Infection and Drug Resistance

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Establishment of a Machine Learning-Based Predictive Model for *Klebsiella pneumoniae* Liver Abscess

Haoran Li[1], Yan Yu[1], Xi Chen[2], Qingqing Sun[1], Xiumin Li[1], Qiujing Shang[1], Minghua Ying[1], Xiulin Liu[1], Jing Meng[3], Lele Bian[4], Shanshan Wu[1], Yuejuan Gao[1]

[1]Senior Department of Oncology, Chinese PLA General Hospital, Beijing, 100039, People's Republic of China; [2]Department of Ultrasound, Tianjin Third Central Hospital, Tianjin, 300170, People's Republic of China; [3]Outpatient Department Comprehensive Treatment Room, Fifth Medical Center of Chinese PLA General Hospital, Beijing, 100039, People's Republic of China; [4]Department of Radiology, Fifth Medical Center of Chinese PLA General Hospital, Beijing, 100039, People's Republic of China

Correspondence: Yuejuan Gao; Shanshan Wu, Senior Department of Oncology, Chinese PLA General Hospital, Beijing, 100039, People's Republic of China, Tel + 86-10-66933330, Email 18601135088@163.com; hyd_wu@163.com

Purpose: To investigate the clinical and ultrasonographic characteristics of pyogenic liver abscess (PLA) caused by *Klebsiella pneumoniae* (K-PLA) and *non-Klebsiella pneumoniae* pathogens, and to develop machine learning models for the differential diagnosis of K-PLA.

Materials and Methods: In this retrospective study, patients clinically diagnosed with PLA and confirmed by ultrasound-guided puncture at the Fifth Medical Center of PLA General Hospital between April 2013 and December 2020 were enrolled. Based on the causative pathogens, patients were categorized into K-PLA and non-K-PLA groups. Baseline data, including ultrasonographic features, clinical characteristics, and laboratory findings, were collected. The Boruta algorithm was employed for feature selection, and four machine learning models-Deep Learning-Fully Connected Neural Network (deeplearning), Distributed Random Forest (drf), Gradient Boosting Machine (gbm), and Generalized Linear Model (glm)—were developed to diagnose K-PLA. The models were validated using 5-fold cross-validation.

Results: A total of 201 patients with bacterial liver abscess were included (median age: 57 years; range: 49–66; 136 males), comprising 134 K-PLA cases and 67 non-K-PLA cases. The Boruta algorithm identified seven significant predictive variables: history of diabetes, history of hepatocellular carcinoma, history of biliary tract disease, history of infectious diseases, duration of fever, body temperature, and alanine aminotransferase (ALT) levels. Using these variables, the four machine learning models were constructed. In the training set, the area under the receiver operating characteristic curve (AUC) for predicting K-PLA was 0.716 (deeplearning), 0.999 (drf), 0.922 (gbm), and 0.718 (glm). In the validation set, the corresponding AUC values were 0.799, 0.763, 0.848, and 0.805, respectively.

Conclusion: This study successfully established four machine learning models for predicting the risk of K-PLA, with the gbm-based model demonstrating the highest diagnostic performance. These models may facilitate early clinical diagnosis and treatment of K-PLA, thereby reducing antibiotic misuse.

Keywords: liver abscess, *Klebsiella pneumoniae*, machine learning, predictive model, misuse of antibiotics

Introduction

Suppurative pyogenic liver abscess (PLA) is a common infectious disease of the liver and biliary system, seriously threatening life and health.[1,2] The global annual incidence rate has risen from 0.86‰ to 1.6‰,[3,4] and the mortality rate is as high as 10%-30%.[5,6] The current mainstream treatment regimens for liver abscess include basic anti-infection therapy and minimally invasive drainage. Anti-infection therapy involves empirical antibiotic use and pathogen-targeted therapy. The former has a high risk of antibiotic abuse and resistance, leading to the failure of the initial plan, while the latter may delay the medication time due to the long duration of pathogen culture, making it difficult to control the condition in a timely manner. Therefore, early diagnosis and treatment are key factors influencing prognosis.[7–11]

Received: 7 June 2025
Accepted: 12 September 2025
Published: 24 September 2025

Studies have shown that PLA may be caused by a single microorganism or multiple microorganisms, among which *Klebsiella* pneumoniae accounts for 88%.[3–6] The antibiotic sensitivity and subsequent treatment of K-PLA are different from those of PLA caused by other pathogenic bacteria. In particular, K-PLA is more prone to migratory infection and has a high fatality rate.[12] Therefore, if K-PLA is diagnosed early, targeted antibiotics can be used earlier to avoid infections in other parts of the body, which is of positive significance for the prognosis of patients. The gold standard for diagnosing PLA is pathogen examination. However, as it is an invasive surgery and the bacterial culture period is long or the result may be negative, there is no unified conclusion on whether the pathogenic bacteria of PLA can be identified through clinical manifestations and imaging features.[9]

Based on this background, this study analyzed the discriminating features of K-PLA and constructed a machine learning prediction model, aiming to achieve early identification before the results are issued, thereby guiding precise medication, reducing antibiotic abuse, and optimizing the clinical decision-making path. The result report is as follows.

Materials and Methods

Study Participants

Patients admitted to the Fifth Medical Center of PLA General Hospital from April 2013 to July 2021 were retrospectively analyzed. Inclusion criteria: (1) The bacterial culture results of blood and pus were positive and the pathogenic bacteria were clear; (2) Clinical manifestations: fever, chills, abdominal pain; (3) Ultrasound examination in our hospital with complete image data; Exclusion criteria: (1) patients diagnosed as amebic liver abscess or infectious liver cyst; (2) patients whose medical records were not clearly documented or who did not receive any treatment; (3) Pregnancy. A total of 201 patients with PLA were enrolled (Figure 1). This study was approved by the Ethics Committee of the Fifth Medical Center of the PLA General Hospital.

Study Methods

Patient data were retrospectively retrieved from the ultrasound diagnostic reporting system of the Fifth Medical Center of PLA General Hospital. Demographic information and grayscale ultrasound images (in both JPG and DICOM formats) were extracted for further analysis. By accessing the hospital's electronic medical record system, we collected

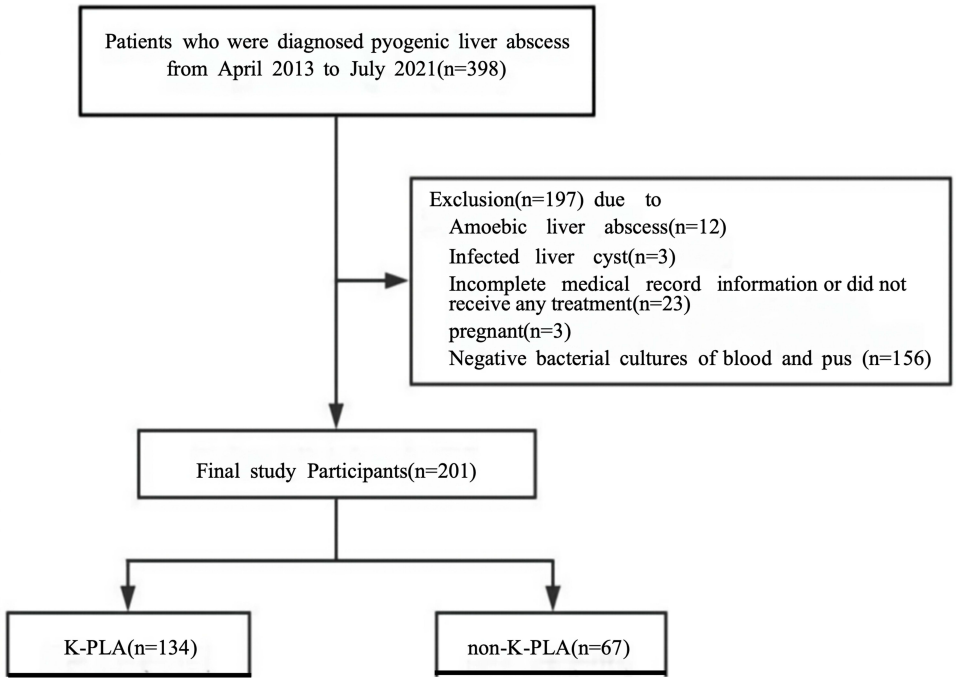

Figure 1 Flowchart showing the patient enrollment process.

microbiological culture results of abscess fluid and the volume of initial drainage for each patient. In this study, pus bacteria culture and drug sensitivity test in traditional microbiological techniques were used to detect bacteria in pus. Abscess fluid sampling was performed under ultrasound or CT guidance, wherein approximately 2 mL of purulent material was aspirated using a 5 mL syringe and immediately transferred into a sterile throat swab culture tube for microbiological analysis at the hospital's clinical laboratory. Aerobic culture of the abscess fluid was conducted using the VITEK 2 automated system (bioMérieux, France) for bacterial identification and antibiotic susceptibility testing, with dedicated reagents and instruments. Further clinical data, including demographic characteristics, clinical features, and laboratory test results, were systematically recorded and analyzed. Based on pathogen identification, patients were stratified into two groups: the *Klebsiella* pneumoniae liver abscess (K-PLA) group (n=134) and the non-*Klebsiella* pneumoniae liver abscess (non-K-PLA) group (n=67). Two board-certified radiologists, each with over five years of experience in abdominal imaging, independently reviewed the ultrasound images in a blinded manner. Discrepancies in interpretation were resolved through consensus discussion. The imaging analysis included assessment of abscess location, number, size, internal architecture (cystic vs solid components), echotexture, and clinical staging.

Data Collection

Collection of patients with clinical data including age, gender, basic diseases (diabetes, high blood pressure, biliary diseases), the history of infectious diseases (hepatitis b viral hepatitis, viral hepatitis c, HIV/AIDS, syphilis, tuberculosis, echinococcosis), body temperature, duration of fever on admission, history of surgery and the history of antibiotic use, initial laboratory values (including peripheral blood leukocytes, Neutrophil, platelet, serum albumin, aspartate amino-transferase, alanine aminotransferase, alkaline phosphatase, γ-transpeptidase, urea nitrogen, creatinine, C-reactive protein and serum calcitonin) and pathogenic bacteria (*Klebsiella* pneumoniae, *Escherichia coli, Pseudomonas aeruginosa, Staphylococcus aureus, Streptococcus*, etc) confirmed by etiological examination. The initial value was the first measurement obtained within 24 hours of discovery. The microbiology of liver abscess was defined as the detection of microorganisms from pus and/or blood cultures of liver abscess. In addition, the first drainage volume of the liver abscess was collected as well as the color of the drainage fluid. The analysis indicators of ultrasound images included: (1) location (left lobe, right lobe and both); (2) number (single or multiple); (3) size (maximum diameter of the largest lesion); (4) septations within the abscess (unilocular and multiocular); (5) structure (mainly cystic: the proportion of cystic part≥80%, solid: the proportion of solid part≥80%, cystic and solid: the proportion of cystic and solid part < 80%); (6) presence of internal gas bubbles; (7) presence of echogenic debris (poor sound transmission in the liquid part, with floating dotted echoes); (8) presence of variable calcification; (9) presence of blood flow signals; (10) margin of the lesion (irregular or indistinct and smooth); (11) abscess wall thickness (thick wall≥2mm, thin wall < 2mm); (12) Staging of abscess (pre-stage, formation stage, absorption stage).

Statistical Analysis

Statistical analysis was performed using Empower Stats4.0, R (version 4.2.2) and MSTATA (www.mstata.com). The clinical and ultrasound features of K-PLA and non-K-PLA were compared. Categorical variables were compared between groups using the X^2 test or Fisher's exact test, and continuous variables were compared between groups using the *t* test or the Mann–Whitney *U*-test. Independence, normal distribution, and homogeneity of variance between the factors with differences were tested. Taking the pathogen species as the outcome variable, the important risk factors were screened by Boruta method (The Boruta algorithm has strong anti-interference ability and rejects false positive events. Capture complex correlations and be able to identify nonlinear/interactions that are ignored by traditional statistics. It has strong adaptability to small-sample high-dimensional medical data). K-PLA was diagnosed in advance based on four machine learning algorithms: deeplearning, drf, gbm and glm. The 5-fold cross validation method was used to validate the above model, and the ROC curve, accuracy, sensitivity, specificity, negative predictive value and positive predictive value were used to evaluate the predictive efficacy and clinical versatility of the model (see the Supplementary Material for source code). $P<0.05$ was considered statistically significant.

Results

Detection of Pathogenic Bacteria

A total of 11 types of bacteria were detected in 201 pus specimens, and the bacterial culture results were positive. The common pathogenic bacteria causing PLA are: *Klebsiella* pneumoniae (66.7%), *Staphylococcus* (9.5%), *Escherichia coli* (8.0%), *Streptococcus* (5.5%), *Enterococcus faecium* (2.0%), *Pseudomonas aeruginosa* (1.5%), *maltophila* (1.0%), *Burkholderia onioniae* (1.0%), *Morgenella* (0.5%), *Enterobacter cloacae* (0.5%) And *Propionibacterium ulcerans* (0.5%). It can be seen that *Klebsiella* pneumoniae dominates (66.7%, 134/201), significantly higher than other pathogens (such as *Escherichia coli* 8.0%, *Staphylococcus* 9.5%). This distribution validates the presupposition hypothesis of the study - that K. pneumoniae is the main pathogen of current bacterial liver abscess, providing an epidemiological basis for the subsequent construction of targeted prediction models.

General Information and Screening of Important Variables of the Two Groups of Patients

There were 136 males and 35 females in 201 PLA patients (Table 1). The mean age was (55.5±13.2) years (range, 8–81 years). The prevalence of diabetes was higher (48.5% vs 23.9%, p<0.001), supporting the hypothesis that "diabetes is a key risk factor for K-PLA". The duration of fever was shorter (median 11 days vs 20 days, p=0.017), suggesting that K-PLA had a more acute onset. The history of biliary tract diseases/infections was less (9.7% vs 29.9%, p<0.001),

Table 1 Patient Clinical Baseline Characteristics

Characteristic	Pathogenic Bacteria			p-value
	Overall, N = 201	Non-klebsiella Pneumoniae, N = 67	Klebsiella Pneumoniae, N = 134	
Age(Y)	57 (49, 66)	58 (48, 65)	57 (49, 66)	0.729
Gender				0.070
Male	136 (67.7%)	51 (76.1%)	85 (63.4%)	
Female	65 (32.3%)	16 (23.9%)	49 (36.6%)	
Stage				0.469
Early formative period	8 (4.0%)	2 (3.0%)	6 (4.5%)	
Period of formation	160 (79.6%)	51 (76.1%)	109 (81.3%)	
Period of absorption	33 (16.4%)	14 (20.9%)	19 (14.2%)	
Fever and or chill	183 (91.0%)	58 (86.6%)	125 (93.3%)	0.116
Temperature (°C)	39.4 (38.8, 40.0)	39.7 (38.5, 41.0)	39.1 (39.0, 40.0)	0.297
Duration of fever (d)	14 (7, 30)	20 (10, 44)	11 (7, 21)	0.017
Texture of pus				0.015
Thin	33 (16.4%)	17 (25.4%)	16 (11.9%)	
Thick	168 (83.6%)	50 (74.6%)	118 (88.1%)	
Diabetes mellitus	81 (40.3%)	16 (23.9%)	65 (48.5%)	<0.001
History of alcoholism	26 (12.9%)	7 (10.4%)	19 (14.2%)	0.457
Infectious diseases	33 (16.4%)	20 (29.9%)	13 (9.7%)	<0.001
Hypertension	47 (23.4%)	16 (23.9%)	31 (23.1%)	0.906
History of surgical procedures	80 (39.8%)	34 (50.7%)	46 (34.3%)	0.025
Cholecystectomy	29 (14.4%)	13 (19.4%)	16 (11.9%)	0.156
Biliary tract disease	33 (16.4%)	20 (29.9%)	13 (9.7%)	<0.001
Liver cancer	18 (9.0%)	14 (20.9%)	4 (3.0%)	<0.001
History of antibiotic use	142 (70.6%)	47 (70.1%)	95 (70.9%)	0.913
Volume of pus obtained at initial aspiration				0.353
<10mL	126 (62.7%)	45 (67.2%)	81 (60.4%)	
≥10mL	75 (37.3%)	22 (32.8%)	53 (39.6%)	

(*Continued*)

Table 1 (Continued).

Characteristic	Pathogenic Bacteria			p-value
	Overall, N = 201	Non-klebsiella Pneumoniae, N = 67	Klebsiella Pneumoniae, N = 134	
GRAN(%)	6.8 (4.5, 9.5)	7.5 (4.5, 10.5)	6.6 (4.6, 9.2)	0.405
PLT(10^9/L)	260 (143, 370)	226 (126, 346)	281 (160, 372)	0.073
ALB(g/L)	31 (28, 34)	30 (25, 34)	31 (28, 35)	0.106
AST(U/L)	31 (21, 54)	34 (21, 67)	31 (21, 47)	0.265
ALT(U/L)	31 (21, 49)	34 (18, 67)	30 (21, 45)	0.960
ALP(U/L)	169 (121, 241)	193 (135, 262)	166 (119, 215)	0.024
GGT(U/L)	118 (71, 184)	126 (68, 180)	117 (71, 186)	0.626
BUN(mmol/L)	4.3 (3.1, 6.0)	4.5 (3.4, 6.9)	4.2 (3.0, 5.9)	0.119
Cr(μmol/L)	69 (55, 82)	71 (54, 83)	68 (56, 82)	0.459
CRP(mg/L)	85 (40, 149)	81 (42, 156)	87 (41, 141)	0.932
PCT(ng/mL)	1 (0, 6)	1 (0, 5)	1 (0, 8)	0.649

Note: Value are presented as number (%), mean±standard deviation. Normally distributed data were tested using independent samples t-tests, and nonparametric tests were used for nonnormal data.

indicating that the infection route was different from that of non-K-PLA (non-biliary origin). These differences indicate that K-PLA has a unique clinical phenotype, which can be preliminarily identified by conventional clinical indicators. The ultrasound imaging characteristics of K-PLA patients (n=134) and non-K-PLA patients (n=67) were compared (Table 2). The Boruta algorithm is adopted to analyze the important attributes of the feature factors. The results showed that seven characteristic variable attributes, namely diabetes, body temperature, duration of fever, history of infectious diseases, history of biliary tract diseases, history of liver cancer, and ALT, were confirmed as important, while the

Table 2 Patient Ultrasonographic Features

Characteristic	Pathogenic Bacteria			p-value
	Overall, N = 201	Non-klebsiella Pneumoniae, N = 67	Klebsiella Pneumoniae, N = 134	
Liver background				0.001
Normal	119 (59.2%)	40 (59.7%)	79 (59.0%)	
Fatty liver	29 (14.4%)	2 (3.0%)	27 (20.1%)	
Liver fibrosis	53 (26.4%)	25 (37.3%)	28 (20.9%)	
Location				0.919
Right lobe	122 (60.7%)	42 (62.7%)	80 (59.7%)	
Left lobe	25 (12.4%)	8 (11.9%)	17 (12.7%)	
Both	54 (26.9%)	17 (25.4%)	37 (27.6%)	
Number				0.816
Single	152 (75.6%)	50 (74.6%)	102 (76.1%)	
Multiple	49 (24.4%)	17 (25.4%)	32 (23.9%)	
Septations within the abscess				0.068
None	60 (29.9%)	20 (29.9%)	40 (29.9%)	
Single	86 (42.8%)	35 (52.2%)	51 (38.1%)	
Multiple	55 (27.4%)	12 (17.9%)	43 (32.1%)	
Structure of the abscess				0.833
Cystic	83 (41.3%)	26 (38.8%)	57 (42.5%)	
Cystic solid	58 (28.9%)	21 (31.3%)	37 (27.6%)	
Solidity	60 (29.9%)	20 (29.9%)	40 (29.9%)	

(*Continued*)

https://doi.org/10.2147/IDR.S545440

Table 2 (Continued).

Characteristic	Pathogenic Bacteria			p-value
	Overall, N = 201	Non-klebsiella Pneumoniae, N = 67	Klebsiella Pneumoniae, N = 134	
The cystic part was echogenic				0.069
Low	138 (68.7%)	40 (59.7%)	98 (73.1%)	
Heterogeneity of mass	62 (30.8%)	27 (40.3%)	35 (26.1%)	
High	1 (0.5%)	0 (0.0%)	1 (0.7%)	
Internal gas bubble	21 (10.4%)	9 (13.4%)	12 (9.0%)	0.328
Echogenic debris	41 (20.4%)	16 (23.9%)	25 (18.7%)	0.386
The echo of the solid part				0.316
Low	32 (15.9%)	13 (19.4%)	19 (14.2%)	
Heterogeneity of mass	128 (63.7%)	44 (65.7%)	84 (62.7%)	
High	41 (20.4%)	10 (14.9%)	31 (23.1%)	
Variable calcification	58 (28.9%)	21 (31.3%)	37 (27.6%)	0.582
Blood flow signals	112 (55.7%)	33 (49.3%)	79 (59.0%)	0.192
Margin of lesion				0.212
Not clear	72 (35.8%)	20 (29.9%)	52 (38.8%)	
Clarity	129 (64.2%)	47 (70.1%)	82 (61.2%)	
Abscess wall thickening				0.866
Without wall	59 (29.4%)	19 (28.4%)	40 (29.9%)	
Thin wall	18 (9.0%)	7 (10.4%)	11 (8.2%)	
Thick wall	124 (61.7%)	41 (61.2%)	83 (61.9%)	
Whether the inner wall is worm eaten	96 (47.8%)	32 (47.8%)	64 (47.8%)	>0.999

attributes of two characteristic variables, BUN and ALB, were not determined. The attributes of the remaining 35 characteristic variables are not significant (Table 3, Figure 2A and B). A history of diabetes indicates the host's immune deficiency status. High body temperature and prolonged fever suggest acute invasive infection. Elevated ALT indicates

Table 3 Variable Selection with Boruta Algorithm

Vars	Mean Importance	Median Importance	Min Importance	Max Importance	Norm Hits	Decision
Liver.cancer	10.17974410	10.29213577	7.23217766	13.4800562	1.00000000	Confirmed
Duration.of.fever	7.67484542	8.00196933	3.29088334	11.4090887	0.96969697	Confirmed
Temperature	6.88207222	7.01887966	3.34409734	10.3972684	0.95959596	Confirmed
Biliary.tract.disease	6.64484001	6.69715729	3.45203911	9.2917249	0.95959596	Confirmed
ALT	5.14006060	5.28687440	1.35262604	8.6174280	0.83838384	Confirmed
Diabetes.mellitus	4.99686957	4.89170327	1.39223329	8.6943611	0.83838384	Confirmed
Infectious.diseases	4.48649970	4.53875285	1.15901385	7.3148654	0.85858586	Confirmed
ALB	2.78190565	2.86074218	0.09390123	5.2894017	0.56565657	Tentative
BUN	2.55955325	2.35578941	0.51351151	5.4568618	0.49494949	Tentative
AST	1.85003950	1.96025420	−0.52695292	3.7207693	0.22222222	Rejected
WBC	1.74557842	1.91605601	0.49630685	3.3353449	0.00000000	Rejected
GRAN	1.48711699	1.30665206	−1.73953867	3.9330178	0.17171717	Rejected
Liver.background	1.19515283	1.26327342	−0.85266492	3.7102045	0.08080808	Rejected
Gender	1.25367939	1.19996459	0.02692810	2.9778495	0.01010101	Rejected
History.of.surgical.procedures	0.65994990	0.90456524	−1.20041172	2.1438720	0.00000000	Rejected
PLT	0.31276208	0.85717604	−2.88849192	2.4991867	0.01010101	Rejected

(*Continued*)

Table 3 (Continued).

Vars	Mean Importance	Median Importance	Min Importance	Max Importance	Norm Hits	Decision
Septations.within.the.abscess	0.89683420	0.65244347	−0.16860403	2.8975737	0.00000000	Rejected
Fever.and.or.chill	0.50762631	0.57834817	−1.80626218	2.5335669	0.01010101	Rejected
ALP	0.72400080	0.56197353	−0.92232399	2.8416624	0.00000000	Rejected
Age	0.23282309	0.28983920	−1.56781169	2.3744611	0.00000000	Rejected
Cholecystectomy	0.31159434	0.27752062	−1.16595806	3.0703899	0.00000000	Rejected
The.echo.of.the.solid.part	0.03159109	0.25363988	−2.08216849	1.8726990	0.00000000	Rejected
The.cystic.part.was.echogenic	0.29565799	0.21269595	−2.61459400	2.0135278	0.00000000	Rejected
Abscess.size	0.04676174	0.17224303	−1.36314113	0.9646902	0.00000000	Rejected
Echogenic.debris	0.15094340	0.15959532	−1.81230535	1.9317192	0.00000000	Rejected
Margin.of.lesion	0.10375469	0.09335512	−1.33820788	2.0194257	0.00000000	Rejected
Texture.of.pus	0.35591192	0.09156745	−0.87974280	2.3460773	0.00000000	Rejected
Volume.of.pus.obtained.at.initial.aspiration	0.02333087	−0.02378114	−1.31228246	1.4522941	0.00000000	Rejected
History.of.alcoholism	−0.04192639	−0.04754143	−1.52355973	1.0509147	0.00000000	Rejected
Stage	0.34369202	−0.09299704	−0.68639356	2.7863405	0.00000000	Rejected
Location	−0.08924603	−0.16942749	−0.93840297	1.5425655	0.00000000	Rejected
PCT	−0.10268954	−0.19760289	−1.18271448	0.9555708	0.00000000	Rejected
History.of.antibiotic.use	−0.29176428	−0.26147130	−2.21588317	1.1646631	0.00000000	Rejected
Structure.of.the.abscess	−0.39549530	−0.31509364	−2.25820321	1.2821337	0.00000000	Rejected
Hypertension	−0.11395809	−0.32676367	−1.85775228	1.8047861	0.00000000	Rejected
Number	−0.49678890	−0.34388108	−2.13809692	0.6769289	0.00000000	Rejected
Cr	−0.14630977	−0.34853169	−1.47412442	1.4939316	0.00000000	Rejected
Internal.gas.bubble	−0.41354315	−0.36537226	−1.65396852	1.1817010	0.00000000	Rejected
Blood.flow.signals	−0.18024845	−0.41475490	−1.51154671	1.8089374	0.00000000	Rejected
CRP	−0.70500226	−0.51717563	−2.93636417	1.0666076	0.00000000	Rejected
Variable.calcification	−0.36759225	−0.70384565	−1.48479764	0.9671697	0.00000000	Rejected
Abscess.wall.thickening	−0.67164235	−0.73276623	−2.49768115	0.4867499	0.00000000	Rejected
Whether.the.inner.wall.is.worm.eaten	−0.67224780	−0.80915114	−2.44559885	1.4688717	0.00000000	Rejected
GGT	−0.97168153	−1.06329041	−2.70772220	0.7957722	0.00000000	Rejected

the degree of liver cell damage. The absence of a history of biliary tract diseases/infections suggests non-biliary transmission routes. All the ultrasound signs were confirmed as important factors, indicating that the discriminatory value of ultrasound images is limited. However, this negative result still advanced the research purpose and clarified that a single ultrasound sign is insufficient to identify the pathogen, and a comprehensive judgment should be made in combination with clinical variables.

Construction and Validation of the Prediction Model

The seven important feature variables screened out by the Boruta algorithm were introduced into the prediction model. The K-PLA early diagnosis and prediction model was established by using machine learning algorithms such as deeplearning, drf, gbm, and glm. The model was validated using five-level cross-validation and the ROC curve. The results show that the AUC values of the training set are 0.716, 0.999, 0.922 and 0.718 respectively, and those of the validation set are 0.799, 0.763, 0.848 and 0.805 respectively (Figure 3). The accuracies of the training sets were 0.826, 0.757, 0.835 and 0.757 respectively, and those of the validation sets were 0.733, 0.733, 0.826 and 0.802 respectively (Table 4). A high-precision prediction tool has been successfully established, addressing the pain point of clinical delayed diagnosis. Among them, the machine learning model based on the gbm algorithm has the best performance and can identify K-PLA at an early stage, guiding the precise use of antibiotics.

https://doi.org/10.2147/IDR.S545440

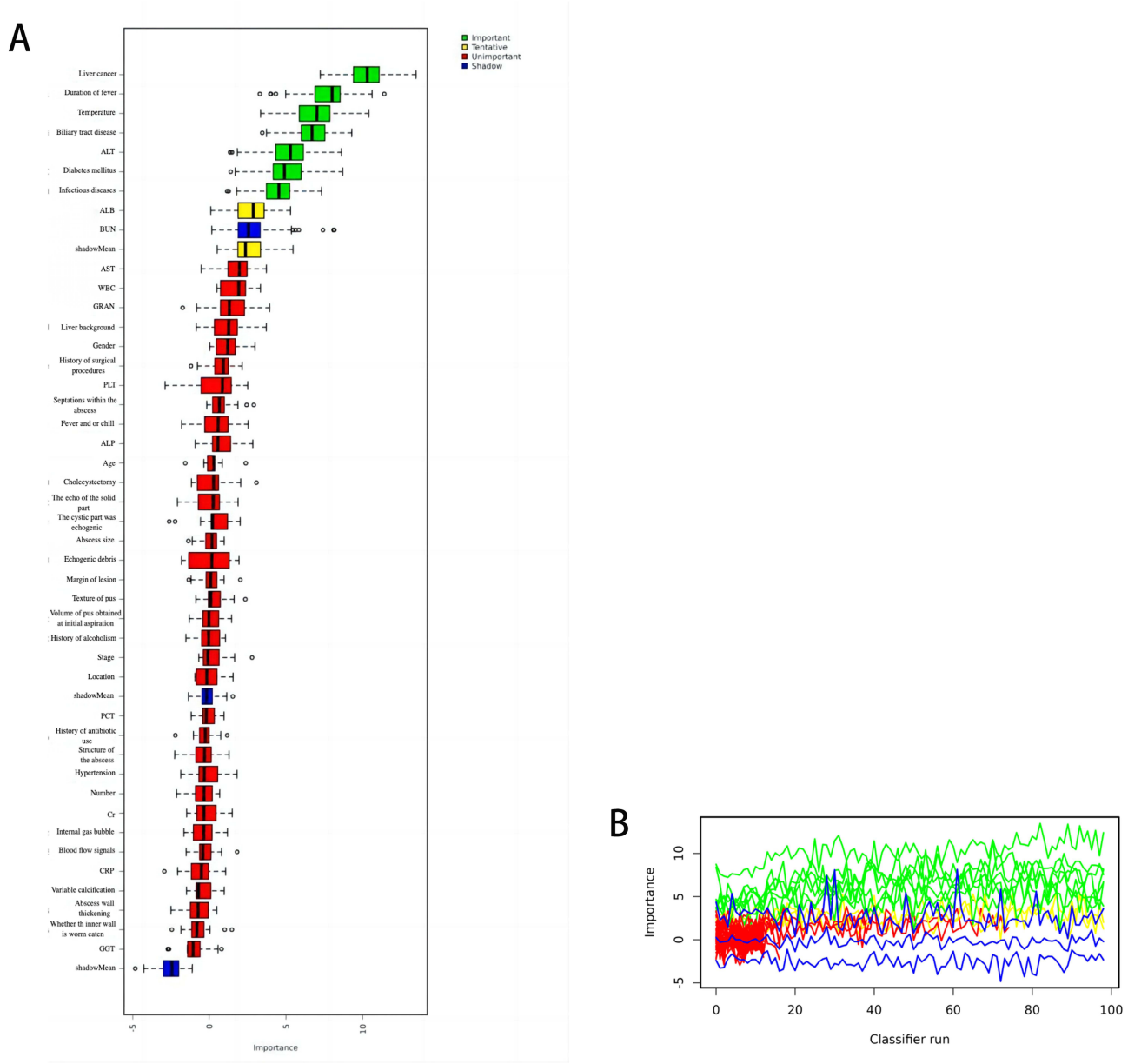

Figure 2 (A) Ranking of clinical variables for predicting perioperative complication by Boruta algorithm.The plot demonstrates boxplot of important attributes in color green, tentative attributes in yellow,non-important attributes in red, and shadow attributes in blue box, respectively.The vertical axis lists the name of each variable, and the horizontal axis is the Z-value. (**B)** History graph of each decision of accepting or rejecting by Random Forest in Boruta algorithm.The accepted attributes (green) have distinctly higher importance than the other attributes.

Discussion

In this study, four machine learning models were successfully established to predict K-PLA, among which the GBM-based model had a high diagnostic efficiency (AUC=0.85). While previous studies have reported similar prediction tools, our approach advances the field in two key ways: one is comprehensive variable integration: Unlike existing models that rely primarily on CT radiomics[13–16] or limited clinical parameters,[17] we included multidimensional data covering clinical history, laboratory markers, and ultrasound features, significantly expanding the set of discriminative features. The second is enhanced clinical utility: our model achieves comparable or better performance (eg, AUC 0.85 vs 0.78–0.83 in Wang et al[13]) while eliminating the need for expensive imaging modalities and relying only on history and laboratory tests. Recently, prediction tools for K-PLA have begun to emerge, but there are very few machine learning prediction models, and the performance of different algorithms and models is uneven, which still has some key limitations. Feng et al[17] developed a random forest model

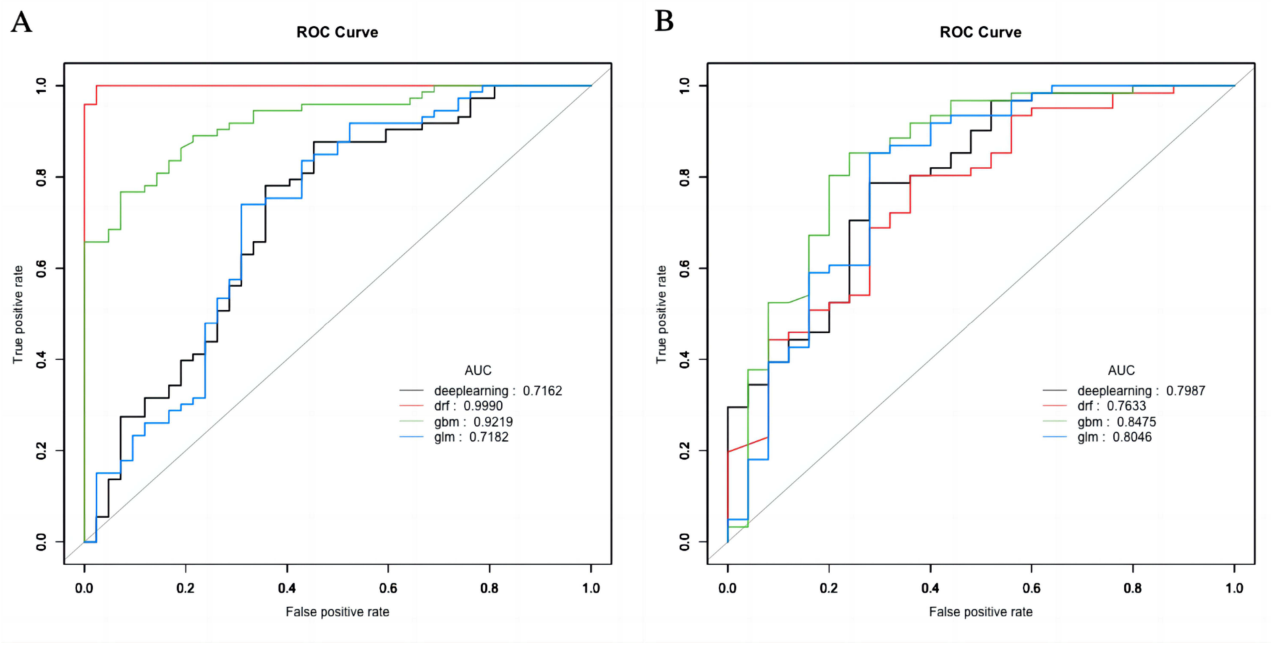

Figure 3 (**A**) ROC curves of the four machine learning models in the training set. (**B**) ROC curves of the four machine learning models in the test set.

using only diabetes-related variables (AUC=0.82), but its applicability to the general population remains unproven. Wang et al[13] established a CT-based radiomics nomogram (AUC=0.83) with good accuracy, but its clinical application is limited by cost and accessibility. Similarly, Gu et al[17] combined clinical and radiomics features (AUC=0.85) but required specialized imaging analysis. In contrast, our model not only uniquely integrates ultrasound features and clinical laboratory data; New predictors such as higher body temperature and shorter duration of fever were also identified; Demonstrated robust performance without relying on CT/MRI - providing a practical solution for resource-limited environments. However, unlike Lan et al[18] who validated their RF algorithm on genomic data, our study lacked external validation, a limitation shared by several predecessors.[13–17]

In our study, both K-PLA patients and non-K-PLA patients, mostly male, will have systemic infectious symptoms, most of them will take antibiotics by themselves before going to the hospital for treatment, and they are in the stage of abscess formation when they go to the hospital, which may be caused by the patients' lack of attention to their own health status, leading to the delay and progress of the disease. Increasing the difficulty and cost of follow-up treatment. This study found that compared with non-K-PLA, diabetic patients are more likely to suffer from liver abscess caused by single *Klebsiella pneumoniae*, which is consistent with the research results of Giorgio A. In the Chinese population,

Table 4 Performance Indicators of Four Machine Learning Models

Model	ACC	SEN	SPE	NPV	PPV	AUC
deeplearning trainning set	0.826	0.987	0.525	0.955	0.796	0.716
drf trainning set	0757	1.000	0.333	1.000	0.723	0.999
gbm trainning set	0.835	0.932	0.667	0.849	0.829	0.922
glm trainning set	0.757	0.918	0.476	0.769	0.753	0.718
deeplearning validation set	0.733	0.864	0.444	0.600	0.773	0.799
drf validation set	0.733	0.984	0.120	0.750	0.732	0.763
gbm validation set	0.826	0.912	0.600	0.750	0.849	0.848
glm validation set	0.802	0.934	0.480	0.750	0.814	0.805

Abbreviations: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; PPV, Positive predictive value; NPV, Negative predictive value; deeplearning, Deeplearning - Fully connected Neural Network; drf, Random forest; gbm, Gradient Intensifier; glm, Generalized Linear Model.

https://doi.org/10.2147/IDR.S545440

Klebsiella pneumoniae infection is closely related to diabetes,[19,20] because diabetes is a metabolism-related endocrine disease. Poor blood glucose control can selectively impair the phagocytic function of neutrophils of K1 and K2 capsular serotypes in *Klebsiella pneumoniae*, which can lead to secondary immune function reduction, and the increase and ectopic colonization of *Klebsiella pneumoniae* in the intestine can induce PLA.[19] This study shows that K-PLA patients have less history of infectious diseases, biliary tract diseases and liver cancer than non-K-PLA patients. Domestic scholars have found that the treatment bacteria of liver abscess patients with a history of infectious diseases (such as HIV infection) are mostly non-*Klebsiella pneumoniae*,[21] and different pathogenic bacteria spectrum may be related to the transmission route of infectious diseases.[22] A positive history of biliary tract disease can distinguish between hematogenous (*Klebsiella pneumoniae* predominated) and biliary abscess (mixed flora).[2,23] A positive history of liver cancer is associated with tumor-induced immunosuppression and the creation of *Klebsiella pneumoniae* niches by biliary obstruction.[24,25] In addition, patients with *Klebsiella pneumoniae* liver abscess are older, have a higher body temperature, and have a relatively shorter duration of fever, which have hardly been mentioned in previous studies. More relevant evidence is needed to support these characteristics. From the perspective of pathophysiological mechanism, a higher body temperature may reflect the strong endotoxin production of hypervirulent strains that triggers a strong pyrogen response.[26] The short duration of fever suggests that community-acquired K-PLA has the characteristics of acute onset rather than an indolent biliary abscess.[20,22] In terms of laboratory indicators, compared with non-*Klebsiella pneumoniae* PLA patients, *Klebsiella pneumoniae* PLA patients mostly have high ALT levels, which may indicate that hypermucoid strains directly attack hepatocytes, but there are inconsistencies between different studies.[25,26] Although the p-values in Table 1 were not significant, the Boruta algorithm retained ALT as the key predictor, probably due to the nonlinear associations captured by the ML model. Temperature differences lacked significance but were prioritized by feature selection, suggesting clinical relevance beyond traditional statistics. Of note, the negative association between K-PLA and history of biliary disease/infection highlights a unique etiology: *Klebsiella pneumoniae* utilizes intestinal translocation (promoted by diabetes) rather than biliary or iatrogenic routes.[22,23] This is consistent with our finding of thicker pus viscosity in K-PLA, which is a potential marker of the hypermucinous phenotype. In terms of ultrasound images, most of the patients in the two groups had normal liver background without underlying liver disease. Most of the lesions were single unilocular cystic structures located in the right lobe of the liver, with clear boundaries and thick abscess walls. In addition, Joyce's study[27] found that *Klebsiella pneumoniae* liver abscess was mostly solid in texture and complicated with fatty liver background, while this study did not find these characteristics, which may be caused by regional differences and population characteristics. Joyce's study population was distributed in Hong Kong, and the included population was more than those who went to the hospital for ultrasound examination in the early stage of abscess. However, our study population was distributed in the mainland of China, and most of the included population were in the abscess formation stage when they underwent ultrasound examination, and rarely came to the hospital in the pre-abscess or absorption stage. Therefore, this study did not conclude that *Klebsiella pneumoniae* liver abscess was mostly solid in texture and complicated with fatty liver background.

Although our model shows promise, there are limitations that should be noted: first, the single-center design limits generalability compared to the multi-center CT model.[14,17] In the future, we will expand the sample size, collect multicenter data, and supplement the external validation set to improve the accuracy and generality of the prediction model. (2) The heterogeneity of abscess stage (79.6% of formation stage) may affect the extraction of ultrasound features, which is inconsistent with Giorgio's solid appearance.[26] Therefore, to reduce the above bias, future prospective studies should be conducted to collect more samples before abscess formation and at the absorption stage. The clinical characteristics and ultrasound imaging findings of PLA patients were analyzed, and a stratified prediction model was established to further improve the prediction accuracy. Current studies have shown that PLA caused by different pathogens has different CT or MRI imaging findings.[2,25,28,29] Based on this finding, future attempts could be made to classify pathogens by different pathogenic bacteria or according to antibiotic sensitivity. To further investigate the ultrasonographic features of PLA caused by non-*Klebsiella pneumoniae*.

In conclusion, this study found that higher body temperature, short duration of fever, diabetes mellitus, history of liver cancer, no history of infectious diseases, no biliary tract disease, and high ALT level may be risk factors for K-PLA. A machine learning prediction model for K-PLA was established, which has good predictive ability and clinical

practicality. It is helpful for clinicians to pre-diagnose K-PLA and use targeted antibiotics as early as possible to improve the prognosis of patients. Due to the fact that the results of bacteriological culture often take 3 to 5 days to be clear, empirical broad-spectrum antibiotics are generally used in clinical practice, which may lead to the risk of antibiotic abuse and bacterial resistance. The above prediction model can diagnose K-PLA in advance, and also reduce the occurrence of adverse events such as antibiotic abuse to a certain extent. The detection of predictors in the prediction model only relies on clinical history and laboratory tests, which is more economical, convenient and time-sensitive than CT and MRI examinations, and can reduce the national economic burden to a certain extent.

Abbreviations

PLA, pyogenic liver abscess; K-PLA, *Klebsiella pneumoniae* pyogenic liver abscess; drf, Random Forest; gbm, Gradient Boosting Machine; glm, Generalized Linear Model; LASSO, least absolute shrinkage and selection operator; AUC, area under the curve; CT, computed tomography; MRI, magnetic resonance imaging.

Ethics Approval and Consent to Participate

This retrospective study was carried out using the opt-out method for the case series of our hospital. The study was approved by the Ethics Committee of the Fifth Medical Center of Chinese PLA General Hospital and was conducted in accordance with the 1964 Helsinki Declaration (approval number: KY-2024-10-157-1) and its later amendments or comparable ethical standards. Informed consent was waived by our Institutional Review Board because of the retrospective nature of our study.

Author Contributions

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

Disclosure

Haoran Li and Yan Yu are co-first authors for this study. Yuejuan Gao and Shanshan Wu are co-corresponding authors for this study. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Li S, Yu S, Peng M, et al. Clinical features and development of Sepsis in Klebsiella pneumoniae infected liver abscess patients: a retrospective analysis of 135 cases. *BMC Infect Dis*. 2021;21(1):597. doi:10.1186/s12879-021-06325-y
2. Boeckmans J, Rombaut M, Demuyser T, et al. Infections at the nexus of metabolic-associated fatty liver disease. *Arch Toxicol*. 2021;95(7):2235–2253. doi:10.1007/s00204-021-03069-1
3. Tsai FC, Huang YT, Chang LY, et al. Pyogenic liver abscess as endemic disease, Taiwan. *Emerg Infect Dis*. 2008;14(10):1592–1600. doi:10.3201/eid1410.071254
4. Lo JZ, Leow JJ, Ng PL, et al. Predictors of therapy failure in a series of 741 adult pyogenic liver abscesses. *J Hepato-Bil-Pan Sci*. 2014;22(2):156–165. doi:10.1002/jhbp.174
5. Wang H, Ren Y, Chang Z, et al. The increased recurrence rate of liver abscess caused by extended-spectrum β-lactamase-producing Klebsiella pneumoniae. *Eur J Clin Microbiol*. 2020;39(7):1315–1320. doi:10.1007/s10096-020-03848-1
6. Mavilia MG, Molina M, Wu GY. The evolving nature of hepatic abscess: a review. *J Clin Transl Hepato*. 2016;4(2):158–168. doi:10.14218/JCTH.2016.00004
7. Luo M, Yang T, Tan B, et al. Distribution of common pathogens in patients with pyogenic liver abscess in China: a meta-analysis. *Eur J Clin Microbiol Infect Dis*. 2016;35(10):1557–1565. doi:10.1007/s10096-016-2712-y
8. Lin YT, Wang FD, Wu PF, et al. Klebsiella pneumoniae liver abscess in diabetic patients: association of glycemic control with the clinical characteristics. *BMC Infect Dis*. 2013;13:56. doi:10.1186/1471-2334-13-56
9. Jun JB. Klebsiella pneumoniae Liver Abscess. *Infect Chemother*. 2018;50(3):210–218. doi:10.3947/ic.2018.50.3.210
10. Pei N, Liu X, Jian Z, et al. Genome sequence and genomic analysis of liver abscess caused by hypervirulent Klebsiella pneumoniae. 3 BIOTECH. *3 Biotech*. 2023;13(3):76. doi:10.1007/s13205-023-03458-6
11. Qian C, Zhang S, Mengxin X, et al. Genetic and phenotypic characterization of multidrug-resistant klebsiella pneumoniae from liver abscess. *Microbiol Spectr*. 2023;11(1):e0224022. doi:10.1128/spectrum.02240-22

12. Lee JH, Jang YR, Ahn SJ, et al. A retrospective study of pyogenic liver abscess caused primarily by Klebsiella pneumoniae vs. non-Klebsiella pneumoniae: CT and clinical differentiation. *Abdom Radiol*. 2020;45(9):2669–2679. doi:10.1007/s00261-019-02389-2
13. Wang H, Guo Y, Yan B, et al. Development and validation of a prediction model based on clinical and CT features for invasiveness of K. pneumoniae liver abscess. *Eur Radiol*. 2022;32(9):6397–6406. doi:10.1007/s00330-022-08740-4
14. Gu L, Ai T, Ye Q, et al. Development and validation of a clinical-radiomics nomogram for the early prediction of Klebsiella pneumoniae liver abscess. *Ann Med*. 2024;56(1):2413923. doi:10.1080/07853890.2024.2413923
15. Yuan T, Zhong T, Song J. Vascular penetration sign: dual-phase enhanced CT manifestations of atypical liver abscess caused by Klebsiella pneumoniae. *Eur Radiol*. 2025;35(8):4685–4691. doi:10.1007/s00330-025-11460-0
16. Lee JJ, Hong SB, Lee NK, et al. Characteristics of computed tomography for identifying patients at high risk of endogenous endophthalmitis due to klebsiella pneumoniae-related pyogenic liver abscess. *J Clin Med*. 2022;11(15). doi:10.3390/jcm11154376
17. Feng C, Di J, Jiang S, et al. Machine learning models for prediction of invasion Klebsiella pneumoniae liver abscess syndrome in diabetes mellitus: a singled centered retrospective study. *BMC Infect Dis*. 2023;23(1):284. doi:10.1186/s12879-023-08235-7
18. Lan P, Shi Q, Zhang P, et al. Core genome allelic profiles of clinical klebsiella pneumoniae strains using a random forest algorithm based on multilocus sequence typing scheme for hypervirulence analysis. *J Infect Dis*. 2020;221(Suppl 2):S263–S271. doi:10.1093/infdis/jiz562
19. Wang JH, Liu YC, Lee SS, et al. Primary liver abscess due to Klebsiella pneumoniae in Taiwan. *Clin Infect Dis*. 1998;26(6):1434–1438. doi:10.1086/516369
20. Lardière-Deguelte S, Ragot E, Amroun K, et al. Hepatic abscess: diagnosis and management. *J Visc Surg*. 2015;152(4):231–243. doi:10.1016/j.jviscsurg.2015.01.013
21. Yoon JH, Kim YJ, Jun YH, et al. Liver abscess due to Klebsiella pneumoniae: risk factors for metastatic infection. *Scand J Infect Dis*. 2013;46(1):21–26. doi:10.3109/00365548.2013.851414
22. Aston SJ. Pneumonia in the developing world: characteristic features and approach to management. *Respirology*. 2017;22(7):1276–1287. doi:10.1111/resp.13112
23. Chen YC, Lin CH, Chang SN, et al. Epidemiology and clinical outcome of pyogenic liver abscess: an analysis from the National Health Insurance Research Database of Taiwan, 2000-2011. *J Microbiol Immunol*. 2014;49(5):646–653. doi:10.1016/j.jmii.2014.08.028
24. Li F, Zhang H, Xu Y, et al. Clinical and CT comparative study of invasive and non-invasive Klebsiella pneumoniae liver abscesses. *Clin Radiol*. 2022;78(1):40–46. doi:10.1016/j.crad.2022.08.145
25. Alsaif HS, Venkatesh SK, Chan DS, et al. CT appearance of pyogenic liver abscesses caused by klebsiella pneumoniae. *Radiology*. 2011;260(1):129–138. doi:10.1148/radiol.11101876
26. Liu Y, Wang JY, Jiang W. An increasing prominent disease of klebsiella pneumoniae liver abscess: etiology, diagnosis, and treatment. *Gastroent Res Pract*. 2013;258514. doi:10.1155/2013/258514
27. Giorgio A, Tarantino L, Mariniello N, et al. Pyogenic liver abscesses: 13 years of experience in percutaneous needle aspiration with US guidance. *Radiology*. 1995;195(1):122–124. doi:10.1148/radiology.195.1.7892451
28. Ribeiro Da Costa R, Andres A, Huttner B. Pyogenic liver abscesses. *Rev Med Suisse*. 2020;16(708):1822–1826. PMID: 32997454.
29. Lee CC, Chen CY, Chen FH, et al. Septic metastatic endophthalmitis from Klebsiella pneumoniae liver abscess: CT and MR imaging characteristics--report of three cases. *Radiology*. 1998;207(2):411–416. doi:10.1148/radiology.207.2.9577489