Original Research Article

A machine learning model for predicting abnormal liver function induced by a Chinese herbal medicine preparation (Zhengqing Fengtongning) in patients with rheumatoid arthritis based on real-world study

Ze Yu [a], Fang Kou [a], Ya Gao [b], Fei Gao [c], Chun-ming Lyu [d],*, Hai Wei [a],*

[a] *Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*
[b] *Department of Pharmacy, Fuwai Hospital, Chinese Academy of Medical Sciences, Beijing 100037, China*
[c] *Beijing Medicinovo Technology Co. Ltd., Beijing 100071, China*
[d] *Experiment Center for Science and Technology, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*

ARTICLE INFO

ABSTRACT

Objective: Rheumatoid arthritis (RA) is a systemic autoimmune disease that affects the small joints of the whole body and degrades the patients' quality of life. Zhengqing Fengtongning (ZF) is a traditional Chinese medicine preparation used to treat RA. ZF may cause liver injury. In this study, we aimed to develop a prediction model for abnormal liver function caused by ZF.
Methods: This retrospective study collected data from multiple centers from January 2018 to April 2023. Abnormal liver function was set as the target variable according to the alanine transaminase (ALT) level. Features were screened through univariate analysis and sequential forward selection for modeling. Ten machine learning and deep learning models were compared to find the model that most effectively predicted liver function from the available data.
Results: This study included 1,913 eligible patients. The LightGBM model exhibited the best performance (accuracy = 0.96) out of the 10 learning models. The predictive metrics of the LightGBM model were as follows: precision = 0.99, recall rate = 0.97, F1_score = 0.98, area under the curve (AUC) = 0.98, sensitivity = 0.97 and specificity = 0.85 for predicting ALT < 40 U/L; precision = 0.60, recall rate = 0.83, F1_score = 0.70, AUC = 0.98, sensitivity = 0.83 and specificity = 0.97 for predicting $40 \leq$ ALT < 80 U/L; and precision = 0.83, recall rate = 0.63, F1_score = 0.71, AUC = 0.97, sensitivity = 0.63 and specificity = 1.00 for predicting ALT $\geq$ 80 U/L. ZF-induced abnormal liver function was found to be associated with high total cholesterol and triglyceride levels, the combination of TNF-α inhibitors, JAK inhibitors, methotrexate + nonsteroidal anti-inflammatory drugs, leflunomide, smoking, older age, and females in middle-age (45–65 years old).
Conclusion: This study developed a model for predicting ZF-induced abnormal liver function, which may help improve the safety of integrated administration of ZF and Western medicine.
Please cite this article as: Yu Z, Kou F, Gao Y, Lyu CM, Gao F, Wei H. A machine learning model for predicting abnormal liver function induced by a Chinese herbal medicine preparation (Zhengqing Fengtongning) in patients with rheumatoid arthritis based on real-world study. J Integr Med. 2025; 23 (1): 25–35.

1. Introduction

Rheumatoid arthritis (RA) is a systemic autoimmune disease that often affects small joints throughout the body, leading to joint pain, tenderness, swelling and morning stiffness, and other symptoms [1,2]. The incidence of RA, which is one of the most common

* Corresponding authors.
E-mail addresses: chunming83g@126.com (C.M. Lyu), wei_hai@hotmail.com (H. Wei).

inflammatory arthritis types, is estimated to be approximately 0.4% in China [3]. Zhengqing Fengtongning (ZF) is a modern traditional Chinese medicine preparation refined from the alkaloids extracted from the Chinese herbal medicine Qingfengteng (*Sinomenium acutum* [Thunb.] Rehd. et Wils. var. cinereum Rehd. et Wils) [4]. Sinomenine is the main active ingredient of ZF and is commonly used to treat RA and other rheumatic diseases in China in combination with Western medicine, such as methotrexate (MTX), iguratimod (IGU) and leflunomide (LEF) [5]. RA is a chronic disease that patients generally require long-term treatment, which may induce some adverse drug reactions (ADRs).

The common ADRs caused by sinomenine include bloating, diarrhea, low white blood cell count, and whole body or local rash [6,7]. Abnormal liver function is a rare but potentially severe ADR that can lead to life-threatening consequences. Currently, few studies have explored the prediction of abnormal liver function in patients using ZF. Adverse reactions to a drug can be affected by many factors, such as gender, age and drug combination, and the severity of ADRs may vary in different patients [8]. Therefore, it is necessary to predict abnormal liver function in patients at individualized levels.

In current medical research, real-world evidence (RWE) from clinical practice is an important data source [9]. RWE can be collected from electronic medical records, laboratory data, and follow-up records during and/or after diagnosis and treatment. Compared with clinical trials, studies based on RWE are more relevant to clinical practice and can be applied in individualized programs with rich evidence and good external validity [10,11]. Machine learning and deep learning algorithms, such as extreme gradient boosting (XGBoost), categorical boosting (CatBoost), light gradient boosting machine (LightGBM), gradient boosting decision tree (GBDT), and TabNet [12–16], can be used to interpret sophisticated and multidimensional data from RWE. These algorithms are adept at data mining, feature extraction, and model development. By processing large amounts of data, data-driven estimates can be evaluated from multiple variables, and nonlinear variable relationships can be identified, enabling high-precision prediction of clinical outcomes [17,18]. The use of machine learning and deep learning in clinical medicine research has seen increasing interest in recent years. For instance, a GBDT model was shown to have higher precision than global trigger tools for predicting adverse drug events in 1,746 pediatric patients (44% *vs* 13.3%) [19]; an ensemble model using five machine learning algorithms achieved greater R^2 in predicting vancomycin concentration, compared with the population pharmacokinetic model (0.6 *vs* 0.3) [20]; and a XGBoost model was able to construct a personalized medicine model of Shengmai injection [21]. With the larger sample size of input data, machine learning and deep learning models can be continually optimized to achieve better performance and practicality. We believe that taking machine learning and deep learning-based models from development to real-world application is worth further research.

This study aimed to develop a real-world database of RA from multiple centers to explore the important factors that are linked to abnormal liver function following administration of ZF and to construct a predictive model for abnormal liver function caused by ZF. For patients at high risk of abnormal liver function, the incidence of severe outcomes could be reduced by timely and effective adjustment of the medication regimen.

2. Methods

2.1. Study design and population

This was a retrospective study. We enrolled patients at multiple Grade-A tertiary hospitals in China, including the First Affiliated Hospital of Soochow University, Nanfang Hospital of Southern Medical University, and Hebei General Hospital, from January 2018 to April 2023. We also enrolled patients at the Second Affiliated Hospital of Naval Medical University (Shanghai Changzheng Hospital) from June 1, 2022 to May 31, 2023 for external validation. The target variable was the alanine transaminase (ALT) level. According to the Common Terminology Criteria for Adverse Events (CTCAE v 4.0) scale and practical clinical experience of the physicians, ALT < 40 U/L was considered to indicate normal liver function, while 40 ≤ ALT < 80 U/L was considered to indicate abnormal liver function, and ALT ≥ 80 U/L was considered to indicate severely abnormal liver function [22]. A total of 38 features associated with ZF-induced abnormal liver function were chosen based on clinical knowledge and previous literature for screening and were subsequently used to develop a prediction model of ZF-induced abnormal liver function.

The inclusion criteria were as follows: (1) patients diagnosed with RA; (2) patients aged ≥ 18 years; (3) patients receiving treatment for RA; and (4) outpatients or inpatients with at least 12 weeks of follow-up records. RA treatments include: MTX, IGU, LEF, MTX + ZF, IGU + ZF, LEF + ZF, MTX + LEF, MTX + nonsteroidal anti-inflammatory drugs (NSAIDs), MTX + total glucosides of paeony, MTX + Kunxian Capsule, MTX + Biqi Capsule, MTX + LEF + ZF, MTX + NSAIDs + ZF, Janus kinase (JAK) inhibitor + ZF, JAK inhibitor + MTX, JAK inhibitor + IGU, tumor necrosis factor-α (TNF-α) inhibitor + ZF, TNF-α inhibitor + MTX, and TNF-α inhibitor + IGU. The exclusion criteria were as follows: (1) patients had other rheumatic diseases; (2) patients were pregnant or breastfeeding; (3) patients did not use ZF; and (4) patients lacked primary liver function data or other primary research data.

This study was approved by the medical ethics committees of the First Affiliated Hospital of Soochow University (No. 2023-268), Nanfang Hospital of Southern Medical University (No. NFEC-2023-438) and Hebei General Hospital (No. 2023-324). The study was conducted in accordance with the *Helsinki Declaration* of 1964, study data were fully deidentified, and confidential information was deleted from the patients in accordance with the Council for International Organizations of Medical Sciences/World Health Organization *International Ethical Guidelines for Health-related Research Involving Humans* (2016) [23]. As this was a large retrospective study based on electronic medical record data and patient information was fully deidentified, informed consent exemptions were approved by all the ethics committees.

2.2. Data collection and processing

The following data were collected from the patients' electronic medical records: demographic information (such as age and gender), daily ZF dose, smoking status, presence or absence of cardiovascular disease (CVD), drug combinations, assay indices (such as erythrocyte sedimentation rate [ESR], C-reactive protein [CRP], rheumatoid factor [RF], triglyceride [TG], and total cholesterol [TC]), and the disease activity score in 28 joints based on CRP level.

Variables with a missing data rate greater than 50% were eliminated. The missing data rate is a measure of the proportion of missing values in a dataset. The variables with highly imbalanced sample sizes, characterized by a ratio of positive samples to the total number of samples less than 0.1, were deleted. The classification variables such as gender, smoking and CVD were coded by one-hot encoding, and gender/age groups, ZF doses, and drug combinations were coded by ordinal encoding as follows (shown in Table 1).

2.3. Feature selection

First, a univariate analysis was conducted for all 38 features associated with ZF-induced abnormal liver function. Significance was tested by Kruskal-Wallis test or chi-square test, and those

Table 1
Variable encoding.

Item	Coded by ordinal encoding						
	0	1	2	3	4	5	6
Gender	Male	Female					
Age/gender group	Young male	Young female	Middle-aged male	Middle-aged female	Elderly male	Elderly female	
ZF daily dose	60 mg	120 mg	180 mg	240 mg			
Drug combination	MTX + ZF	IGU + ZF	LEF + MTX + ZF	LEF + ZF	MTX + NSAIDs + ZF	TNF-α inhibitor + ZF	JAK inhibitor + ZF
ALT level	ALT < 40 U/L	40 U/L $\leq$ ALT < 80 U/L	ALT $\geq$ 80 U/L				

ALT: alanine transaminase; IGU: iguratimod; JAK: Janus kinase; LEF: leflunomide; MTX: methotrexate; NSAID: nonsteroidal anti-inflammatory drug; TNF: tumor necrosis factor; ZF: Zhengqing Fengtongning.

with a *P* value < 0.05 were considered important variables to ZF-induced abnormal liver function. The sequential forward selection (SFS) algorithm based on XGBoost was subsequently applied for further screening [24]. The selection process involved adding one feature to the feature subset each time via the SFS algorithm, iteratively generating a new model, and calculating the model performance. The model performance was expressed by the F1_score, which is a comprehensive evaluation index of precision and recall rate, where a higher F1_score indicates a more robust model. Iteration stopped when the F1_score of the model built with a certain feature subset reached the highest value. Important variables were included in this feature subset with the minimum variable set and optimum performance.

2.4. Model development and interpretation

As shown in Fig. 1, the train_test_split function in the scikit-learn library was used to randomly divide the dataset into a training set and a testing set at a ratio of 4:1. To generate an unbiased and reliable prediction model, a 6-fold cross-validation was conducted in the training set, and the model was optimized by parameter tuning with the F1_score as the evaluation index. With respect to the training set, prediction models for ZF-induced abnormal liver function were developed by 10 algorithms, and their predictive performance was compared. These algorithms, which include logistic regression, random forest, artificial neural network, support vector machine, k-nearest neighbor, CatBoost, LightGBM, XGBoost, GBDT and TabNet, have good predictive ability as common algorithm types. The development principles of the ten machine learning and deep learning models are shown in Supplementary File 1. The predictive performance of the algorithms was evaluated on the testing set, and the algorithm with the best performance was selected to develop the final prediction model. The parameters of the 10 models are displayed in Table S1.

Subsequently, importance scores for the selected variables were calculated by the algorithm with the best predictive performance. The importance of a variable refers to the degree to which each variable in the model contributes to improving the predictive power of the whole model. Afterward, Shapley additive explanation (SHAP) was applied to interpret the impacts of important vari-
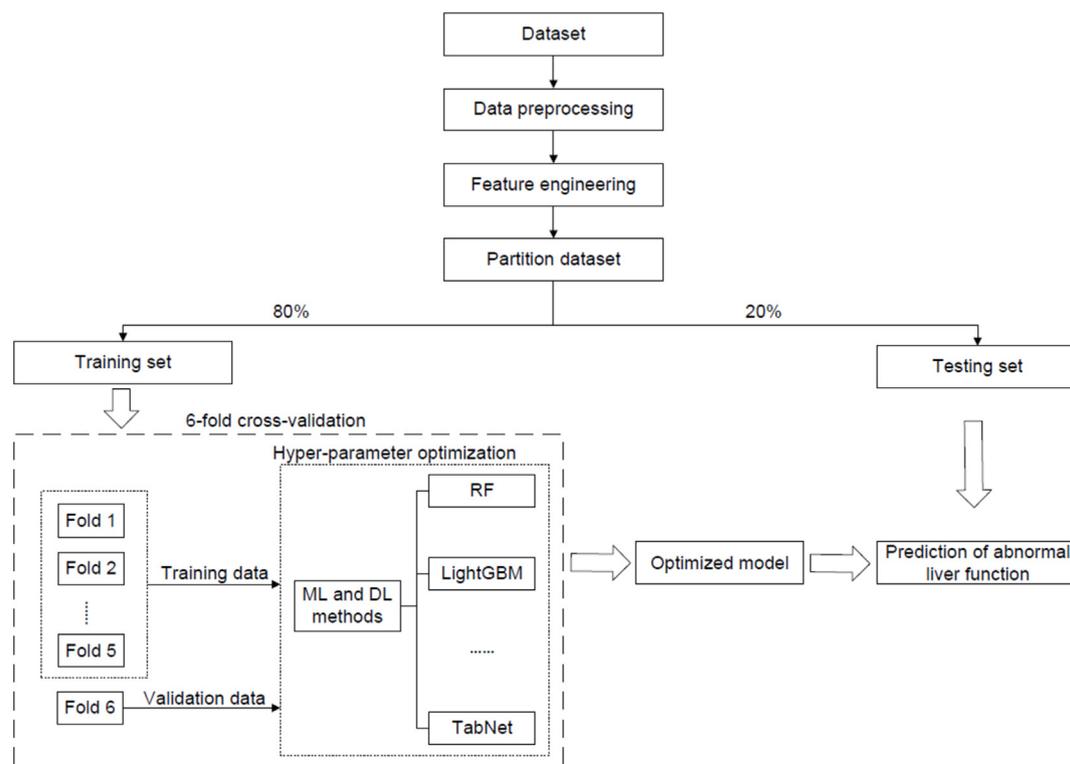

Fig. 1. Workflow of data processing and modeling. DL: deep learning; LightGBM: light gradient boosting machine; ML: machine learning; RF: random forest.

ables on the model output [25]. Eventually, a confusion matrix was used to calculate the precision and recall rate and to visualize the performance of the prediction model in the testing set.

2.5. External validation and model application

A confusion matrix was used to show the predictive performance of the external validation. Eventually, in order to facilitate the application of the model, we established a web application that can provide decision support for generating predictions of ZF-induced abnormal liver function for individual patients.

2.6. Statistical analysis

In the univariate analysis, continuous variables were analyzed using the Kruskal-Wallis test via SciPy library (version 1.7.3). Disordered categorical variables were analyzed by the chi-square test. All the statistical analyses were conducted using two-sided tests, and the test statistics and corresponding P values are reported. A P value < 0.05 was considered to indicate statistical significance.

To evaluate the predictive performance of the models, evaluation metrics, including precision, recall rate, F1_score, accuracy, AUC, sensitivity and specificity, were calculated. The F1_score is used to measure the merits and defects of the model, and a higher F1_score indicates better model performance. The specific formulae for the evaluation metrics are as follows:

Precision = TP/(TP + FP).
Recall rate = TP/(TP + FN).
F1_score = 2 × TP/(2 × TP + FP + FN).
Accuracy = (TP + TN)/(TP + FN + FP + TN).
Sensitivity = TP/(TP + FN).
Specificity = TN/(TN + FP).

TP: true positive, indicating the number of positive cases that were correctly identified; TN: true negative, indicating the number of negative cases that were correctly identified; FP: false positive, indicating the number of negative cases identified as positive; FN: false negative, indicating the number of positive cases identified as negative.

A confusion matrix was constructed with the Matplotlib (version 3.8.2) package. All the machine learning and deep learning algorithm experiments were run on Windows 10 with an Intel® Core™ i5-10400F CPU @ 2.90 GHz 12 CPUs and 512 GB of memory. The data analysis was conducted using Python 3.9.12 and IBM SPSS Statistics 22.

3. Results

3.1. Baseline information

The workflow of patient inclusion is displayed in Fig. 2. A total of 65,007 patients with RA were identified. Among them, 56,701 patients were aged ≥ 18 years. Of these, 46,138 patients had received RA drug treatments, and 4017 of these patients had at least 12 weeks of follow-up records. Patients with other rheumatic diseases (34 patients), pregnant or lactating women (9 patients), patients not using ZF (1964 patients), or patients lacking major research data (97 patients) were excluded. Ultimately, a total of 1913 patients were included in this study. Among them, there were 916 patients, 692 patients and 305 patients collected from the First Affiliated Hospital of Soochow University, Southern

Hospital of Southern Medical University and Hebei General Hospital, respectively.

According to a ratio of 4:1, 1530 patients were included in the training set and 383 patients were included in the testing set. The demographic and clinical characteristics of the patients in the dataset are displayed in Table 2. There were 1371 (71.67%) males and 542 (28.33%) females. The median age of the individuals in the whole dataset was 58 years (interquartile range [IQR] 50–68 years). Among the gender_age groups, middle-aged males had the highest proportion (37.11%), and young females had the lowest proportion (4.81%). The daily doses of ZF were divided into four subgroups: 60 mg (n [%] = 108 [5.65%]), 120 mg (n [%] = 900 [47.05%]), 180 mg (n [%] = 192 [10.04%]), and 240 mg (n [%] = 713 [37.27%]). There were 599 (31.31%) patients who smoked and 371 (19.39%) who had CVD. For treatment plans, the number (%) treated with MTX + ZF, IGU + ZF, LEF + MTX + ZF, LEF + ZF, MTX + NSAIDs + ZF, TNF-α inhibitor + ZF, and JAK inhibitor + ZF was 468 (24.46%), 253 (13.23%), 263 (13.75%), 217 (11.34%), 220 (11.50%), 254 (13.28%), and 238 (12.44%), respectively.

There were 176 patients with abnormal liver function, their demographic and clinical characteristics are displayed in Table S2. Among them, 126 were males and 50 were females. The median age was 68.5 years (IQR 56–75 years). Among the gender_age groups, elderly males had the highest proportion (56.25%). In the daily dose subgroups, the dose of 120 mg was most commonly used (51.70%). About 49% of the patients smoked and 12.50% had CVD. MTX + ZF was the most commonly used treatment plan (31.25%).

3.2. Variable selection

The results of univariate analysis are shown in Table 3. A total of 22 significant variables were selected with P < 0.05, including age, gender_age, CVD status, treatment plan, ZF daily dose, CRP, RF, ESR, hemoglobin, prealbumin, uric acid, total protein, albumin, γ-glutamyl transferase, neutrophil absolute value, mean platelet volume, mean corpuscular volume, mean corpuscular hemoglobin concentration, mean corpuscular hemoglobin, packed cell volume, TC and TG.

After that, the 22 variables were further screened through the SFS method. Based on the XGBoost models developed with 1 to 22 variables, the F1_score of each XGBoost model was obtained (Fig. 3). As the number of included variables was increased, the F1_score of the model continued to increase and reached a maximum value when the model included 6 variables (F1_score = 0.73). As a concise and accurate model was pursued, a minimal variable set that can achieve good model performance was adopted. The first 6 variables were gender_age, treatment plan, age, TG, smoking status, and TC.

The F1_scores of the different models based on 22 and 6 variables are illustrated in Fig. 4 and Table S3. According to Fig. 4A, in the models based on 22 variables obtained by univariate analysis, the LightGBM model shows the best result with a mean F1_score of 0.77 and a standard deviation (Std) of 0.07. In Fig. 4B, for the models based on the 6 variables obtained by SFS, the LightGBM model yields the best results, with a mean F1_score of 0.78 and an Std of 0.07. A comparison of the models developed with 6 variables and 22 variables revealed that the model developed with 6 variables had a higher F1_score, which was both concise and robust. Therefore, the 6 variables were used for modeling in the testing set. The difference of the target variable and important influencing variables between training set and testing set is illustrated in Table S4. It can be seen that variables between the two datasets did not show a significant difference (P value > 0.05).
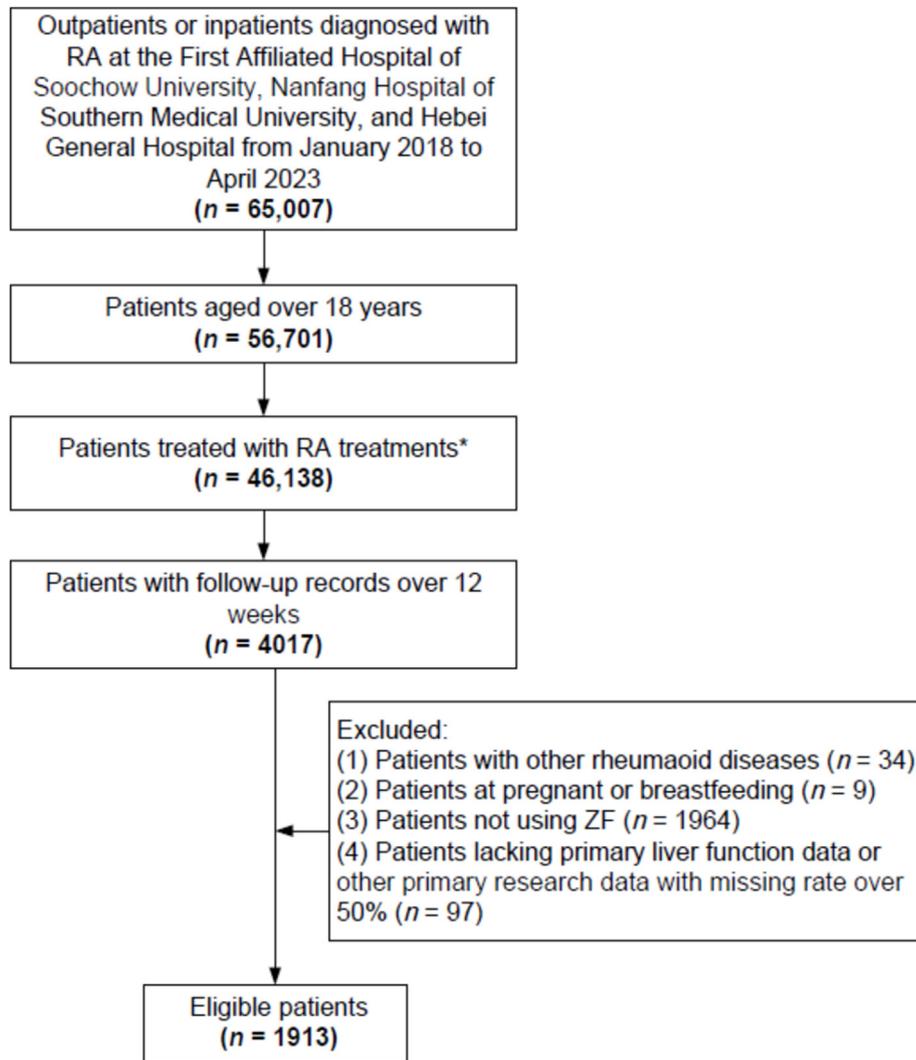
Fig. 2. Workflow of patient enrollment. RA: rheumatoid arthritis; ZF: Zhengqing Fengtongning.

3.3. Model performance and interpretation

The predictive performance of the 10 models in the testing set is illustrated in Table 4. Among them, the LightGBM model shows the best performance. The model has the following predictive metrics: precision = 0.99, recall rate = 0.97, F1_score = 0.98, AUC = 0.98, sensitivity = 0.97 and specificity = 0.85 for predicting ALT < 40 U/L; precision = 0.60, recall rate = 0.83, F1_score = 0.70, AUC = 0.98, sensitivity = 0.83 and specificity = 0.97 for predicting 40 ≤ ALT < 80 U/L; and precision = 0.83, recall rate = 0.63, F1_score = 0.71, AUC = 0.97, sensitivity = 0.63 and specificity = 1.00 for predicting ALT ≥ 80 U/L. The accuracy is 0.96 for the whole LightGBM model.

Based on the LightGBM model, the importance scores of the 6 variables used for modeling are presented in Fig. S1 and Table S5. The importance scores of TC, TG, treatment plan, age, gender_age and smoking status were 388, 292, 232, 202, 155 and 122, respectively. The importance score indicates the degree of importance of features in the model. A higher score means greater predictive power of the variable on the ZF-induced abnormal liver function. The results reveal that TC was the most predictive variable of ZF-induced abnormal liver function, followed by TG, treatment plan, age, gender_age and smoking status. The importance scores of the 22 variables are displayed in Table S6.

For the 6 important variables, each variable's impact on the model output was demonstrated via SHAP analysis (Figs. S2–S7). The SHAP scatter plots indicate monotonic relationships between variables and the occurrence of abnormal liver function. Specifically, the risk of abnormal liver function increases with increasing TC and TG levels, and greater age. For diverse treatment plans, the risk of abnormal liver function was most likely to increase when TNF-α inhibitor + ZF, JAK inhibitor + ZF, MTX + NSAIDs + ZF, or LEF + ZF was used. For the age_gender groups, the risk of abnormal liver function was greatest in elderly patients, followed by middle-aged females. Additionally, the risk of abnormal liver function was greater in smoking patients than non-smoking ones.

The testing set consisted of 383 patients, among which 357 had ALT < 40 U/L, 18 had 40 ≤ ALT < 80 U/L, and 8 had ALT ≥ 80 U/L (Fig. S8). Based on the final LightGBM model, the risk of abnormal liver function was predicted, and the precision and recall rate were calculated. Precision is the ability of the model to correctly predict specific outcomes, while recall rate is the ability of the model to make correct predictions for all positive cases, that is, not to miss positive cases. Three hundred and fifty-two (352) patients were predicted to belong to the ALT < 40 U/L group. Of these 348 were correctly classified, 4 patients were incorrectly classified, and 9 patients were mis-classified, resulting in a precision of 99% and a

Table 2
Baseline information for the whole dataset.

Category	Variable	Median (IQR) or n [%]	Miss rate
Demographic information	Age (year)	58.00 (50.00–68.00)	0.00%
	Gender		0.00%
	Male	1371 (71.67%)	
	Female	542 (28.33%)	
	Gender_age		0.00%
	Young male	202 (10.56%)	
	Young female	92 (4.81%)	
	Middle-aged male	710 (37.11%)	
	Middle-aged female	260 (13.59%)	
	Elderly male	459 (23.99%)	
	Elderly female	190 (9.93%)	
ZF information	Daily dose of ZF		0.00%
	60 mg	108 (5.65%)	
	120 mg	900 (47.05%)	
	180 mg	192 (10.04%)	
	240 mg	713 (37.27%)	
Smoking history	Smoking	599 (31.31%)	0.00%
CVD history	CVD	371 (19.39%)	0.00%
Medication information	Treatment plan		0.00%
	MTX + ZF	468 (24.46%)	
	IGU + ZF	253 (13.23%)	
	LEF + MTX + ZF	26 (13.75%)	
	LEF + ZF	217 (11.34%)	
	ZF + MTX + NSAIDs	220 (11.50%)	
	TNF-α inhibitor + ZF	254 (13.28%)	
	JAK inhibitor + ZF	238 (12.44%)	
Assay index	CRP (mg/L)	3.16 (0.83–8.93)	0.00%
	RF (IU/mL)	100.31 (28.55–199.32)	0.00%
	ESR (mm/h)	25.00 (15.61–41.34)	0.00%
	GLB (g/L)	28.10 (25.60–30.60)	22.74%
	HB (g/L)	127.00 (119.00–135.75)	6.85%
	PA (mg/L)	244.00 (215.00–276.00)	0.63%
	UA (μmol/L)	287.55 (230.70–345.50)	36.64%
	TP (g/L)	71.20 (68.10–74.00)	0.58%
	ALB (g/L)	42.90 (40.80–44.90)	22.74%
	γ-GT (U/L)	16.40 (14.20–24.40)	0.58%
	NEUT (10^9/L)	3.90 (2.92–5.09)	6.95%
	MONO (10^9/L)	0.40 (0.31–0.51)	6.95%
	BC (10^9/L)	0.03 (0.02–0.04)	6.95%
	EC (10^9/L)	0.09 (0.06–0.16)	6.95%
	MPV (fL)	10.00 (9.20–10.80)	7.27%
	LF (10^9/L)	1.54 (1.20–1.95)	6.95%
	RDW (%)	13.20 (12.70–14.20)	6.85%
	MCV (fL)	91.20 (88.10–94.40)	6.85%
	MCHC (g/L)	327.00 (320.00–334.00)	6.85%
	MCH (pg)	29.90 (28.70–31.30)	6.85%
	HCT (L/L)	0.39 (0.36–0.41)	6.85%
	RBC (10^{12}/L)	4.27 (3.99–4.58)	6.85%
	PDW (%)	15.70 (11.70–16.30)	7.27%
	PCT (%)	0.22 (0.19–0.26)	7.42%
	PLT (10^9/L)	222.00 (187.00–270.25)	6.85%
	LDLC (mmol/L)	2.22 (1.71–2.71)	62.00%
	TC (mmol/L)	3.14 (1.70–4.41)	0.00%
	TG (mmol/L)	1.1 (0.87–1.52)	15.17%
	Cr (μmol/L)	60.10 (51.70–76.30)	48.41%
	HDLC (mmol/L)	0.65 (0.37–0.94)	4.29%
Other examinations	DAS28-CRP	4.22 (3.69–5.43)	0.00%

ALB: albumin; BC: basophil count; Cr: creatinine; CRP: C-reactive protein; CVD: cardiovascular disease; DAS28-CRP: disease activity score in 28 joints based on CRP level; EC: eosinophil count; ESR: erythrocyte sedimentation rate; γ-GT: γ-glutamyl transferase; GLB: globulin; HB: hemoglobin; HCT: hematocrit; HDLC: high-density lipoprotein cholesterol; IGU: iguratimod; IQR: interquartile range; JAK: Janus kinase; LDLC: low-density lipoprotein cholesterol; LEF: leflunomide; LF: lymphocyte count; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; MCV: mean corpuscular volume; MONO: monocyte count; MPV: mean platelet volume; MTX: methotrexate; NEUT: neutrophil count; NSAID: nonsteroidal anti-inflammatory drug; PA: prealbumin; PCT: plateletocrit; PDW: platelet distribution width; PLT: platelet count; RBC: red blood cell count; RDW: red blood cell distribution width; RF: rheumatoid factor; TC: total cholesterol; TG: triglyceride; TNF: tumor necrosis factor; TP: total protein; UA: uric acid; ZF: Zhengqing Fengtongning.

recall rate of 97%. Twenty-five patients were predicted to be in the $40 \leq \text{ALT} < 80$ U/L group. Of these, 15 were correctly classified, 10 were incorrectly classified, and 3 patients were mis-classified; the precision and recall rate were 60% and 83%, respectively. Six patients were correctly predicted to be in the ALT $\geq$ 80 U/L group, 1 patient was incorrectly classified, and 3 patients were mis-classified; the precision and recall rate were 83% and 63%, respectively. The results of precision and recall rate indicate that the

occurrence of abnormal liver function was predicted accurately and the model had a good effect to meet the clinical needs.

3.4. External validation

The external validation set contained 106 patient cases. Among them, there were 99 patients with ALT < 40 U/L, 5 patients with $40 \leq \text{ALT} < 80$ U/L, and 2 patients with ALT $\geq$ 80 U/L (Fig. S9).

Table 3
Univariate analysis.

Variable	Statistics	P value	Variable	Statistics	P value
Age	62.649	< 0.001	BC	4.599	0.100
Gender	0.240	0.886	MONO	2.140	0.343
Gender_age	145.348	< 0.001	EC	4.151	0.125
Smoking	33.161	< 0.001	MPV	6.906	0.032
CVD	6.549	0.0378	LF	4.051	0.132
Treatment plan	150.466	< 0.001	RDW	4.276	0.118
Daily dose of ZF	11.305	0.079	MCV	7.558	0.023
CRP	7.008	0.030	MCHC	16.983	< 0.001
RF	7.352	0.025	MCH	16.699	< 0.001
ESR	8.657	0.013	PCV	16.895	< 0.001
DAS28-CRP	0.936	0.626	RBC	3.844	0.146
GLB	2.138	0.343	PDW	1.585	0.453
HB	19.355	< 0.001	PCT	4.514	0.105
PA	8.882	0.012	PLT	3.083	0.214
UA	6.909	0.032	LDLC	2.871	0.238
TP	8.591	0.014	TC	220.098	< 0.001
ALB	10.409	0.005	TG	14.704	0.001
γ-GT	74.655	< 0.001	Cr	0.464	0.793
NEUT	10.482	0.005	HDLC	0.724	0.696

ALB: albumin; BC: basophil count; Cr: creatinine; CRP: C-reactive protein; CVD: cardiovascular disease; DAS28-CRP: disease activity score in 28 joints based on CRP level; EC: eosinophil count; ESR: erythrocyte sedimentation rate; γ-GT: γ-glutamyl transferase; GLB: globulin; HB: hemoglobin; HCT: hematocrit; HDLC: high-density lipoprotein cholesterol; IGU: iguratimod; JAK: Janus kinase; LDLC: low-density lipoprotein cholesterol; LEF: leoflunomide; LF: lymphocyte count; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; MCV: mean corpuscular volume; MONO: monocyte count; MPV: mean platelet volume; MTX: methotrexate; NEUT: neutrophil count; NSAID: nonsteroidal anti-inflammatory drug; PA: prealbumin; PCT: plateletocrit; PDW: platelet distribution width; PLT: platelet count; RBC: red blood cell count; RDW: red blood cell distribution width; RF: rheumatoid factor; TC: total cholesterol; TG: triglyceride; TNF: tumor necrosis factor; TP: total protein; UA: uric acid; ZF: Zhengqing Fengtongning.
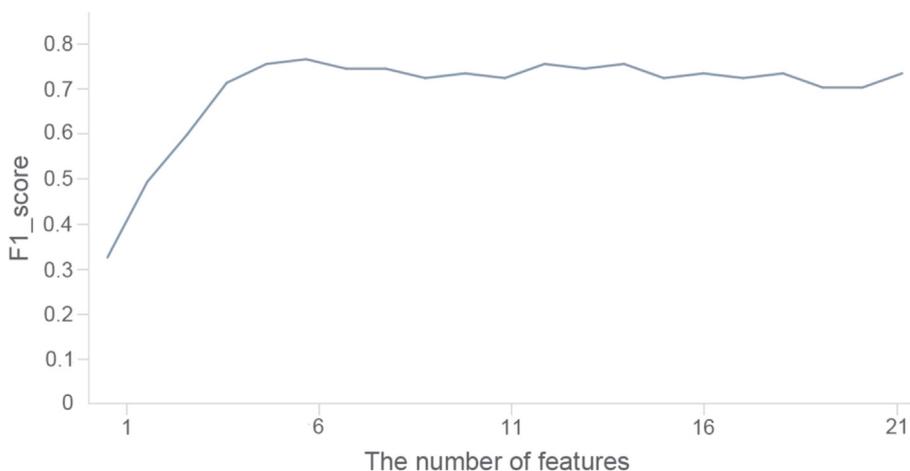

Fig. 3. The F1_score corresponding to the number of ranked features in sequential forward selection.

We predicted the levels of liver function using the model and calculated summary statistics for the model from the confusion matrix. One hundred and one (101) patients were predicted to belong to the ALT < 40 U/L group. Ninety-nine were correctly classified, 2 were incorrectly classified, and no patients were mis-classified. The model had a precision of 98% and a recall rate of 100% for this group. Three patients were correctly predicted to belong to the 40 ≤ ALT < 80 U/L group, and 2 patients were mis-classified, resulting in a precision of 100% and a recall rate of 60%. Two patients were correctly predicted to belong to the ALT ≥ 80 U/L group, and no patients were mis-classified to this group, resulting in a precision of 100% and a recall rate of 100%. The results of precision and recall rate indicate that the occurrence of abnormal liver function was predicted accurately and comprehensively in the validation dataset.

3.5. Model application

We established a web application for predicting ZF-induced abnormal liver function based on the LightGBM model (Fig. 5).

By using the web application, users can predict the risk of ZF-induced abnormal liver function by entering the values of the six important variables on the web page. After the model is applied to the data, the result page will display the predicted liver function value for the submitted data.

4. Discussion

RA is one of the most common immune diseases; currently, it cannot be cured and requires lifelong treatment [26]. Anti-rheumatic drugs are often used in clinical practice to relieve symptoms and improve quality of life, but their long-term use can easily cause ADRs [27,28]. ZF was launched in China in 2000 and has been widely used in the treatment of RA [29]. With the widespread clinical use of traditional Chinese medicine, attention has been given to the ADRs associated with ZF [30–34]. Abnormal liver function was not mentioned in the instructions for ZF as a potential ADR, but a review of 11 studies containing 956 participants revealed that abnormal liver function is one of the most common ADRs
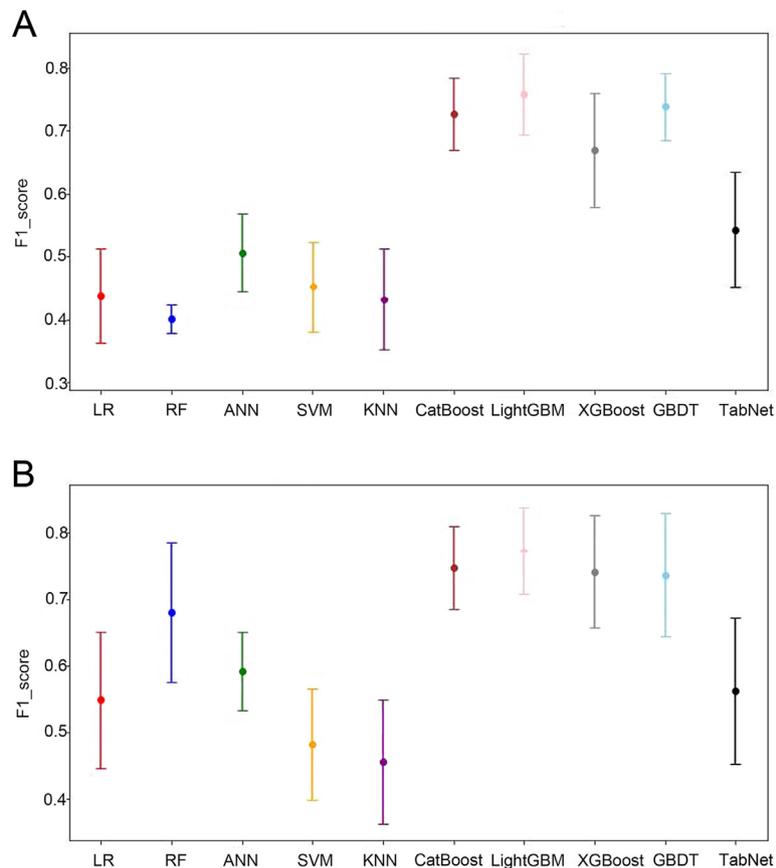
Fig. 4. F1-score of different models based on 22 variables (A) and 6 variables (B) with the mean ± standard deviation. ANN: artificial neural network; CatBoost: categorical boosting; GBDT: gradient boosting decision tree; KNN: k-nearest neighbor; LightGBM: light gradient boosting machine; LR: logistic regression; RF: random forest; SVM: support vector machine; XGBoost: extreme gradient boosting.

associated with ZF [29,35]. Abnormal liver function may lead to severe clinical outcomes, especially in long-term treatment in RA patients. Therefore, it is necessary to predict the risk of abnormal liver function in RA patients using ZF. It is crucial to recognize the factors influencing the ADRs of ZF and to adopt early prevention measures for safe drug use. However, at present, studies based on RWE on the safety of ZF are lacking. This study focused on RA patients treated with ZF and collected RWE from their clinical records. We identified several important variables that contribute to the risk of abnormal liver function and used a machine learning technique to develop a predictive model.

In this study, multiple variables were shown to have remarkable impacts on the risk of abnormal liver function. According to the results of the SHAP analysis, the risk of abnormal liver function increased with increasing TC and TG levels. TC and TG levels are strongly associated with liver function [36]. The liver is the main organ that metabolizes TC and TG. If the function of the liver is abnormal, the synthesis and clearance of TC and TG may be affected, resulting in an increase in blood level of TC and TG [36]. Thus, abnormal liver function may be more likely to occur in patients with elevated TC and TG levels when using ZF treatment; this knowledge may facilitate timely adjustments to treatment regimen based on the prediction results, such as reducing the dose of ZF. Furthermore, the TNF-α inhibitor + ZF, JAK inhibitor + ZF, MTX + NSAIDs + ZF, and LEF + ZF treatment regimens increased the risk of abnormal liver function. Because of the immunomodulatory and anti-inflammatory effects of ZF, it is commonly used in combination with other drugs for treating RA in the clinic. Drug interactions caused by the combination of drugs in treatment plans may either change the strength or the property of the drug effect,

thus affecting the effectiveness and safety of the drug and lead to ADRs [37,38]. In the clinical setting, patients using TNF-α inhibitor + ZF, JAK inhibitor + ZF, MTX + NSAIDs + ZF, and LEF + ZF treatment regimens need to be aware of their risk of developing abnormal liver function. Due to limited conditions, we did not further study the underlying reasons of the effect of drug combination on ZF-induced abnormal liver function, which may be related to drug-drug interaction and should be deeply explored in future study. Additionally, the risk of abnormal liver function was most likely to increase in smoking patients. Smoking increases the production of proinflammatory cytokines (interleukin [IL-1], IL-6 and TNF-α), which cause liver cell damage and may decrease liver function [39]. Thus, smoking patients need to take extra precautions against abnormal liver function during ZF treatment. Wu et al. [40] found that based on 4,064 patients, the occurrence of ADRs associated with ZF was related to gender, specific age group, and drug combination. The authors demonstrated that the risk of ADRs associated with ZF was greater in females and people aged 31 to 50 years [40]. In this study, we investigated the relationships between gender, age and ZF-induced abnormal liver function in diverse gender_age groups. Notably, abnormal liver function is more likely to occur in elderly individuals than in young individuals. Among middle-aged (45–65 years old) patients, the risk of abnormal liver function is greater in females than in males. This may be due to decreases in the physiological activity of various systems and reduced ability to metabolize drugs in the elderly [41]. Moreover, the physiological status of females is different from that of males; females are more sensitive to drugs, have lower tolerance, and typically take a more active role in managing their health [8]. Therefore, the model constructed

Table 4
Model performance.

Model	ALT group	Precision	Recall rate	F1-score	AUC	Sensitivity	Specificity	Accuracy
LR								0.94
	0	0.95	0.99	0.97	0.91	0.99	0.35	
	1	0.33	0.11	0.17	0.86	0.11	0.99	
	2	0.80	0.50	0.62	0.99	0.50	1.00	
RF								0.95
	0	0.95	1.00	0.98	0.98	1.00	0.31	
	1	0.57	0.22	0.32	0.97	0.22	0.99	
	2	1.00	0.13	0.22	0.99	0.13	1.00	
ANN								0.93
	0	0.95	0.99	0.97	0.95	0.99	0.31	
	1	0.14	0.06	0.08	0.92	0.06	0.98	
	2	0.75	0.38	0.50	0.99	0.38	1.00	
SVM								0.94
	0	0.94	1.00	0.97	0.95	1.00	0.12	
	1	0.00	0.00	0.00	0.92	0.00	1.00	
	2	0.75	0.38	0.50	0.99	0.38	1.00	
KNN								0.92
	0	0.96	0.97	0.97	0.90	0.97	0.42	
	1	0.25	0.28	0.26	0.81	0.28	0.96	
	2	0.00	0.00	0.00	0.87	0.00	1.00	
CatBoost								0.95
	0	0.98	0.97	0.98	0.98	0.97	0.69	
	1	0.55	0.61	0.58	0.97	0.61	0.98	
	2	0.86	0.75	0.80	0.99	0.75	1.00	
LightGBM								0.96
	0	0.99	0.97	0.98	0.98	0.97	0.85	
	1	0.60	0.83	0.70	0.98	0.83	0.97	
	2	0.83	0.63	0.71	0.97	0.63	1.00	
XGBoost								0.94
	0	0.95	1.00	0.97	0.98	1.00	0.23	
	1	0.67	0.22	0.33	0.97	0.22	0.99	
	2	0.00	0.00	0.00	0.99	0.00	1.00	
GBDT								0.95
	0	0.99	0.96	0.97	0.98	0.96	0.81	
	1	0.50	0.78	0.61	0.97	0.78	0.96	
	2	0.71	0.63	0.67	0.97	0.63	0.99	
TabNet								0.93
	0	0.96	0.99	0.97	0.94	0.99	0.38	
	1	0.27	0.17	0.21	0.92	0.17	0.98	
	2	0.75	0.38	0.50	0.92	0.38	1.00	

Label "0" indicates patients with ALT < 40 U/L, Label "1" indicates patients with $40 \leq$ ALT < 80 U/L, and Label "2" indicates patients with ALT $\geq$ 80 U/L. ALT: alanine transaminase; ANN: artificial neural network; AUC: area under the curve; CatBoost: categorical boosting; GBDT: gradient boosting decision tree; KNN: k-nearest neighbor; LightGBM: light gradient boosting machine; LR: logistic regression; RF: random forest; SVM: support vector machine; XGBoost: extreme gradient boosting.

in this study was more beneficial for elderly patients and middle-aged female patients. The model predictions allow patients to be aware of potential risks to their liver function when using ZF to help manage RA.

This study compared 10 machine learning and deep learning algorithms. After comparing the predictive metrics, LightGBM was chosen for modeling the risk of developing abnormal liver function. LightGBM is an improved version of the GBDT framework that involves fast, distributed and high-performance methods based on decision trees [13,42]. LightGBM uses three techniques, gradient-based one-side sampling, exclusive feature bundling, and histogram analysis, to require fewer samples, fewer features, and less memory [13,42]. LightGBM has been optimized for parallel computing, supporting feature parallelism and data parallelism, and has been optimized for each parallel mode to reduce traffic [13,42]. In clinical applications, machine learning and deep learning techniques can be applied to develop models with strong generalizability and good robustness, especially in real-world studies [43–45]. With increasing data input, the model performance will become more stable, which is a characteristic of machine learning and deep learning techniques. In short, the LightGBM can build a robust and accurate model relying on fewer samples and less traffic that can be used well for prediction and analysis of large datasets.

To our knowledge, this is the first model for predicting ZF-induced abnormal liver function based on multicenter RWE, which involves grouping predictions according to different ALT levels to provide prompt decision support for individuals. Multicenter studies involving patients from numerous hospitals can increase the utilization of clinical resources and the ability to mine large datasets, resulting in better generalization of the study outcomes. The prediction model was also externally validated in a different medical center to verify the model robustness. Furthermore, we explored the important variables influencing abnormal liver function through machine learning and deep learning techniques, which are well-suited to high-dimensional data mining. Third, we compared different variable sets to achieve the best predictive performance and ultimately obtained a concise and accurate model. Lastly, for further application, we established an easy-to-use online prediction website for clinicians. In brief, we proposed a model to provide accurate prediction of ZF-induced abnormal liver function, which can be easily used in clinical practice for medication guidance.

The limitations of real-world studies are evident. Since these studies were nonexperimental, the outcomes were inevitably influenced by various confounding factors. Due to the large amount of real-world data from different centers, the technical levels, equipment conditions and patient backgrounds at each research

Fig. 5. The web application for predicting ZF-induced abnormal liver function. The web can be accessed through the following website: https://119.3.220.198/work/pec1.html. ALT: alanine transaminase; MTX: methotrexate; RA: rheumatoid arthritis; ZF: Zhengqing Fengtongning.

center may introduce bias and diversity, thus affecting the homogeneity of the clinical data. Herein, all hospitals involved in this study adhere to the "Management Measures for the Mutual Recognition of Medical Institution Examination and Inspection Results," ensuring rigorous quality control standards. These standards are in line with national-level quality assessment criteria, facilitating the mutual recognition of test results among hospitals. In addition, some medical records data, such as uric acid, albumin, low-density lipoprotein cholesterol and creatinine levels, were missing at high rates, mostly due to recording omittance, which may have affected the final outcome. Also, some variables, such as liver-related diseases, metabolic disorders, course of treatment, use of other drugs, and AST levels were not included. We expect to enhance the training of medical staff in the appropriate maintenance of medical records in order to include more comprehensive data in follow-up studies. In the future, a prospective study should be conducted to provide evidence that supports our findings.

5. Conclusion

This study prepared a new prediction model for abnormal liver function in RA patients using ZF and provides guidance for safe drug use. Real-world data were explored in detail. Machine learning and deep learning techniques were introduced in the process of data mining and modeling. Among the 10 algorithms we tested, the LightGBM exhibited superior performance in terms of prediction metrics and model stability. This model may help clinicians predict and prevent the occurrence of abnormal liver function caused by ZF in real-time in clinical practice.

CRediT authorship contribution statement

ZY designed the study, analyzed data, and wrote the draft manuscript. **YG** and **FK** collected data. **HW** provided medical guidance. FG provided methodological guidance. **CL** provided pharmacological guidance. All authors reviewed the manuscript.

Declaration of competing interests

The authors declare the following personal relationships, which may be considered potential competing interests: Fei Gao was employed by Beijing Medicinovo Technology Co., Ltd.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.joim.2024.12.001.

References

[1] Auréal M, Machuca-Gayet I, Coury F. Rheumatoid arthritis in the view of osteoimmunology. Biomolecules 2020;11(1):48.
[2] Ehrlich GE. Kelley's textbook of rheumatology. JAMA 2009;302(8):902–3.
[3] Tian XP, Zeng XF. Enhance treat-to-target strategy by standardize the diagnosis and treatment of rheumatoid arthritis. Zhonghua Nei Ke Za Zhi 2018;57(4):240–1 [Chinese with abstract in English].
[4] Wei HH, Zhao XX, Cheng L, Li SS, Song YL, Han TF, et al. Progress and the prospect of sinomenine to treat rheumatoid arthritis. Zhonghua Zhong Yi Yao Za Zhi 2013;28(10):3001–5 [Chinese with abstract in English].
[5] Li RC, Li J. System review of Zhengqing Fengtongning and methotrexate on treatment of rheumatoid arthritis. Zhongguo Shi Yan Fang Ji Xue Za Zhi 2012;18(11):1–4 [Chinese with abstract in English].
[6] Xu M, Liu L, Qi C, Deng B, Cai X. Sinomenine versus NSAIDs for the treatment of rheumatoid arthritis: a systematic review and meta-analysis. Planta Med 2008;74(20):1423–9.
[7] Teng P, Liu HL, Zhang L, Feng LL, Huai Y, Deng ZS, et al. Synthesis and biological evaluation of novel sinomenine derivatives as anti-inflammatory agents. Eur J Med Chem 2012;50(12):63–74.
[8] Sun YC, Liu LL, Shi TL, Wang CW, Zhang L, Su D, et al. Retrospective analysis of gender differences in ADR of 7690 cases of reported adverse drug reactions. Anhui Yi Xue 2015;36(4):436–40 [Chinese with abstract in English].

[9] Blonde L, Khunti K, Harris SB, Meizinger C, Skolnik NS. Interpretation and impact of real-world clinical data for the practicing clinician. Adv Ther 2018;35 (11):1763–74.

[10] Palacios M. The quality of research with real-world evidence. Colomb Med (Cali) 2019;50(3):140–1.

[11] Robson C. Real world research. 3rd ed. New York: John Wiley & Sons Ltd; 2011.

[12] Chen TQ, Guestrin C. XGBoost: a scalable tree boosting system. In: KDD'16: proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. San Francisco: Association for Computing Machinery; 2016. p. 785–94.

[13] Ke G, Meng Q, Finley T, Wang TF, Chen W, Ma WD. LightGBM: a highly efficient gradient boosting decision tree. In: von Luxburg U, Guyon I, Bengio S, Wallach H, Fergus R, editors. NIPS'17: proceedings of the 31st international conference on neural information processing systems. Long Beach: Curran Associates, Inc.; 2017. p. 3149–57.

[14] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. Catboost: unbiased boosting with categorical features. In: Bengio S, Wallach HM, Larochelle H, Grauman K, Cesa-Bianchi N, editors. NIPS'18: proceedings of the 32nd international conference on neural information processing systems. Red Hook: Curran Associates Inc.; 2018. p. 6629–49.

[15] Zhang J, Mucs D, Norinder U, Svensson F. LightGBM: an effective and scalable algorithm for prediction of chemical toxicity-application to the tox21 and mutagenicity data sets. J Chem Inf Model 2019;59(10):4150–8.

[16] Arik SO, Pfister T. TabNet: attentive interpretable tabular learning. ArXiv; 1908.07442. [2024-01-10] https://doi.org/10.48550/arXiv.1908.07442.

[17] Guo W, Yu Z, Gao Y, Lan XQ, Zang YN, Yu P, et al. A machine learning model to predict risperidone active moiety concentration based on initial therapeutic drug monitoring. Front Psychiatry 2021;12:711–868.

[18] Ma P, Liu RX, Gu WR, Dai Q, Gan Y, Cen J, et al. Construction and interpretation of prediction model of teicoplanin trough concentration via machine learning. Front Med (Lausanne) 2022;9:808–969.

[19] Yu Z, Ji HH, Xiao JW, Wei P, Song L, Tang TT, et al. Predicting adverse drug events in Chinese pediatric inpatients with the associated risk factors: a machine learning study. Front Pharmacol 2021;12:659099.

[20] Huang XH, Yu Z, Bu SH, Lin ZY, Hao X, He WJ, et al. An ensemble model for prediction of vancomycin trough concentrations in pediatric patients. Drug Des Devel Ther 2021;15:1549–59.

[21] Ma J, Yu Z, Chen T, Li P, Liu Y, Chen JH, et al. The effect of Shengmai injection in patients with coronary heart disease in real world and its personalized medicine research using machine learning techniques. Front Pharmacol 2023;14:1208621.

[22] U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute. Common terminology criteria for adverse events (CTCAE), Version 4.0. NIH publication 09-7473. Published May 29, 2009. Revised June 14, 2010. [2024-08-13]. http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE4.03_2010-06-14_QuickReference_5x7.pdf.

[23] van Delden JJ, van der Graaf R. Revised CIOMS international ethical guidelines for health-related research involving humans. JAMA 2017;317(2):135–6.

[24] Hatamikia S, Maghooli K, Nasrabadi AM. The emotion recognition system based on autoregressive model and sequential forward feature selection of electroencephalogram signals. J Med Signals Sens 2014;4(3):194–201.

[25] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: von Luxburg U, Guyon I, Bengio S, Wallach H, Fergus R, editors. NIPS'17: Proceedings of the 31st international conference on neural information processing systems. Long Beach: Curran Associates, Inc.; 2017. p. 4765–74.

[26] Ibfelt EH, Jacobsen RK, Kopp TI, Cordtz RL, Jakobsen AS, Seersholm N, et al. Methotrexateand risk of interstitial lung disease and respiratory failure in rheumatoid arthritis: a nationwide population-based study. Rheumatology (Oxford) 2021;60(1):346–52.

[27] Wang R, Yang J, Liu YX, Guan BX, Feng Q, Wang YH, et al. The effect of acupoint application of sinomenine for rheumatoid arthritis measured by micro-dialysis and UPLC-MS/MS. Evid Based Complement Alternat Med 2019;2019:5135692.

[28] Huang RY, Pan HD, Wu JQ, Zhou H, Li ZG, Qiu P, et al. Comparison of combination therapy with methotrexate and sinomenine or leflunomide for active rheumatoid arthritis: a randomized controlled clinical trial. Phytomedicine 2019;57:403–10.

[29] Wu ZZ, Long LP, Chen X, Chen GJ. Analysis of 193 cases of adverse drug reactions induced by Zhengqingfengtongning sustained release tablets. Zhongguo Yao Wu Jing Jie 2016;13(8). 489–91, 495 [Chinese with abstract in English].

[30] Du MR, Feng FH, Zhang J. Efficacy of Zhengqingfengtongning Tablets and non-steroid anti-inflammatory drugs combination therapy for osteoarthritis: meta-analysis. Zhonghua Lin Chuang Mian Yi He Bian Tai Fan Ying Za Zhi 2015;9 (2):121–5 [Chinese with abstract in English].

[31] Yang J, Zeng H, Zheng F. Comparing curative effects of Zhengqing Fengtongning and chemical medicines on knee osteoarthritis: meta-analysis. Zhong Cheng Yao 2015;37(6):1197–201 [Chinese with abstract in English].

[32] Wang BX, Li X, Gu J, Huang HY, Wang JG, Fan FY, et al. Systematic review of Zhengqing Fengtongning on treatment of rheumatoid arthritis. Hunan Zhong Yi Yao Da Xue Xue Bao 2015;35(3):67–71 [Chinese with abstract in English].

[33] Liang SH, Lan WR, Li YK. Efficacy and safety of Zhengqing Fengtongning for ankylosing spondylitis: a systematic review. Zhongguo Shi Yan Fang Ji Xue Za Zhi 2013;19(21):323–30 [Chinese with abstract in English].

[34] Yang QR, Wang WH, Ma Z, Ren SH. Clinical observation of Zhengqing Fengtongning sustained-release tablet combined with glucocorticoid in treatment of primary micropathological nephropathy. Shi Yong Zhong Yi Yao Za Zhi 2016;32(4):345–6 [Chinese with abstract in English].

[35] Chen XM, Huang RY, Huang QC, Chu YL, Yan JY. Systemic review and meta-analysis of the clinical efficacy and adverse effects of Zhengqing Fengtongning combined with methotrexate in rheumatoid arthritis. Evid Based Complement Alternat Med 2015;2015:910376.

[36] Endo-Umeda K, Makishima M. Liver X receptors regulate cholesterol metabolism and immunity in hepatic nonparenchymal cells. Int J Mol Sci 2019;20(20):5045.

[37] Huang YX, Liu EW. Drug interaction and research methods of combined drug administration. Yao Wu Ping Jia Yan Jiu 2014;37(3):276–9 [Chinese with abstract in English].

[38] Yu X, Zhang HX. Adverse drug reaction case reports should not be ignored drug interaction. Yao Wu Liu Xing Bing Xue Za Zhi 2012;21(6):298–300 [Chinese with abstract in English].

[39] El-Zayadi AR. Heavy smoking and liver. World J Gastroenterol 2006;12 (38):6098–101.

[40] Wu ZZ, Long LQ, Qiu P, Wen YY, Xie ZX, Li LL. Influencing factors of adverse drug reactions induced by Zhengqing Fengtongning sustained release tablets. Zhongguo Yao Ye 2018;27(21):83–5 [Chinese with abstract in English].

[41] Long LP, Zhou YL, Yang GP. Analysis of 309 cases of aged people's serious adverse drug reaction. Zhongguo Yao Ye 2009;6(1):26–30 [Chinese with abstract in English].

[42] Meng Q, Ke GL, Wang TF, Chen W, Liu TY, Ye QW, et al. A communication-efficient parallel algorithm for decision tree. In: Lee D, Sugiyama M, Luxburg U, Guyon I, Garnett R, editors. Advances in neural information processing systems 29: 30th annual conference on neural information processing systems 2016 (Barcelona, Spain). Red Hook: Curran Associates, Inc.; 2016. p. 1279–87.

[43] Mo XL, Chen XJ, Li HW, Li JL, Zeng FL, Chen YL, et al. Early and accurate prediction of clinical response to methotrexate treatment in juvenile idiopathic arthritis using machine learning. Front Pharmacol 2019;10:1155.

[44] Lee HC, Yoon SB, Yang SM, Kim WH, Ryu HG, Jung CW, et al. Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. logistic regression model. J Clin Med 2018;7(11):428.

[45] Kruppa J, Ziegler A, König IR. Risk estimation and risk prediction using machine-learning methods. Hum Genet 2012;131(10):1639–54.