Journal of Hydroinformatics

Journal of Hydroinformatics Vol 26 No 3, 641 doi: 10.2166/hydro.2024.247

Analysis of characteristic index and prediction of river bottom tearing scour in the Yellow River

Longfei Sun[a,b], Yanhui Liu[a,b], Yuanjian Wang[a,b,*], Qinghao Dong[a,c] and Wanjie Zhao[a,d]

[a] Yellow River Institute of Hydraulic Research, Yellow River Conservancy Commission, Zhengzhou 450003, China
[b] Key Laboratory of Lower Yellow River Channel and Estuary Regulation, Ministry of Water Resources, Zhengzhou 450003, China
[c] School of Water Conservancy and Transportation, Zhengzhou University, Zhengzhou 450001, China
[d] College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China
*Corresponding author. E-mail: wangyuanjian_yrcc@aliyun.com

ABSTRACT

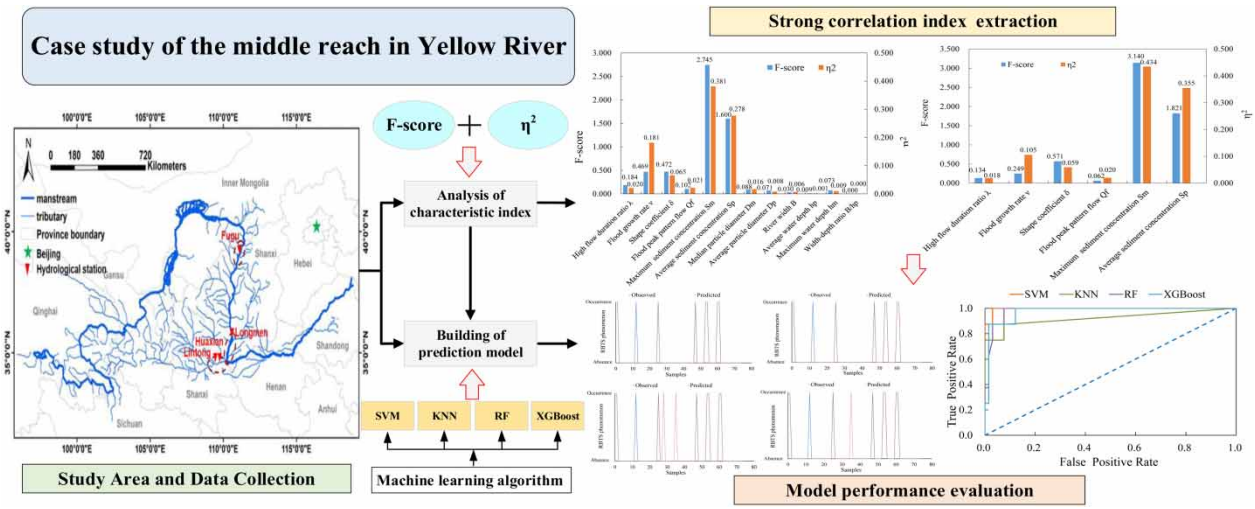
River bottom tearing scour (RBTS) has a strong effect on the scouring and moulding of channel in the Yellow River. Due to the special forming conditions, complex influencing factors, and limited observed data, it is difficult to predict whether RBTS will occur accurately. By collecting and disposing of the hydrodynamic, sediment, and initial boundary data of 246 flood events related to RBTS in three typical reaches of the Yellow River basin, the correlation between different characteristic influencing factors and the occurrence and absence of RBTS were analysed, and prediction models based on machine learning algorithms were constructed. The results showed that under the existing data conditions, the maximum sediment concentration S_m, average sediment concentration S_p, flood growth rate ν, and shape coefficient δ were the four key indices to more easily distinguish whether RBTS will occur. The support vector machine algorithm model had the best performance results and exhibited higher accuracy and precision in predicting its occurrence compared with other models under given water and sediment conditions. The method proposed in this study provides a new method for accurately predicting RBTS in the Yellow River.

Key words: characteristic index, machine learning, prediction model, river bottom tearing scour

HIGHLIGHTS

- The maximum sediment concentration, average sediment concentration, flood growth rate, and shape coefficient are the four key indices for distinguishing whether RBTS will occur.
- Prediction models of RBTS based on machine learning algorithms were built.
- The SVM model showed the best predictive performance in the case study of the middle reach of the Yellow River.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Researchers have focused on the erosion and deposition patterns of river channels both domestically and abroad. In China, as the second largest river, the Yellow River has the typical characteristics of less water and more sediment, and the changes in its main and tributary channels have naturally aroused the interest of numerous researchers (Giglou *et al.* 2016; Bi *et al.* 2019; Chen *et al.* 2021; Medel *et al.* 2022). In general, the river cohesive sediment will be gradually consolidated and compacted, affected by the long-term water flow, and form stable structures with strong ability of resistance against erosion. When it encounters a huge flood, there may cause channel scour and sediment transport if the hydrodynamic forces surpass the particle resistance forces (Slaa *et al.* 2013; Bosa *et al.* 2018; Das *et al.* 2019). In the areas of Longmen and Lintong of the Weihe River, there is a special phenomenon of large-scale and long-distance erosion occurring in a short period due to high-sediment floods, causing the riverbed sediment layer to be uplifted from the river bottom and exposed in blocks or patches on the water surface. The block area can reach several square metres or even tens of square metres. Subsequently, the sediment is carried away by the flowing water in a short period, resulting in the riverbed being scoured deeply by several or even ten metres during a single flood, which is called river bottom tearing scour (RBTS; Figure 1) (Kuang *et al.* 2000; Cao *et al.* 2006; Van Maren *et al.* 2009a; Jiang *et al.* 2010).

The occurrence of RBTS needs certain conditions. The channel deposition is usually quite serious before its occurrence, and the riverbed sediment has become dense and a certain thickness. Moreover, the clay blocks composed of fine cohesive sediment particles has formed after a period of consolidation, surrounded by granular coarse sand. On the basis of the formation of clay blocks, there are also required adequate hydrodynamic conditions. High-sediment concentration flood from upstream with the strong ability of carrying sand can improve the viscosity force of water flow, and intensify the impact of sediment particles on the silted riverbed, which can not only provide sufficient energy for the upward lifting of sediment at the river bottom, but also continuously carry the sediment broken by the impact to the downstream. Therefore, when the high-sediment concentration flood occurs, as the discharge and velocity increases, the main channel appears strongly eroded, which may result in the RBTS phenomenon (Van Maren *et al.* 2009b; Wang *et al.* 2009; Tabarestani & Zarrati 2015; He & Yan 2019; Anand *et al.* 2021).

RBTS in the Yellow River has a strong effect on the scouring and moulding of the channel and may cause channel erosion and deposition, main channel migration, and constantly changing slip position of engineering, which easily causes river engineering to collapse and brings tremendous pressure to flood defence (Jiang *et al.* 2015). Therefore, it can provide an important basis for dealing with the occurrence of RBTS phenomenon and ensuring the safety of river engineering by studying the influencing factors of its occurrence conditions and achieving accurate prediction, which has positive practical significance for river management.

Presently, research on RBTS in the Yellow River has mainly focused on its genetic mechanisms and mechanical analysis. From a mechanical point of view, RBTS can be divided into three stages: separation, sliding, and turning of the settlement layer of the riverbed (Gou 2004). Based on the analysis of the original test data, combined with the method of the generalised model test, the main conditions of the RBTS are considered to be in the form of sediment, water and sediment conditions,

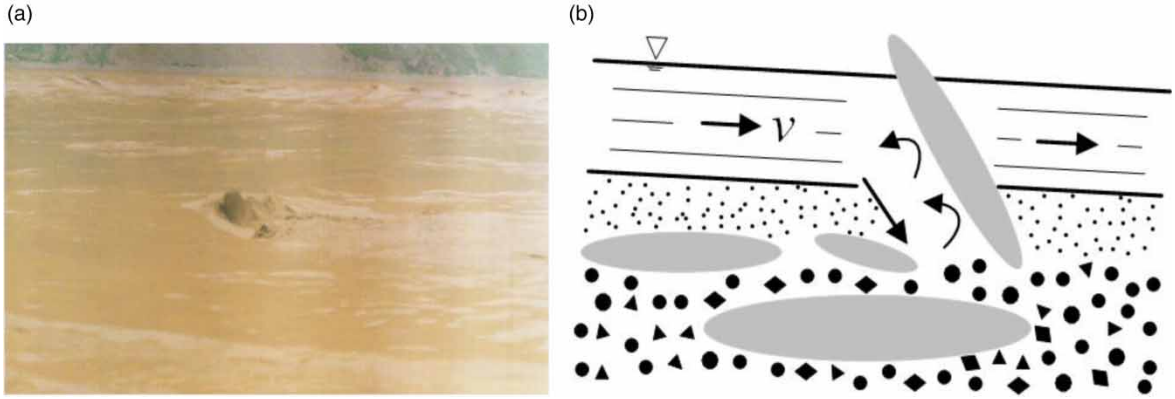


Figure 1 | The actual scenario (a) and schematic diagram (b) of the RBTS phenomenon.

characteristics of the previous channel boundary, and the instantaneous lifting force caused by the different propagation speeds of the pulsating pressure wave on the upper and lower surfaces of the clay block. A critical criteria for determining the occurrence of the phenomenon was established (Jiang *et al.* 2010; Zhang *et al.* 2022b). A relationship model of the clay particle size distribution and shear strength of the RBTS reach in the Yellow River was established and solved (Dong *et al.* 2012). Furthermore, according to the characteristics of the sediment layer, some researchers proposed the concept of an elastic viscous layer of the riverbed and calculated the critical instability force of the cemented block using the material mechanics cross-section method and the principle of minimum potential energy (Zhang & Hu 2013). However, existing research on the discrimination conditions of RBTS occurrence is not uniform, and some assumptions and simplification methods have been used to obtain the discriminant index derived from empirical formulas or parameters. The reliability of the index lacks sufficient verification and applicability, and its accuracy is limited.

With the continuous development of artificial intelligence and data mining technology, intelligent algorithms such as machine learning, which use automatic data analysis and modelling to predict multifactor and nonlinear system problems, have been widely used in several scientific fields and become effective tools for data analysis. These methods have also been applied to related research on the prediction about the evolution of the Yellow River's and other rivers' bed. For instance, Xia *et al.* (2023) used remote sensing image, cross-sectional terrain, water and sediment data to construct a random forest (RF) prediction model for thalweg migration in the wandering reach of the lower Yellow River, and the average accuracy of different river sections was up to 80%. Hu *et al.* (2023) proposed a new method for monitoring river suspended sediment concentrations (SSCs) that combines the remote sensing technique and light gradient-boosting machine method, and it showed accuracy and robustness in the experiment of the lower Yellow River. Yan *et al.* (2023) developed a solution to discriminate river patterns based on rough set theory, and compared with several competitive machine learning methods including support vector machine (SVM), extreme gradient-boosting (XGBoost), and deep neural networks (DNNs). Though the proposed methods displayed good performance with the advantages of interpretability, simple modelling, and fewer training samples, the results showed that the XGBoost even had the highest performance as one of the black box models. Li *et al.* (2021) used 10-m Sentinel-2 multispectral instrument (MSI) imagery and digital elevation model (DEM) data, as well as RF algorithm to extract bankfull river widths on the upper Yellow River. Moreover, Ahmed *et al.* (2019) used remote sensing and unsupervised machine learning techniques to quickly forecast regions that are subject to future river sediment deposition. Xu *et al.* (2022) built an integrated model combining a numerical fluid-flow and sediment model with the long short-term memory (LSTM) module to analyse the detailed process of the bed-steadying discharge (Q_S) in the Middle Huaihe River. Ren *et al.* (2020) adopted RF to identify the most influential factors and develop the model for categorising and mapping the spatial distributions of riverbed substrate grain size of the Hanford Reach located in the Columbia River Basin. In these relevant researches, multiple input parameters usually need to be considered, and machine learning methods can be used to quickly respond to driving factors and construct models. The models have shown good predictive performance. However, there are few reports on using machine learning to solve the prediction problem of RBTS in the Yellow River. Complex nonlinear relationships exist between each influencing factor for the RBTS, and machine learning algorithms can be considered to overcome the problem of establishing complex relationships between multiple factors. Therefore, the study on accurate prediction models and application effect evaluations of machine learning algorithms for RBTS prediction still needs addressing.

Due to the special formation conditions, complicated influencing factors, and limited observed data of RBTS, predicting the occurrence of RBTS is difficult. To solve the above problems, this study establishes a prediction model for the RBTS phenomenon based on the machine learning algorithms to provide theoretical and technical support for its accurate prediction and effective prevention and control.

2. METHODS

In this study, the observed water, sediment, and initial boundary data related to the occurrence of RBTS in typical reaches of the Yellow River Basin under different flood events were collected, and the corresponding characteristic influencing factors were calculated according to three categories: hydrodynamic, sediment, and initial boundary. Second, the key indices of strong correlation were extracted based on the calculation method of the degree of influence for evaluating the characteristic influencing factors on the dependent variables. Subsequently, different machine learning classification algorithms were used to establish a prediction model considering different characteristic indices. Finally, the prediction effects of different models

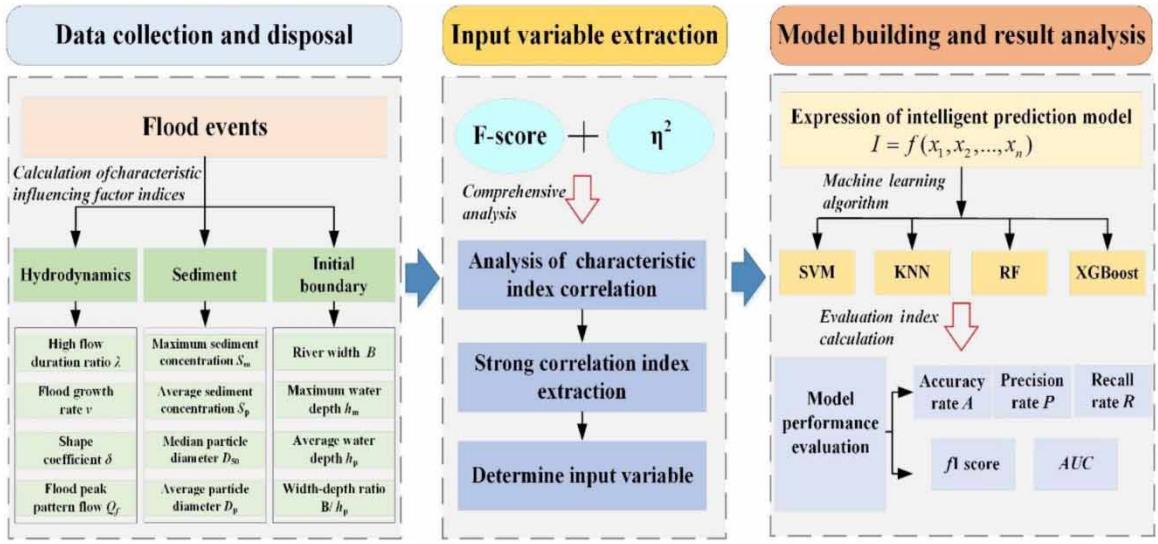


Figure 2 | Overall framework of the methodology adopted in this study.

were evaluated, and the best prediction model was obtained to accurately predict the RBTS phenomenon in the Yellow River. The overall framework is illustrated in Figure 2.

2.1. Characteristic influencing factor calculation

Based on the mechanism of the RBTS phenomenon and existing studies, the characteristic influencing factor indices selected in this study were mainly divided into hydrodynamic, sediment, and initial boundary (Dong *et al.* 2011; Li *et al.* 2017; Liu 2018; Liu *et al.* 2021; Zhang *et al.* 2022a).

2.1.1. Hydrodynamic factor

(1) High flow duration ratio λ

λ is an index that reflects the flood peak shape and can be expressed as shown in Equation (1):

$$\lambda = T_1/T \tag{1}$$

where T_1 refers to the duration of the flow that is greater than 75% of the peak flow, and T refers to the entire duration of the flood.

(2) Flood growth rate v

v is an index that reflects the growth rate of the rise process of flood, and can be expressed as shown in Equation (2):

$$v = \frac{Q_m - Q_{t0}}{(t_1 - t_0) \times 24} \tag{2}$$

where Q_m refers to the peak discharge, Q_{t0} refers to the flow at the beginning of the flood, and t_1 and t_0 are the corresponding times of the peak discharge and beginning of the flood, respectively.

(3) Shape coefficient δ

δ is an index that reflects the form of the flood process throughout the entire period, and can be expressed as shown in Equation (3):

$$\delta = \frac{W}{Q_m T \times 0.36} \tag{3}$$

where W is the flood volume, Q_m is the peak discharge, and T is the flood duration.

(4) Flood peak pattern flow Q_f

Q_f is an index that reflects the average discharge under statistical conditions, and can be expressed as shown in Equation (4):

$$Q_f = \bar{Q}\frac{Q_m - Q_n}{Q_m + Q_n} \tag{4}$$

where Q_m is the peak discharge, Q_n is the minimum flow, and $\bar{Q}$ is the average discharge.

2.1.2. Sediment factor

(1) Maximum sediment concentration S_m

S_m is an index that reflects the maximum value of the sediment concentration during an entire flood and can be expressed as shown in Equation (5):

$$S_m = \max\{S_i\} \tag{5}$$

where S_i is the daily average sediment concentration.

(2) Average sediment concentration S_p

S_p is an index that reflects the average value of the sediment concentration during an entire flood and can be expressed as shown in Equation (6):

$$S_p = \frac{1}{T}\sum_{i=1}^{T} S_i \tag{6}$$

where S_i is the daily average sediment concentration and T is the entire duration of the flood.

(3) Median particle diameter D_m and average particle diameter D_p

D_m is the corresponding particle size when the cumulative particle size distribution percentage reaches 50% and D_p is the average value of particle size obtained by using standardised sampling methods on the day that is closest to the start date of the flood.

2.1.3. Initial boundary factor

(1) River width B

B is an index that reflects the width of the channel shape before the flood and can be expressed as shown in Equation (7):

$$B = L_2 - L_1 \tag{7}$$

where L_1 and L_2 are the starting distances of the left and right sides corresponding to the water level elevation respectively.

(2) Maximum water depth h_m

h_m is an index that reflects the extreme value of the water depth before a flood and can be expressed as shown in Equation (8):

$$h_m = \max\{z_w - z_i\} \tag{8}$$

where z_w is the elevation of the water level and z_i is the elevation of the river bottom at different locations in a section.

(3) Average water depth h_p

h_p is an index that reflects the average value of the water depth before a flood and can be expressed as shown in Equation (9):

$$h_p = \frac{1}{n}\sum_{i=1}^{n}(z_w - z_i) \tag{9}$$

where n is the number of measurement points at different positions in a section.

(4) Width-depth ratio B/h_p

B/h_p is a comprehensive index that reflects the initial boundary conditions of the river channel before the flood.

2.2. Key indicator extraction with strong correlation

In this study, due to the occurrence and absence of the RBTS phenomenon is a discrete variable with only two types of results and the indices of the characteristic influencing factors are continuous data variables, the conventional correlation analysis method (such as correlation coefficient R^2) is not applicable to reflect the degree of correlation between continuous and discrete variables well. For the correlation analysis between discrete and continuous variables, the F-score and η^2 index were combined to evaluate the degree of influence of different factors on the dependent variable, which could provide better threshold standards to select key factors for subsequent modelling.

(1) F-score

The larger the F-score value, the greater the influence of this characteristic influencing factor on the dependent variable. Therefore, it is easier for this characteristic factor to distinguish whether the RBTS phenomenon occurs. The F-score is defined as (Chen & Lin 2006):

$$F(i) = \frac{(\bar{X}_i^+ - \bar{X}_i)^2 + (\bar{X}_i^- - \bar{X}_i)^2}{\sum_{k=1}^{n_+}(x_{k,i}^+ - \bar{X}_i^+)^2/(n_+ - 1) + \sum_{k=1}^{n_-}(x_{k,i}^- - \bar{X}_i^-)^2/(n_- - 1)} \tag{10}$$

where $\bar{X}_i$, $\bar{X}_i^+$, and $\bar{X}_i^-$ are the average values of the ith characteristic index in the dataset of the total, occurrence, and absence of RBTS, respectively; $x_{k,i}^+$ and $x_{k,i}^-$ are the kth values of the ith characteristic index in the dataset of the occurrence and absence of RBTS, respectively; and n_+ and n_- are the number of occurrences and absences of RBTS, respectively.

(2) η^2

η^2 is an index used to evaluate the degree of influence of independent variables on dependent variables in continuous-discrete variables. According to empirical criteria, the correlation is small, medium, and large when $\eta^2 \in [0.01, 0.06)$, $\eta^2 \in [0.06, 0.14)$, and $\eta^2 \in [0.14, \infty)$, respectively(Chen 2020), and η^2 is defined as:

$$\eta^2 = \sum_{k\in I} n_k(\bar{X}_k - \bar{X})^2 / \sum_{k\in I}\sum_{i=1}^{n_k}(X_{ik} - \bar{X})^2 \tag{11}$$

where k is the group of the RBTS phenomenon, $I = \{\text{occurrence, absence}\}$, n_k is the number of samples in a group, $\bar{X}_k$ is the average value of the characteristic index in the k group, $\bar{X}$ is the average value of the characteristic index in all groups, and X_{ik} is the ith value of the characteristic index in the k group.

Based on the above calculation formula of the F-score and η^2, the correlation between the index of characteristic influencing factors and the occurrence or absence of the RBTS phenomenon was analysed. The extraction steps of key indices with strong correlation are as follows:

(1) For collected data of each flood, calculate the indices of characteristic influencing factors in hydrodynamics, sediment, and initial boundary;
(2) Calculate the F-score and η^2 using Equations (10) and (11);
(3) According to the empirical results of $\eta^2$and the control criteria of the F-score, select an appropriate threshold [F-score] to distinguish the key indices with strong correlations from others with weak correlations;

(4) The key indices with strong correlations are selected as the input variables of the subsequent prediction model.

2.3. Basic principle of machine learning algorithms

The prediction models of the RBTS phenomenon were built based on four machine learning algorithms, including SVM, k-nearest neighbour (KNN), RF, and XGBoost. These four algorithms belong to different types and have good competitiveness and performance, which is more conducive to obtaining the optimal prediction model.

(1) SVM

The SVM algorithm is a supervised learning method, and has the advantage of better prediction accuracy owing to its ability to constitute a decision rule that generalises well in a high-dimensional feature space. The training goal was to determine an optimal hyperplane that could effectively segment different categories of data in the training set (Yang *et al.* 2013; Aburomman & Reaz 2017; Huang *et al.* 2023). For a given training set of m pairs of data points $\{x_i, y_i\}_{i=1}^{m}$, where $x_i \in R^n$, $y_i \in \{-1, +1\}$, x_i and y_i are the feature vector and category label of the ith sample, respectively, and it can be equivalent to the following optimisation problem:

$$\begin{aligned} &\min_{w,b} \frac{1}{2}\|w\|^2 + c\sum_{i-1}^{l} \xi_i \\ &\text{s.t.} y_i(w \cdot x_i + b) \geq 1 - \xi_i, i = 1, 2, \ldots, m \\ &\xi_i > 0, i = 1, 2, \ldots, m \end{aligned} \tag{12}$$

where w is a normal vector of the hyperplane, b is a bias, ξ_i is the "slack variable" and is a hinge loss function, and c is a regularisation constant that controls the trade-off between the classification margin and misclassification cost.

(2) KNN

The KNN algorithm is a classical and effective algorithm for classification recognition that has been widely applied in several domains owing to its methodological simplicity, nonparametric working principle, and easy implementation.

Its basic principle is to compare the current new test samples to be classified with other training sample data, select the k most similar data samples according to the distance size, and predict the category of sample data to be classified according to the labels of k most similar data samples using the majority voting principle (Wang *et al.* 2020; Kim 2021). The distances can be calculated as shown in Equation (13):

$$d(x_i, y_i) = \sqrt{(x - y_i)^T (x - y_i)} \tag{13}$$

where x_i and y_i are the ith data points of samples X and Y, respectively.

(3) RF

The RF algorithm is an integrated classification method based on multiple decision trees built from data subspaces. It can achieve high accuracy in classifying high-dimensional data and has the characteristics of trivial parallelisation, high speed, and strong generalisation ability. It uses a random method to generate a decision tree by introducing attribute selection in the training process and finally classifies by summarising the results of the decision tree (Mantas *et al.* 2019; Wang *et al.* 2022). The basic steps of the RF algorithm classification are:

(1) Randomly select N training samples from the original dataset and obtain k independent training sample subsets by k round extraction with samples that are returned;
(2) Different decision tree models can be constructed by selecting features from the original input variables, and the feature with the best value is used to split the nodes;
(3) Each decision tree returns a classification value that is a vote for that class. Based on the different decision tree results, the final classification result is determined using integrated voting.
(4) XGBoost

XGBoost is optimised and improved based on the gradient progressive regression tree algorithm, which is a typical ensemble model with excellent learning effects and high computational performance. The basic idea is to use multiple decision trees

as base classifiers. In each training session, the residual of the last predicted result is added to a new base classifier for learning. By constantly adding new decision trees to minimise the loss value, multiple base classifiers are weighted and integrated into a strong estimator for predictive analysis to improve the model accuracy and obtain the final classification results (Gu *et al.* 2022; Pan *et al.* 2022).

The objective function of the XGBoost algorithm can be expressed as shown in Equation (14):

$$Obj(\theta) = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{j=1}^{m} \Omega(f_j) \tag{14}$$

where n is the total amount of sample data; y_i and $\hat{y}_i$ are the observed and predicted values of the sample, respectively; and $L(y_i, \hat{y}_i)$ is the loss function that reflects the difference between them. $\Omega(f_j)$ is the complexity of the model and f_j is the model of the jth tree.

2.4. Prediction model construction and effect evaluation based on machine learning

For the key feature influencing factor indicators extracted with strong correlation, the machine learning algorithms were applied to establish a prediction model that comprehensively considered different feature indicators for whether RBTS occurs. The methods for establishing the prediction model and evaluating its effects are as follows:

(1) The key indices of strong correlation determined by the F-score and η^2 method analysis, and the occurrence and absence of the RBTS phenomenon were used to build the expression of the prediction model as input and output, respectively:

$$I = f(x_1, x_2, \ldots, x_n) \tag{15}$$

where I is the dependent variable of whether the RBTS phenomenon occurs, and $x_1, x_2, \ldots, x_n$ refers to the extracted key influencing factor indices.

(2) Normalised pre-processing for the selected sample data was performed, and the calculation formula was as follows:

$$\omega' = \frac{\omega - \bar{\omega}}{\sigma} \tag{16}$$

where ω' is the normalised data, ω is the original data, $\bar{\omega}$ is the average of the original data, and σ is the standard deviation of the original data.

(3) The data were divided according to a certain proportion, and the training and test samples were determined.

(4) The training samples were placed into different machine learning classification algorithms for learning and the parameter combinations of the algorithms were adjusted and optimised to establish prediction models that comprehensively considered different characteristic indices of RBTS.

(5) The test data were placed into the established model for calculations to obtain the predicted result and then compared with the observed situation. The prediction accuracy of the model was evaluated according to five indices: accuracy rate A, precision rate P, recall rate R, $f1$ score, and AUC value (the area under the receiver operating characteristic (ROC) curve) (Fu *et al.* 2023). The calculation formulas for the first four evaluation indices are:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$P = \frac{TP}{TP + FP} \tag{18}$$

$$R = \frac{TP}{TP + FN} \tag{19}$$

$$f1 = \frac{2PR}{P + R} \tag{20}$$

where TP and TN are the correctly predicted results of the occurrence and absence of the RBTS phenomenon,

respectively; *FP* and *FN* are the incorrectly predicted results of the occurrence and absence of the RBTS phenomenon, respectively.

(6) The preferred machine learning prediction model was determined.

3. STUDY AREA AND DATA COLLECTION

The length of the main stream of the Yellow River is 5,464 km. According to the geographical, geological, and hydrological conditions of the formation and development of the basin, the main stream is divided into three sections: upper, middle, and lower reaches. The middle reaches of the Yellow River extend from Hekouzhen in Shanxi to Taohuayu in Henan Province. Its length and basin area are 1,206 km and 344,000 km^2, respectively. Moreover, most tributaries in the middle reaches of the Yellow River are located in the Loess Plateau area with loose soil, and the major tributaries are the Weihe and Fenhe rivers. The reaches of Fugu, Longmen to Tongguan, and Lintong to Huaxian of the Weihe River in the middle reaches of the Yellow River were the main streams with more RBTS phenomena, and selected as the study area (Figure 3).

A total of 246 floods obtained from four hydrological stations, Fugu, Longmen, Lintong, and Huaxian, in three different reaches of the Yellow River Basin were selected as the research objects to conduct data analysis (Figure 3). Furthermore, hydrological data on discharge and sediment of the floods from 1954 to 2017 were collected from measured data gauged at different hydrological stations. Based on the specific occurrence time of the RBTS phenomenon, the flood events of the RBTS absence were selected in the before and after adjacent years to ensure the relevant data more effectively reflected the comparison between occurrence and absence of the RBTS. The main data information is presented in Table 1. All measured data were provided by the Yellow River Conservancy Commission in Zhengzhou, China. In addition, due to lack of existing large cross-sectional data, the water depth and river width data of the nearest time before the occurrence and absence of the RBTS floods was used to reflect the initial boundary conditions and distinguish from hydrodynamic and sediment factors during the flood processes. No further adjustments were made to the datasets, as this would have introduced additional scale-dependent issues requiring further correction.

4. RESULTS AND DISCUSSION

4.1. Analysis of correlation degree of different characteristic influencing factors

According to the research methods in Sections 2.1 and 2.2, various characteristic influencing factor indices were calculated based on the measured data of different flood events. Notably, there were no data of particle size in the sediment factors for

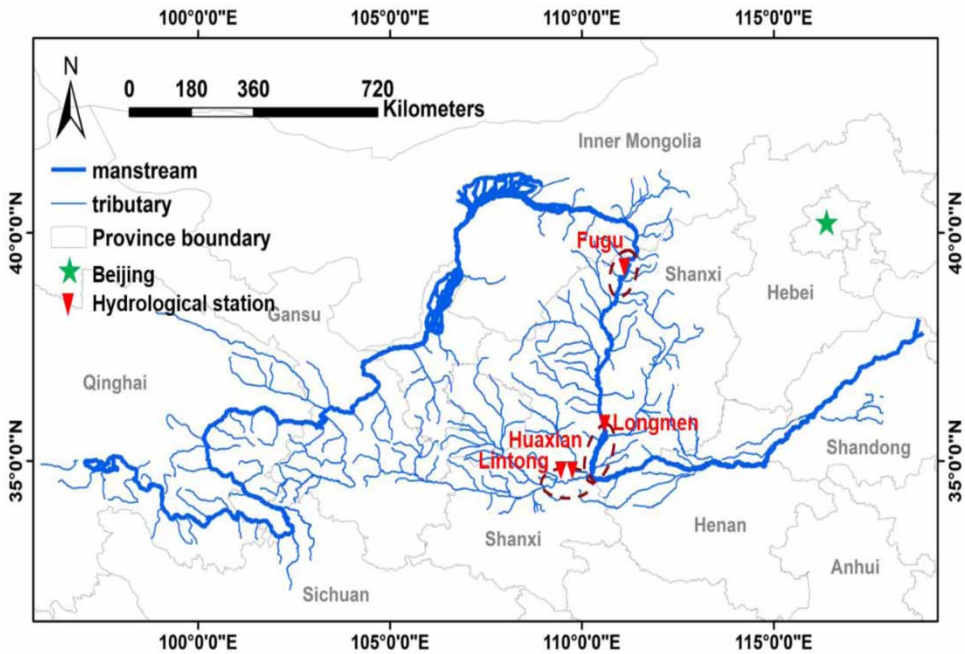


Figure 3 | Location map showing the study reaches.

Table 1 | Main data information statistics of the study reach during the RBTS floods

Study reach	Whole study period	Occurrence time	Q_m (m^3/s)	S_m (kg/m^3)	Occurrence of events	Absence of events	Total
Fugu	1986/03/26-1990/12/03	1988/06/12-1988/06/25	506	98.6	1	19	20
		1954/08/29-1954/09/09	10,800	500			
		1964/07/03-1964/07/09	5,260	433			
		1966/07/17-1966/07/24	3,450	667			
		1969/07/22-1966/08/06	3,210	668			
		1970/07/31-1970/08/15	4,670	702			
Longmen	1954/08/29-2017/08/02	1977/07/05-1977/07/13	6,450	485	11	126	137
		1977/08/01-1977/08/09	7,910	603			
		1993/07/04-1993/07/18	546	336.5			
		1995/07/14-1995/07/21	2,310	397.8			
		2002/07/04-2002/07/07	1,220	788			
		2017/07/24-2017/08/02	2,990	217			
Lintong	1964/05/24-1979/09/28	1964/07/16-1964/07/25	2,410	562	6	40	46
		1964/08/12-1964/08/17	1,990	613			
		1966/07/26-1966/08/01	3,260	589			
		1970/08/01-1970/08/09	2,250	555			
		1975/07/25-1975/07/28	1,350	553			
		1977/07/05-1977/07/10	4,120	609			
Huaxian	1964/05/24-1978/09/17	1964/07/16-1964/07/25	2,640	491	6	37	43
		1964/08/13-1964/08/17	2,550	598			
		1966/07/26-1966/08/01	4,660	573			
		1970/08/02-1970/08/09	2,150	423.8			
		1975/07/25-1975/07/29	962	531.3			
		1977/07/06-1977/07/10	3,610	656.5			

some of whole 246 floods in this study. Therefore, to investigate the influence of all indices, 179 floods containing all index factor data were analysed first, and there were 15 occurrences and 164 absence events of the RBTS phenomenon. A normalised box diagram is shown in Figure 4.

The dispersion of each characteristic index in the occurrence of the RBTS phenomenon was smaller than that in its absence. In addition to the high flow duration ratio λ and shape coefficient δ, the average of the other indices in the occurrence category was higher than that in the absence. This indicates that the occurrence of the RBTS phenomenon has the main feature of high-sediment flood, with a faster growth rate of flood, and higher sediment concentration.

In addition, regarding correlation degree, the correlation of the maximum sediment concentration S_m and the average sediment concentration S_p was higher, while the correlation of initial boundary factors and the median particle diameter D_m and average particle diameter D_p was lower. The results indicate that under the existing data conditions, initial boundary factors and particle size are not important influencing factors for the occurrence of the RBTS phenomenon, whereas sediment content is the key characteristic index.

Figure 5(a) shows the F-score and η^2 of all the characteristic influencing factors. Although the results and importance order of each index calculated by the two methods differed, overall, the F-score and η^2 of the maximum sediment concentration S_m and the average sediment concentration S_p in the sediment factors and the flood growth rate v, and shape coefficient δ in the hydrodynamics factors ranked in the top four among all the characteristic indices. Meanwhile, the F-score and η^2 value of these indices were above 0.469 and 0.065, respectively (according to the empirical results that belong to the above medium correlation), indicating that these four indices had a greater impact on the RBTS phenomenon and could easily distinguish whether RBTS would occur. In particular, the two indices with the highest F-score and η^2 value were the maximum sediment concentration S_m and the average sediment concentration S_p, which further indicated that sediment concentration was the key factor under the influence of the RBTS phenomenon and more sensitive to its occurrence. The calculation results also reflect the relationship between the turbulence intensity of the flow and the gravity action of the clay block in the riverbed, which is the physical mechanism behind the RBTS phenomenon. The starting scour of the clay block requires

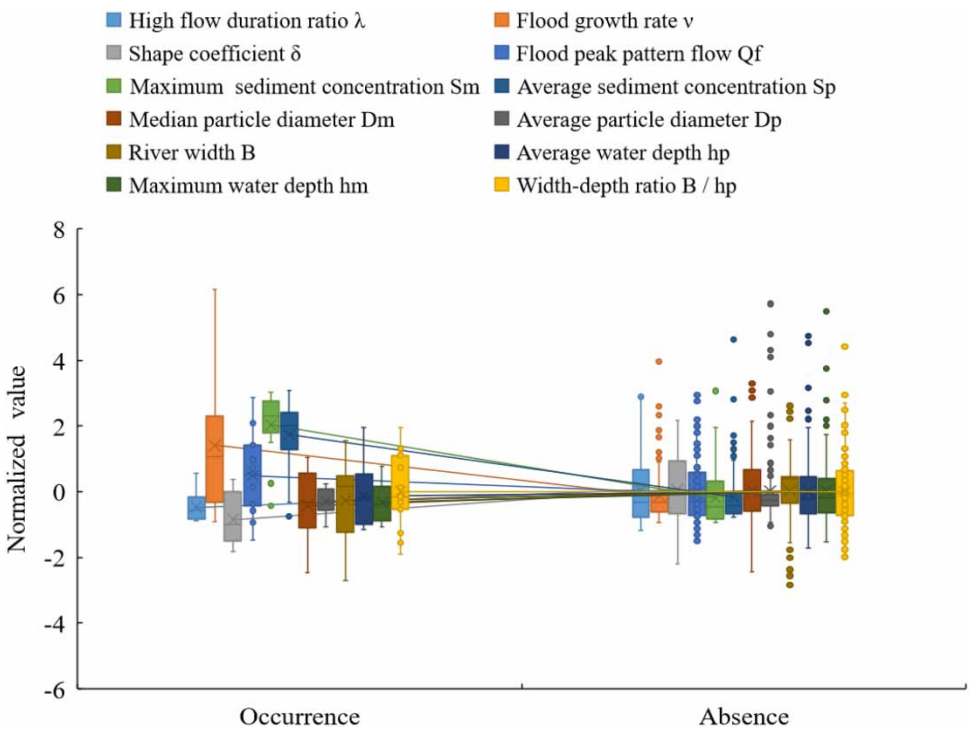


Figure 4 | Box chart distribution of each characteristic influencing factor index.

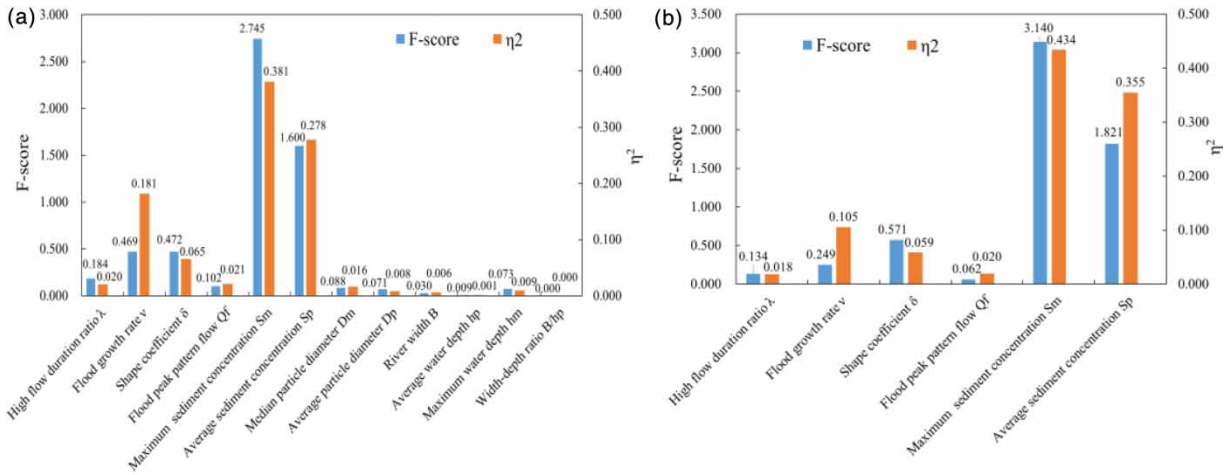


Figure 5 | *F*-score and η^2 of all indices (a) and the main indices (b) of characteristic influencing factors under all flood fields.

sufficient hydrodynamic conditions, including flow and its pulsation intensity and flow structure near the bottom. Its efficiency also meets the law of rise erosion and fall deposition; that is, in the process of a flood, erosion of the riverbed typically occurs during the rise stage, and conversely, siltation (Kothyari & Jain 2008; Li *et al.* 2017). This is consistent with the calculation results of the flood growth rate v, and shape coefficient δ. In addition, the increase in sediment inflow in the upper reaches improves the viscous force and sediment transport capacity of the flow and increases the collision effect of sediment particles on the silted bed surface, which further improves the erosion strength of viscous sediment under the condition of consolidation and density of the riverbed (Thompson & Amos 2004; Lichtman 2017). Therefore, high-sediment concentration is also an important condition for the occurrence of RBTS.

4.2. Strong correlation index extraction

According to the analysis results in Section 4.1, under the existing data conditions, the initial boundary factors and particle size indices had a low correlation with whether RBTS occurred, and their *F*-score values were almost an order of magnitude different from other indicators, which showed that they were not the key factors. Therefore, considering that the observed data samples of the occurrence of RBTS in the statistics were insufficient, the initial boundary factors and particle size indices were excluded to improve the generalisation of the subsequent prediction model, and all data of 246 floods including 24 occurrence and 222 absence events of the RBTS phenomenon were used for analysis. The new *F*-score and η^2 results corresponding to the different indices are shown in Figure 5(b).

After increasing the amount of data, the results of the *F*-score and η^2 values changed again. However, the maximum sediment concentration S_m, average sediment concentration S_p, flood growth rate v, and shape coefficient δ obtained from the analysis still ranked in the top four. Although the *F*-score of the flood growth rate v decreased, the η^2 of each index remained basically above 0.06 (still remained above the medium correlation), reflecting that these key index factors were minimally affected by the change in the amount of data and could still ensure a certain stability of strong correlation with the increase in the amount of data. In particular, the *F*-score and η^2 of the maximum sediment concentration S_m and average sediment concentration S_p increased with an increase in the amount of data, further indicating that sediment concentration had the most critical influence on the occurrence of the RBTS phenomenon.

4.3. Data processing and model training

According to the mechanism of the RBTS phenomenon, the starting of the clay blocks needs sufficient hydrodynamic and sediment conditions, in fact, the occurrence of RBTS is a competition between the water flow force and impact resistance of clay block. Therefore, it is appropriate and necessary to consider both hydrodynamic and sediment factor combining the analysis results of Sections 4.1 and 4.2. Comprehensively considering the calculation value of the *F*-score and η^2 method, [*F*-score] = 0.25 was taken as the threshold value to contain valid water and sediment information and reduce the number of input variables to the greatest extent. Four key indices were extracted as the input variables including the maximum sediment concentration S_m, average sediment concentration S_p, flood growth rate v, and shape coefficient δ, and the occurrence and absence of RBTS were used by the binary numbers of '1' and '0' to represent respectively the output variable. All observed data of 246 floods were used to build the prediction models of RBTS. It should be noted that although the sample number was only 246, it was a classification problem with only two results. Therefore, the accuracy of classification results can still be guaranteed when there is obvious distinction between the samples corresponding to the two results. Moreover, the extraction of strong correlation indices was specially carried out in this paper, which further improved the differentiation between the occurrence and absence of the RBTS, so the rationality and stability of the results could be considered acceptable.

Because there were few occurrence data samples, a segmentation proportion of 2:1 was adopted to ensure the generalisation of the model. Overall, all data were divided into training and test sets according to a proportion of 7:3. A total of 172 samples were used for training and learning, and the remaining 74 were used for prediction. The data composition is presented in Table 2.

4.4. Prediction results evaluation of different algorithm models

The data in Section 4.3 were substituted into the SVM, KNN, RF, and XGBoost classification algorithms successively to establish prediction models. The comparison results between the predicted and observed values of different algorithm models were obtained (Figure 6).

Table 2 | Information of the model data

Algorithm type	Input variable	Output variable	Training set	Test set
SVM RF KNN XGBoost	Flood growth rate v, shape coefficient δ, maximum sediment concentration S_m, average sediment concentration S_p	Occurrence, Absence	Occurrence: 16 Absence: 156 Total: 172	Occurrence: 8 Absence: 66 Total: 74

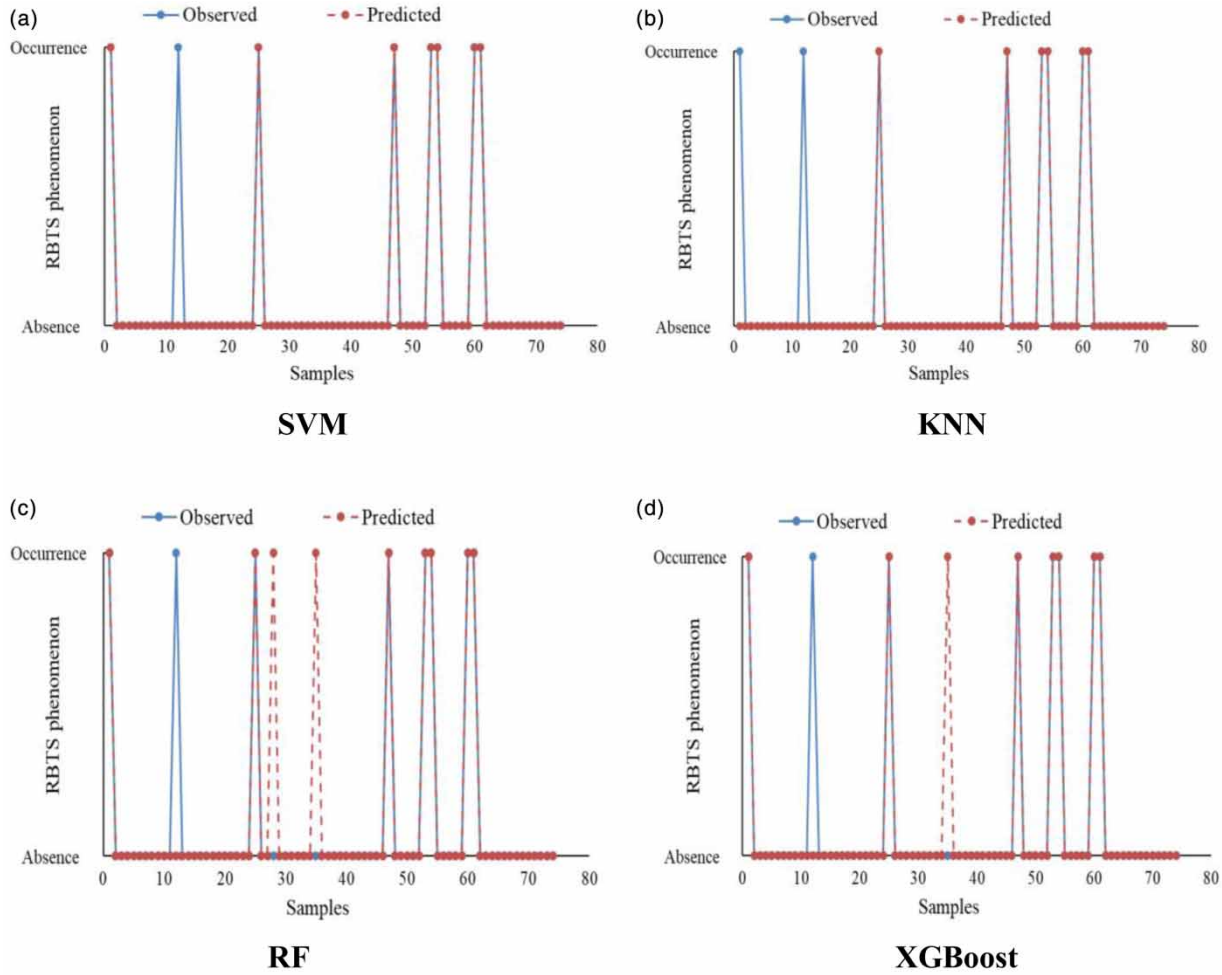


Figure 6 | Comparison between the predicted and observed results of different algorithm models.

Among the 74 test sample data, all different algorithm models could accurately predict most of the results regardless of whether RBTS occurred, indicating the effectiveness of applying machine learning algorithms to predict the RBTS phenomenon. The SVM algorithm successfully predicted seven occurrences and 66 absence events. Conversely, the prediction accuracy of the KNN algorithm model decreased in terms of occurrence events, and it only successfully predicted six of them but ensured the accurate prediction of all absence events. The prediction results of the RF and XGBoost algorithm models both successfully predicted seven occurrence events, whereas there were 2 and 1 absence cases incorrectly predicted as occurrences respectively, and the prediction accuracy of the absence of the phenomenon was relatively low. In addition, the four models all predicted incorrectly of the second occurrence of the RBTS phenomenon. According to the statistical results, the maximum and average sediment concentration indices S_m and $S_p$of the second occurrence were both lower than those of others (Figure 7). The correlation between the sediment concentration and the occurrence of RBTS was the highest among all indices, and the higher the sediment concentration, the easier the occurrence of RBTS. So it was considered that the lower sediment concentration might result in the inaccurate predictions of the models for RBTS.

The evaluation results of prediction performance are shown in Table 3. Among the final model evaluation indices, the SVM algorithm model had the highest values of each index, and its accuracy rate A, precision rate P, recall rate R, and $f1$ score under the occurrence events were 0.986, 1.000, 0.875, and 0.933, respectively. By contrast, among the other three algorithm models, the overall accuracy rate A, precision rate P, and $f1$ score of the RF algorithm model were lower than those of the KNN and XGBoost algorithm models, which were 0.959, 0.778, and 0.824, respectively. However, though the precision rate

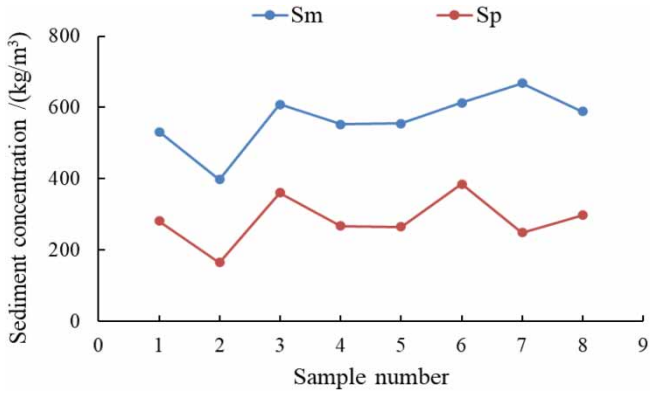


Figure 7 | The maximum and average sediment concentration of the eight tested cases of the RBTS.

Table 3 | Comparison of prediction performance among different algorithm models

		Evaluation index			
Algorithm type	**Predicted results**	**Accuracy rate *A***	**Precision rate *P***	**Recall rate *R***	***f*1 score**
SVM	Occurrence	0.986	1.000	0.875	0.933
	Absence		0.985	1.000	0.992
KNN	Occurrence	0.973	1.000	0.750	0.857
	Absence		0.971	1.000	0.985
RF	Occurrence	0.959	0.778	0.875	0.824
	Absence		0.985	0.970	0.977
XGBoost	Occurrence	0.973	0.875	0.875	0.875
	Absence		0.985	0.985	0.985

P index of the KNN algorithm model was higher than that of the XGBoost algorithm model, its recall rate *R* and *f*1 score were only 0.750 and 0.857, respectively.

To reflect the prediction performance of the different classification models more directly, the ROC curves of each model were further drawn (Figure 8). The larger the area of the ROC curve (*AUC* index), the more accurate the classification and the better the prediction performance of the model.

Among the prediction results of the different algorithm models, the ROC curve obtained by the SVM algorithm model had the largest area surrounded by the X-axis. The *AUC* values of the SVM, KNN, RF, and XGBoost models were 0.996, 0.922, 0.980, and 0.975, respectively. Therefore, the prediction performance of the model obtained using the SVM algorithm was the best with respect to prediction of the RBTS phenomenon.

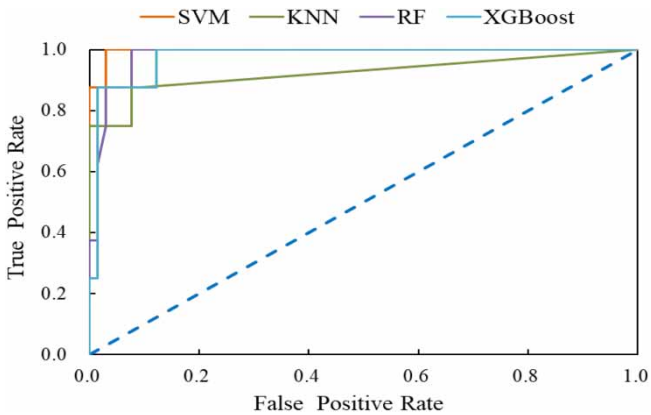


Figure 8 | ROC curves of different algorithm models.

5. CONCLUSION

RBTS is a special erosion phenomenon of high-sediment concentration floods in the Yellow River that has a considerable impact on river channel engineering. Owing to the small amount of observed data and relatively complex influencing factors, it is difficult to accurately predict whether it will occur. Therefore, the observed data of 246 floods in three different typical reaches of the Yellow River basin, including Fugu, Longmen, and Weihe, were used in this study, and the prediction models were built. The primary research results and conclusions of this study are summarised as follows:

(1) The total data of 179 floods with all the index factors were analysed, and the F-score and η^2 of the maximum sediment concentration S_m, average sediment concentration S_p, flood growth rate v, and shape coefficient δ ranked in the top four among all characteristic indices. The F-score and η^2 values of these indices were above 0.469 and 0.065, respectively, indicating that the four indices are the key factors to more easily distinguish whether RBTS will occur.

(2) After increasing the amount of data, although the new results of the F-score and η^2 changed, the top four indices remained the same, and each index was above the medium correlation, indicating that these key index factors could still ensure a certain stability of strong correlation with the change in the amount of data.

(3) Four key indices were extracted as the input variables, and the occurrence and absence of RBTS were the output variables. Prediction models were constructed based on the SVM, KNN, RF, and XGBoost classification algorithms. After comparison and analysis, the accuracy rate A, precision rate P, recall rate R, $f1$ score, and AUC of the SVM algorithm model under occurrence events were 0.986, 1.000, 0.875, 0.933, and 0.996, respectively, which were the highest among the four models and showed that the SVM algorithm had higher accuracy and precision in predicting the RBTS phenomenon.

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (42041004, U2243601, U2243241, U2243215), the Major Science and Technology Project of Ministry of Water Resources (SKR-2022021, SKS-2022088), the Science and Technology Development Fund of the Yellow River Institute of Hydraulic Research (HKF202111), and the Special Fund of Basic Scientific Research Business Expenses of Central Public Welfare Scientific Research Institutes (HKY-JBYW-2024-13).

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Aburomman, A. A. & Reaz, M. B. I. 2017 A novel weighted support vector machines multiclass classifier based on differential evolution for intrusion detection systems. *Information Sciences* **414**, 225–246. http://dx.doi.org/10.1016/j.ins.2017.06.007.

Ahmed, N., Mahmud, S., Elahi, M. M. L., Ahmed, S. & Sujauddin, M. 2019 Forecasting river sediment deposition through satellite image driven unsupervised machine learning techniques. *Remote Sensing Applications: Society and Environment* **13**, 435–444. https://doi.org/10.1016/j.rsase.2018.12.011.

Anand, A., Beg, M. & Kumar, N. 2021 Experimental studies and analysis on mobilization of the cohesionless sediments through alluvial channel: a review. *Civil Engineering Journal* **7** (5), 915–936. http://dx.doi.org/10.28991/cej-2021-03091700.

Bi, N. S., Sun, Z. Q., Wang, H. J., Wu, X., Fan, Y. Y., Xu, C. L. & Yang, Z. S. 2019 Response of channel scouring and deposition to the regulation of large reservoirs: A case study of the lower reaches of the Yellow River (Huanghe). *Journal of Hydrology* **568**, 972–984. https://doi.org/10.1016/j.jhydrol.2018.11.039.

Bosa, S., Petti, M. & Pascolo, S. 2018 Numerical modelling of cohesive bank migration. *Water* **10**, 961. doi:10.3390/w10070961.

Cao, Z. X., Pender, G. & Carling, P. 2006 Shallow water hydrodynamic models for hyperconcentrated sediment-laden floods over erodible bed. *Advances in Water Resources* **29**, 546–557. doi:10.1016/j.advwatres.2005.06.011.

Chen, J. J. 2020 *BIM-assisted UAV Safety Inspection and Automated Visual Recognition of Hazards for Water Diversion Channel*. Tianjin University, Tianjin.

Chen, Y. W. & Lin, C .J. 2006 Combining SVMs with various feature selection strategies. In: Feature Extraction. Studies in Fuzziness and Soft Computing (Guyon, I., Nikravesh, M., Gunn, S. & Zadeh, L. A., eds.). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-35488-8_13.

Chen, X. J., An, Y. Q., Zhang, Z. H. & Hu, C. H. 2021 Equilibrium relations for water and sediment transport in the Yellow River. *International Journal of Sediment Research* **36**, 328–334. https://doi.org/10.1016/j.ijsrc.2020.07.006.

Das, V. K., Roy, S., Barman, K., Debnath, K., Chaudhuri, S. & Mazumder, B. S. 2019 Investigations on undercutting processes of cohesive river bank. *Engineering Geology* **252**, 110–124. https://doi.org/10.1016/j.enggeo.2019.03.004.

Dong, Z. D., Ji, Z. W., Hu, H. H. & Wang, D. W. 2011 Experimental study on influence of discharge on transformation of river patterns. *Hydro-Science and Engineering* (4), 46–51. doi:10.16198/j.cnki.1009-640x.2011.04.011. (in Chinese).

Dong, W. S., Jiang, X. F., He, X. F. & Zai, Y. Y. 2012 Study on the clay mechanical characteristics of RBTS in the Yellow River. *Applied Mechanics and Materials* **212–213**, 108–112. doi:10.4028/www.scientific.net/AMM.212-213.108.

Fu, X. Y., Chen, Y. Y., Yan, J. R., Chen, Y. M. & Xu, F. 2023 BGRF: A broad granular random forest algorithm. *Journal of Intelligent and Fuzzy Systems* **44**, 8103–8117. doi:10.3233/JIFS-223960.

Giglou, A. N., Sahebari, A. J., Shakibaeinia, A. & Borghei, S. M. 2016 An experimental study of sediment transport in channel confluences. *International Journal of Sediment Research* **31**, 87–96. http://dx.doi.org/10.1016/j.ijsrc.2014.08.001.

Gou, Y. Y. 2004 Calculation and analysis of the bottom tearing scouring phenomenon in flood with hyperconcentration in the Yellow River. *Advances in Water Science* **15** (2), 156–159. doi:10.14042/j.cnki.32.1309.2004.02.005. (in Chinese).

Gu, Z. Y., Cao, M. C., Wang, C. G., Yu, N. & Qing, H. Y. 2022 Research on mining maximum subsidence prediction based on genetic algorithm combined with XGBoost model. *Sustainability* **14**, 10421. https://doi.org/10.3390/su141610421.

He, L. & Yan, M. 2019 Coupling analysis of connectivity between coarse sediment source areas and deposition area in the Yellow River, China. *Polish Journal of Environmental Studies* **28** (2), 503–513. doi:10.15244/pjoes/83615.

Hu, J. L., Miao, C. Y., Zhang, X. P. & Kong, D. X. 2023 Retrieval of suspended sediment concentrations using remote sensing and machine learning methods: A case study of the lower Yellow River. *Journal of Hydrology* **627**, 130369. https://doi.org/10.1016/j.jhydrol.2023.130369.

Huang, Q. H., Wang, C., Ye, Y., Wang, L. & Xie, N. G. 2023 Recognition of EEG based on Improved Black Widow Algorithm optimized SVM. *Biomedical Signal Processing and Control* **81**, 104454. https://doi.org/10.1016/j.bspc.2022.104454.

Jiang, E. H., Li, J. H., Zhao, L. J. & Cao, Y. T. 2010 Research and verification on discriminating index for bottom tearing scour. *Journal of Hydraulic Engineering* **41** (6), 727–731. doi:10.13243/j.cnki.slxb.2010.06.011. (in Chinese).

Jiang, E. H., Cao, Y. T., Zhang, Q., Li, J. H., Yuan, M. J. & Liu, Y. L. 2015 Pattern of channel adjustment due to bottom-block scour on the Yellow River. *Advances in Water Science* **26** (4), 509–516. doi:10.14042/j.cnki.32.1309.2015.04.007. (in Chinese).

Kim, K. 2021 Normalized class coherence change-based kNN for classification of imbalanced data. *Pattern Recognition* **120**, 108126. https://doi.org/10.1016/j.patcog.2021.108126.

Kothyari, U. C. & Jain, R. K. 2008 Influence of cohesion on the incipient motion condition of sediment mixtures. *Water Resources Research* **44**, W04410. https://doi.org/10.1029/2007WR006326.

Kuang, S. F., Xu, Y. N. & Liang, Z. Y. 2000 An investigation of ripping up the riverbed. *Journal of Hydrodynamics* **1**, 27–34.

Li, J., Xia, J. Q., Zhou, M. R. & Deng, S. S. 2017 Variation in reach-scale thalweg-migration intensity in a braided reach of the lower Yellow River in 1986–2015. *Earth Surface Processes and Landforms* **42** (13), 1952–1962. https://doi.org/10.1002/esp.4154.

Li, D., Wang, G., Qin, C. & Wu, B. S. 2021 River extraction under bankfull discharge conditions based on Sentinel-2 Imagery and DEM Data. *Remote Sensing* **13**, 2650. https://doi.org/10.3390/rs13142650.

Lichtman, I. D. 2017 *Combined effects of hydrodynamics and cohesive clay on bedform morphology and migration on sandy tidal flats*. Bangor University, United Kingdom.

Liu, J. 2018 Study on the calculation method of shape coefficient generalization of designed flood process in a small watershed without data in eastern Liaoning Province. *Water Resources Planning and Design* **5**, 76–79. doi:10.3969/j.issn.1672-2469.2018.05.024. (in Chinese).

Liu, Y. H., Wang, Y. J. & Jiang, E. H. 2021 Stability index for the planview morphology of alluvial rivers and a case study of the lower Yellow River. *Geomorphology* **389**, 107853. https://doi.org/10.1016/j.geomorph.2021.107853.

Mantas, C. J., Castellano, J. G., García, S. M. & Abellán, J. 2019 A comparison of random forest based algorithms: random credal random forest versus oblique random forest. *Soft Computing* **23**, 10739–10754. https://doi.org/10.1007/s00500-018-3628-5.

Medel, I. D., Stubblefield, A. P. & Shea, C. 2022 Sedimentation and erosion patterns within anabranching channels in a lowland river restoration project. *International Journal of River Basin Management* **20** (3), 399–409. https://doi.org/10.1080/15715124.2020.1809435.

Mechanical analysis on ripping up bottom scour in the Yellow River. In: Proceedings of the 35th IAHR World Congress (Wang, Z. Y., Li, H. W., Gao, J. Z. & Cao, S. Y., eds.). Tsinghua University Press, Beijing, China, pp. 3839–3847.

Pan, S. W., Zheng, Z. C., Guo, Z. & Luo, H. N. 2022 An optimized XGBoost method for predicting reservoir porosity using petrophysical logs. *Journal of Petroleum Science and Engineering* **208**, 109520. https://doi.org/10.1016/j.petrol.2021.109520.

Ren, H. Y., Hou, Z. S., Duan, Z. R., Song, X. H., Perkins, W. A., Richmond, M. C., Arntzen, E. V. & Scheibe, T. D. 2020 Spatial mapping of riverbed grain-size distribution using machine learning. *Frontiers in Water* **2**, 551627. doi:10.3389/frwa.2020.551627.

Slaa, S. T., He, Q., Maren, D. S. V. & Winterwerp, J. C. 2013 Sedimentation processes in silt-rich sediment systems. *Ocean Dynamics* **63**, 399–421. doi:10.1007/s10236-013-0600-x.

Tabarestani, M. K. & Zarrati, A. R. 2015 Sediment transport during flood event: A review. *International Journal of Environmental Science and Technology* **12**, 775–788. doi:10.1007/s13762-014-0689-6.

Thompson, C. E. L. & Amos, C. L. 2004 Effect of sand movement on a cohesive substrate. *Journal of Hydraulic Engineering* **130** (11), 1123–1125. https://doi.org/10.1061/(ASCE)0733-9429(2004)130:11(1123).

Van Maren, D. S., Winterwerp, J. C., Wang, Z. Y. & Pu, Q. 2009a Suspended sediment dynamics and morphodynamics in the Yellow River, China. *Sedimentology* **56**, 785–806. doi:10.1111/j.1365-3091.2008.00997.x.

Van Maren, D. S., Winterwerp, J. C., Wu, B. S. & Zhou, J. J. 2009b Modelling hyperconcentrated flow in the Yellow River. *Earth Surface Processes and Landforms* **34**, 596–612. doi:10.1002/esp.1760.

Wang, Z. Y., Qi, P. & Melching, C. S. 2009 Fluvial hydraulics of hyperconcentrated floods in Chinese rivers. *Earth Surface Processes and Landforms* **34**, 981–993. doi:10.1002/esp.1789.

Wang, Y. D., Pan, Z. B. & Pan, Y. W. 2020 A training data Set cleaning method by classification ability ranking for the k-Nearest neighbor classifier. *IEEE Transactions on Neural Networks and Learning Systems* **31** (5), 1544–1556. doi:10.1109/TNNLS.2019.2920864.

Wang, S. Z., Zhang, Z. F., Geng, S. S. & Pang, C. Y. 2022 Research on optimization of random forest algorithm based on spark. *Computers, Materials & Continua* **71** (2), 3721–3731. doi:10.32604/cmc.2022.015378.

Xia, J. Q., Cheng, Y. F. & Zhou, M. R. 2023 River regime evolution and simulation technique in wandering section of the lower Yellow River. *China Water Resources* **20**, 20–25. (in Chinese).

Xu, J., Zhang, C. X., Wang, L. L., Zhu, H., Tang, H. W. & Avital, E. J. 2022 Variation of dominant discharge along the riverbed based on numerical and deep-learning models: A case study in the middle Huaihe River, China. *Journal of Hydrology* **612**, 128225. https://doi.org/10.1016/j.jhydrol.2022.128285.

Yan, C. D., Li, Z. W., Boota, M. W., Zohaib, M., Liu, X., Shi, C. L. & Xu, J. K. 2023 River pattern discriminant method based on Rough Set theory. *Journal of Hydrology: Regional Studies* **45**, 011285. https://doi.org/10.1016/j.ejrh.2022.101285.

Yang, X. W., Yu, Q. Z., He, L. F. & Guo, T. J. 2013 The one-against-all partition based binary tree support vector machine algorithms for multi-class classification. *Neurocomputing* **113**, 1–7. http://dx.doi.org/10.1016/j.neucom.2012.12.048.

Zhang, Y. F. & Hu, T. J. 2013 Mechanical analysis on ripping up bottom scour in the Yellow River. In: *Proceedings of the 35th IAHR World Congress* (Wang, Z. Y., Li, H. W., Gao, J. Z. & Cao, S. Y., eds.). Tsinghua University Press, Beijing, China, pp. 3839-3847.

Zhang, F., Zhang, Y. Y., Chen, J. X., Zhai, X. Y. & Hu, Q. F. 2022a Performance of multiple machine learning model simulation of process characteristic indicators of different flood types. *Progress in Geography* **41** (7), 1239–1250. doi:10.18306/dlkxjz.2022.07.008. (in Chinese).

Zhang, J. L., Lu, J., Gao, X., Zhu, C. H. & Liu, J. X. 2022b Study on mechanism of 'River bottom tearing'of hyper-Concentrated floods in the Yellow River. *Yellow River* **44** (10), 1–5. (in Chinese).

First received 25 October 2023; accepted in revised form 5 February 2024. Available online 29 February 2024