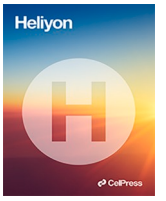
Heliyon 10 (2024) e38422

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Risk prediction model for pneumothorax or pleural effusion after microwave ablation in patients with lung malignancy

Zihang Wang [a,b,1], Yufan Liu [a,b,1], Xiaowen Cao [a,b], Miaoyan Liu [a,b], Li Wang [c,**], Lou Zhong [a,*]

[a] Department of Thoracic Surgery, Affiliated Hospital of Nantong University, Nantong, China
[b] Medical School of Nantong University, Nantong University, NanTong, China
[c] Research Center for Intelligent Information Technology, Nantong University, Nantong, China

ARTICLE INFO

Keywords:
Machine learning
Pneumothorax
Pleural effusion
Microwave ablation
Lung malignancy

ABSTRACT

Background: Although microwave ablation (MWA) has been shown to be an effective treatment for lung malignancies (LM), there is no effective way to predict pneumothorax or pleural effusion after MWA so that timely measures can be taken to prevent it.
Methods: This study comprised LM patients undergoing MWA at Affiliated Hospital of Nantong University from January 2013 to September 2023. Patients before May 2023 constituted the training set (n = 340), while data from May to September served as the test set (n = 58). Unformatted and formatted data extracted from electronic medical records (EMR) were utilized for model construction. Predictors for pneumothorax or pleural effusion were determined through univariate analysis and backward stepwise regression in the training set. Six ML algorithms were employed to create four models based on the research timeframe. Evaluation of the four models was performed using receiver operating characteristic (ROC) analysis, area under the ROC curve (AUC), and 10-fold cross validation.
Findings: A total of 398 patients (216 aged 70 or above, 271 males) were included, with 23.37 % (93/398) experiencing pneumothorax and 33.42 % (133/398) developing pleural effusion. Across all four predictive models, Logistic Regression (LR) demonstrated optimal predictive performance in the test set, with AUC values of 0.727 for Model I, 0.876 for Model II, 0.895 for Model III, and 0.807 for Model IV.
Interpretation: ML models effectively predict post-MWA pneumothorax or pleural effusion.

1. Introduction

Lung cancer is one of the most common cancers and one of the leading causes of cancer deaths worldwide [1]. Surgical resection is the standard treatment for patients with early-stage non-small cell lung cancer (NSCLC) [2]. Unfortunately, approximately 20–40 % of NSCLC patients are not candidates for radical resection due to comorbidities such as cardiopulmonary insufficiency [3] and need to be replaced with localized treatments such as stereotactic body radiation therapy (SBRT), radiofrequency ablation (RFA) or microwave

* Corresponding author.
** Corresponding author.
E-mail addresses: wangli@ntu.edu.cn (L. Wang), tdfyzl@ntu.edu.cn (L. Zhong).
[1] Equal contribution.

https://doi.org/10.1016/j.heliyon.2024.e38422
Received 1 February 2024; Received in revised form 19 August 2024; Accepted 24 September 2024
Available online 29 September 2024

ablation (MWA) [4]. Thermal ablation, as a precise and minimally invasive therapeutic strategy, has been increasingly used to treat early-stage lung cancer. MWA exhibits superior attributes compared to RFA in the context of lung cancer therapy, characterized by expedited ablation durations and the potential to establish larger ablation zones [5]. These inherent advantages contribute to the potential enhancement of treatment efficacy, consequently rendering MWA a prevalent modality in clinical applications.

Pneumothorax is the most common complication of pulmonary ablation, occurring in approximately 30 % of patients. However, more than half of these pneumothorax patients do not require any treatment. Only a minority of patients require chest tube placement and surgical intervention [6,7]. The pneumothorax is not a symptomatic condition, and some patients do not have any symptoms. Some patients with pneumothorax are asymptomatic, but it may worsen over time, affecting lung recuperation and eventually leading to impaired ventilation or even respiratory failure. Pleural effusion is one of the common complications after thermal ablation, with the incidence of pleural effusion ranging from 2.9 to 21.7 % in patients treated with MWA [8,9]. Although most patients with pleural effusion are asymptomatic, persistent pleural effusion may affect the elastic retractive force of the lungs, leading to atelectasis, ventilation dysfunction, hypoxia and even respiratory failure. Li et al. constructed and validated a prediction model for pneumothorax and pleural effusion after MWA in lung cancer patients by nomogram [10]. However, no study has constructed a machine learning model to predict the risk of pneumothorax and pleural effusion after MWA.

In recent years, Artificial Intelligence (AI) has been implemented in the medical and healthcare fields [11]. Machine learning (ML) is an important technology in the field of AI with wide applications in computer science, and its core idea is to continuously improve the model to increase the prediction accuracy based on the data-based training process, while continuously expanding the application areas by using techniques such as deep learning and reinforcement learning. The discipline involves several fields such as statistics, optimization theory, and computer science, and aims to enable computer systems to learn and adapt to different environments in an automated manner in order to better solve problems in practice. Machine learning algorithms create models that can make predictions or decisions without being explicitly programmed to perform the task. The functionality of disease diagnosis is important for its application in cancer-related diagnosis and treatment, as well as for proper retrospective analysis [12,13] ML methods are used to build predictive models that are tested and trained to obtain appropriate algorithmic models for fast and accurate diagnosis, prediction, and monitoring of diseases. ML methods also help physicians to develop treatment plans for patients [14]. The application of ML has recently received much attention in the field of cancer, and several studies have been conducted to predict the prognosis of cancer using AI. Jinya Liu et al. revealed the effect of copper death-related genes on melanoma prognosis by constructing machine

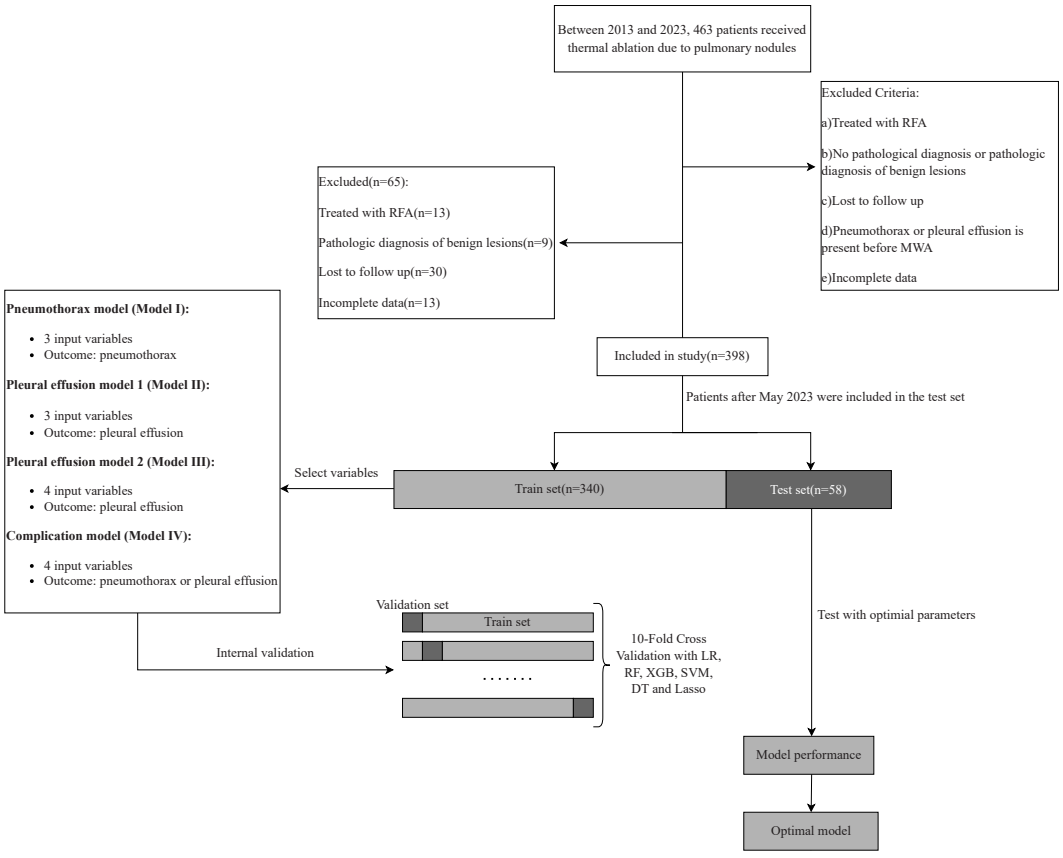


Fig. 1. Flowchart of this study. RFA, Radiofrequency ablation; MWA, Microwave ablation; LR, Logistic regression; RF, Random forest; XGB, Extreme gradient boosting; SVM, Support vector machines; DT, Decision tree; Lasso, Least absolute shrinkage and selection operator.

learning models [15]. Li Chaofan et al. revealed the risk factors of breast cancer brain metastasis (BCBM), constructed a prediction model and identified the potential prognostic factors of BCBM patients through ML methods [16] Fumihiko Kinoshita et al. constructed a prognostic model for patients with NSCLC after surgical resection with higher predictive accuracy compared with traditional TNM staging through ML methods [17]. This shows that ML plays an important role in clinical practice and can provide more efficient and accurate guidance for clinical diagnosis and treatment.

In this investigation, we employed Logistic Regression (LR), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Support Vector postoperative pneumothorax or pleural effusion in LM patients undergoing Machines (SVM), Decision Tree (DT), and Least absolute shrinkage and selection operator (Lasso) as ML models to forecast the occurrence ofMWA. Our objective was to construct an improved predictive model for AI. The performance of the four algorithms was assessed through receiver operating characteristic (ROC) analysis, area under the ROC curve (AUC) and 10-fold cross validation. Notably, LR exhibited superior predictive capabilities across all models in the test set. Specifically, LR demonstrated AUC values of 0.727, 0.876, 0.895, and 0.807 for Model I, Model II, Model III and Model IV, respectively.

2. Methods

2.1. Patient criteria

Patient data originated from Affiliated Hospital of Nantong University, gathered in adherence to the established guidelines sanctioned by the institution. The study received ethical approval from the Ethical Review Committee of Affiliated Hospital of Nantong University, aligning with the principles outlined in the Declaration of Helsinki. Given the retrospective nature of this study, informed consent was waived. Our inclusion criteria involved screening individuals diagnosed or suspected of LM who underwent MWA at Affiliated Hospital of Nantong University between January 2013 and September 2023. A total of 398 eligible patients were identified, with 340 patients before May 2023 forming the training cohort and the remaining 58 patients from May to September constituting the test cohort (Fig. 1). This study included patients meeting the following criteria: (a) confirmed or suspected cases of LM treated with MWA, and (b) individuals with a definitive histopathological examination confirming the malignancy diagnosis. Exclusion criteria encompassed: (a) treated with RFA; (b) no pathological diagnosis or pathologic diagnosis of benign lesions; (c) loss to follow up; (d) pneumothorax or pleural effusion is present before MWA; (e) incomplete data.

2.2. Diagnosis and preoperative preparation

The diagnosis of pneumothorax, pleural effusion, and pulmonary emphysema relied on computed tomography (CT) scans, interpreted by two radiologists with over 5 years of experience each. Prior to MWA, all LM patients underwent chest enhanced CT scan, which were reviewed by two radiologists, each possessing over 5 years of experience in thoracic radiology. The assessments aimed to evaluate tumor location, size, density, presence of lymph nodes, and distant metastasis. All laboratory tests were conducted within 1–3

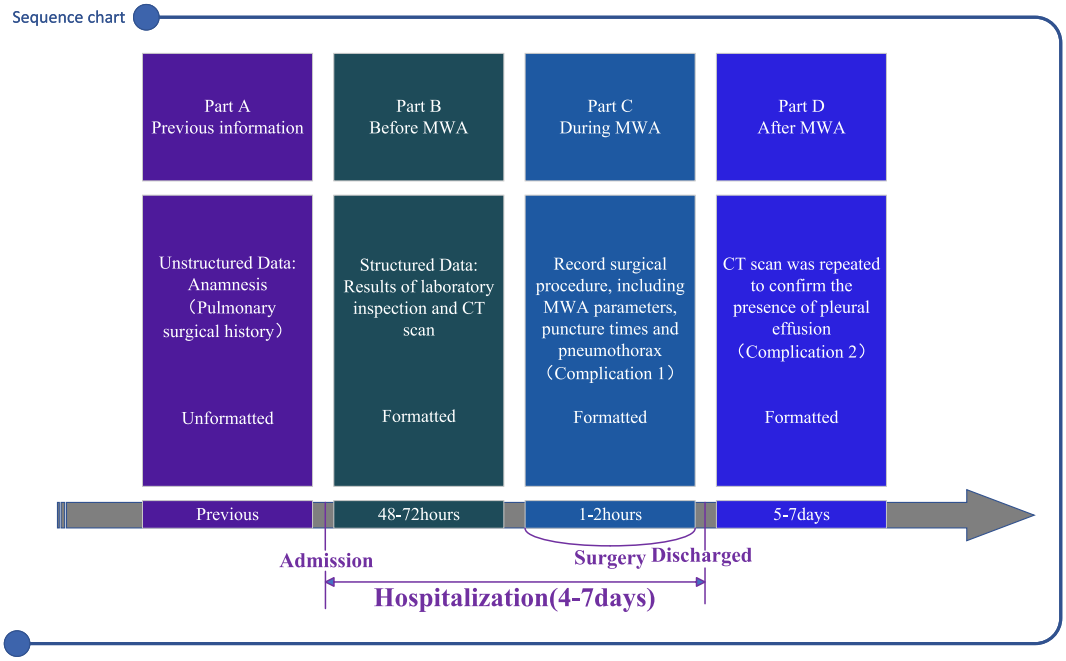


Fig. 2. Sequence chart of MWA. MWA, Microwave ablation.

days before MWA.

2.3. MWA and postoperative follow-up

All MWA procedures were carried out by two experienced radiologists with over 5 years of CT-guided interventional expertise. Utilizing the MTC-3C microwave therapy device (Vison Medical) operating at 2450 ± 10 % MHz with a maximum output power of 100W, a water-cooled 18G (diameter) × 180 mm (effective length) microwave needle was employed for CT-guided MWA. Treatment plans, including patient positioning, puncture sites, optimal needle trajectories, and the number of ablation needles, were determined through CT examinations. Patients were positioned according to tumor location, and local anesthesia using lidocaine was administered. The ablation needle was gradually inserted along the planned trajectory, confirmed by CT, for tumor ablation with adjustments to power and duration as needed. Surgery concluded with a 5–10 mm margin beyond the lesion boundary in the ablation zone. Immediate post-ablation chest CT scans assessed the ablation zone and detected pneumothorax. Chest tubes were inserted for moderate to severe pneumothorax, and the procedure ended upon lung surface re-expansion. For patients without histopathological examination, synchronous coaxial sheath biopsy was performed if necessary. Follow-up outpatient visits with CT scans were conducted 5–7 days after MWA to confirm lesion ablation status and check for pleural effusion.

2.4. Research timeframe

Our study involved a hospitalization period of 4–7 days for patients (Fig. 2). On the first day of admission, unformatted data, encompassing current and past medical history, and personal information from the Electronic Medical Record (EMR), was collected (Part A). This unformatted data, including various formats like office documents, text, images, XML, and HTML, poses challenges in organization and formatting [18]. Within 48–72 h before surgery, laboratory tests and enhanced CT scans were performed to gather clinically relevant information, defined as formatted data (Part B). Formatted data refers to information represented in relational databases, organized in two-dimensional tables with consistent attributes and tuples [19]. The MWA procedure, lasting 1–2 h, included an intraoperative CT scan to confirm pneumothorax, classified as complication 1. Moderately severe pneumothorax was promptly treated with closed chest drainage, while mild cases were observed without intervention. After MWA, intraoperative parameters (Part C) such as ablation duration, maximum power, and puncture times were recorded as formatted data. Follow-up included outpatient CT scans 1–3 days after surgery and 5–7 days after discharge to confirm pleural effusion, defined as complication 2 (Part D). The study aimed to establish eleven prediction models based on data type, including four using formatted data, three using unformatted data, and four combining both. By comparing the performance of these models, we identified the optimal model for predicting postoperative complications in LM patients undergoing MWA.

2.5. Data preprocessing

In this study, we employed Named Entity Recognition (NER) in Python (version 3.7) to extract unformatted data, capturing pertinent medical history and symptoms from EMR. The NER system utilized the LERT-BiLSTM-CRF model, combining the language-informed LERT pre-training model with BiLSTM and CRF layers. LERT employed a pre-training strategy involving three types of linguistic features (word position in the sentence, NER label, and word lexicality) [20]. BiLSTM utilized the vector representation of LERT to extract richer contextual features, and the CRF layer determined the label for each word based on these contextual features. To address the challenge of excessively long sentences in electronic medical records, which could hinder training efficacy, we segmented sentences exceeding 128 characters into more manageable units. Subsequently, we conducted masked language model training tasks to enhance the efficiency of named entity recognition. Extensive experiments were conducted on medical NER tasks across multiple hospitals. For real EMR with varying formats and insufficient standardization, we standardized input sentences by segmenting them based on punctuation. Presently, the optimal test set achieved an F1 value of 90.67 % and an Accuracy value of 94.44 %. Variables with more than 20 % missing data were excluded, and missing values were imputed using the mean substitution method. Through NER, we extracted a total of eight unformatted data elements, including patient symptoms such as cough and choke, as well as past medical histories such as hypertension, emphysema, bullae, pneumonia, chemotherapy history, and lung surgery history.

2.6. Machine learning methods

Logistic Regression (LR) is a statistical modeling technique commonly used for binary classification tasks. It analyzes the relationship between a dependent binary variable and one or more independent variables by estimating the probability of occurrence of an event [21]. LR is widely employed in various fields, such as medicine and ML, due to its simplicity, interpretability, and ability to provide probabilistic predictions.

Random Forest (RF) is an integrated learning method that consists of multiple decision trees. Each tree is trained based on randomly selected features and samples, and then predicted by voting or averaging. This method can effectively reduce the risk of overfitting, and also has good performance when dealing with large-scale data [22].

Extreme Gradient Boosting (XGBoost) is a powerful ML algorithm known for its efficiency and accuracy, particularly in formatted/tabular data scenarios. It belongs to the gradient boosting family and utilizes an ensemble of decision trees to make predictions [23]. XGBoost is designed to optimize both computational speed and model performance, incorporating techniques like regularization, parallel processing, and tree pruning. It has become a popular choice in various competitions and real-world applications for its ability

to handle complex relationships within data and avoid overfitting.

Support Vector Machine (SVM) is a ML algorithm commonly used for classification and regression tasks. SVM works by finding a hyperplane that best separates data points into different classes, maximizing the margin between the classes. It is effective in high-dimensional spaces and is versatile in handling linear and non-linear relationships. SVM can employ different kernel functions to transform the input data into higher-dimensional spaces, allowing it to capture complex patterns and relationships [24]. SVM has been widely applied in various fields, including image recognition, text classification, and bioinformatics, due to its robustness and generalization capabilities.

Decision Tree (DT) is a popular supervised learning algorithm used for both classification and regression tasks. It models decisions and their possible consequences as a tree-like structure [25]. Each internal node represents a decision based on an attribute, each branch represents the outcome of a decision, and each leaf node represents a final decision or classification. Decision Trees are easy to interpret, as they provide a clear and visual representation of decision-making paths. However, they can be prone to overfitting, especially with complex datasets, which can be mitigated by techniques like pruning.

Least Absolute Shrinkage and Selection Operator (Lasso) is a type of linear regression that incorporates regularization to prevent overfitting by adding a penalty equal to the absolute value of the magnitude of coefficients [26]. This penalty helps shrink some coefficients to zero, effectively performing variable selection. Lasso is particularly useful when dealing with high-dimensional data, where it can simplify the model by selecting only the most important features. This makes it a valuable tool for feature selection in predictive modeling.

2.7. Statistical analysis

Categorical variables were delineated by frequencies and percentages, while continuous variables were expressed as mean ± standard deviation (SD). IBM SPSS version 27.0 was employed for statistical analyses. In the training set, univariate logistic regression was initially conducted, and variables with $p < 0.05$ were chosen for subsequent inclusion in logistic backward stepwise regression for multivariate analysis. The final model incorporated variables with the smallest Akaike Information Criterion (AIC) values, where a lower AIC signifies a more parsimonious model, indicative of superior fit [27]. The study identified potential predictors of pneumothorax and pleural effusion, encompassing demographics, tests, ablation factors, and radiologic features across 22 parameters. Six ML algorithms (LR, RF, XGB, SVM, DT and Lasso) were employed to predict the risk of pneumothorax and pleural effusion following MWA in LM patients. The R Programming Language (version 4.3.0) facilitated the implementation of the ML modeling process. After constructing the prediction models, 10-fold cross validation and assessment of the area under the ROC curve (AUC) in the test set were conducted to verify the reliability of the models.

Table 1
Data resource and types in this study.

Data resource	Data type	Variable
Part A	Unformatted	Cough
		Choke
		Hypertension
		Bullae
		Pneumonia
		History of chemotherapy
		Emphysema
		History of lung surgery
Part B	Formatted	Sex
		Age
		Diameter
		Site
		Laterality
		D-Dimer
		Albumin
		WBC
		RBC
		Hb
		PLT
Part C	Formatted	Maximum power
		Ablation duration
		Puncture times
Part D	Formatted	Pneumothorax (complication 1)
		Pleural effusion (complication 2)

*Complication.
WBC, White Blood Cell Count; RBC, Red Blood Cell Count; Hb, Hemoglobin; PLT, Platelet Count.
*Complication includes pneumothorax or pleural effusion.

3. Results

3.1. Baseline characteristics

A total of 398 LM patients treated with MWA were included in this study. Of these, 340 patients were assigned to the training set, while 58 patients were assigned to the test set. We extracted eight unformatted data using NER. All the data and their types in this study are shown in Table 1. The baseline characteristics of the two sets is shown in Table 2. Significant differences were observed between the two sets in terms of tumor size, maximum power, ablation duration, pneumothorax, pleural effusion, D-dimer, albumin, hemoglobin, choke, bullae, history of chemotherapy and history of lung surgery ($p < 0.05$). Tumor size, maximum power, and ablation duration were significantly smaller in the test set than in the training group, and there were more patients with abnormal D-dimer, lung surgery history and bullae. In addition, in the training set, a lower percentage of patients had pneumothorax or pleural effusion, but more patients had abnormal albumin, hemoglobin and choke. The remaining variables did not differ significantly between the two sets. Pneumothorax occurred in 93 patients (23.4 %) and pleural effusion in 133 patients (33.4 %) after MWA, for a total of 176 patients with complications of pneumothorax or pleural effusion (44.2 %). The baseline tables for the grouping of pneumothorax or pleural effusion are shown in Supplement Table S1, S2, and S3.

3.2. Variable filtering

Initially, univariate logistic regression was conducted, and variables with $p < 0.05$ were selected for inclusion in the logistic backward stepwise regression for multivariate analysis. The final model incorporated variables with the smallest AIC values. A lower AIC signifies a more parsimonious model, indicating superior performance with fewer inputs [27]. We defined the pneumothorax prediction model as Model I, the pleural effusion prediction model (ABC→Pleural effusion) as Model II, the pleural effusion prediction model (ABC + Pneumothorax→Pleural effusion) as Model III, and the pneumothorax or pleural effusion prediction model as Model IV. For formatted data in Model I, variables screened for minimum AIC included sex, age, and puncture times. In Model II, age, ablation duration, and puncture times were considered. Model III involved age, ablation duration, pneumothorax, and puncture times. Model IV included age, sex, and puncture times. Unformatted data for Model I considered emphysema and cough. While no suitable unformatted data variables were identified for Model II and III. Model IV included emphysema. The combined formatted and unformatted data screening resulted in Model I with variables: emphysema, cough, and puncture times. Model II included age, ablation duration, and puncture times. Model III considered age, ablation duration, pneumothorax, and puncture times. Model IV involved age, sex, puncture times, and emphysema (Table 3, Supplement Table S4-S7).

3.3. Performance of machine learning algorithms

Pneumothorax and pleural effusion served as outcome measures, with identified factors designated as variables in the respective prediction models. Various ML algorithms (LR, RF, XGB, SVM, DT, Lasso) were applied to the training set, validated through 10-fold cross validation, and performance assessed (Supplement Table S8, Supplement Figure S1). The 10-fold cross validation AUC_{mean} for each model is presented in Table 4 and Supplement Figure S2. In training set, RF performed best for Model I using formatted data ($AUC_{mean} = 0.673$); RF using unformatted data was most effective for Model I ($AUC_{mean} = 0.652$); and LR outperformed others for Model I when using a combination of formatted and unformatted data ($AUC_{mean} = 0.724$). Models built solely with formatted or unformatted data performed less effectively. Thus, the final Model I was constructed using a combination of formatted and unformatted data. In test set, LR maintained superiority among the six ML algorithms in ROC curve analysis (AUC = 0.727), solidifying its selection as the final Model I. The AUC values for the six ML algorithms of each model in the test set are presented in Table 5 and Supplement Figure S3. Similarly, for Model II, three formatted data variables (age, ablation duration, puncture times) were identified, and unformatted data did not yield significant factors. Considering the sequential occurrence of pneumothorax and pleural effusion, pneumothorax was analyzed as a potential factor, revealing a significant correlation. Consequently, two pleural effusion prediction models were developed: Model II, without considering pneumothorax, suitable for preoperative prediction, and Model III, considering pneumothorax, suitable for immediate postoperative prediction. In training set, LR performed best for both Model II ($AUC_{mean} = 0.650$) and Model III ($AUC_{mean} = 0.690$). In test set, LR of Model II displayed superior ROC curve analysis (AUC = 0.876), and LR of Model III also outperformed other ML algorithms (AUC = 0.895). Model III demonstrated superior predictive performance compared to Model II. Therefore, immediate postoperative prediction of pleural effusion with Model III is recommended, although preoperative prediction with Model II remains effective. For Model IV, LR using both formatted and unformatted data performed best in training set ($AUC_{mean} = 0.671$). In test group, LR maintained superiority among the six ML algorithms in ROC curve analysis (AUC = 0.807). Hence, LR was chosen as the final predictor for Model IV.

3.4. Model development and validation

Based on the findings from the aforementioned study, we constructed four nomograms, each dedicated to predicting specific outcomes: pneumothorax, pleural effusion and the occurrence of either pneumothorax or pleural effusion (Fig. 3). These nomograms employed a cumulative scoring system based on individual parameter scores obtained before and/or during MWA to ascertain the likelihood and percentage risk of pneumothorax or pleural effusion. Calibration curves demonstrated consistent accuracy, affirming the reliability of the nomogram in predicting post-MWA pneumothorax or pleural effusion (Fig. 4). Subsequently, the model

Table 2
Patient demographics and baseline characteristics.

Variable	Total (n = 398)	Train set (n = 340)	Test set (n = 58)	Statistic	P
Sex, n (%)				$\chi^2 = 0.207$	0.649
Female	127 (31.91)	107 (31.47)	20 (34.48)		
Male	271 (68.09)	233 (68.53)	38 (65.52)		
Age, n (%)				$\chi^2 = 1.630$	0.202
<70	182 (45.73)	151 (44.41)	31 (53.45)		
≥70	216 (54.27)	189 (55.59)	27 (46.55)		
Tumor size, n (%)				$\chi^2 = 21.012$	<0.001
<2.5 cm	205 (51.51)	159 (46.76)	46 (79.31)		
≥2.5 cm	193 (48.49)	181 (53.24)	12 (20.69)		
Maximum power, n (%)				$\chi^2 = 49.813$	<0.001
<50W	174 (43.72)	124 (36.47)	50 (86.21)		
≥50W	224 (56.28)	216 (63.53)	8 (13.79)		
Ablation duration, n (%)				$\chi^2 = 10.544$	0.001
<6min	176 (44.22)	139 (40.88)	37 (63.79)		
≥6min	222 (55.78)	201 (59.12)	21 (36.21)		
Puncture times, n (%)				–	0.380
1	211 (53.02)	181 (53.24)	30 (51.72)		
2	162 (40.7)	140 (41.18)	22 (37.93)		
3	18 (4.52)	13 (3.82)	5 (8.62)		
4	7 (1.76)	6 (1.76)	1 (1.72)		
Pneumothorax, n (%)				$\chi^2 = 8.042$	0.005
No	305 (76.63)	269 (79.12)	36 (62.07)		
Yes	93 (23.37)	71 (20.88)	22 (37.93)		
Pleural effusion, n (%)				$\chi^2 = 3.973$	0.046
No	265 (66.58)	233 (68.53)	32 (55.17)		
Yes	133 (33.42)	107 (31.47)	26 (44.83)		
[a]Complication, n (%)				$\chi^2 = 2.344$	0.126
No	222 (55.78)	195 (57.35)	27 (46.55)		
Yes	176 (44.22)	145 (42.65)	31 (53.45)		
Site, n (%)				$\chi^2 = 1.689$	0.194
Middle and lower lobe	168 (42.21)	139 (40.88)	29 (50.00)		
Upper lobe	230 (57.79)	201 (59.12)	29 (50.00)		
Laterality, n (%)				$\chi^2 = 2.410$	0.121
Left	189 (47.49)	156 (45.88)	33 (56.90)		
Right	209 (52.51)	184 (54.12)	25 (43.10)		
D-Dimer, n (%)				$\chi^2 = 67.695$	<0.001
Abnormal	97 (24.37)	58 (17.06)	39 (67.24)		
Normal	301 (75.63)	282 (82.94)	19 (32.76)		
Albumin, n (%)				$\chi^2 = 5.457$	0.019
Abnormal	88 (22.11)	82 (24.12)	6 (10.34)		
Normal	310 (77.89)	258 (75.88)	52 (89.66)		
WBC, n (%)				$\chi^2 = 0.290$	0.590
Abnormal	79 (19.85)	69 (20.29)	10 (17.24)		
Normal	319 (80.15)	271 (79.71)	48 (82.76)		
RBC, n (%)				$\chi^2 = 0.333$	0.564
Abnormal	87 (21.86)	76 (22.35)	11 (18.97)		
Normal	311 (78.14)	264 (77.65)	47 (81.03)		
Hb, n (%)				$\chi^2 = 12.826$	<0.001
Abnormal	129 (32.41)	122 (35.88)	7 (12.07)		
Normal	269 (67.59)	218 (64.12)	51 (87.93)		
PLT, n (%)				$\chi^2 = 1.933$	0.164
Abnormal	41 (10.3)	38 (11.18)	3 (5.17)		
Normal	357 (89.7)	302 (88.82)	55 (94.83)		
Cough, n (%)				$\chi^2 = 1.397$	0.237
No	267 (67.09)	232 (68.24)	35 (60.34)		
Yes	131 (32.91)	108 (31.76)	23 (39.66)		
Choke, n (%)				$\chi^2 = 8.781$	0.003
No	341 (85.68)	284 (83.53)	57 (98.28)		
Yes	57 (14.32)	56 (16.47)	1 (1.72)		
Bullae, n (%)				$\chi^2 = 5.636$	0.018
No	320 (80.4)	280 (82.35)	40 (68.97)		
Yes	78 (19.6)	60 (17.65)	18 (31.03)		
Pneumonia, n (%)				$\chi^2 = 0.969$	0.325
No	273 (68.59)	230 (67.65)	43 (74.14)		
Yes	125 (31.41)	110 (32.35)	15 (25.86)		
History of chemotherapy, n (%)				$\chi^2 = 5.135$	0.023
No	348 (87.44)	292 (85.88)	56 (96.55)		
Yes	50 (12.56)	48 (14.12)	2 (3.45)		

(*continued on next page*)

Table 2 (*continued*)

Variable	Total (n = 398)	Train set (n = 340)	Test set (n = 58)	Statistic	P
Emphysema, n (%)				$\chi^2 = 1.568$	0.210
No	262 (65.83)	228 (67.06)	34 (58.62)		
Yes	136 (34.17)	112 (32.94)	24 (41.38)		
History of lung surgery, n (%)				$\chi^2 = 8.085$	0.004
No	282 (70.85)	250 (73.53)	32 (55.17)		
Yes	116 (29.15)	90 (26.47)	26 (44.83)		

WBC, White Blood Cell Count; RBC, Red Blood Cell Count; Hb, Hemoglobin; PLT, Platelet Count.
[a] Complication: pneumothorax or Pleural effusion.

Table 3
Definition of the four models and their final predictors included in this research.

Model	Definition	Outcome	Predictors
Pneumothorax prediction model (ABC→Pneumothorax)	Model I	Pneumothorax	Puncture times Emphysema Cough
Pleural effusion prediction model (ABC→Pleural effusion)	Model II	Pleural effusion	Age Puncture times Ablation duration
Pleural effusion prediction model (ABC + Pneumothorax→Pleural effusion)	Model III	Pleural effusion	Age Puncture times Ablation duration Pneumothorax
Pneumothorax or pleural effusion prediction model (ABC→Pneumothorax or Pleural effusion)	Model IV	Pneumothorax or Pleural effusion	Age Sex Puncture times Emphysema

Table 4
AUC_{mean} of 10-Fold Cross Validation with six ML algorithms in four prediction models.

Method	Model I	Model II	Model III	Model IV
LR	0.724	0.650	0.690	0.671
RF	0.714	0.628	0.667	0.645
XGB	0.713	0.620	0.656	0.634
SVM	0.630	0.615	0.616	0.637
DT	0.640	0.590	0.634	0.594
Lasso	0.693	0.577	0.603	0.613

LR, Logistic regression; RF, Random forest; XGB, Extreme gradient boosting; SVM, Support vector machines; DT, Decision tree; Lasso, Least absolute shrinkage and selection operator.

Table 5
AUC value in test set of six ML algorithms in four prediction models.

Method	Model I	Model II	Model III	Model IV
LR	0.727	0.876	0.895	0.807
RF	0.679	0.729	0.827	0.728
XGB	0.688	0.778	0.808	0.744
SVM	0.594	0.598	0.713	0.621
DT	0.703	0.856	0.858	0.661
Lasso	0.670	0.838	0.872	0.762

LR, Logistic regression; RF, Random forest; XGB, Extreme gradient boosting; SVM, Support vector machines; DT, Decision tree; Lasso, Least absolute shrinkage and selection operator.

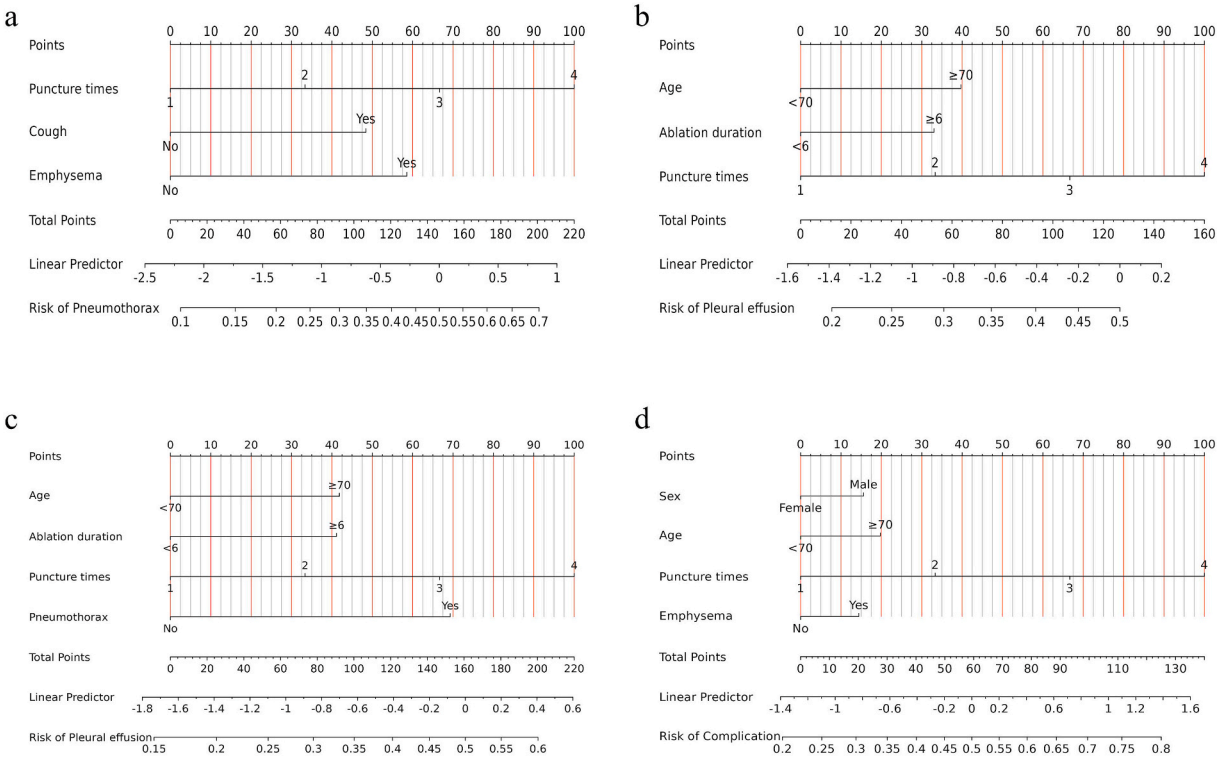


Fig. 3. (a) Nomogram used to predict pneumothorax after MWA. (b) Nomogram used to predict pleural effusion after MWA preoperatively. (c) Nomogram used to predict pleural effusion after MWA in the immediate postoperative period. (d) Nomogram used to predict pneumothorax or pleural effusion after MWA. To use the nomograms, individual patient values are located on the axis of each variable, and a line is drawn upward to identify the number of points obtained for each variable. The sum of these points is located on the total points axis and a line is plotted down to the risk axis to determine the likelihood of developing a pneumothorax or pleural effusion.

underwent further analysis using Decision Curve Analysis (DCA). The DCA curves highlighted the model's superiority over models involving either all treated or untreated patients (Fig. 5).

3.5. Evaluation of model performance

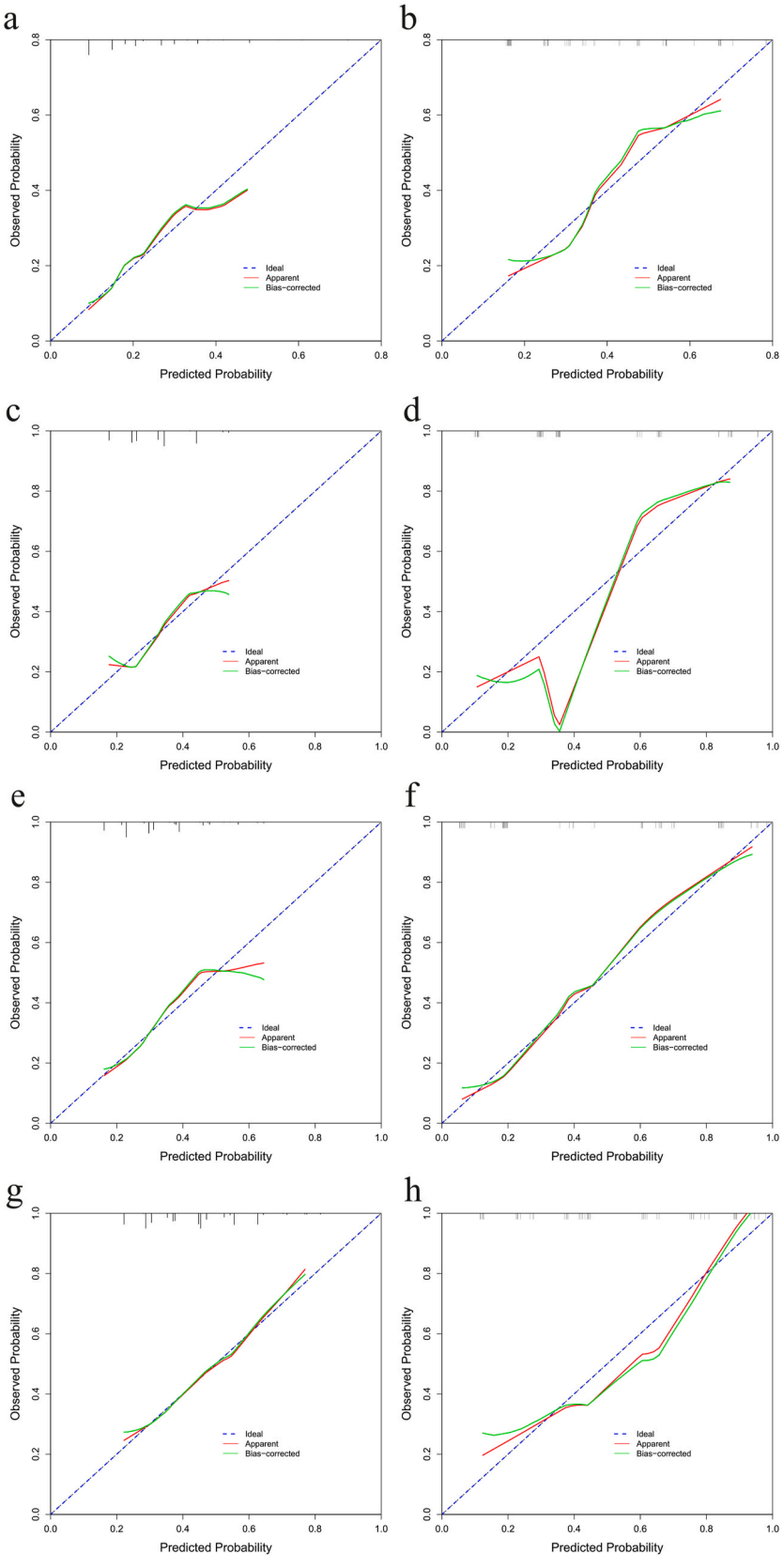
Based on the maximum Youden index, the optimal cutoff value for Model I was determined to be 0.25, for Model II 0.3, for Model III 0.3, and for Model IV 0.5. This indicates that, during preoperative assessment, if the risk of pneumothorax exceeds 25 %, there is a significant likelihood that the patient will develop pneumothorax during the procedure. Similarly, if the risk of pleural effusion exceeds 30 %, there is a high probability that the patient will develop pleural effusion postoperatively. The confusion matrices for each model are shown in Fig. 6, and the model performance metrics are presented in Table 6. Model I, II, III, and IV all demonstrated high sensitivity, specificity, and F1 scores.

Additionally, this study found that the false negative rates for Model I and III were 0.23 and 0.19, respectively, indicating a low rate of missed diagnoses when using Model I and Model III to predict pneumothorax and pleural effusion, respectively. Since pneumothorax and pleural effusion are common complications following MWA, failure to detect and address these conditions promptly may lead to worsening over time, impairing lung re-expansion and potentially resulting in ventilatory dysfunction or even respiratory failure. Therefore, Model I and III can assist clinicians in early warning and detection of pneumothorax and pleural effusion through imaging, reducing the likelihood of missed diagnoses and preventing adverse postoperative events due to delayed follow-up.

Furthermore, the false positive rates for the four models were 0.11, 0.12, 0.19, and 0.04, respectively, indicating a low misdiagnosis rate. This suggests that the models rarely predict pneumothorax or pleural effusion in patients who do not actually have these conditions, thereby avoiding unnecessary treatments such as the injection of porcine fibrin glue and more frequent imaging follow-up, ultimately reducing the patient's additional medical costs and psychological burden.

3.6. Impact of variables on prediction performance

We investigated the importance of factors in LR algorithm across four distinct prediction models. In Model I, the factors were ranked in the following order of importance: Emphysema, Puncture times, and Cough, with Emphysema serving as the most crucial predictor (Fig. 7a). For Model II, the relative importance of variables was ranked as follows: Puncture times > Age > Ablation duration



(caption on next page)

Fig. 4. Calibration plots of four nomograms. (a) calibration plot of Model I using train set; (b) calibration plot of Model I using test set; (c) calibration plot of Model II using train set; (d) calibration plot of Model II using test set; (e) calibration plot of Model III using train set; (f) calibration plot of Model III using test set; (g) calibration plot of Model IV using train set; (h) calibration plot of Model IV using test set.

(Fig. 7b). Model III exhibited the following relative importance of variables: Pneumothorax > Puncture times > Age > Ablation duration (Fig. 7c). In model IV, the order of importance for factors was determined as Puncture times > Age > Sex > Emphysema (Fig. 7d).

3.7. Clinical application

Our team has already applied the research findings in clinical practice. During preoperative assessments, if a patient's risk of pneumothorax exceeds 25 %, we prepare a chest drainage tube and porcine fibrin glue in advance. If a minor pneumothorax occurs during the procedure, with lung compression less than 20 %, we continue to observe without any intervention. However, if a moderate to massive pneumothorax occurs, with lung compression of 20 % or more, we immediately place a chest drainage tube. At the end of the procedure, porcine fibrin glue is injected through a cannula into the puncture tract to seal the lung and puncture site, preventing the occurrence or progression of pneumothorax. If a patient's risk of pleural effusion exceeds 30 %, we perform a chest X-ray on the day after the procedure and during the first follow-up after discharge to confirm whether pleural effusion has developed. If significant pleural effusion is detected, we perform an ultrasound-guided thoracentesis to drain the fluid, significantly reducing the risk of postoperative complications such as pulmonary infection and atelectasis. Preoperative assessment using the nomogram developed in this study allows us to rapidly identify high-risk patients and create individualized MWA treatment plans and postoperative care strategies, helping patients recover quickly and reducing hospital stay duration.

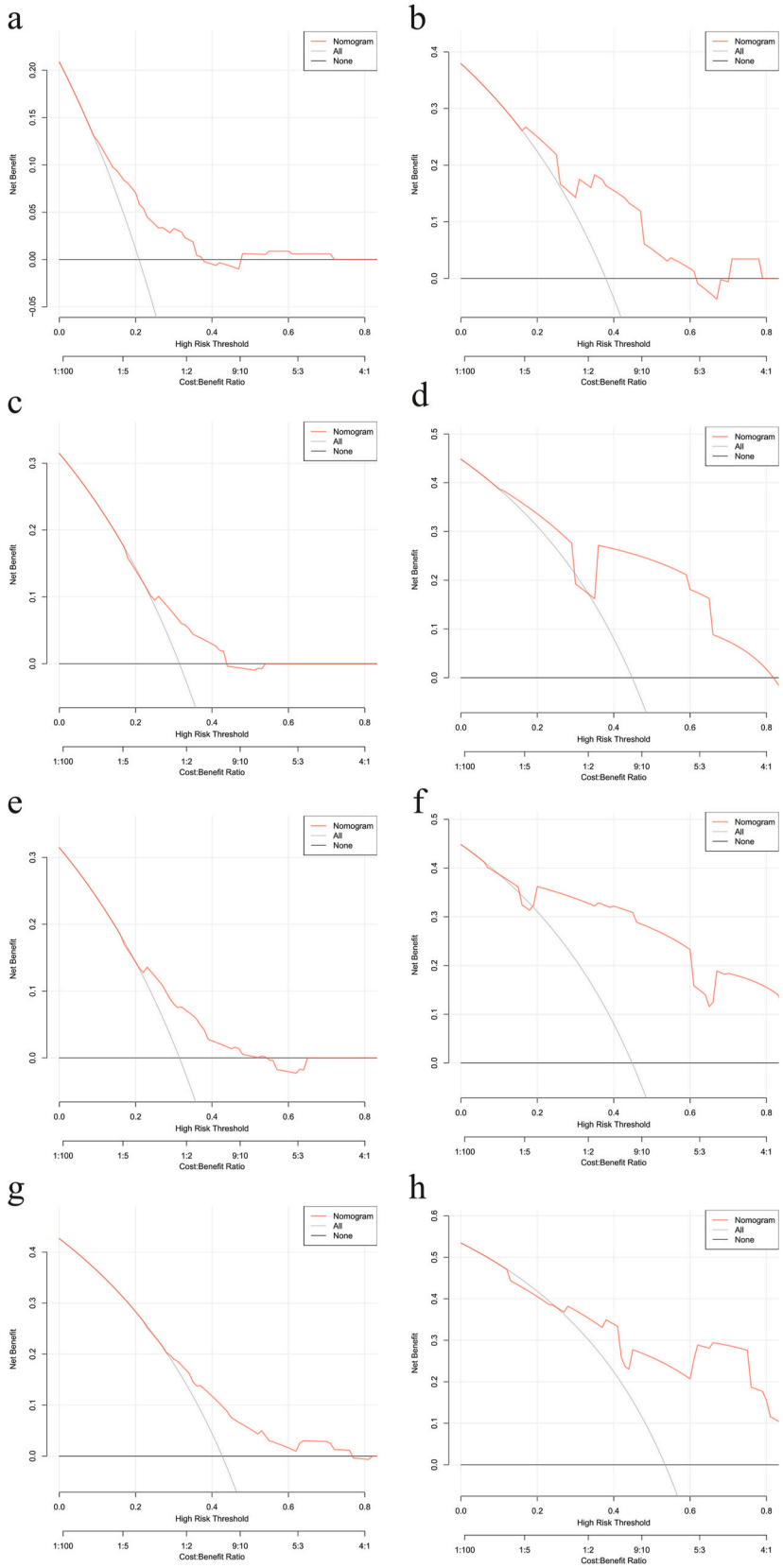
Below is a case example. During the preoperative assessment, the patient's total score using the pneumothorax nomogram developed in this study was 140, indicating a 48 % risk of pneumothorax, with a high likelihood of intraoperative occurrence. As shown in Fig. 8a, intraoperative CT confirmed the preoperative assessment, and the patient developed pneumothorax during the procedure. After the procedure, we injected porcine fibrin glue through a cannula and placed a chest drainage tube, as shown in Fig. 8b. On the first postoperative day, no more air was detected in the chest drainage tube, and the water column fluctuation ceased, so the drainage tube was removed, and the patient was discharged on the second postoperative day. Using the two pleural effusion nomograms developed in this study, the patient's total scores were 106 and 185, indicating risks of 44 % and 56 %, respectively, both suggesting a high likelihood of postoperative pleural effusion. A chest X-ray one week after the procedure confirmed these findings, showing a small amount of left-sided pleural effusion, as seen in Fig. 8c. This case demonstrates that preoperative assessment using the nomograms developed in this study can help us quickly identify high-risk patients, create individualized MWA treatment plans and postoperative care strategies, and reduce the patient's hospital stay from the original 7 days–5 days.

4. Discussion

Thermal ablation stands as an effective therapeutic strategy for patients with LM, exhibiting 3- and 5-year survival rates of 35–80 % and 25–60 %, respectively, following RFA in non-small cell lung cancer patients [28,29]. While numerous predictive models for malignancies rely on multivariate regression or correlation analysis, often necessitating independent and linear variables [30,31], we expanded our approach beyond traditional methods. In addition to conventional univariate and multivariate analyses, we employed various ML algorithms, widely applied in healthcare data analysis, to construct predictive models for postoperative pneumothorax or pleural effusion following MWA in LM patients. Our investigation revealed LR model's superior performance, it performed best among the four models. This superior performance may be attributed to LR's ability to provide stable and rapid models when data is relatively small and features are limited. Additionally, while Models II and III did not identify unformatted factors related to pleural effusion, Models I and IV, which combined formatted and unformatted data, achieved better results. This improvement is likely due to the use of NER for extracting unformatted data, allowing for the inclusion of more factors in the model. Ultimately, we identified LR model as the optimal predictor for postoperative pneumothorax and pleural effusion in LM patients undergoing MWA.

Significantly, our study distinguished between unformatted and formatted data, revealing that the unformatted data variables, emphysema and cough emerged as substantial predictors in Model I. To enhance clarity and intuitiveness, we employed a sequence chart for the four prediction models, a novel approach not previously reported in the literature. In Model III, pneumothorax was incorporated as a potential factor based on the sequence of events, and univariate and multivariate analyses confirmed a significant association between pneumothorax and pleural effusion. The importance analysis of Model III highlighted pneumothorax as the foremost factor, indicating a robust association. Consequently, our study underscores a strong link between pneumothorax and the development of pleural effusion post-MWA.

Prior investigations have indicated a significant correlation between multiple predictive factors and pneumothorax [32–35]. Emphysema is characterized by a permanent enlargement of airspaces distal to terminal fine bronchioles, accompanied by the destruction of alveolar walls [36]. Our study highlights emphysema as a pivotal predictive factor for post-MWA pneumothorax. Among 136 patients with pulmonary emphysema, pneumothorax occurred in 50 cases (50/136 36.7 %) following MWA, and univariate analysis underscored the substantial association between emphysema and pneumothorax. Furthermore, within LR model, emphysema ranked as the most critical among the three factors. The underlying mechanisms are as follows: First, emphysema patients often develop pulmonary bullae, which are air-filled cystic structures within the lungs. These bullae have thin and fragile walls, and when



(caption on next page)

Fig. 5. Decision curve analysis (DCA) of four nomograms. (a) DCA of Model I using train set; (b) DCA of Model I using test set; (c) DCA of Model II using train set; (d) DCA of Model II using test set; (e) DCA of Model III using train set; (f) DCA of Model III using test set; (g) DCA of Model IV using train set; (h) DCA of Model IV using test set.

punctured by a needle, the likelihood of pneumothorax increases. Second, emphysema is a form of chronic obstructive pulmonary disease (COPD) characterized by destruction of alveolar walls and enlargement of alveolar spaces. These pathological changes reduce the elasticity and increase the brittleness of lung tissue, making it more prone to rupture and resulting in pneumothorax. Third, chronic inflammation and tissue repair processes in emphysema patients lead to alterations in the pleural structure, making it irregular and less elastic, thereby increasing the risk of pleural rupture and pneumothorax.

Moreover, our investigation revealed a robust correlation between the number of surgical punctures and both pneumothorax and pleural effusion. Among 187 patients with two or more puncture times, 29.4 % (55/187) experienced pneumothorax, and 45.4 % (85/187) developed pleural effusion post-MWA. Puncture times emerged as a highly significant predictor across all four prediction models, consistently ranking among the top two in importance for each model. The potential mechanisms include: During MWA, the ablation needle creates a larger pleural puncture, which may be further enlarged by lung movement induced by breathing. Increased puncture frequency leads to a higher number of pleural defects and their enlargement, thereby raising the risk of pneumothorax. Additionally, each puncture operation can rupture lung tissue or the visceral pleura, causing air to leak from the alveoli into the pleural cavity,

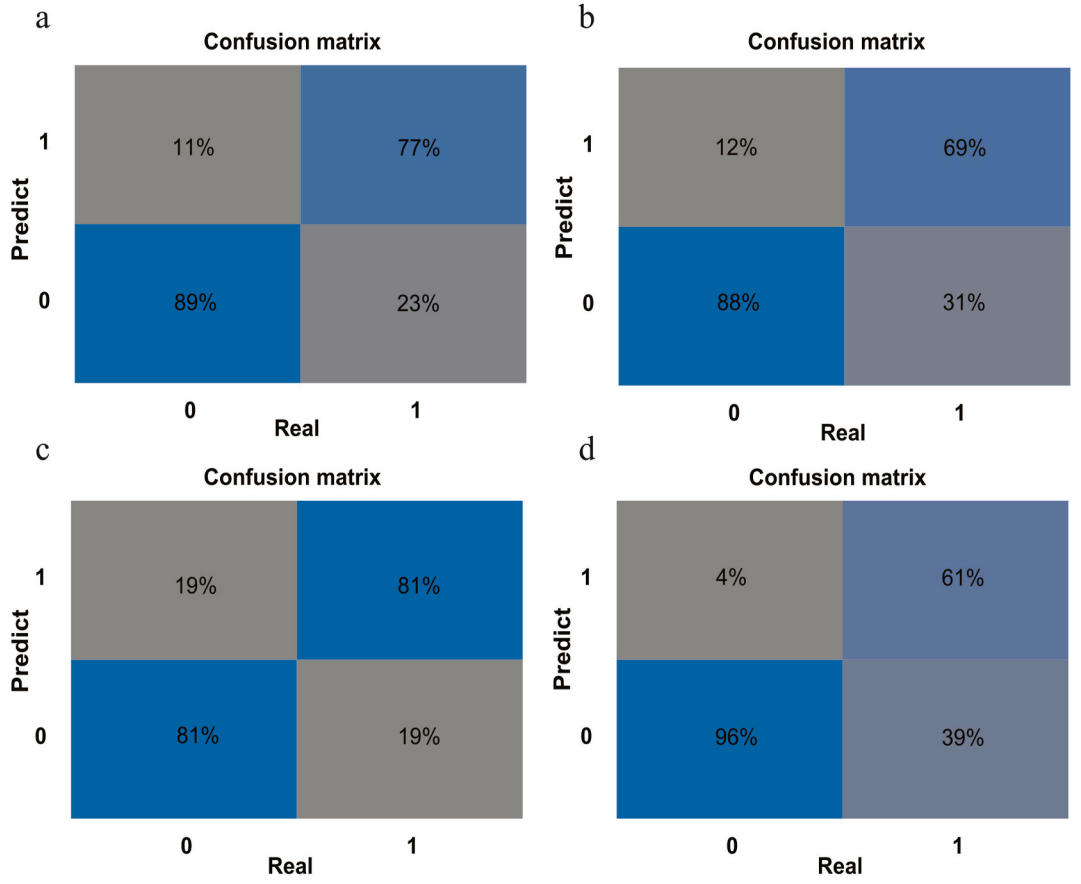


Fig. 6. The confusion matrices of (a) Model I; (b) Model II; (c) Model III; (d) Model IV.

Table 6
Evaluation indicators of the four models.

Model	Accuracy	Precision	Recall	F1 Score	Sensitivity	Specificity	FPR	FNR
I	0.845	0.81	0.773	0.791	0.81	0.89	0.11	0.23
II	0.793	0.818	0.692	0.75	0.818	0.88	0.12	0.31
III	0.81	0.778	0.808	0.792	0.778	0.81	0.19	0.19
IV	0.776	0.95	0.613	0.745	0.613	0.96	0.04	0.39

FPR, False Positive Rate; FNR, False Negative Rate.

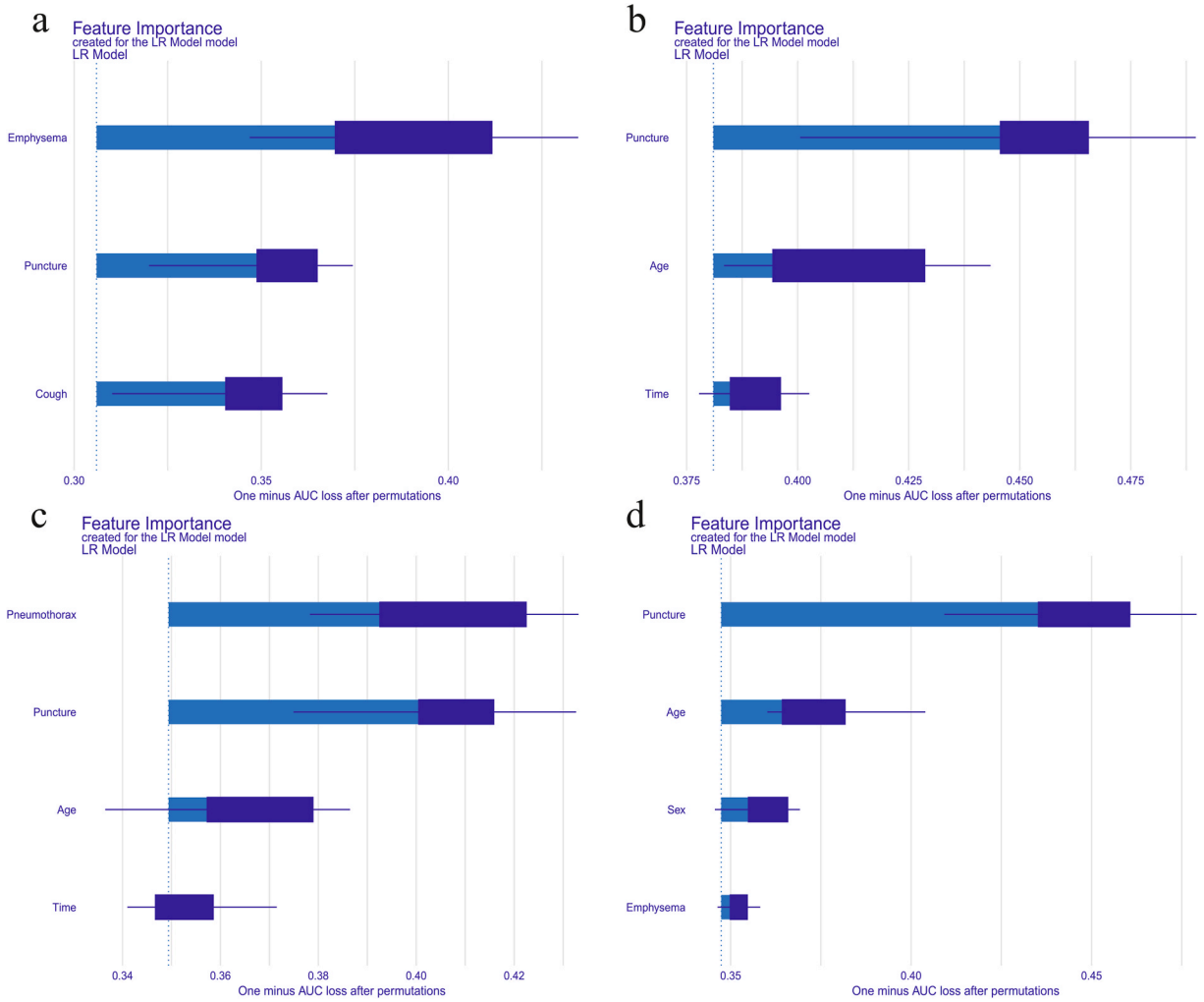


Fig. 7. The relative importance of variables for (a) Model I; (b) Model II; (c) Model III; (d) Model IV.

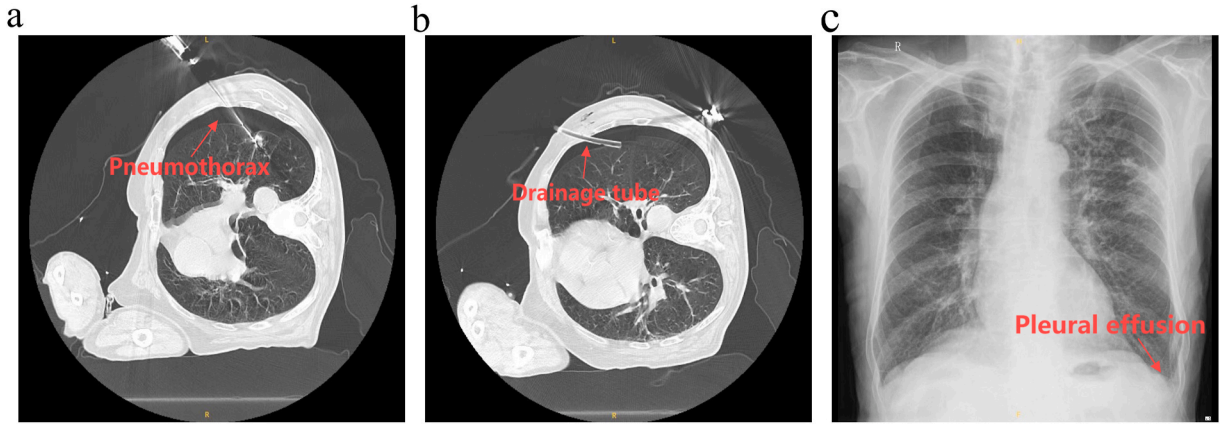


Fig. 8. Case Illustration. (a) Intraoperative occurrence of pneumothorax in the patient, as indicated by the arrow; (b) Postoperative placement of a chest tube, as shown by the arrow; (c) Chest X-ray one week post-surgery indicating left-sided pleural effusion, as marked by the arrow.

leading to pneumothorax. With more punctures, the cumulative risk of pneumothorax also increases. Lung punctures can provoke local inflammatory responses, leading to increased fluid exudation. If the inflammatory response is severe, it may result in the formation or exacerbation of pleural effusion. Prior studies have underscored puncture times as a pivotal risk factor for aseptic pleurisy [37], potentially escalating the likelihood of pleural effusion. Additionally, punctures may damage blood vessels, causing bleeding and leading to hemothorax or mixed pleural effusion. As the number of punctures increases, the probability of bleeding rises, subsequently increasing the risk of pleural effusion.

Thermal injury to the pleura can compromise its elasticity and reparative capacity, resulting in pleurisy and subsequent pleural effusion. In this study, 55.8 % (222/398) underwent MWA with ablation duration $\geq$6 min. Of these, 38.2 % (85/222) developed pleural effusion post-MWA, underscoring the significance of ablation time as a predictor for pleural effusion. The underlying mechanism likely involves the cumulative microwave energy imparted to the lung parenchyma with increasing ablation duration, leading to elevated pleural temperature and a heightened risk of pleurisy. Although higher maximum ablation power may similarly elevate pleural temperature and pose an increased pleurisy risk, it was not included in the final models for the two pleural effusion predictions in this study. Subsequent studies are warranted for further validation.

In this study, we established LR models for predicting post-MWA pneumothorax, pleural effusion, and the combined occurrence. Data from the training set were used for model development, and the test set data were employed for model validation. The AUC values demonstrated sufficient predictive accuracy in both sets, underscoring the high accuracy of LR. This study has several limitations, which, despite their presence, can serve as guidance for future improvements. First, as a retrospective study, there may be selection bias in the patient cohort. The majority of patients who underwent MWA at our hospital were in the middle to late stages of LM and were not candidates for thoracoscopic surgery. Consequently, the study sample may not be representative, limiting the generalizability of the model to a broader population, which could affect the external validity of the results. Second, although a validation cohort was established and demonstrated good predictive performance, external validation with data from other datasets is necessary. Additionally, deep learning methods were not employed for modeling and prediction in this study. In the future, we plan to conduct prospective randomized controlled trials to validate these findings while exploring other variables that may enhance predictive accuracy. Furthermore, we intend to collaborate with other centers to conduct multi-center studies, incorporating more patient data and variable information to reduce the selection bias inherent in single-center data. In subsequent research, we will also utilize deep learning techniques to integrate multimodal data, thereby improving the accuracy of predictive models. However, acknowledging these limitations, our study affirms that ML-based predictive models effectively anticipate the likelihood of pneumothorax and pleural effusion in patients undergoing MWA, considering factors like Sex, Age, Emphysema, Cough, Puncture times, and Ablation duration. LR outperformed in all predictive models according to ROC analysis, followed by the generation of a nomogram aimed at assisting clinicians in identifying patients with postoperative pneumothorax or pleural effusion for timely radiological review to monitor complications.

Data availability statement

Data will be made available on request.

Funding

This work was supported by the Natural Science Research Project of Nantong Science and Technology Bureau (No. MS12020029).

CRediT authorship contribution statement

Zihang Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Data curation. **Yufan Liu:** Data curation. **Xiaowen Cao:** Data curation. **Miaoyan Liu:** Data curation. **Li Wang:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Lou Zhong:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors of this manuscript declare no relationships with any companies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e38422.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 71 (3) (2021) 209–249.

[2] P.E. Postmus, K.M. Kerr, M. Oudkerk, S. Senan, D.A. Waller, J. Vansteenkiste, C. Escriu, S. Peters, Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up, Ann. Oncol. 28 (suppl_4) (2017) iv1–iv21.
[3] D.S. Ettinger, D.E. Wood, D.L. Aisner, W. Akerley, J.R. Bauman, A. Bharat, D.S. Bruno, J.Y. Chang, L.R. Chirieac, M. DeCamp, et al., NCCN Guidelines® insights: non-small cell lung cancer, version 2.2023, J Natl Compr Canc Netw 21 (4) (2023) 340–350.
[4] M.J. Baine, R. Sleightholm, B.K. Neilsen, D. Oupický, L.M. Smith, V. Verma, C. Lin, Stereotactic body radiation therapy versus nonradiotherapeutic ablative procedures (Laser/Cryoablation and electrocautery) for early-stage non-small cell lung cancer, J Natl Compr Canc Netw 17 (5) (2019) 450–458.
[5] V. Aufranc, G. Farouil, M. Abdel-Rehim, P. Smadja, M. Tardieu, S. Aptel, A. Guibal, Percutaneous thermal ablation of primary and secondary lung tumors: comparison between microwave and radiofrequency ablation, Diagn Interv Imaging 100 (12) (2019) 781–791.
[6] T.T. Healey, B.T. March, G. Baird, D.E. Dupuy, Microwave ablation for lung neoplasms: a retrospective analysis of long-term results, J Vasc Interv Radiol 28 (2) (2017) 206–211.
[7] M.T. Tsakok, M.W. Little, G. Hynes, R.S. Millington, P. Boardman, F.V. Gleeson, E.M. Anderson, Local control, safety, and survival following image-guided percutaneous microwave thermal ablation in primary lung malignancy, Clin. Radiol. 74 (1) (2019) 80.e19–80.e26.
[8] F.J. Wolf, D.J. Grand, J.T. Machan, T.A. Dipetrillo, W.W. Mayo-Smith, D.E. Dupuy, Microwave ablation of lung malignancies: effectiveness, CT findings, and safety in 50 patients, Radiology 247 (3) (2008) 871–879.
[9] A. Zheng, X. Wang, X. Yang, W. Wang, G. Huang, Y. Gai, X. Ye, Major complications after lung microwave ablation: a single-center experience on 204 sessions, Ann. Thorac. Surg. 98 (1) (2014) 243–248.
[10] S. Xu, J. Qi, B. Li, Z.X. Bie, Y.M. Li, X.G. Li, Risk prediction of pneumothorax in lung malignancy patients treated with percutaneous microwave ablation: development of nomogram model, Int J Hyperthermia 38 (1) (2021) 488–497.
[11] W. Li, J. Wang, W. Liu, C. Xu, W. Li, K. Zhang, S. Su, R. Li, Z. Hu, Q. Liu, et al., Machine learning applications for the prediction of bone cement leakage in percutaneous vertebroplasty, Front. Public Health 9 (2021) 812023.
[12] G. Cammarota, G. Ianiro, A. Ahern, C. Carbone, A. Temko, M.J. Claesson, A. Gasbarrini, G. Tortora, Gut microbiome, big data and machine learning to promote precision medicine for cancer, Nat. Rev. Gastroenterol. Hepatol. 17 (10) (2020) 635–648.
[13] K.Y. Ngiam, I.W. Khor, Big data and machine learning algorithms for health-care delivery, Lancet Oncol. 20 (5) (2019) e262–e273.
[14] R.D. Nindrea, T. Aryandono, L. Lazuardi, I. Dwiprahasto, Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: a meta-analysis, Asian Pac J Cancer Prev 19 (7) (2018) 1747–1752.
[15] J.Y. Liu, L.P. Liu, Z. Li, Y.W. Luo, F. Liang, The role of cuproptosis-related gene in the classification and prognosis of melanoma, Front. Immunol. 13 (2022) 986214.
[16] C. Li, M. Liu, Y. Zhang, Y. Wang, J. Li, S. Sun, X. Liu, H. Wu, C. Feng, P. Yao, et al., Novel models by machine learning to predict prognosis of breast cancer brain metastases, J. Transl. Med. 21 (1) (2023) 404.
[17] F. Kinoshita, T. Takenaka, T. Yamashita, K. Matsumoto, Y. Oku, Y. Ono, S. Wakasu, N. Haratake, T. Tagawa, N. Nakashima, et al., Development of artificial intelligence prognostic model for surgically resected non-small cell lung cancer, Sci. Rep. 13 (1) (2023) 15683.
[18] A. Swaminathan, I. Lopez, W. Wang, U. Srivastava, E. Tran, A. Bhargava-Shah, J.Y. Wu, A.L. Ren, K. Caoili, B. Bui, et al., Selective prediction for extracting unstructured clinical data, J. Am. Med. Inform. Assoc. 31 (1) (2023) 188–197.
[19] S. Liu, X. Wang, Y. Hou, G. Li, H. Wang, H. Xu, Y. Xiang, B. Tang, Multimodal data matters: language Model pre-training over structured and unstructured electronic health records, IEEE J Biomed Health Inform 27 (1) (2023) 504–514.
[20] Y. Cui, W. Che, S. Wang, T. Liu, LERT: A Linguistically-Motivated Pre-trained Language Model, 2011 *arxiv:221105344[csCL,csLG]*.
[21] H.M. Castro, J.C. Ferreira, Linear and logistic regression models: when to use and how to interpret them? J. Bras. Pneumol. 48 (6) (2023) e20220439.
[22] J. Hu, S. Szymczak, A review on longitudinal data analysis with random forest, Brief Bioinform 24 (2) (2023).
[23] X. Song, J. Zhu, X. Tan, W. Yu, Q. Wang, D. Shen, W. Chen, XGBoost-based feature learning method for mining COVID-19 novel diagnostic markers, Front. Public Health 10 (2022) 926069.
[24] G. Wu, R. Zheng, Y. Tian, D. Liu, Joint ranking SVM and binary relevance with robust low-rank learning for multi-label classification, Neural Netw 122 (2020) 24–39.
[25] Fryan LH. Al, M.I. Shomo, M.B. Alazzam, M.A. Rahman, Processing decision tree data using internet of things (IoT) and artificial intelligence technologies with special reference to medical application, BioMed Res. Int. 2022 (2022) 8626234.
[26] Z. Li, M.J. Sillanpää, Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection, Theor. Appl. Genet. 125 (3) (2012) 419–435.
[27] Q. Liu, M.A. Charleston, S.A. Richards, B.R. Holland, Performance of Akaike information criterion and Bayesian information criterion in selecting partition models and mixture models, Syst. Biol. 72 (1) (2023) 92–105.
[28] J. Palussiere, P. Lagarde, A. Aupérin, F. Deschamps, F. Chomy, T. de Baere, Percutaneous lung thermal ablation of non-surgical clinical N0 non-small cell lung cancer: results of eight years' experience in 87 patients from two centers, Cardiovasc. Intervent. Radiol. 38 (1) (2015) 160–166.
[29] T. Hiraki, H. Gobara, T. Iguchi, H. Fujiwara, Y. Matsui, S. Kanazawa, Radiofrequency ablation for early-stage nonsmall cell lung cancer, BioMed Res. Int. 2014 (2014) 152087.
[30] D. Arefan, R.M. Hausler, J.H. Sumkin, M. Sun, S. Wu, Predicting cell invasion in breast tumor microenvironment from radiological imaging phenotypes, BMC Cancer 21 (1) (2021) 370.
[31] Z. Liu, M. Mi, X. Li, X. Zheng, G. Wu, L. Zhang, A lncRNA prognostic signature associated with immune infiltration and tumour mutation burden in breast cancer, J. Cell Mol. Med. 24 (21) (2020) 12444–12456.
[32] A.M. Moussa, E. Ziv, S.B. Solomon, J.C. Camacho, Microwave ablation in primary lung malignancies, Semin. Intervent. Radiol. 36 (4) (2019) 326–333.
[33] T.J. Vogl, N.A. Nour-Eldin, M.H. Albrecht, B. Kaltenbach, W. Hohenforst-Schmidt, H. Lin, B. Panahi, K. Eichler, T. Gruber-Rouh, A. Roman, Thermal ablation of lung tumors: focus on microwave ablation, Röfo 189 (9) (2017) 828–843.
[34] X. Ye, W. Fan, H. Wang, J. Wang, Z. Wang, S. Gu, W. Feng, Y. Zhuang, B. Liu, X. Li, et al., Expert consensus workshop report: guidelines for thermal ablation of primary and metastatic lung tumors (2018 edition), J Cancer Res Ther 14 (4) (2018) 730–744.
[35] M.S. Kim, H.P. Hong, S.Y. Ham, D.H. Koo, D.Y. Kang, T.Y. Oh, Complications after 100 sessions of cone-beam computed tomography-guided lung radiofrequency ablation: a single-center, retrospective experience, Int J Hyperthermia 37 (1) (2020) 763–771.
[36] M. Mascalchi, M. Luconi, Lung cancer screening, emphysema, and COPD, Chest 159 (5) (2021) 1699–1700.
[37] M. Kashima, K. Yamakado, H. Takaki, H. Kodama, T. Yamada, J. Uraki, A. Nakatsuka, Complications after 1000 lung radiofrequency ablation sessions in 420 patients: a single center's experiences, AJR Am. J. Roentgenol. 197 (4) (2011) W576–W580.