

RESEARCH

Open Access



Revolutionizing Wilson disease prognosis: a machine learning approach to predict acute-on-chronic liver failure

Zhihong Rao¹, Wenming Yang^{1,2,3*}, Yulong Yang¹, Yue Yang¹, Yongsheng Han^{6,7}, Xiang Li¹, Siyang Lu¹, Yuchen Li¹, Hu Xi⁴, Ke Diao⁵, Shuzhen Fang¹, Wei He¹ and Sihuan Zhu⁸

Abstract

Background and objectives Wilson disease (WD), an inherited copper metabolism disorder, is a cause of acute-on-chronic liver failure (ACLF), posing life-threatening risks due to rapid progression. This study aimed to develop a machine learning (ML)-based model to predict ACLF risk in WD patients.

Methods We retrospectively analyzed 3692 WD patients (Leipzig score ≥ 4) from The First Affiliated Hospital of Anhui University of Chinese Medicine (2014–2024), including 104 ACLF and 104 non-ACLF cases. The original data set was randomly divided into the training and test cohorts in a ratio of 7:3. Demographic, biochemical, and ultrasound data were collected. Six ML algorithms (LR, SVM, KNN, ExtraTrees, XGBoost, LightGBM) were applied to construct a predictive model, with SHAP explaining feature importance.

Results The XGBoost model achieved optimal performance (AUC: 0.998, accuracy: 0.968). Key predictors included TBA, APTT, diagnosis age, onset age, Hb. Elevated TBA, APTT and diagnosis age correlated with higher ACLF risk, while reduced onset age and Hb indicated poorer outcomes. Additional parameters (TT, Cl^- , CER and hepatic imaging features) also contributed modestly to predictions.

Conclusions The ML-based model effectively predicts WD-ACLF risk, with XGBoost demonstrating superior performance. TBA, APTT, diagnosis age, onset age and Hb emerged as critical biomarkers, offering actionable insights for early clinical intervention.

Keywords Wilson disease, Acute-on-chronic liver failure, Machine learning, XGBoost, SHAP

*Correspondence:

Wenming Yang
yangwm8810@126.com

¹Department of Neurology, The First Affiliated Hospital of Anhui University of Chinese Medicine, 117 Meishan Road, Hefei 230031, Anhui, China

²Center for Xin'an Medicine and Modernization of Traditional Chinese Medicine, Institute of Health and Medicine Hefei Comprehensive National Science Center, Hefei, Anhui, China

³Key Laboratory of Xin'an Medicine, Ministry of Education, Hefei, Anhui, China

⁴Wangjing Hospital, China Academy of Chinese Medical Sciences, Beijing, China

⁵Resource Center for Chinese Medicinal Materials, China Academy of Chinese Medical Sciences, Beijing, China

⁶Institute of Neurology, Anhui University of Chinese Medicine, Hefei, Anhui, China

⁷Wannan Medical College, Wuhu, Anhui, China

⁸Anhui University of Chinese Medicine, Hefei, Anhui, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Wilson disease (WD) is an autosomal recessive disorder caused by a pathogenic variant in the ATP7B gene encoding a copper transporter protein. The main mechanism is the impairment of hepatic cuprocyanin synthesis and biliary copper excretion [1]. Its most common pathological changes include pathologic deposition of copper ions in the liver, basal ganglia and kidney, and this imbalance in metal homeostasis can induce progressive tissue damage [2]. About 62–85% of WD patients may manifest symptoms of liver disease, with highly heterogeneous clinical outcomes. Although early chelating agents can strongly improve the prognosis of most patients, 8–15% of WD patients may develop acute-on-chronic liver failure (ACLF), a complication with a 28-day mortality rate of approximately 85–95%, recognized as the leading cause of WD-related deaths [3]. A decade-long retrospective study (2014–2024) conducted at the First Affiliated Hospital of Anhui University of Chinese Medicine showed a 2.8% incidence of WD-ACLF among 3692 WD patients, which presented similar epidemiological characteristics to HBV-ACLF. However, few therapeutic options for WD-ACLF are available. Although liver transplantation can improve the survival outcome of patients, its application is limited by donor shortage and the risk of immunosuppression. Traditional scoring systems such as model for end-stage liver disease or Child–Pugh are not specific enough for WD-ACLF, which makes prognosis prediction difficult. In addition, WD-ACLF presents with a rapid onset and a short diagnostic window period. Therefore, a biomarker-driven early warning assessment model for WD-ACLF was constructed to guide the application of interventions through a multidimensional clinical management strategy, thereby improve patient survival outcomes and minimize the need for liver transplantation.

Machine learning (ML) techniques can facilitate risk stratification of complex diseases by integrating multimodal data and dynamic time-series analysis. In the field of neurodegenerative diseases, deep learning architectures have transitioned from the traditional diagnostic approaches to the precision medicine paradigm. For example, in Parkinson's disease and Alzheimer's disease, ML models have significantly improved the accuracy of early diagnosis and the feasibility of individualized treatment by integrating multimodal data (e.g., imaging, genomics, and clinical indicators) [4, 5]. A previous study performed ML-based risk stratification for WD cirrhosis [6]. Other studies on ACLF have mainly focused on the viral and alcoholic liver disease aspects, and no model has been developed to facilitate prediction of the development of WD-ACLF, which delays the initiation of clinical intervention and increases the mortality rate of WD-ACLF.

In this study, we integrated artificial intelligence algorithms with multimodal clinical data (covering demographic baseline, hepatobiliary, pancreatic and splenic ultrasound, blood and urine tests, etc.) to identify the most important indicators associated with WD-ACLF and to construct predictive models for the occurrence of WD-ACLF. This dynamic stratification model addresses inherent constraints of conventional static scoring systems through precise identification of high-risk cohorts, facilitating targeted interventions to reduce mortality and improve quality of life in Wilson disease.

Methods

Data collection and processing

In this study, a retrospective cohort analysis was conducted using a multidimensional data collection approach to obtain data of neurological inpatients at the First Affiliated Hospital of Anhui University of Chinese Medicine from an electronic medical record system. The search involved the use of terms such as “Hepatolenticular degeneration” or “Wilson disease”, combined with the International Classification of Diseases (ICD), with the relevant ICD-10 code of E83.0, and the time window was from January 2014 to December 2024. A total of 3692 patients with WD were enrolled in the study, among which 104 patients with WD-ACLF were included. To achieve number matching, 104 patients without ACLF were randomly selected and assigned to the control group. Twenty cases of WD-ACLF and 25 cases of WD-nonACLF hospitalised at the Affiliated Hospital of the Institute of Neurology, Anhui University of Traditional Chinese Medicine, between January 2022 and December 2024, were used to validate the model's external validity. All patients were given a unique hospitalization identification code to ensure the independence and traceability of case data. The study was reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Anhui University of Chinese Medicine (No. 2024AH-13), and was performed according to the requirements of the Declaration of Helsinki by employing the triple desensitization (ID substitution, date offset, and blurring of sensitive information). The ethics committee waived the requirements of informed consent and clinical trial registration. To mitigate the impact of invalid variables on model calculations, relevant literature was identified to select statistically significant variables in each study, and variables that were closely associated with disease from the institution's existing programmes were combined. Some baseline data of the patients, liver and kidney function, blood routine, coagulation function, ceruloplasmin (CER), 24-h urine copper and liver, gallbladder and spleen ultrasound, etc., which were finally selected as clinical observation indicators. Liver and kidney function, blood routine, coagulation function, CER and other indicators were

examined on the day the disease appeared or the following day.

Case selection criteria

The WD inclusion criteria were based on the EASL 2012 Clinical Practice Guidelines [7], as well as a Leipzig Scale assessment, which was independently performed by two neurologists of associate title or above, with a total score of ≥ 4 on the scale. Liver diseases caused by factors such as hepatitis, schistosomiasis infection and alcohol poisoning were excluded from the study. At the same time, basic consistency in treatment modalities was maintained for the included patients. This included the use of zinc gluconate tablets and chelators, while irregular use was excluded.

Research design and predictive variables

The endpoint indicator of our study was the occurrence of ACLF, diagnosed based on the APASL 2019 version of the diagnostic criteria [8]. The ACLF group was further stratified into the ACLF and non-ACLF groups following the ACLF diagnosis within 72 h of admission. Moreover, baseline parameters, laboratory indicators, and hepatobiliary, pancreatic, and splenic ultrasound were used as predictors.

Data preprocessing

During the preprocessing stage of data, outliers in the dataset were screened and excluded, and missing values were estimated. Regarding the variables subtypes, we used the median filling method for estimation, while for continuous variables, we randomly selected values near the mean to be filled within the range of variance. To improve accuracy and reliability of the estimated data, the estimation error was controlled within 3% by comparing the key statistical indicators such as mean, median and standard deviation before and after estimation. To further verify the robustness of the estimated data, a ML cross-validation methodology was employed to ensure the scientific nature of the data processing and enhance the reliability of the prediction model through multiple iterations of validation.

The original data set was randomly divided into training and test datasets at the ratio of 7:3 (See supplementary Table 1), in which all samples in the training dataset were employed to construct the model and perform parameter optimization, while the test dataset was used as independent validation samples to objectively determine the actual predictive capacity and generalizability of the model. This division was adopted to ensure that the model could fully learn the data features during the training process, and the independent test dataset validation effectively avoids the model overfitting

phenomenon, eventually enhancing the reliability and scientific rigor of the evaluation results.

Establishments of the clinical model

In this study, a multi-stage modelling strategy was employed to construct the WD-ACLF risk prediction model. Initially, we screened the predictors through step-wise regression: one-factor logistic regression for initial screening, followed by multifactor regression modelling. Next, we integrated the logistic regression (LR), support vector machine (SVM), K-nearest neighbors (KNN), Extremely randomized trees (Extra Trees), eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) six algorithms to develop the composite model. The selected algorithms encompass diverse modeling approaches to capture complex data patterns and enhance predictive accuracy. Specifically: SVM excel in high-dimensionality and small-sample scenarios due to their maximum-margin classification principle. LightGBM/XGBoost leverage decision tree ensembles for superior performance on heterogeneous data, with accelerated training through histogram-based optimization. KNN provides a non-parametric, interpretable baseline method. Random Forest and Extra-Trees fortify model resilience by aggregating multiple decorrelated decision trees. This comprehensive evaluation protocol ensures robust identification of optimal WD-ACLF predictors. Using stratified randomization, the dataset was partitioned 7:3 into training (model development) and testing (independent performance validation) subsets. Parameter tuning was performed using the five-fold cross-validation in the modelling phase, and the final model discriminative efficacy was determined based on the AUC value of the ROC curve. To identify the key variables, the contribution of each indicator to the prediction results was quantified using a feature correlation matrix, which visualizes the strength of the role of key biomarkers. This multidimensional validation system enhanced the model's goodness-of-fit on the training dataset and validated its clinical utility through independent testing. Leveraging SHapley Additive exPlanations (SHAP) values, we quantified the marginal contribution of individual predictors to model outputs and visualized non-linear feature-target relationships through dependency plots, overcoming limitations of correlation-based interpretation.

Model evaluation indicators

The optimal model configuration after hyper-parameter optimization was employed for the training, and efficacy validation was implemented in an independent validation cohort. The model evaluation system consisted of six dimensions of clinical efficacy: sensitivity (Sn) reflects the ability to detect diseases, specificity (Sp) measures the accuracy of identifying healthy samples, positive

predictive value (PPV) and negative predictive value (NPV) characterize the clinical credibility of diagnostic results. The AUC is calculated to comprehensively assess the discriminative power of the classifier, while accuracy (Acc) serves as a global performance benchmark. The mathematical expression formula for each index are shown below:

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

$$S_n = \frac{TN}{TN + FP}$$

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

$$NPV = \frac{TN}{TN + FN}$$

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

Statistics

In this study, we performed data analysis using Python statistical modelling tool library statsmodels (V0.13.2), and statistically significant associations between the variables of interest were determined at $P < 0.05$. For numerical variables, data distribution characteristics were first tested for normality, and the unpaired t-test

was employed if the assumption of normality was met. Variables that did not show normal distribution were analyzed using the Mann–Whitney U-test and presented as the median (interquartile spacing). For categorical variables, the Pearson chi-square test was applied when the expected frequency was equal to or greater than 5, and the Fisher exact probability method was utilized when the cell expectation was less than 5.

Results

Clinical features of patients

The flow chart of this study is shown in Fig. 1. A total of 568 patients with WD were enrolled in the study, among which 104 patients had ACLF. In addition, 104 patients without ACLF were randomly selected and assigned to the control group, 111 (53.3%) were male patients, 97 (46.6%) were female patients, the male to female ratio was not statistically significant ($P > 0.05$). The first symptoms were divided into hepatic and cerebral types, with 148 patients (71.1%) having the hepatic type and 60 patients (28.8%) with the cerebral type. In the WD-ACLF group, 75 patients (72.1%) had the hepatic type and 29 patients (27.8%) had the cerebral type. Compared with the WD non-ACLF group, patients in the WD-ACLF group were significantly older age at the onset age (18.5 years vs. 11.5 years, $P < 0.0001$) and at the diagnosis age (20 years vs. 12 years, $P < 0.0001$), with a higher proportion of males (54.8% vs. 45.1%), and a longer disease duration (11.5 years vs. 8 years, $P < 0.0001$). In addition, there were more patients with abnormal hepatic volume

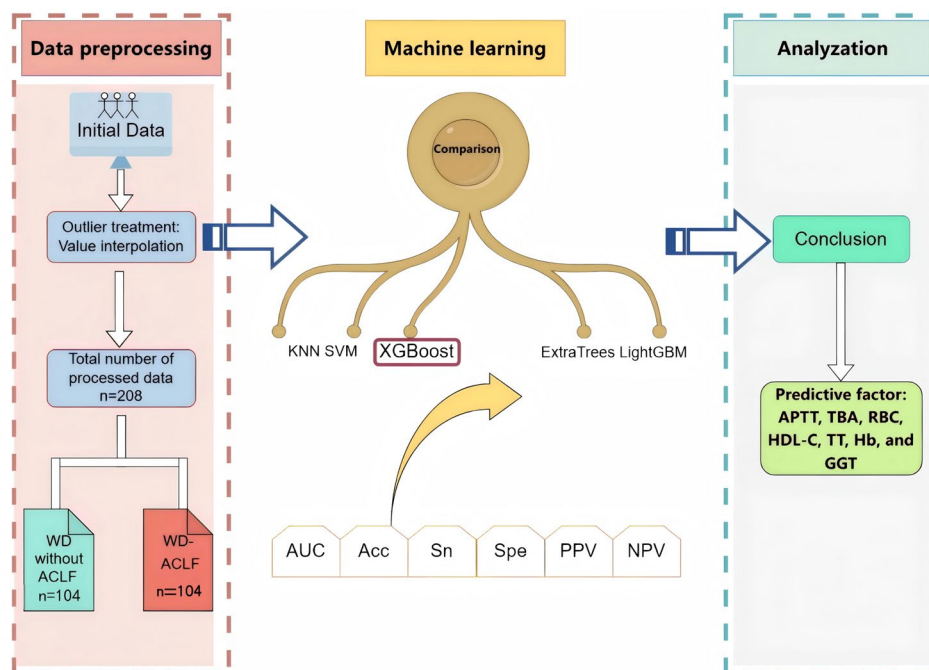


Fig. 1 The flowchart of this study

in the WD-ACLF group (52.8% vs. 0.9%, $P < 0.0001$), and hepatic capsule was significantly different between the groups (81.7% vs. 13.4%, $P < 0.0001$). The distribution of Hepatic Parenchymal Echogenicity was significantly different among the groups (100% vs. 77.8%, $P < 0.0001$), and most of the patients had splenomegaly (96.2% vs. 38.5%, $P < 0.0001$), exhibited a widened Portal Vein Diameter (11.63 vs. 10.29, $P < 0.0001$). Notably, the ALB (28.12 vs. 42.44, $P < 0.0001$); RBC (2.96 vs. 4.53, $P < 0.0001$), Hb (97.5 vs. 130.5, $P < 0.0001$), and PLT (86.5 vs. 203, $P < 0.0001$) were significantly decreased, probably due to the anemia caused by liver failure. The concentrations of PT (26 vs. 32, $P < 0.0001$), FBG (1.14 vs. 2.29, $P < 0.0001$) were decreased, APTT (42.9 vs. 11.3, $P < 0.0001$) and TT (23.55 vs. 20.3, $P < 0.0001$) were increased, indicating that the coagulation function was affected. Further analysis revealed that the liver function markers (e.g., DBIL, IBIL, TBA, ALT, AST, GGT, LDH, and ALP) were significantly elevated, but ALP, TG, TC HDL-C, and LDL-C were significantly decreased. Liver fibrosis markers (PIIINP, LN, HA, CIV) were significantly increased, while the 24-h urine copper concentration was elevated (696 vs. 657.7, $P = 0.0071$) and CER was also increased (0.079 vs. 0.022, $P < 0.0001$). The detailed parameters are shown in Table 1. These variable can be used to predict the risk of WD-ACLF (See supplementary Table 2).

Model performance

The reliability of the constructed model was tested in a rigorously designed cohort. Notably, there was no gender difference in ACLF patients and non-ACLF patients, indicating that the interference of confounding factors was balanced between the groups. Based on multimodal clinical data (covering baseline parameters, laboratory indicators, hepatobiliary-pancreatic-spleen ultrasound, etc.), the predictive efficacy of six algorithms (LR, Extra Trees, LightGBM, SVM, XGBoost, and KNN) was compared using an integrated learning framework. Analysis of the experimental data showed that the LR test dataset had an AUC of 0.962 and an accuracy of 0.937, indicating that the LR model could effectively discriminate between the target classes. The AUC and accuracy of the training dataset were close to 1, and while the sensitivity of test dataset was 0.935 and specificity was 0.937, with a PPV and NPV of 0.935, indicating that the model was balanced in the positive and negative class prediction, avoiding the risk of misjudgment due to bias. We further observed that the XGBoost algorithm showed optimal overall performance in the independent validation set: as shown in Table 2, the test P values of XGBoost in the training and test datasets are 0.968 and 0.631 (threshold $P > 0.05$), respectively, indicating that its predicted probability is highly consistent with the actual event distribution and has excellent calibration performance. The

integrated calibration curve indicated that the predicted values were in good agreement with the observed values between the training and testing cohorts of the XGBoost algorithm. These data suggested that the XGBoost model exhibited optimal overall performance in the training and test cohorts, and the calibration curves in both cohorts is shown in Fig. 2a, b. The area under the receiver operating characteristic curve of the working characteristics of the subjects was 0.998 (95% confidence interval (CI) 0.993–1.000) (Fig. 2c, d), the Acc was 0.968 (95% CI 0.953–1.015), the Sn and Sp were 0.968 and 0.969, respectively, while the positive and negative predictive values were 0.968 and 0.969, respectively (see Table 3 for details). The decision curve analysis (DCA) was adopted to assess each model. As shown in Fig. 2e, f, DCA revealed that XGBoost had the best performance, with a significantly higher net benefit compared with the other methods in the range of medium and high-risk thresholds. This confirmed that XGBoost could effectively integrate non-linear associations between copper metabolic abnormalities and organ function indicators when dealing with multidimensional medical data, providing a high-precision tool for early warning detection of WD-ACLF. Furthermore, to verify the model's general applicability, external validation was conducted in this study. This demonstrated that XGBoost still performed optimally, with an area under the curve (AUC) of 0.974 (95% CI 0.928–1.000), an Acc of 0.956, a Sn of 0.920 and a Sp of 1.000, as well as a PPV and NPV of 1.000 and 0.909 (Fig. 3 and Table 4).

Identification of predictive factors for the WD-ACLF prediction model

Based on the SHAP interpretability framework, the feature contribution of XGBoost classifiers was analyzed from the global and individual perspectives. The wider the distribution of the region, the greater its impact (detailed information is shown in Fig. 4A). Then, the average absolute value of SHAP value of each feature was calculated by combining the degree of influence of each feature in the machine learning model on the prediction results and plotted as a bar chart (detailed information is shown in Fig. 4B). By quantifying the contribution of the features, the influence ranking of each feature in the machine learning model was visually demonstrated, which provided an important basis for the model's interpretability. Based on Gini impurities, the most important predictors in the prediction model were found to be TBA, APTT, diagnosis age, onset age, Hb.

Discussion

In this retrospective case–control study design, a ML prediction model for WD-ACLF was constructed for the first time, which underscored the utility of the integration

Table 1 Comparison of clinical characteristics between the two groups of WD patients

Feature name	WD-ACLF	WD non-ACLF	P
N	104	104	
Sex (man/woman)	57/47	54/50	0.6767
Initial presentation type (Cerebral type/Hepatic type)	29/75	31/73	0.7595
Age	33	19	< 0.0001***
median [IQR], year	[27, 40]	[14, 32]	
Onset age	18.5	11.5	< 0.0001***
median [IQR], year	[13, 29]	[5.75, 21]	
Diagnosis age	20	12	< 0.0001***
median [IQR], year	[14, 30]	[6, 21]	
Disease duration	11.5	8	< 0.0001***
median [IQR], year	[3.75, 19]	[4.75, 13]	
Ascites	17	0	< 0.0001***
median [IQR], mm	[0, 48.5]	[0]	
Hepatic volume (Normal/Abnormal)	49/55	103/1	< 0.0001***
Hepatic capsule (Smoothing/Non-smoothing)	19/85	90/14	< 0.0001***
Coarse and hyperechoic liver parenchyma (Yes/No)	103/1	10/94	0.0053**
Distribution of Hepatic Parenchymal Echogenicity (Regularity/Irregularity)	0/104	23/81	< 0.0001***
Splenomegaly (Yes/No)	100/4	40/64	< 0.0001***
Portal Vein Diameter (mm)	11.63 ± 2.03	10.29 ± 2.08	< 0.0001***
ALB (g/L)	28.12 ± 6.05	42.44 ± 3.86	< 0.0001***
LDL-C (mmol/L)	1.67 ± 0.84	2.57 ± 0.58	< 0.0001***
RBC (10 ¹² /L)	2.96 ± 0.7	4.53 ± 0.57	< 0.0001***
PT	26	32	< 0.0001***
median [IQR], s	[19.68, 39.58]	[30.3, 33.7]	
APTT	42.9	11.3	< 0.0001***
median [IQR], s	[25.85, 65.55]	[10.8, 11.9]	
TT	23.55	20.3	< 0.0001***
median [IQR], s	[20.38, 25.13]	[19.58, 20.9]	
FBG	1.14	2.29	< 0.0001***
median [IQR], g/L	[0.86, 1.58]	[2.02, 2.59]	
Cr	61.45	52.75	0.0008***
median [IQR], umol/L	[50.43, 77.05]	[45, 64.1]	
UA	134.5	287	< 0.0001***
median [IQR], umol/L	[90.5, 196.25]	[220.25, 362]	
DBIL	70.2	2.5	< 0.0001***
median [IQR], umol/L	[48.43, 143.88]	[1.9, 3.3]	
IBIL	70.86	9.1	< 0.0001***
median [IQR], umol/L	[48.93, 107.7]	[6.9, 12.3]	
TBA	120.7	6.55	< 0.0001***
median [IQR], umol/L	[61.55, 193.85]	[4.48, 11]	
ALT	48	30.75	0.0016**
median [IQR], U/L	[31.75, 75.4]	[17.7, 72.8]	
AST	85	27.95	< 0.0001***
median [IQR], U/L	[54.8, 119.78]	[21.15, 45.2]	
GGT	68.5	29	< 0.0001***
median [IQR], U/L	[36.75, 163]	[18, 51]	
ALP	139	120.5	0.2636
median [IQR], U/L	[124, 194.25]	[90, 244]	
LDH	259	183.5	< 0.0001***
median [IQR], U/L	[210.25, 335.5]	[154.75, 212]	
TG	0.63	1.13	< 0.0001***
median [IQR], mmol/L	[0.46, 0.86]	[0.81, 1.61]	
TC	2.06	4.285	< 0.0001***
median [IQR], mmol/L	[1.39, 3.15]	[3.68, 4.92]	
HDL-C	0.38	1.255	< 0.0001***
median [IQR], mmol/L	[0.24, 0.66]	[1.11, 1.43]	

Table 1 (continued)

Feature name	WD-ACLF	WD non-ACLF	P
HCY	18.15	10.2	< 0.0001***
median [IQR], umol/L	[11.75, 27.18]	[7.78, 12.9]	
Na ⁺	136.5	139.9	< 0.0001***
median [IQR], mmol/L	[131.95, 139]	[138.88, 140.58]	
Cl ⁻	107.4	106.2	0.1226
median [IQR], mmol/L	[103.3, 135.85]	[105, 108]	
Hb	97.5	130.5	< 0.0001***
median [IQR], g/L	[86, 111.25]	[121.75, 143]	
PLT	86.5	203	< 0.0001***
median [IQR], 10 ⁹ /L	[43.75, 136.5]	[132, 283]	
RDW-CV	17.8	13.2	< 0.0001***
median [IQR], %	[15.9, 19.8]	[12.16, 13.8]	
CER	0.08	0.02	< 0.0001***
median [IQR], g/L	[0.05, 0.1]	[0.02, 0.05]	
PIIINP	19.91	12.59	< 0.0001***
median [IQR], ng/ml	[16.55, 32.71]	[10.11, 17.09]	
LN	208	126.5	< 0.0001***
median [IQR], ng/ml	[166.08, 358.28]	[82.46, 169.86]	
HA	896.5	93.95	< 0.0001***
median [IQR], ng/ml	[194.58, 1382.81]	[64.92, 147.36]	
CIV	237.9	65.55	< 0.0001***
median [IQR], ng/ml	[95.44, 395.17]	[52.11, 85.06]	
WB-Cu	4.88	4.88	< 0.0001***
median [IQR], umol/L	[4.88, 9.28]	[1.92, 4.88]	
24 h urine copper	696	657.7	0.0071**
median [IQR], ug/24 h	[512.78, 1440.46]	[431.2, 924.5]	

N, number; ALB, albumin; LDL-C, low-density lipoprotein cholesterol; RBC, red blood cell; PT, prothrombin time; APTT, activated partial thromboplastin time; TT, thrombin time; FBG, fibrinogen; Cr, creatinine; UA, uric acid; DBIL, direct bilirubin; IBIL, indirect bilirubin; TBA, total bile acid; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, γ-Glutamyl Transferase; ALP, alkaline phosphatase; LDH, lactate dehydrogenase; TG, Triglycerides; TC, Total Cholesterol; HDL-C, High-Density Lipoprotein Cholesterol; HCY, homocysteine; Na⁺, Serum Sodium; Cl⁻, Serum Chloride; Hb, hemoglobin; PLT, platelet count; RDW-CV, Red Cell Distribution Width-Coefficient of Variation; CER, Ceruloplasmin; PIIINP, Procollagen III N-terminal Peptide; LN, Laminin; HA, Hyaluronic Acid; CIV, Collagen Type IV; WB-Cu, Whole Blood Copper

*P < 0.05; **P < 0.01; ***P < 0.001

Table 2 Hosmer_lemeshow_test in the training and test dataset

Model	Train P value	Test P value
LR	1.0	0.0
KNN	0.972	0.024
ExtraTrees	0.019	0.056
XGBoost	0.968	0.631
LightGBM	0.002	0.009
rbf_SVM	0.940	0.237
linear_SVM	0.995	0.010
sigmoid_SVM	0.395	0.008
poly_SVM	0.705	0.581

of multidimensional clinical data and algorithmic innovation level in clinical decision-making.

The predictive performance of six ML algorithms (LR, ExtraTrees, LightGBM, SVM, XGBoost, and KNN), was compared and results revealed that the XGBoost model had excellent clinical applicability, with an AUC value of 0.998 (95% CI 0.993–1.000), a classification accuracy of 96.8%, a sensitivity and specificity of over 96%, a positive score of over 1,000, and a balanced distribution of positive/negative predictive values exceeding 96%.

Its good performance may be due to its unique technical architecture, i.e., the gradient boosting framework effectively captures the nonlinear associations among clinical parameters by integrating weak classifiers; the regularization strategy combines the L1/L2 penalty terms to precisely control the model's complexity, and the adaptive missing value processing mechanism significantly improves the compatibility of non-completeness of clinical data. Notably, the ExtraTrees cannot effectively identify novel predictors due to its over-reliance on traditional liver disease indicators such as ALT and TBIL (feature contribution > 60%), a phenomenon that reveals the complex association between algorithmic feature selection preferences and disease heterogeneity, providing important insights that will guide future algorithm optimization.

Using the feature attribution analysis of Shapley's additive interpretation algorithm, a multidimensional prediction system consisting of coagulation function, bile metabolism and hematological indexes, etc. was established for the first time, which demonstrated that TBA, APTT, diagnosis age, onset age and Hb were the core

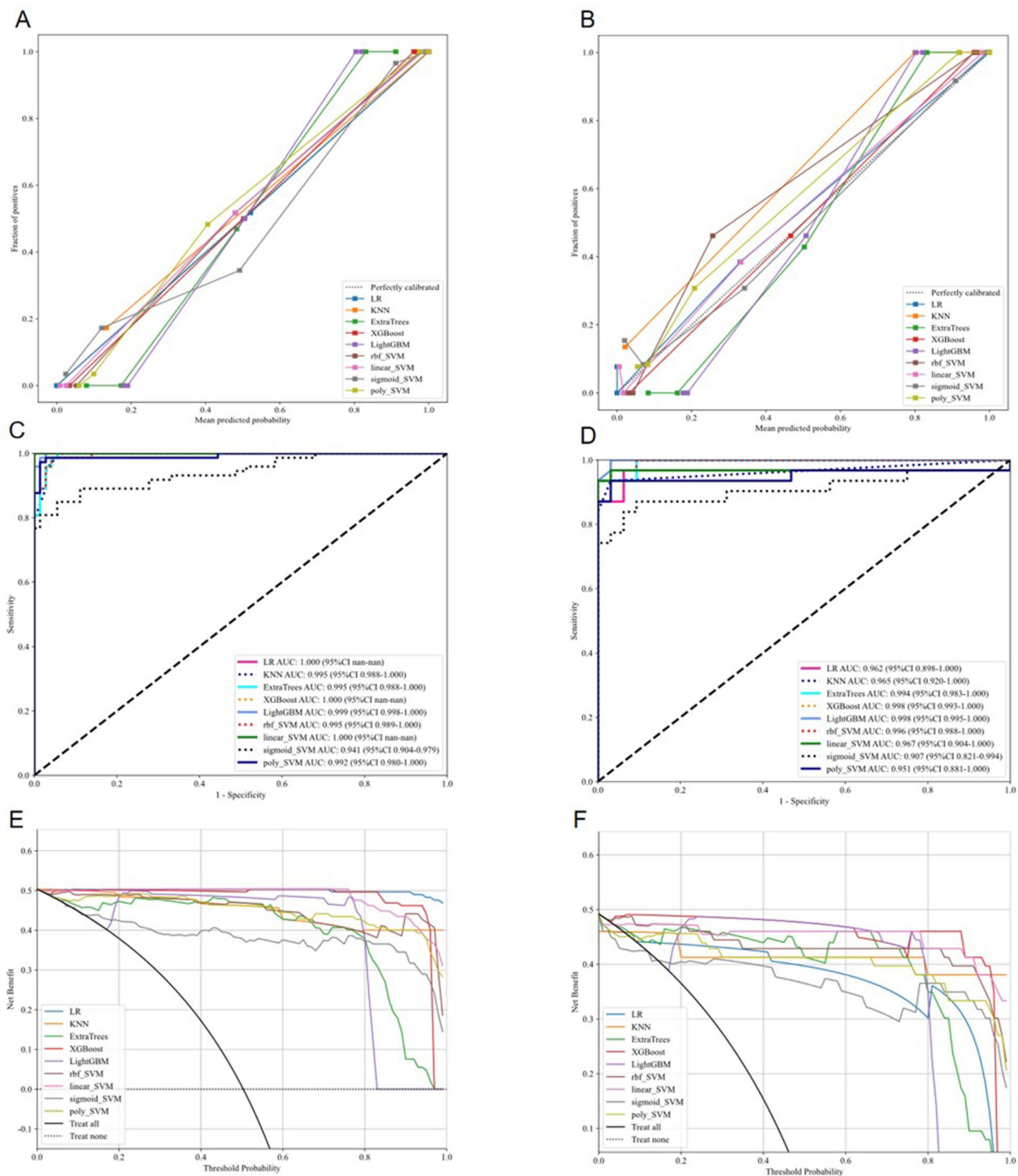


Fig. 2 The calibration curves, AUC and decision curve of the training and test cohorts. **a** The calibration curves of the training cohort. **b** The calibration curves of the test cohort. **c** The AUC of the training cohort. **d** The AUC of the test cohort. **e** Decision curve for the training cohort. **f** Decision curve for the test cohort

predictor combinations. Specifically, the TBA, APTT and diagnosis age values were significantly and positively associated with increased risk of developing WD-ACLF, whereas onset age and Hb were negatively correlated with WD-ACLF. In addition, APTT and TBA were identified

as the core drivers, and the elevated APTT values in the WD-ACLF group than in the control group may be related to the fact that copper toxicity disrupts coagulation homeostasis through a dual mechanism, i.e., direct inhibition of hepatic synthesis of coagulation factors II/

Table 3 Machine learning modeling analysis in the train and test dataset

Cohort	Model	AUC	AUC 95% CI	Acc	Acc 95% CI	Sen	Spe	PPV	NPV
Train	LR	1.000	1.0000–1.0000	0.993	1.0000–1.0000	0.986	1.000	1.000	0.986
Test	LR	0.962	0.8980–1.0000	0.937	0.8323–0.9772	0.935	0.937	0.935	0.937
Train	KNN	0.995	0.9879–1.0000	0.959	0.9262–0.9910	0.945	0.972	0.972	0.946
Test	KNN	0.965	0.9197–1.0000	0.921	0.8539–0.9874	0.839	1.000	1.000	0.865
Train	Extra Trees	0.995	0.9884–1.0000	0.966	0.9358–0.9952	0.973	0.958	0.959	0.972
Test	Extra Trees	0.994	0.9834–1.0000	0.952	0.8539–0.9874	0.903	1.000	1.000	0.914
Train	XGBoost	1.000	1.0000–1.0000	0.993	0.9796–1.0066	0.986	1.000	1.000	0.986
Test	XGBoost	0.998	0.9932–1.0000	0.968	0.9533–1.0150	0.968	0.969	0.968	0.969
Train	LightGBM	0.999	0.9978–1.0000	0.979	0.9561–1.0025	0.973	0.986	0.986	0.973
Test	LightGBM	0.998	0.9948–1.0000	0.968	0.9533–1.0150	0.968	0.969	0.968	0.969
Train	poly_SVM	0.992	0.9797–1.0000	0.972	0.9077–0.9820	0.959	0.986	0.986	0.959
Test	poly_SVM	0.951	0.8814–1.0000	0.937	0.8539–0.9874	0.903	0.969	0.966	0.912
Train	rbf_SVM	0.995	0.9886–1.0000	0.966	0.9168–0.9866	0.973	0.958	0.959	0.972
Test	rbf_SVM	0.996	0.9882–1.0000	0.952	0.8763–0.9967	0.935	0.969	0.967	0.939
Train	linear_SVM	1.000	1.0000–1.0000	0.993	1.0000–1.0000	0.986	1.000	1.000	0.986
Test	linear_SVM	0.967	0.9035–1.0000	0.952	0.9250–1.0115	0.935	0.969	0.967	0.939
Train	sigmoid_SVM	0.941	0.9036–0.9792	0.890	0.8304–0.9351	0.795	0.986	0.983	0.826
Test	sigmoid_SVM	0.907	0.8207–0.9938	0.873	0.8113–0.9665	0.839	0.906	0.897	0.853

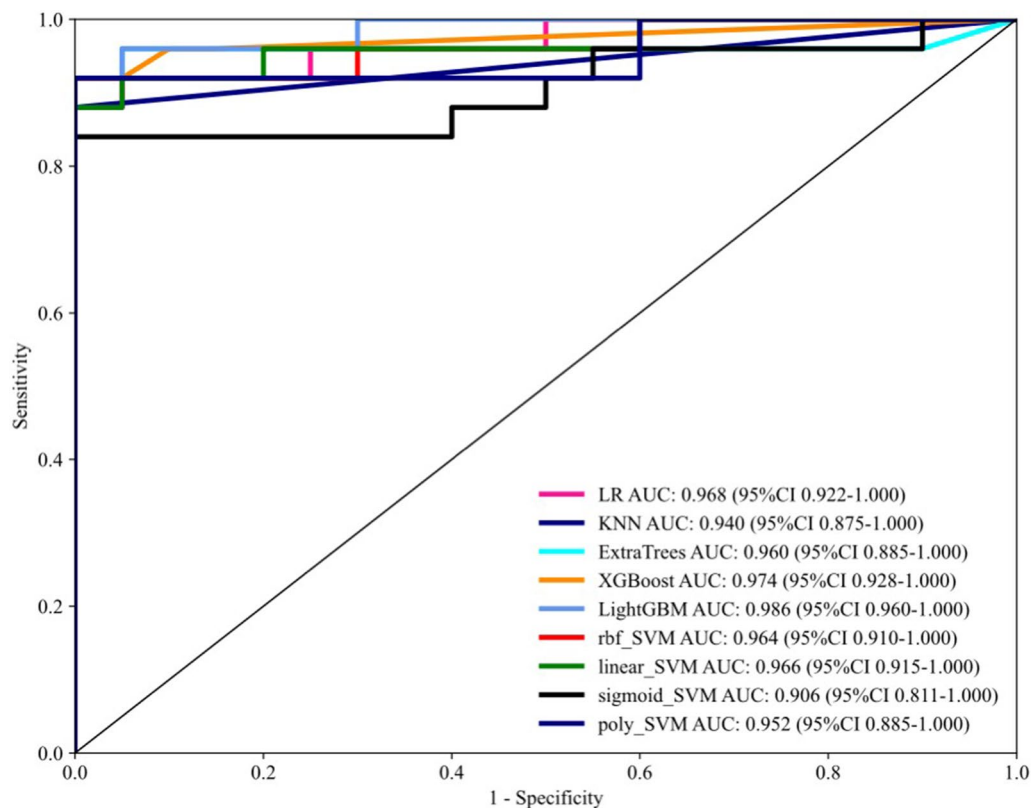


Fig. 3 Test dataset for external validation of AUC cohorts

VII/IX, and degradation of vascular endothelial glycocalyx. APTT, as a commonly used coagulation test, can be used to assess the hepatic impairment and bleeding risk. It has been reported that APTT influences the prognosis of many diseases. For example, in COVID-19 patients, aspartate aminotransferase to platelet ratio index (APRI)

combined with APTT showed good prediction potential for in-hospital mortality [9]. TBA, an indicator of liver function, is commonly used to assess the metabolic function and cholestasis of the liver. In HBV-associated acute liver failure, TBA was linked with the risk of short-term mortality among patients, and when combined with

Table 4 Machine learning modeling analysis in the test dataset for external validation

Cohort	Model	AUC	AUC 95% CI	Acc	Acc 95% CI	Sen	Spe	PPV	NPV
Test	LR	0.968	0.9218–1.0000	0.933	0.880	1.000	1.000	Test	LR
Test	KNN	0.940	0.8750–1.0000	0.933	0.880	1.000	1.000	Test	KNN
Test	ExtraTrees	0.960	0.8854–1.0000	0.956	0.920	1.000	1.000	Test	ExtraTrees
Test	XGBoost	0.974	0.9282–1.0000	0.956	0.920	1.000	1.000	Test	XGBoost
Test	LightGBM	0.986	0.9602–1.0000	0.956	0.920	1.000	1.000	Test	LightGBM
Test	rbf_SVM	0.964	0.9102–1.0000	0.956	0.920	1.000	1.000	Test	rbf_SVM
Test	linear_SVM	0.966	0.9147–1.0000	0.933	0.880	1.000	1.000	Test	linear_SVM
Test	sigmoid_SVM	0.906	0.8107–1.0000	0.911	0.840	1.000	1.000	Test	sigmoid_SVM
Test	poly_SVM	0.952	0.8845–1.0000	0.956	0.920	1.000	1.000	Test	poly_SVM

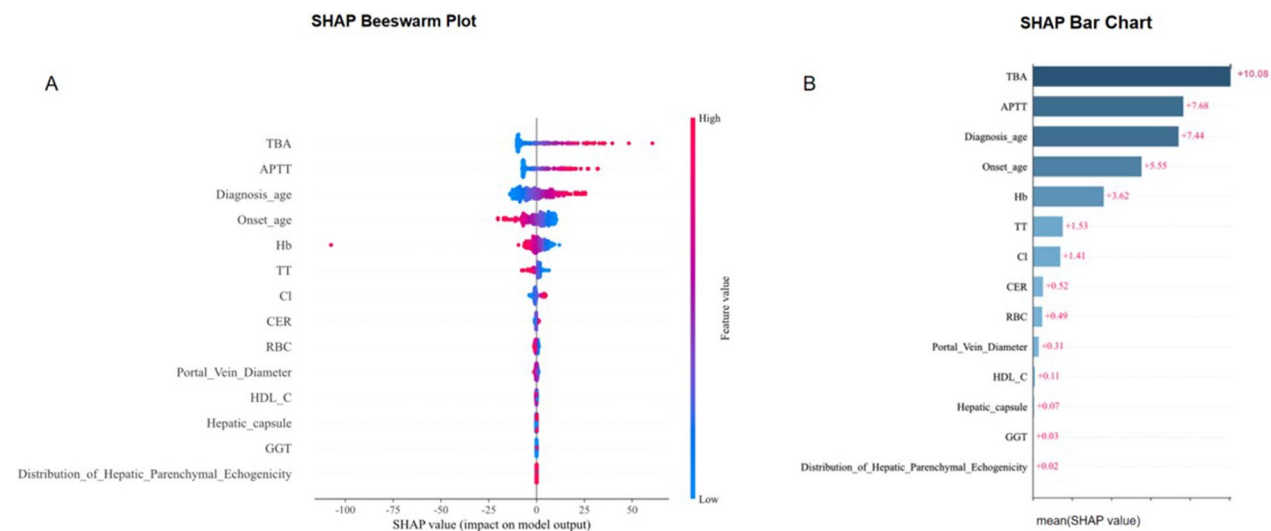


Fig. 4 Interpretability analysis of WD-ACLF predictors via SHAP value. In Fig. 3A, the horizontal coordinates indicate the extent to which each feature contributes to the model output, and the magnitude of the feature value is indicated by colour (red indicates a higher feature value, blue indicates a lower feature value, and purple indicates adjacency to the mean)

other indicators, it improved the prediction accuracy [10], which is consistent with our findings, in which the TBA of patients with liver failure was significantly higher than that of patients without liver failure ($P > 0.001$). This observation was associated with the inhibition of the hepatocyte regeneration pathway by bile acids mediated by the FXR receptor. The decrease in Hb and RBC accompanied by increase of RDW-CV suggests that copper accumulation can damage the structural integrity of erythrocyte membranes via lipid peroxidation, to impair erythrocyte survival and function [11, 12]. Therefore, clinicians can apply this model in clinical practice. When the TBA, APTT and diagnosis age are higher and the onset age and Hb are lower, clinicians need to pay attention to this and intervene early to prevent further development into WD-ACLF.

In addition, we found that patients with WD-ACLF exhibited a unique clinical phenotype. Analysis of age-related factors showed that delayed age of onset and age at diagnosis and prolonged disease duration were significant predictors of ACLF development, suggesting that cumulative effects of copper toxicity in disease

progression. Elevated CER in WD-ACLF patients, contrary to conventional diagnostic criteria for WD, likely reflects synergistic effects of acute-phase inflammation, hepatocyte damage, and adaptive copper metabolic shifts. Multidimensional analysis of the indicators confirmed the existence of the triad of “liver injury-fibrosis-metabolic imbalance” in WD-ACLF [13]. Significantly elevated AST in hepatic impairment mainly reflects hepatocellular necrosis, elevated GGT may be associated with cholestasis or oxidative stress, whereas elevated LDH is mostly a non-specific manifestation of multisystemic injury, respectively [14–16]. Abnormal elevation of liver fibrosis markers PIIINP, LN, HA, and CIV reveals accelerated extracellular matrix remodeling process [17]. Moreover, metabolic disorders were characterized by significant homocysteine accumulation accompanied with decreased ALB and lipid metabolism markers (e.g. TG, TC), which indicated multi-systemic dysfunction. In this study, we found that elevated HDL-C was associated with a lower risk of liver failure, which may be attributed to improved reverse cholesterol transport and the anti-inflammatory effect of ApoA1. Data indicates

that HDL-C levels are strongly associated with survival in patients with chronic liver failure, and low levels of HDL-C may be associated with poor prognosis [18]. In addition, ultrasonographic characterization revealed that the risk of ACLF was significantly higher in patients with widened portal vein diameter, uneven hepatic capsule, abnormal hepatic volume, and ascites, suggesting that the structural abnormalities of the liver preceded the changes in biochemical indicators, providing a basis for early warning detection.

In this study, we constructed the WD-ACLF prediction system that resolve the nonlinear interaction effect of predictors. The predictive model multidimensional prediction system constructed in this study not only improves early identification of WD-ACLF, but also provides a guideline that will guide the development of targeted therapeutic strategies by revealing the multiorgan interaction mechanism. The combination of key indicators such as TBA, APTT, diagnosis age, onset age, Hb and other features can more effectively help to stratify patients and improve the clinical management of liver disease.

Notwithstanding the clinical relevance of this study, several notable constraints merit consideration. First, the retrospective design precluded serial measurements during hospitalization, limiting dynamic assessment of clinical trajectory and prognostic divergence across subgroups. Second, the absence of objective quantification of hepatic morphology and function (e.g., histopathology or non-invasive fibrosis diagnostics like FibroScan-CAP) may confound outcome interpretation. Future prospective longitudinal cohorts should incorporate frequent multimodal assessments—including standardized serological panels, imaging biomarkers, and validated clinical scales—to longitudinally track treatment response and disease progression.

Conclusion

In conclusion, the WD-ACLF prediction model constructed using the XGBoost algorithm showed excellent potential for clinical application, outperforming that of traditional prediction systems. However, this study still has limitations. For instance, some data are missing, the number of included clinical cases is limited, and there is a lack of prospective research design. Therefore, prospective multicentre cohort validation and dynamic modeling of treatment time window should be conducted to accelerate the transformation of this prediction system into a clinical decision support system, and enhance precision medicine for rare diseases.

Abbreviations

WD	Wilson disease
ACLF	Acute-on-chronic liver failure
ML	Machine learning

LR	Logistic regression
SVM	Support vector machine
KNN	K-nearest neighbors
Extra Trees	Extremely randomized trees
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
SHAP	SHapley Additive exPlanations

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-06987-1>.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

We gratefully acknowledge the Department of Laboratory Medicine and the Information Center of the First Affiliated Hospital of Anhui University of Chinese Medicine for providing patients information and data support.

Author contributions

ZR drafted the initial manuscript and contributed to the revisions. SL, HX, KD, YL, SF and WH conducted data acquisition, performed systematic investigations, and executed statistical analyses. YY and YLY contributed to the development of methodological framework. XL and WY conceived and designed the study, supervised the conceptualization process, and critically revised the manuscript for intellectual content. All authors rigorously reviewed the final draft and provided unanimous approval for submission.

Funding

The authors acknowledge financial support for this investigation, including research execution, authorship activities, and publication costs. This work was funded through the following grants: the Regional Innovation and Development Joint Fund of NSFC (No. U22A20366), the Anhui Province Traditional Chinese Medicine Science and Technology Research Special Project (202303a07020004) and the Anhui Province Clinical Medical Research Transformation Special Project (2022042951070-2006).

Data availability

All datas in this study should be requested from the corresponding authors.

Declarations

Ethics approval and consent to participate

Ethical clearance for this retrospective investigation was formally granted by the Institutional Review Board of the First Affiliated Hospital of Anhui University of Chinese Medicine (Protocol ID: 2024AH-13). The ethics committee waived the requirements of informed consent and clinical trial registration.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 18 June 2025 / Accepted: 9 August 2025

Published online: 24 September 2025

References

1. Tanzi RE, Petrukhin K, Chernov I, et al. The Wilson disease gene is a copper transporting ATPase with homology to the Menkes disease gene. *Nat Genet.* 1993;5(4):344–50. <https://doi.org/10.1038/ng1293-344>.
2. Walshe JM. Clinical investigations standing committee of the Association of Clinical B, Wilson's disease: the importance of measuring serum

- caeruloplasmin non-immunologically. *Ann Clin Biochem.* 2003;40(Pt 2):115–21. <https://doi.org/10.1258/000456303763046021>.
3. Devarbhavi H, Reddy VV, Singh R. Wilson disease presenting with acute on chronic liver failure: a single-center experience of outcome and predictors of mortality in 68 patients. *J Clin Exp Hepatol.* 2019;9(5):569–73. <https://doi.org/10.1016/j.jceh.2019.02.006>.
 4. Fisher CK, Smith AM, Walsh JR, et al. Machine learning for comprehensive forecasting of Alzheimer's disease progression. *Sci Rep.* 2019;9(1):13622. <https://doi.org/10.1038/s41598-019-49656-2>.
 5. Marianna A, Carlo R, Sarah A, et al. Machine learning can predict mild cognitive impairment in Parkinson's disease. *Front Neurol.* 2022;13:1010147. <https://doi.org/10.3389/fneur.2022.1010147>.
 6. Chen K, Wan Y, Mao J, et al. Liver cirrhosis prediction for patients with Wilson disease based on machine learning: a case-control study from southwest China. *Eur J Gastroenterol Hepatol.* 2022;34(10):1067–73. <https://doi.org/10.1097/MEG.0000000000002424>.
 7. European Association for Study of Liver. EASL clinical practice guidelines: Wilson's disease. *J Hepatol.* 2012;56(3):671–85.
 8. Sarin SK, Choudhury A, Sharma MK, et al. Acute-on-chronic liver failure: consensus recommendations of the Asian Pacific association for the study of the liver (APASL): an update. *Hepatol Int.* 2019;13(4):353–90. <https://doi.org/10.1007/s12072-019-09946-3>.
 9. Madian A, Eliwa A, Abdalla H, et al. Aspartate transferase-to-platelet ratio index-plus: a new simplified model for predicting the risk of mortality among patients with COVID-19. *World J Gastroenterol.* 2022;28(16):1671–80. <https://doi.org/10.3748/wjg.v28.i16.1671>.
 10. Wan Z, Wu Y, Yi J, et al. Combining serum cystatin C with total bilirubin improves short-term mortality prediction in patients with HBV-related acute-on-chronic liver failure. *PLoS ONE.* 2015;10(1): e0116968. <https://doi.org/10.1371/journal.pone.0116968>.
 11. Tariq T, Karabon P, Irfan FB, et al. National trends and outcomes of nonautoimmune hemolytic anemia in alcoholic liver disease: analysis of the nationwide inpatient sample. *J Clin Gastroenterol.* 2021;55(3):258–62. <https://doi.org/10.1097/MCG.0000000000001383>.
 12. Kozelskaya AI, Panin AV, Khlusov IA, et al. Morphological changes of the red blood cells treated with metal oxide nanoparticles. *Toxicol In Vitro.* 2016;37:34–40. <https://doi.org/10.1016/j.tiv.2016.08.012>.
 13. Członkowska A, Litwin T, Dusek P, et al. Wilson disease. *Nat Rev Dis Primers.* 2018;4(1): 21. <https://doi.org/10.1038/s41572-018-0018-3>.
 14. Xu W, Huang C, Fei L, et al. Dynamic changes in liver function tests and their correlation with illness severity and mortality in patients with COVID-19: a retrospective cohort study. *Clin Interv Aging.* 2021;16:675–85. <https://doi.org/10.2147/CIA.S303629>.
 15. Kim OJ, Kim S, Park EY, et al. Exposure to serum perfluoroalkyl substances and biomarkers of liver function: the Korean national environmental health survey 2015–2017. *Chemosphere.* 2023;322: 138208. <https://doi.org/10.1016/j.chemosphere.2023.138208>.
 16. Liu R, Huang Y, Jia Z, et al. Elevated serum AST and LDH levels are associated with infant death in premature babies with neonatal leukemoid reaction: a cohort study. *Transl Pediatr.* 2022;11(12):1920–7. <https://doi.org/10.21037/tip-22-543>.
 17. Wei XL. The relationship between serum visfatin and the progress of chronic viral hepatitis B cirrhosis. *Eur Rev Med Pharmacol Sci.* 2017;21(2):297–301.
 18. Trieb M, Rainer F, Stadlbauer V, et al. HDL-related biomarkers are robust predictors of survival in patients with chronic liver failure. *J Hepatol.* 2020;73(1):113–20. <https://doi.org/10.1016/j.jhep.2020.01.026>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.