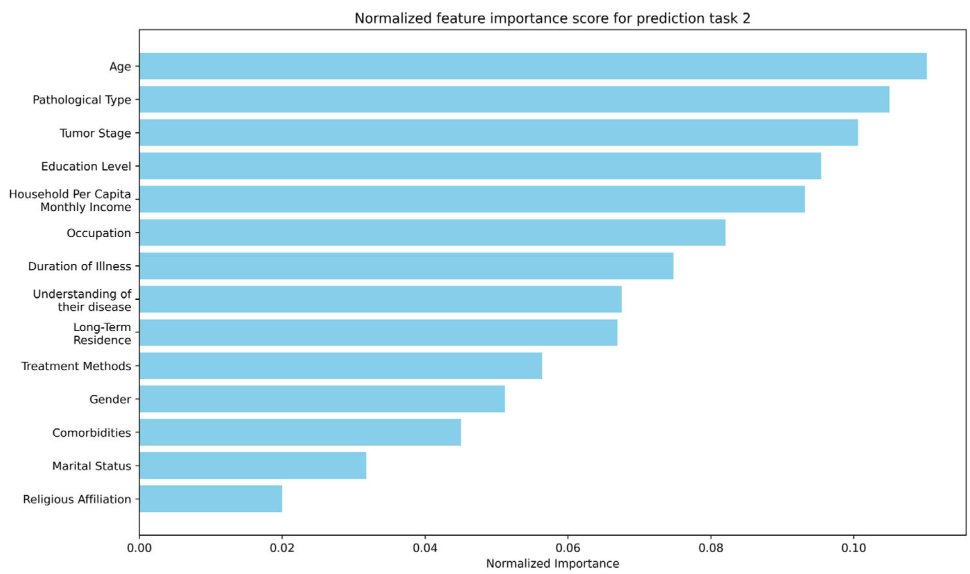
This study showed that the rate of unmet health systems and information needs among hospitalized patients with lung cancer was the highest, which is consistent with the results of related studies [57, 58]. The reason for this may be that most patients are in the initial treatment stage and have a low level of education. As a result, they have a pressing need for extensive information to reduce their uncertainty about the disease. Patients urgently seek guidance on disease-related knowledge, particularly regarding disease progression, treatment options, and self-care. Therefore, it is suggested that medical staff should attach importance to the assessment and satisfaction of patients' health information needs, and carry out various forms of targeted education during hospitalization, such as organizing departmental lectures, encouraging peer sharing, and doing a good job of follow-up.

In addition, as shown in Table 3, the relative demand for patient care and support reached 75.20%, ranking second among all the evaluated dimensions. Farley's [59] investigation revealed that patients desire more opportunities to choose specialized doctors and hospitals and want medical staff to acknowledge their feelings. This study supports these findings. The primary reasons for this unmet need may include limitations in medical resources and energy of medical staff. Lintz et al. [60] reported that 33% of cancer patients' care and support needs were not met during treatment. In

Fig. 5 Visualization of feature importance for random forest prediction tasks



(a) Task 1



(b) Task 2

the early stages of diagnosis, patients often experience significant psychological anxiety and doubt, and they long for effective treatment and emotional recognition and concern from medical staff.

In the dimension of physical and daily living needs, the demand accounted for 74.00%. Further investigation revealed that lung cancer inpatients' concerns in this dimension varied with different treatment methods: surgical patients primarily focused on alleviating postoperative pain, managing drainage tubes, and performing respiratory exercises, while chemotherapy patients were mainly concerned about how to reduce the occurrence of adverse reactions such as fatigue, nausea, hair loss, and leukopenia. A comparative study on the supportive care needs of breast cancer patients showed that in Hong Kong, female patients were mainly concerned about health information, while in Germany, female patients were more in need of physical and psychological support [61]. Compared with Western women, Chinese women reported more health information needs and physical needs, but relatively fewer psychological needs. Therefore, cancer supportive care services need to consider more differences in cultural and health service contexts to reduce unmet needs.

The sexuality (34.8%) dimension scored the lowest among all dimensions, which is similar to the research results of Au et al. [32]. On the other hand, some research indicates that discussions about sexuality are often avoided [62] or approached with caution in China due to cultural sensitivities [63], this may lead to biased results. While most cancer patients reported sexuality to be a lower priority compared to other needs [64], it remains important for healthcare professionals to offer timely sexual education and address any related misconceptions.

Machine learning model analysis

Effect of feature normalization on results

During the data collection process, as shown in Table 1, all variables except age were mutually independent categorical features. Since feature normalization is primarily applicable to continuous variables [65], and age was the only continuous variable in our dataset, we applied min-max normalization to age, while the remaining features were encoded using one-hot encoding. The experimental results, illustrated in Figs. 6 and 7, show that applying min-max normalization to age, while using one-hot encoding for the other features, resulted in worse model performance compared to the approach where all features, including age, were encoded using one-hot encoding. Specifically, for Task 1, the MSE results of the various machine learning models increased after the normalization of the features. For Task 2, the ROC-AUC scores of the various machine learning models decreased with normalized features. This may be explained by inconsistency in feature representation, which could have caused the model to allocate unequal attention to different types of features during training, thereby impairing overall performance. When age was discretized and subsequently one-hot encoded—ensuring a consistent representation format across all features—the model was able to learn more stable relationships between features and achieved improved predictive performance.

Results and interpretability analysis

In this study, six different ML models were employed to develop predictive models for supportive care needs based on demographic and disease characteristics. Among the models, the RF model performed the best. For predicting the total supportive care needs score for lung cancer inpatients, the RF model achieved an MAE of 4.45 and an R^2 of 0.57. For predicting the dimension with the highest level of supportive care needs, the RF model achieved an accuracy of 88.42%, Macro-F1 score of 87.49%, and ROC-AUC of 0.9061. Feature importance analysis based on the RF model indicated that, for predicting the total supportive care needs score, the most important features were education level, tumor stage, pathological type, household per capita monthly income, and occupation. These features highlight the critical role of patients' economic status, disease severity, and individual background in influencing overall supportive care needs. For

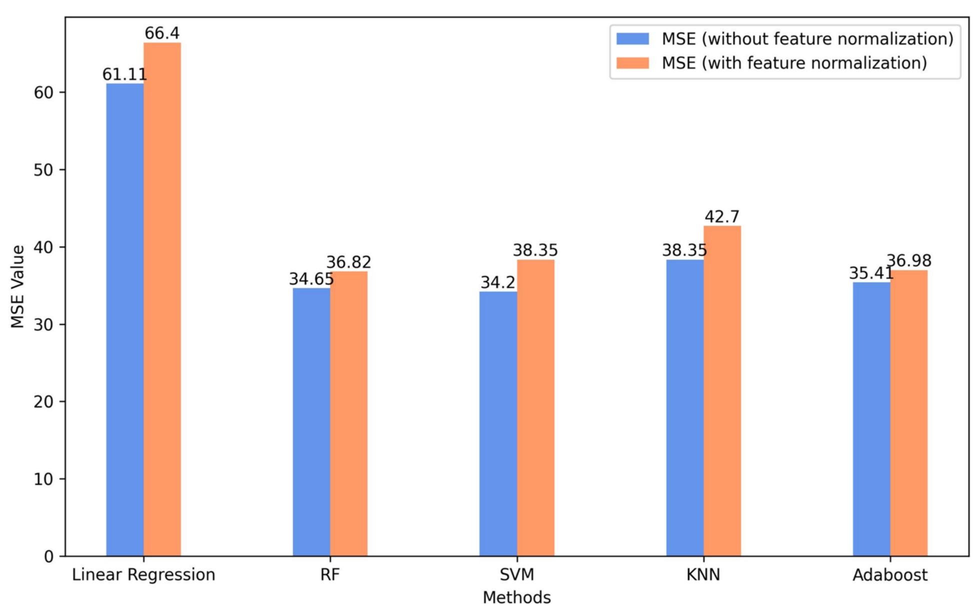
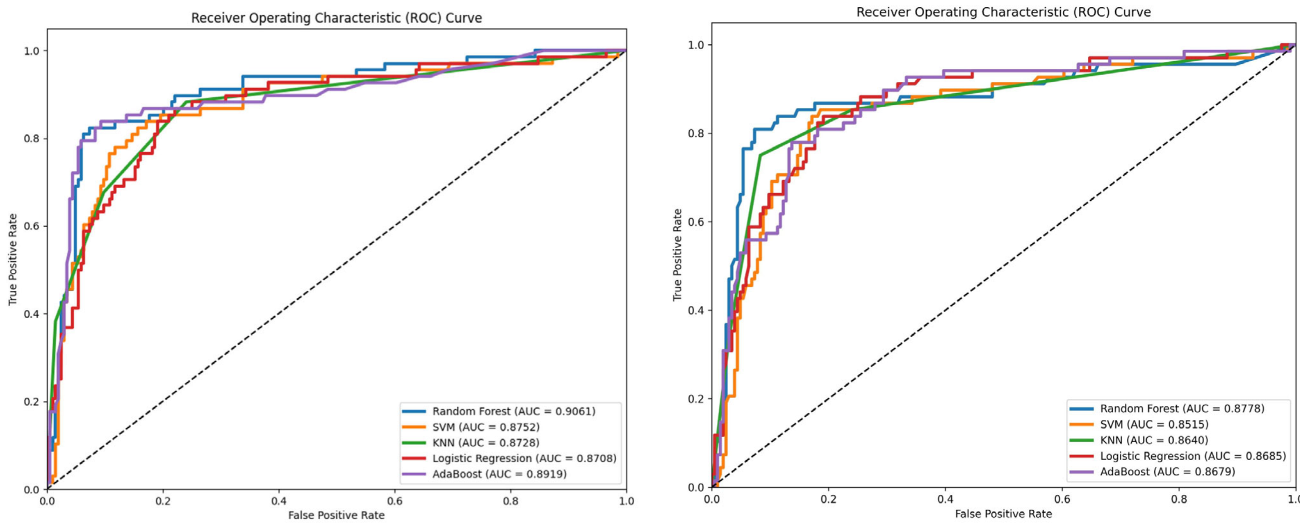


Fig. 6 MSE results of different models in Task 1 (blue: without feature normalization; orange: with feature normalization)



(a) Results without Feature Normalization (b) Results with Feature Normalization

Fig. 7 ROC curves of different models in Task 2: **a** without feature normalization and **b** with feature normalization

predicting the dimension with the highest level of supportive care needs, age, pathological type, tumor stage, education level, and household per capita monthly income were identified as the most important predictive features. The results reveal significant differences in supportive care needs based on age, education level, occupation, tumor stage, and ahousehold per capita monthly income among lung cancer patients. These factors' impact on supportive care needs scores aligns with the key features identified by the RF model, reflecting the model's effectiveness and reliability in both prediction tasks. This also provides important predictive insights for personalized supportive care. From a clinical perspective, the results presented in Fig. 3 and Table 4 indicate ML models have the potential to predict patients' supportive care needs in various aspects of lung cancer treatment. Such predictions may assist medical staff in formulating more personalized care plans, thereby improving treatment outcomes and enhancing patients' quality of life [66]. Additionally, these models can help identify high-risk patients and detect early signs of disease progression.

Conclusions

In conclusion, hospitalized lung cancer patients generally exhibit significant needs in the areas of health systems and information, patient care and support, as well as physical and daily living. Supportive care needs are influenced by various factors. Univariate analysis in this study revealed significant differences in supportive care needs scores based on age, education level, occupation, tumor stage, and household per capita monthly income ($P < 0.05$). Further multivariate linear regression analysis indicated that education level and household per capita monthly income ($P < 0.05$) are the primary factors affecting supportive care needs.

To more efficiently predict supportive care needs, this study developed predictive models based on demographic and disease characteristics. Among the various ML methods, the RF model demonstrated the best performance, achieving an MAE of 4.45 for predicting the total supportive care needs score for lung cancer inpatients, with an accuracy of 88.42%, F1 score of 87.49%, and ROC-AUC of 0.9061 for predicting the dimension with the highest level of supportive care needs. These results not only highlight the reliability and accuracy of the model in predictions but also provide crucial insights for developing personalized care plans, contributing to the optimization of treatment outcomes and quality of life for lung cancer patients.

Relevance to clinical practice and limitations

The results of this study provide a robust foundation for assessing, predicting, and delivering supportive care needs for hospitalized lung cancer patients. Considering the diversity and complexity of patients' needs, medical staff should effectively identify and evaluate these needs to ensure the provision of the best personalized care services. Additionally, the application of ML in the nursing field is progressively transforming traditional care models. By analyzing patients' medical data, ML algorithms can accurately predict patients' needs and disease development trends, helping medical staff to develop more personalized and effective care plans. This not only improves the accuracy and efficiency of nursing services, but also enables the early detection of potential health problems, thereby enabling timely intervention and treatment.

This study has certain limitations in interpreting the data. Firstly, as a cross-sectional study, many of the evaluation metrics are subjective and may be influenced by life events, thus potentially impacting the results and failing to capture changes in supportive care needs over time. Secondly, the study was conducted solely among hospitalized lung cancer patients in one tertiary hospital, which limits the generalizability of the findings and conclusions to a broader population. Regarding the results of our machine learning models, it is important to note that model evaluation was conducted solely using cross-validation, without an independent hold-out set or external validation. This may lead to overfitting and an overestimation of model performance. Therefore, these conclusions should be interpreted with caution. Further validation using larger, independent datasets and testing in different clinical settings is necessary to confirm the robustness and generalizability of these findings.

Acknowledgements We thank all participants involved in this study. We also express our gratitude to the medical staff at the Seventh Clinical Medical College of Guangzhou University of Chinese Medicine. Additionally, we acknowledge the support from the Sanming Project of Medicine in Shenzhen (No. SZZYSM202206014).

Author contributions Z.C. contributed significantly to the conception of the study, data collection and analysis, and the preparation and writing of the manuscript; S. Q., J. Z., and C. L. assisted in the collection and analysis of data; J.Z. and B.H. contributed to the conception of the study and reviewed the manuscript. All authors have read and approved the manuscript.

Funding This study was supported by the Sanming Project of Medicine in Shenzhen (No. SZZYSM202206014), the National Key Clinical Specialty (TCM) Construction Project (2013-239) and Shenzhen Baoan District Medical and Health Research Project (2022JD237).

Data availability No datasets were generated or analysed during the current study.

Declarations

Ethical approval This study was approved by the Ethics Committee of Shenzhen Bao'an Chinese Medicine Hospital at Guangzhou University of Chinese Medicine (No. KY-2024-021-01). All methods were conducted in accordance with the Declaration of Helsinki.

Consent to participate Informed consent was obtained from all participants involved in the study.

Competing interests The authors declare no competing interests.

References

1. Organizatio G (2022) World health statistics 2022: monitoring health for the SDGs, sustainable development goals. World Health Organization, Geneva
2. Deo S, Sharma J, Kumar S (2022) GLOBOCAN 2020 report on global cancer burden: challenges and opportunities for surgical oncologists. Ann Surgical Oncol 29(11):6497–6500
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer J Clinicians 68(6):394–424
4. Fang Y, Li Z, Chen H, Zhang T, Yin X, Man J, Yang X, Lu M (2023) Burden of lung cancer along with attributable risk factors in china from 1990 to 2019, and projections until 2030. J Cancer Res Clinical Oncol 149(7):3209–3218
5. Kent EE, Rowland JH, Northouse L, Litzelman K, Chou W-YS, Shelburne N, Timura C, O'Mara A, Huss K (2016) Caring for caregivers and patients: research and clinical priorities for informal cancer caregiving. Cancer 122(13):1987–1995
6. Park J, Jung W, Lee G, Kang D, Shim YM, Kim HK, Jeong A, Cho J, Shin DW (2023) Unmet supportive care needs after non-small cell lung cancer resection at a tertiary hospital in Seoul, South Korea. In: Healthcare, MDPI, vol. 11, pp 2012
7. Chan DN-S, Choi K-C, Chan CW-H, Wan RW-M, Mak SS-S, Wong EM-L, So WK-W et al (2012) Supportive care needs and health-related quality of life among Chinese lung cancer survivors. Adv Lung Cancer 1(02):5
8. Zhu L, Yao J, Schroevers MJ, Zhang H, Xie J, Liu A, Fleer J, Ranchor AV, Song Z (2018) Patterns of unmet supportive needs and relationship to quality of life in Chinese cancer patients. Psycho-Oncol 27(2):600–606
9. Liao Y-C, Liao W-Y, Shun S-C, Yu C-J, Yang P-C, Lai Y-H (2011) Symptoms, psychological distress, and supportive care needs in lung cancer patients. Supportive Care Cancer 19:1743–1751
10. Puts M, Papoutsis A, Springall E, Tourangeau A (2012) A systematic review of unmet needs of newly diagnosed older cancer patients undergoing active cancer treatment. Supportive Care Cancer 20:1377–1394
11. McFarland DC, Miller AH, Nelson C (2021) A longitudinal analysis of inflammation and depression in patients with metastatic lung cancer: associations with survival. Biol Res Nursing 23(3):301–310
12. Paterson C, Primeau C, Howard N, Xiberras P, Pillay B, Crowe H (2020) Experiences of unmet need and access to supportive care for men receiving androgen deprivation therapy for prostate cancer: a bi-national study. In: Seminars in oncology nursing, Elsevier, vol. 36, pp 151049
13. Weir HK, Sloan M, Kreiger N (1994) The relationship between cigarette smoking and the risk of endometrial neoplasms. Int J Epidemiol 23(2):261–266

14. Hui D (2014) Definition of supportive care: does the semantic matter? Current Opinion Oncol 26(4):372–379
15. Sanders SL, Bantum EO, Owen JE, Thornton AA, Stanton AL (2010) Supportive care needs in patients with lung cancer. Psycho-oncology 19(5):480–489
16. Emanuel LL, Alpert HR, Emanuel EE (2001) Concise screening questions for clinical assessments of terminal care: the needs near the end-of-life care screening tool. J Palliative Med 4(4):465–474
17. Jarrett N, Scott I, Addington-Hall J, Amir Z, Brearley S, Hodges L, Richardson A, Sharpe M, Stamataki Z, Stark D et al (2013) Informing future research priorities into the psychological and social problems faced by cancer survivors: a rapid review and synthesis of the literature. Eur J Oncol Nursing 17(5):510–520
18. Hashiguchi TCO, Oderkirk J, Slawomirski L (2022) Fulfilling the promise of artificial intelligence in the health sector: let's get real. Value Health 25(3):368–373
19. Tack C (2019) Artificial intelligence and machine learning| applications in musculoskeletal physiotherapy. Musculoskeletal Sci Pract 39:164–169
20. Chaturvedi P, Jhamb A, Vanani M, Nemade V (2021) Prediction and classification of lung cancer using machine learning techniques. In: IOP Conference Series: Materials Science and Engineering, IOP Publishing, vol. 1099, pp 012059
21. Huang T, Le D, Yuan L, Xu S, Peng X (2023) Machine learning for prediction of in-hospital mortality in lung cancer patients admitted to intensive care unit. Plos one 18(1):0280606
22. Lynch MR, Tran MT, Parikh SM (2018) Pgc1α in the kidney. Am J Physiol-Renal Physiol 314(1):1–8
23. Nguyen M, Corbin CK, Eulalio T, Ostberg NP, Machiraju G, Marafino BJ, Baiocchi M, Rose C, Chen JH (2021) Developing machine learning models to personalize care levels among emergency room patients for hospital admission. J Am Med Inf Assoc 28(11:2423–2432
24. Hercus C, Hudaib A-R (2020) Delirium misdiagnosis risk in psychiatry: a machine learning-logistic regression predictive algorithm. BMC Health Services Res 20:1–7
25. Raineri AM, Grotto RL, Fioravanti G, Casale S, Rotella F, Fei L, Villanucci A, Amunni G, Faravelli C (2021) Assessing psychological needs in female cancer patients: a tailored model. Acta Bio Medica: Atenei Parmensis 92(Suppl 2):2021005
26. Nunez J-J, Leung B, Ho C, Ng RT, Bates AT (2024) Predicting which patients with cancer will see a psychiatrist or counsellor from their initial oncology consultation document using natural language processing. Commun Med 4(1):69
27. Chien S-C, Chang Y-H, Yen C-M, Chen Y-E, Liu C-C, Hsiao Y-P, Yang P-Y, Lin H-M, Lu X-H, Wu I-C et al (2023) Predicting long-term care service demands for cancer patients: a machine learning approach. Cancers 15(18):4598
28. Zhang A, Xing L, Zou J, Wu JC (2022) Shifting machine learning for healthcare from development to deployment and from models to data. Nat Biomed Eng 6(12):1330–1345
29. Li Z, Liu Y (2018) Nursing research methods, People's Medical Publishing House, Beijing, pp 65
30. Wu M (2018) Practical statistical analysis of questionnaires: SPSS operation and application. Chongqing University Press, Chongqing, p 59
31. Boyes A, Girgis A, Lecathelinais C (2009) Brief assessment of adult cancer patients' perceived needs: development and validation of the 34-item Supportive Care Needs Survey (SCNS-sf34). J Evaluation Clinical Pract 15(4):602–606
32. Au A, Lam W, Kwong A, Suen D, Tsang J, Yeo W, Suen J, Ho W, Yau T, Soong I et al (2011) Validation of the Chinese version of the short-form Supportive Care Needs Survey Questionnaire (SCNS-sf34-c). Psycho-Oncol 20(12):1292–1300
33. Zhang Y, Zhang H, Cai H (2023) Unmet supportive care needs of prostate cancer survivors in china: a cross-sectional study. Adv Educ, Human Soc Sci Res 8(1):229–229
34. Neuhaus J, McCulloch C (2011) Generalized linear models. Wiley Interdisciplinary Reviews: Computational Statistics 3(5):407–413
35. Fox J-P, Veen D, Klotzke K (2018) Generalized linear mixed models for randomized responses. Methodology
36. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. IEEE Intell Syst Appl 13(4):18–28
37. Ye Y, Gao J, Shao Y, Li C, Jin Y, Hua X (2020) Robust support vector regression with generic quadratic nonconvex ε-insensitive loss. Appl Math Modell 82:235–251
38. Basak D, Pal S, Patranabis DC et al (2007) Support vector regression. Neural Inf Process-Lett Rev 11(10):203–224
39. Prasath V, Alfeilat HAA, Hassanat A, Lasassmeh O, Tarawneh AS, Alhasanat MB, Salman HSE (2017) Distance and similarity measures effect on the performance of k-nearest neighbor classifier–a review. arXiv:1708.04321
40. Abu Alfeilat HA, Hassanat AB, Lasassmeh O, Tarawneh AS, Alhasanat MB, Eyal Salman HS, Prasath VS (2019) Effects of distance measure choice on k-nearest neighbor classifier performance: a review. Big Data 7(4):221–248
41. Tajmouati S, Wahbi BEL, Bedoui A, Abarda A, Dakkon M (2024). Applying k-nearest neighbors to time series forecasting: Two new approaches. https://doi.org/10.1002/for.3093
42. Bernard S, Heutte L, Adam S (2009) On the selection of decision trees in random forests. In: 2009 International joint conference on neural networks, IEEE, pp 302–307
43. Petkovic D, Altman R, Wong M, Vigil A (2018) Improving the explainability of random forest classifier–user centered approach. In: Pacific symposium on biocomputing. pacific symposium on biocomputing, NIH Public Access, vol. 23, pp 204
44. Breiman L (2001) Random forests. Mach Learn 45:5–32
45. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, Hamprecht FA (2009) A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinf 10:1–16
46. Nembrini S, König IR, Wright MN (2018) The revival of the Gini importance? Bioinformatics 34(21):3711–3718
47. Fitch M (2008) Supportive care framework. Canadian Oncology Nursing Journal/Revue canadienne de soins infirmiers en oncologie 18(1):6–14
48. Zhang T, Hongye H, Qunhui L, Xiaoqing L, Yongxia S, Jingfang H (2019) Supportive care needs of patients with lung cancer in mainland china: a cross-sectional study. J Nursing Res 27(6):52
49. Sarna L, Cooley ME, Brown JK, Williams RD et al (2006) Quality of life and health status of dyads of women with lung cancer and family members. In: Oncology Nursing Forum, Oncology Nursing Society, vol. 33, pp 1109
50. Leiter A, Veluswamy RR, Wisnivesky JP (2023) The global burden of lung cancer: current status and future trends. Nat Rev Clinical Oncol 20(9):624–639
51. Li J, Girgis A (2006) Supportive care needs: are patients with lung cancer a neglected population? Psycho-Oncol: J Psychological. Soc Behavioral Dimensions of Cancer 15(6):509–516
52. Kapoor S, Deppen SA, Paulson AB, Haddad D, Cook JP, Sandler KL (2020) Education level predicts appropriate follow-up of incidental findings from lung cancer screening. J Am College Radiol 17(5):613–619
53. Yun Y, Shon E-J, Yang A-J, Kim S, Kim Y, Chang Y, Lee J, Kim M, Lee H-S, Zo J et al (2013) Needs regarding care and factors associated with unmet needs in disease-free survivors of surgically treated lung cancer. Ann Oncol 24(6):1552–1559

54. Hodgkinson K, Butow P, Hunt G, Wyse R, Hobbs K, Wain G (2007) Life after cancer: couples' and partners' psychological adjustment and supportive care needs. Supportive Care Cancer 15:405–415
55. Mellon S, Northouse LL, Weiss LK (2006) A population-based study of the quality of life of cancer survivors and their family caregivers. Cancer Nursing 29(2):120–131
56. Harrison JD, Young JM, Price MA, Butow PN, Solomon MJ (2009) What are the unmet supportive care needs of people with cancer? a systematic review. Support Care Cancer 17:1117–1128
57. Husain A, Barbera L, Howell D, Moineddin R, Bezjak A, Sussman J (2013) Advanced lung cancer patients' experience with continuity of care and supportive care needs. Supportive Care Cancer 21:1351–1358
58. Yang P, Zhu J, Xiao W, Liu J, Zhou X, Li Q (2016) Supportive care needs of intermediate or advanced cancer patients and its influencing factors. Acad J Chinese Pla Med School 37(5):452–456
59. Farley AC (2014) Health and supportive care needs of surgical lung cancer patients, and the prognostic significance of smoking. PhD thesis, University of Birmingham
60. Lintz K, Moynihan C, Steginga S, Norman A, Eeles R, Huddart R, Dearnaley D, Watson M (2003) Prostate cancer patients' support and psychological care needs: survey from a non-surgical oncology clinic. Psycho-Oncol: J Psychological, Soc Behavioral Dimensions Cancer 12(8):769–783
61. Lam WW, Au AH, Wong JH, Lehmann C, Koch U, Fielding R, Mehnert A (2011) Unmet supportive care needs: a cross-cultural comparison between Hong Kong Chinese and German Caucasian women with breast cancer. Breast Cancer Res Treatment 130:531–541
62. Ji Y, Reiss MJ (2022) Cherish lives? Progress and compromise in sexuality education textbooks produced in contemporary china. Sex Educ 22(4):496–519
63. Zou S, Cao W, Jia Y, Wang Z, Qi X, Shen J, Tang K (2022) Sexual and reproductive health and attitudes towards sex of young adults in china. BMJ Sexual Reproductive Health 48(e1):13–21
64. Evans Webb M, Murray E, Younger ZW, Goodfellow H, Ross J (2021) The supportive care needs of cancer patients: a systematic review. J Cancer Educ 1–10
65. Cabello-Solorzano K, Araujo I, Peña M, Correia L, J. Tallón-Ballesteros A (2023) The impact of data normalization on the accuracy of machine learning algorithms: a comparative analysis. In: International conference on soft computing models in industrial and environmental applications, Springer, ppp 344–353
66. Scott E, Jewell A (2024) Supportive care needs of people with pancreatic cancer: a literature review. Cancer Nursing Pract 23(4)

Authors and Affiliations

Zimeng Chen[1] · Sifan Qiu[1] · Chunmei Luo[1] · Jixiang Zeng[1] · Bifang Huang[1] · Juanxia Zheng[1]

✉ Juanxia Zheng
1209910361@qq.com

Zimeng Chen
zimeng418@foxmail.com

Sifan Qiu
iiauthor@gmail.com

Chunmei Luo
luochunmei@szbazyy.net

Jixiang Zeng
eve_abby@163.com

Bifang Huang
524511878@qq.com

[1] The Seventh Clinical Medicine College of Guangzhou University of Chinese Medicine, Shenzhen, Guangdong 518100, China