Data mining technology for the identification and threshold of governing factors of landslide in the Three Gorges Reservoir area

Liangjie Guo¹ · Fasheng Miao¹,² · Fancheng Zhao¹ · Yiping Wu¹

Accepted: 28 April 2022 / Published online: 2 June 2022 © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Due to the complex geological conditions and external triggering factors, the deformation of landslide disaster often has spatial differences, especially in the Three Gorges Reservoir area. It is of great significance to explore the governing factors and their thresholds in different parts of landslide. In this research, the deformation of Shuping landslide is analyzed. First, 9 hydrological factors are selected for data mining analysis by comprehensive research, including 5 reservoir water factors and 4 rainfall factors. Then, the numerical variables are transformed into discrete variables by two-step clustering method, and the Apriori algorithm is utilized to deal with the classified variables to generate the correlation criterion meeting the minimum confidence, and the correlation criterions between triggering factors and landslide displacement are established. Finally, the threshold of governing factors are mined out by decision tree C5.0 models. The results indicate that governing factors controlling the deformation of different parts of landslide are distinct. Specifically, the rear landslide is jointly controlled by the reservoir water and rainfall. On the contrary, the reservoir level controls the deformation of other parts of Shuping landslide. Generally, the daily drop of water level is the most important factor causing the deformation of Shuping landslide. During the period of low water level (138.951 ~ 147.437 m), once the daily drop of water level exceeded 0.416 m/d, the landslide will show severe deformation. This research reveals that the study of association criterion and threshold is of great significance for landslide deformation analysis. Data mining technology can be better applied to the prediction of the reservoir landslide.

Keywords Data mining technology · Shuping landslide · Governing factors · The Three Gorges Reservoir · Reservoir level · Rainfall

1 Introduction

Geological disasters emerge in endlessly with the increasing frequency of extreme weather (Juang et al. 2019; Gong et al. 2020). As one of the most widely distributed geological disasters, landslides have the characteristics of great destruction, which are widely developed in mountainous and reservoirs areas (Huang et al. 2017; Nicu and Asandulesei et al. 2018). The Three Gorges Reservoir area has imperative economic and social benefits (Zhou et al. 2016; Tang et al. 2019). Since the impoundment of water level increased to 175 m, the Three Gorges Reservoir area, which is surrounded by mountains and water, has become a highly prone area of geological disasters (Miao et al. 2018a; Jiang et al. 2021; Ma et al. 2022). The unique and complex geological conditions and ecological environment have given birth to the occurrence and resurrection of a
myriad of riverine landslides, and these extremely devastating landslides will cause huge loss of life and property and even adverse social impacts (Peng et al. 2018; Zhang et al. 2020a; Wang et al. 2022).

The causes of landslides are miscellaneous and varying from disaster triggering factors (Garano and Guzzetti 2016; Milevski et al. 2019). Moreover, the combination of internal and external factors such as tectonics, rainfall, water fluctuation, and human activities will promote the appearance of landslides to varying degrees (Moreiras 2005; Petley et al. 2007; Sassa et al. 2010). However, determined by its own function, the frequent rise and fall of the reservoir water inevitably weakens the structure of the rock and soil (Miao et al. 2018b; Zhang et al. 2020b). Moreover, coupled with the scouring of the increasingly common large-scale rainstorms, rainfall infiltration will inevitably reduce the stability of the reservoir bank slopes (Gong et al. 2021; Sun et al. 2021). Therefore, it is reasonable to conclude that rainfall and fluctuation of reservoir water level are the key factors affecting the stability of reservoir landslide. (Ma et al. 2017a, 2020; Miao et al. 2021, 2022).

Due to the development of monitoring and statistical science, the theory of data analysis has been applied to many fields of landslide research, such as recognition (Shi et al. 2020; Gong et al. 2022), prediction (Du et al. 2013; Chen et al. 2017; Zhang et al. 2020a, b, c), modeling (Stumpf and Kerle 2011; Vorpahl et al. 2012), susceptibility (Hong et al. 2016; Chen et al. 2018; Zhao and Chen 2019). Compared with the traditional data analysis process, data mining has attracted great attention for its ability in exploring the inherent correlation criteria of data (Ma et al. 2017b, 2018). For example, Tan et al. (2017) used Grey Relational Grade Analysis (GRGA) to determine the main control factors of Zhujiadian landslide at different deformation stages. Yao et al. (2019) used the neighborhood rough set theory to identify the triggering factors leading to the deformation of the Baijiabao landslide, which showed that the surface deformation and the expansion of ground fissures with this stepped displacement pattern were mainly controlled by rainfall and reservoir water level changes. Wu et al. (2021) combined the analysis of variance (ANOVA) and K-means clustering method to jointly analyze the monitoring data of reservoir water fluctuation and rainfall intensity, and revealed the mechanism of rapid deformation of Gapa landslide. Kusak et al. (2021) created a list of landslides that occurred in Karahacili at the end of 2019 and used Apriori algorithm and spatial data mining methods to assess the pre-landslide conditions in the area. Althuwaynee et al. (2021) used the Apriori algorithm based on association rule mining to explore the antecedents of landslides through spatial clustering mode. Summarily, the improvement of monitoring technology has also brought about the generation of massive landslide monitoring data, and data mining has been used in the analysis of landslide triggering factors extensively. Nevertheless, considering the spatial variability of triggering factors in different parts of landslide, there is still a lack of reports on how to accurately identify and distinguish the cause of landslides.

In this study, the deformation of Shuping landslide is analyzed. 9 hydrological factors were selected for data mining analysis by comprehensive research (Fig. 1). Data mining was utilized to deal with the classified variables to generate the correlation criterion meeting the minimum confidence. Each monitoring point selects 10 association rules with 100% confidence level for analysis. And decision tree C5.0 models were established to analyze the threshold of governing factors. The evidence showed that together with its high accuracy in monitoring data research, such a data mining method was expected to be widely applied in data analysis and prediction of accumulation landslides in the reservoir area.

2 Method

2.1 Apriori algorithm

Based on the transformation of the landslide monitoring data types by the two-step clustering algorithm (Fig. 2), the variables obtained from the classification are calculated using the Apriori algorithm to mine the association criteria of the monitoring data. The Apriori algorithm, proposed by Agrawal and Srikant (Agrawal et al. 1993), mainly consists of two steps: (1) Generate the set of frequent items above the minimum support; (2) Generate the association criterion above the minimum confidence within it (Guo et al. 2019). The sketch map of Apriori algorithm is displayed in Fig. 3. The implementation process is to search and generate a frequent itemset \( L_1 \) of length 1, which in turn is used to produce a frequent itemset \( L_2 \) of length 2, and iterate until no higher-order frequent itemset can be obtained.

After generating the simple association criterion from frequent item sets, the set of valid rules is selected according to whether the confidence level is greater than the confidence threshold. Specifically, for each frequent itemset \( L \), calculate the confidence level for each of its non-empty subsets, and if \( C_{L' \rightarrow \left(L-L'\right)} \) is greater than a user-specified confidence threshold:

\[
C_{L' \rightarrow \left(L-L'\right)} = \frac{|T(L)|}{|T(L)|} \geq \text{minconf}
\]

Then the association criterion is generated: \( L' \rightarrow \left(L-L'\right) \).
2.2 Decision tree C5.0

The decision tree, as a prevalent classification algorithm and a decision support method, follows a set of rules and an established tree diagram. In terms of classification and prediction, the nodes of the tree are sample attributes chosen for splitting while the branches are attribute values (Pandya and Pandya 2015). Although there are many other decision tree algorithms, C5.0 is a commonly used one. Videlicet, C5.0 obtains a decision tree model by analyzing and summarizing the properties of a large number of samples and combining them with the principles of information theory (Zhang et al. 2017).

For a given node $n$, it is assumed that $N$ is the entire set of samples, $C$ and $t$ are the set of target variables and the number of classifications.

$$\text{Ent}(N) = -\sum_{i} p(C_i|N) \log_2 p(C_i|N)$$

Here, $p(C_i|N)$ is the relative probability of $C_i$ ($i = 1, 2, 3, \ldots, t$). And then:

$$\text{Ent}(N|T) = \sum_{j} \frac{|T_j|}{|N|} \times \text{Ent}(N)$$

The information gain can be expressed as:

$$\text{Gains}(N, T) = \text{Ent}(N) - \text{Ent}(N|T)$$

The growth of the decision tree can be defined as follows:

$$\text{GainRatio} = \frac{\text{Gains}(N, T)}{\text{Ent}(T)}$$

In addition, determining the optimal splitting point is also essential in decision tree growth. The core idea of the solution has the following steps: (1) arrange the input variables in ascending order; (2) define a number of initial intervals so that each variable value of the input variables falls in a separate interval; (3) calculate the cross-group frequency table of the input variables and the output variables; (4) calculate the chi-square value of each two adjacent groups; (5) obtain the chi-square threshold based on the significance level and degrees of freedom, and merge the interval with the smallest cardinality value; (6) repeat recursive steps (2) ~ (5) until any pair of adjacent groups cannot be merged.

Boosting techniques can improve the robustness of the decision tree C5.0 algorithm, which consists of two phases: modelling and model voting.
In the modelling phase, the boosting technique increases the set of simulated samples by repeatedly sampling the existing weighted samples. If the whole process is iterated \( k \) times and the size of the training sample is \( N \), then the modelling process can be expressed as:

1. **Initialized sample weights:**
   \[
   w_j(i) = \frac{1}{N} (i = 1, 2, \ldots, k; j = 1, 2, \ldots, N)
   \]
   where \( w_j(i) \) denotes the weight of \( j \)th samples in the \( i \)th iteration.

2. **According to** \( w_j(i) \), the training sample set \( T_i \) is formed by taking \( n \) samples from \( T \) with replacement;

3. **The model** \( C_i \) can be obtained from \( T_i \) and calculate the error \( e(i) \) of the model;

4. **End the modelling process when** \( e(i) \) \( > 0.5 \) or \( e(i) = 0 \), otherwise the weight of each sample are updated according to the error, recorded as:
   \[
   w_j(i + 1) = w_j(i) \times \beta(i)
   \]
   \[
   \beta(i) = \frac{e_i}{1 - e_i}
   \]
   Among them, sample weight for misclassification remains the same:
   \[
   w_j(i + 1) = w_j(i)
   \]
   Then, normalized:
   \[
   w_j(i + 1) = \frac{w_j(i + 1)}{\sum_{j=1}^{N} w_j(i + 1)}
   \]

5. **Iterative steps** (2) \( \sim \) (4) to obtain \( k \) models and \( k \) errors.

In the voting phase, the voting process for a new sample set \( X \) can be summarized as:

1. **Determine a prediction value** \( C_i(X) \) and a weighting value for each model \( C_i \) \( (i = 1, 2, 3, \ldots, k) \):
   \[
   W_j(X) = -\log \left( \frac{e_i}{1 - e_i} \right)
   \]
   Consequently, \( k \) models will give a number \( k \) of \( C_i(X) \) and \( W_i(X) \).

2. **Calculate the sum of the weights individually according to the categories, and the category with the largest sum is the final classification result of the set** \( X \). Moreover, by combining cross-validation methods with boosting techniques, the generalization ability of the model can be improved to prevent overfitting of the model.

### 3 Research case

#### 3.1 Geological characteristics of Shuping landslide

Shuping Landslide is generally distributed in the north-south direction, about 47 km away from the Three Gorges Dam, which is located on the right bank of the Yangtze River, and the toe of the landslide is directly to the Yangtze River. The sliding zone is steep in the upper part and gentle in the lower part, which is high in the south and low in the north. Two gentle slope platforms are distributed at the elevations of 170 ~ 200 m and 310 ~ 340 m respectively. Shuping landslide is divided into East and West parts, and the eastern part is called the active area, as shown in Fig. 4. The elevation of the shear outlet is 80 m and the area is about \( 55 \times 10^4 \) m\(^2\), with a total volume of \( 2.89 \times 10^7 \) m\(^3\).

Shuping landslide is mainly composed of sliding mass, sliding zone and sliding bed. The sliding mass is mainly silty clay and broken stone soil. The buried depth of the slip zone is 30 ~ 70 m (average depth 50 m), which is the contact zone between the accumulation layer and the bedrock, and the constituent material is mainly silty clay. The sliding bed is purple red and grayish green medium thick bedded siltstone mixed with mudstone, light gray and grayish yellow medium thick bedded limestone and marl of Badong Formation (T\(_b\)) of Middle Triassic system. The attitude of rock stratum is 170°, and the dip angle is 12 ~ 15°, as shown in Fig. 5.

#### 3.2 Deformation of the landslide

Shuping landslide is a historical landslide accumulation body, which is located in the folded mountain area of Western Hubei. This area has the characteristics of high rainfall intensity, long duration and high rainstorm frequency. Moreover, the terrain of this area is relatively low-lying. Therefore, under the influence of rainfall, the local cracks will accelerate the expansion. With the progress of rainfall, multiple structural planes will form a through sliding surface, and rainfall weakens the cohesion, internal friction angle. Its comprehensive factors promote the continuous deformation of the landslide. Before 1996, there were many arc-shaped tensile cracks in the toe of Shuping landslide, resulting in the forced relocation of more than 60 residents, and the damage was mainly caused by local deformation. Since the completion of the Three Gorges Dam and water storage in June 2003, the front edge of Shuping landslide has been affected by reservoir water and rainfall, and the deformation was mainly local surface collapse. From 2004 to 2007, there were many shear
cracks, tension cracks and intermittent pinnate cracks on the east and west sides of the landslide, which caused cracks in the walls of many buildings, forcing 85 residents of 25 households to move. In August 2008, Shuping landslide was affected by heavy rainfall. The cracks on the eastern boundary of the trailing edge of the landslide continued to develop, and many settlement cracks and small-scale collapse began to appear. From May to July after 2009, the main sliding area of Shuping landslide was formed, and small-scale collapse and loosening occurred around the main sliding area. The original cracks at the landslide boundary continued to extend and expand under the action of external force, and many places were connected (Fig. 6). At present, the landslide is still in the creep deformation stage, and severe deformation or sliding is more likely to occur under extreme conditions.

3.3 Analysis of the monitoring data

A series of devices based on GPS for displacement monitoring had been installed in various parts of the Shuping landslide since 2003. Specially, located in the active area, a total of 8 monitoring devices were arranged, which can be regarded as a very meaningful deployment that long-termly and closely reflected the state of displacement of the Shuping landslide instantaneously. Correspondingly, the
Monitoring data can be selected to represent the monitoring status of the entire Shuping landslide.

Since the reservoir was impounded in 2003, the water level maintained a stable fluctuation between 135 and 140 m. Then in 2007, the height of the reservoir level reached 155 m for the first time and fluctuated between 145 and 155 m periodically. Furthermore, after 2008, the reservoir capacity increased again, with the highest water level reaching about 175 m. Up to now, reservoir water has basically been maintained between 145 and 175 m each year. As portrayed in Fig. 7, it is reasonable to divide the deformation of the Shuping landslide into three phases based on the fluctuation of the Three Gorges Reservoir water, which has a certain hysteresis.

(1) Prophase (until June 2006): The water level of the reservoir was the lowest fluctuated between 135 and 140 m. Correspondingly, the displacement of each monitoring point began to increase slowly. The flood season and the fluctuation of the reservoir water level were consistent with the increment of displacement. However, the emergence of extremely high intensity precipitation (July 2005) did not induce the landslide displacement to increase sharply. Especially for ZG89, ZG90, the displacement of which were less affected. Moreover, SP-6 did not even show any displacement changes.

(2) Metaphase (until June 2008): During this phase, the landslide deformation began to increase steadily in a “step-like” manner. Specifically, the fluctuation of the reservoir level weakened the rock and soil structure inside the landslide, sparking off the displacement to rise rapidly with the sharp drop of

---

**Fig. 6** Deformation photos of the Shuping landslide at different positions

**Fig. 7** Monitoring data of Shuping landslide
reservoir level. Specially, the displacement of monitoring points such as ZG86, ZG88, and SP-2 had increased by more than 1000 mm.

(3) Anaphase (until June 2014): The highest water level of the reservoir rose again during this period, reaching a maximum of 175 m. In addition, the periodic rise and fall of the reservoir water led to the landslide displacement to increase in a “step-like” manner continuously. The data of all monitoring points during this period connoted that the displacements were still increasing at a felicitously steady rate. Obviously, the displacements of ZG85, ZG86, ZG88, SP-2, SP-6 were significantly greater than other monitoring points. In this paper, the data of the monitoring points ZG85, ZG86, ZG88, SP-2, SP-6 are selected as the research object, where monitoring points ZG85 and ZG88 were located at the toe landslide, SP-2 was located in the middle of the landslide, and ZG86 and SP-6 were located at the rear landslide.

3.4 Physical mechanism of Shuping landslide deformation

Shuping landslide is an old landslide with a long history. Its unique landform, formation lithology and geological structure conditions are the internal causes of the formation of the landslide and control the formation and development of the landslide. The fluctuation of the reservoir level induces the revival of the landslide, and the rainfall accelerates the deformation of the landslide. These factors play an important incentive role in the deformation of the landslide. Therefore, the physical mechanism of Shuping landslide deformation can be analyzed from two aspects of internal and external factors.

(1) Internal factors

The occurrence of landslides mainly depends on the control of slidable strata, such as layered rock formations containing weak intercalated surfaces or layers (zones). The Badong Formation (T_{2b}) stratum is a typical slippery stratum spread all over the Three Gorges Reservoir area. Furthermore, the Shuping landslide develops in the Middle Triassic strata, constituted by a series of purple-red, gray-green medium-thick layered siltstone intercalated with mudstone and gray, light gray medium-thick layered mudstone mixed shale. The rock mass of Shuping landslide distributes between soft and hard phases where sandwiched with weak and weathering mudstone layers. Due to the high content of hydrophilic minerals, it is easily softened in contact with water and can be classified as slippery rocks. In short, the weak strata in the Shuping landslide area lays a material foundation for the deformation and failure of the slope. Specifically, the effect of gravity contributes to the creep deformation of the weak base at the lower part of the slope, thus pulling the upper slope to produce tensile cracking deformation, which gradually develops into large-scale deformation. In addition, affected by the regional structure, the fractures in the rock mass of Shuping landslide area are highly developed. The fissures tending to the NNE and NNW directions are the most extended, whose orientation is almost the same as the slope aspect, providing a cutting plane for the formation of the landslide. However, the gently sloping fracture surface outside the dip slope is favorable for the formation of the bottom slip surface. Briefly, these internal geological factors jointly control the formation and development of the Shuping landslide.

(2) External factors

During the period of reservoir level rise up, the slope body is back pressured, which is conducive to the stability of the landslide. Conversely, the drawdown of the reservoir causes the groundwater to drain slowly due to the poor permeability of the slope, so that the decline of the groundwater level lags behind the decline of the reservoir water level, forming a positive drop between the groundwater and the reservoir water level. And the seepage force points to the outside of the slope, which is not conducive to the stability of the landslide. Meanwhile, most of the slopes remain in a state of water-saturated with a relatively large severity, because the decline of the groundwater level in the slopes is relatively lagging behind. Therefore, the stability of the slope decreases sharply and shows a hysteresis effect. In particular, a large water level falling speed will induce a severe landslide deformation rate, and the deformation characteristics of the weakly permeable hysteresis landslide shall appear. Thus, the decline of the reservoir level can be regarded as the main factor leading to the deformation of the Shuping landslide, which is a typical reservoir descending landslide. Besides, rainfall also has a great influence on the deformation of the landslide. The infiltration of rainfall into the slope increases the self-weight of the slope and forms the pore seepage pressure, which softens the properties of the slip zone soil and reduces the mechanical parameters. By comparison, it can be seen that there is an obvious correlation between the deformation of the landslide and the seasonal rainfall. Moreover, the landslide deformation curve has a step every year, and it happens to be in the annual rainfall concentration...
period. It shows that atmospheric rainfall is also one of the main factors for the deformation of the landslide.

It can be seen that the drop of reservoir water level and atmospheric rainfall have stimulated the deformation of Shuping landslide. When these two factors are superimposed on each other, the deformation of the landslide is more likely to occur. Especially, owing to the double influence of the lowering of the water level of the Three Gorges Reservoir from 2007 to 2013 and the rainfall in the flood season after April each year, the landslide deformation increased sharply. It can be inferred that the decline of the reservoir water level reduces the stability of the landslide and intensifies the deformation. Nonetheless, if there is an obvious rainfall process during the decline of the reservoir water level, it will have a more adverse effect on the stability of the landslide and aggravate the deformation of the landslide.

The evidence suggests that the fluctuation of the reservoir water level and rainfall are the main factors affecting the deformation of the landslide. Therefore, we select a total of 9 hydrological factors from two major categories for data mining analysis. Among them, the hydrological factors related to the reservoir water level include: Monthly average water level \((h)\) (m), monthly maximum daily drop of water level \((\Delta h^{-}\text{d})\) (m/day), monthly maximum daily rise of water level \((\Delta h^{+}\text{m})\) (m/day), monthly variation of water level \((\Delta h_{1})\) (m/month), and bi-monthly variation of water level \((\Delta h_{2})\) (m/month). Additionally, the hydrological factors related to rainfall include: Monthly maximum continuous rainfall \((q_{\text{c}})\) (mm), monthly cumulative rainfall \((q_{1})\) (mm), bi-monthly cumulative rainfall \((q_{2})\) (mm), and monthly maximum daily rainfall \((q_{\text{m}})\) (mm), as shown in Table 1. It should be noted that the monthly fluctuation of reservoir water level is the height difference between the highest and the lowest reservoir water level in each month. Monthly maximum effective continuous rainfall refers to the cumulative rainfall of a continuous rainfall process.

### 4 Results

#### 4.1 Clustering results

9 hydrological factors were clustered by two-step algorithm, as shown in Table 2. Specifically, the categories were set to 2 \(\sim\) 10 in the clustering process. And the Euclidean distance was employed to two-step clustering algorithm for distance measurement. In addition, the clustering criterion was subjected to Bayesian Information Criterion (BIC).

The monthly velocity was clustered into two groups: V1 and V2 (Table 3). In addition, Fig. 8 displays the count of clustering results of the monitoring sites. Due to the triggering factors were mainly considered of each part of the landslide in this paper, the second type of clustering result V2 with higher monthly, alternately, ZG85, ZG86, ZG88, SP-2, SP-6 at different position of the landslide were selected for research.

#### 4.2 Association criteria mining

In the association criteria mining process, in order to ensure the quantity and accuracy of association criteria, confidence and support are usually used to control. Confidence indicates the proportion of transactions containing both A and B in transactions containing A (Confidence = \(P(A&B)/P(A)\)). Support indicates the proportion of transactions containing both A and B in all transactions (Support = \(P(A&B)\)). In general, there is no fixed value of minimum support and confidence, which should be given according to the acceptance of training data and scene. In this paper, the parameters of Apriori algorithm were set as follows: confidence thresholds was 80%; support thresholds was 1.5%. Considering that the deformation rate of the V2 stage was in the range of 68.87 to 399.45 mm/day. Thus, we were more inclined to explore and discuss the triggering factors of the landslide deformation at this stage. Therefore, in this paper, we screened and listed several typical deformation association rules in the five monitoring

| Table 1 Hydrological factors of Shuping landslide |
|---------------------------------------------|
| Reservoir water level |
| Monthly average water level \((h)\) (m) |
| Monthly maximum daily drop of water level \((\Delta h^{-}\text{d})\) (m/day) |
| Monthly maximum daily rise of water level \((\Delta h^{+}\text{m})\) (m/day) |
| Monthly variation of water level \((\Delta h_{1})\) (m/month) |
| Bi-monthly variation of water level \((\Delta h_{2})\) (m/month) |
| Rainfall |
| Monthly maximum continuous rainfall \((q_{\text{c}})\) (mm) |
| Monthly cumulative rainfall \((q_{1})\) (mm) |
| Bi-monthly cumulative rainfall \((q_{2})\) (mm) |
| Monthly maximum daily rainfall \((q_{\text{m}})\) (mm) |
points of ZG85, ZG86, ZG88, SP-2, and SP-6 at V2 stage. In addition, elevation of each monitoring points were 190 m, 270 m, 190 m, 220 m, 310 m.

Association rule results of ZG85 were displayed in Table 4. It was reasonable to conclude that reservoir level dominated in the former item of the association rule. Rule 1 was the correlation criterion for slope monitoring data at V2, including the monthly maximum daily rise of water level. Alternatively, when it was in sharply-rise and monthly variation of water level was in sharply-variation, the landslide shall suffer a rapid displacement period. Similarly, Rule 2 indicated that the displacement of month velocity will be aggravated to V2 when the monthly maximum daily drop of water level was in the state of sharply-drop. Then, it can be seen from Rule 6 that monthly/bi-monthly variation of water level played a crucial role in triggering factors of water level. Specifically, the sharply-variation of monthly/bi-monthly variation contributed to the possibility of landslide. Moreover, it was more likely to appear under the influence of factors such as a sharp-rise of water level.

As elucidated in Table 5, the clustering Rules 1 ~ 7 of ZG86 meant that the large deformation stage was more likely to occur in the case of heavy-continuous-rainfall, while reservoir water fluctuations less mattered. Subsequently, rules 8 ~ 10 also indicated that V2 was prone to

Table 2 Clustering results based on the two-step clustering

| Category         | Factors          | Clustering results        | Count |
|------------------|------------------|--------------------------|-------|
| Reservoir water level | $h$             | (160.14 – 174.74) High-Water-Level | 46    |
|                  |                  | (144.21–158.47) Medium-Water-Level | 53    |
|                  |                  | (135.23–138.95) Low-Water-Level | 39    |
| $Ah'_m$          |                  | (0.00–0.63) Slowly-Rise  | 106   |
|                  |                  | (0.73–1.87) Medium-Rise  | 22    |
|                  |                  | (2.15–3.69) Sharply-Rise | 10    |
| $Ah''_m$         |                  | (0.00–0.03) Slowly-Drop  | 104   |
|                  |                  | (0.49–1.70) Sharply-Drop | 34    |
| $Ah_1$           |                  | (0–3.86) Slowly-Variation| 89    |
|                  |                  | (3.95–10.45) Medium-Variation | 35    |
|                  |                  | (11.28–18.25) Sharply-Variation | 14    |
| $Ah_2$           |                  | (0.00–7.19) Slowly-Variation | 85    |
|                  |                  | (7.78–25.56) Sharply-Variation | 53    |
| Rainfall         | $q''_c$          | (1.50–53.30) Light-Continuous-Rainfall | 88    |
|                  |                  | (54.60–110.50) Moderate-Continuous-Rainfall | 38    |
|                  |                  | (125.00–239.40) Heavy-Continuous-Rainfall | 12    |
|                  | $q_1$            | (3.10–160.00) Light-Cumulative-Rainfall | 114   |
|                  |                  | (161.10–443.60) Heavy-Cumulative-Rainfall | 24    |
|                  | $q_2$            | (18.40–174.90) Light-Cumulative-Rainfall | 71    |
|                  |                  | (181.40–385.10) Moderate-Cumulative-Rainfall | 52    |
|                  |                  | (403.30–665.80) Heavy-Cumulative-Rainfall | 15    |
|                  | $q''_m$          | (1.30–25.60) Light-Daily-Rainfall | 10    |
|                  |                  | (26.50–51.30) Moderate-Daily-Rainfall | 44    |
|                  |                  | (51.70–95.80) Medium-Daily-Rainfall | 21    |
|                  |                  | (121.10–160.70) Heavy-Daily-Rainfall | 3     |

Table 3 Clustering results of the monthly velocity

| Monthly velocity ($v$) (mm/month) | Clustering results |
|-----------------------------------|-------------------|
| (- 13.78 ~ 55.10)                 | V1                |
| (68.87 ~ 399.45)                  | V2                |

Fig. 8 Count of clustering results of the monitoring sites
Similar to ZG85, these clustering rules such as Rules (Fig. 5). However, the monitoring data elucidated that ZG88 and ZG85 were at the same elevation sharply-variation, the synergy of other reservoir water when the bi-monthly variation of water level reached deformation rate to V2. In addition, Rule 10 indicated that drop of water level, which will increase the landslide monthly maximum daily was prone to induce the sharply-daily rise of water level was in the state of sharply-rise or daily fall of water level can be interpreted as the continuous monthly rainfall and triggering factors of the reservoir water level.

Table 4 Association criteria of ZG85 monitoring point

| Rule ID | Rules | Support (%) | Confidence (%) |
|---------|-------|-------------|----------------|
| 1       | $\Delta h_m = \text{Sharply-Rise} & \Delta h_1 = \text{Sharply-Variation} \land q_u = \text{Light-Continuous-Rainfall} \rightarrow V2$ | 0.72 | 100 |
| 2       | $\Delta h_m = \text{Sharply-Drop} \land h = \text{Medium-Water-Level} \land q_2 = \text{Light-Cumulative-Rainfall} \rightarrow V2$ | 2.17 | 100 |
| 3       | $\Delta h_m = \text{Sharply-Drop} \land \Delta h'_m = \text{Medium-Rise} \land \Delta h_1 = \text{Sharply-Variation} \rightarrow V2$ | 1.45 | 100 |
| 4       | $\Delta h_m = \text{Sharply-Drop} \land \Delta h'_m = \text{Medium-Rise} \land \Delta h_2 = \text{Sharply-Variation} \rightarrow V2$ | 0.72 | 100 |
| 5       | $\Delta h_m = \text{Sharply-Drop} \land \Delta h_1 = \text{Medium-Variation} \land q_2 = \text{Light-Cumulative-Rainfall} \rightarrow V2$ | 1.45 | 100 |
| 6       | $\Delta h_1 = \text{Sharply-Variation} \land \Delta h'_m = \text{Sharply-Drop} \land \Delta h'_m = \text{Medium-Rise} \rightarrow V2$ | 0.72 | 100 |
| 7       | $\Delta h_1 = \text{Sharply-Variation} \land \Delta h'_m = \text{Sharply-Drop} \land q'_m = \text{Medium-Daily-Rainfall} \rightarrow V2$ | 1.45 | 100 |
| 8       | $\Delta h_1 = \text{Sharply-Variation} \land q'_u = \text{Light-Continuous-Rainfall} \land q'_m = \text{Moderate-Daily-Rainfall} \rightarrow V2$ | 1.45 | 100 |
| 9       | $\Delta h_2 = \text{Sharply-Variation} \land \Delta h'_m = \text{Sharply-Drop} \land \Delta h_1 = \text{Sharply-Variation} \rightarrow V2$ | 1.45 | 100 |
| 10      | $\Delta h_2 = \text{Sharply-Variation} \land \Delta h'_m = \text{Sharply-Drop} \land q_2 = \text{Light-Cumulative-Rainfall} \rightarrow V2$ | 0.72 | 100 |

Table 5 Association criteria of ZG86 monitoring point

| Rule ID | Rules | Support (%) | Confidence (%) |
|---------|-------|-------------|----------------|
| 1       | $q'_u = \text{Heavy-Continuous-Rainfall} \land q_2 = \text{Moderate-Cumulative-Rainfall} \land h = \text{Medium-Water-Level} \rightarrow V2$ | 0.72 | 100 |
| 2       | $q'_u = \text{Heavy-Continuous-Rainfall} \land q'_m = \text{Moderate-Daily-Rainfall} \land h = \text{Medium-Water-Level} \rightarrow V2$ | 0.72 | 100 |
| 3       | $q'_u = \text{Heavy-Continuous-Rainfall} \land q'_m = \text{Medium-Daily-Rainfall} \land \Delta h'_m = \text{Slowly-Rise} \rightarrow V2$ | 2.17 | 100 |
| 4       | $q'_u = \text{Heavy-Continuous-Rainfall} \land \Delta h'_m = \text{Medium-Daily-Rainfall} \land \Delta h'_m = \text{Slowly-Drop} \rightarrow V2$ | 1.45 | 100 |
| 5       | $q'_u = \text{Heavy-Continuous-Rainfall} \land q'_m = \text{Medium-Daily-Rainfall} \land \Delta h_1 = \text{Slowly-Variation} \rightarrow V2$ | 1.45 | 100 |
| 6       | $q'_u = \text{Heavy-Continuous-Rainfall} \land q'_m = \text{Medium-Daily-Rainfall} \land \Delta h_2 = \text{Slowly-Variation} \rightarrow V2$ | 1.45 | 100 |
| 7       | $q'_u = \text{Heavy-Continuous-Rainfall} \land \Delta h_1 = \text{Medium-Variation} \land \Delta h'_m = \text{Slowly-Rise} \rightarrow V2$ | 0.72 | 100 |
| 8       | $q_2 = \text{Sharply-Cumulative-Rainfall} \land \Delta h'_m = \text{Medium-Rise} \land \Delta h_1 = \text{Slowly-Variation} \rightarrow V2$ | 0.72 | 100 |
| 9       | $q_2 = \text{Sharply-Cumulative-Rainfall} \land \Delta h'_m = \text{Medium-Rise} \land \Delta h_2 = \text{Slowly-Variation} \rightarrow V2$ | 0.72 | 100 |
| 10      | $q_2 = \text{Sharply-Cumulative-Rainfall} \land \Delta h_1 = \text{Medium-Variation} \land \Delta h'_m = \text{Slowly-Rise} \rightarrow V2$ | 0.72 | 100 |

appear during bi-monthly-cumulative rainfall was in a heavy state. Therefore, these association rules indicated that the part of the landslide in the middle and rear (Fig. 5) was almost dominated by rainfall.

Association rule results of ZG88 were shown in Table 6. Similar to ZG85, these clustering rules such as Rules 1 ~ 5 and Rules 6 ~ 9 indicated that monthly maximum daily rise of water level was in the state of sharply-rise or monthly maximum daily was prone to induce the sharply-drop of water level, which will increase the landslide deformation rate to V2. In addition, Rule 10 indicated that when the bi-monthly variation of water level reached sharply-variation, the synergy of other reservoir water uplift factors will also lead to the realization of V2. Furthermore, ZG88 and ZG85 were at the same elevation (Fig. 5). However, the monitoring data elucidated that ZG88, which was located on the west landslide, had a larger deformation than ZG85 (Fig. 7). Then, it was possible to believe that maximum daily drop of water level was more likely to cause large displacements among the triggering factors of the reservoir water level.

Table 7 portrayed the association rules of SP-2. Rule 1 can be interpreted as the continuous monthly rainfall and maximum daily rise of water level reached in a heavy state, which induced the displacement and deformation rate in the middle of the landslide to reach the V2 stage. Similarly, Rules 2 ~ 5 can be explained as heavy state of monthly cumulative rainfall can aggravate the deformation of landslide with other triggering factors of rainfall such as monthly variable of water level. Then, Rules 6 ~ 10 meant that under the combined effect of rainfall and reservoir water level fluctuations, such as bi-monthly cumulative rainfall (Heavy-Cumulative-Rainfall) and monthly maximum daily rise of water level (Medium/Sharply-Rise), the landslide was likely to show severe deformation.
Table 7 Association criteria of SP-2 monitoring point

| Rule ID | Rules                                                                 | Support (%) | Confidence (%) |
|---------|------------------------------------------------------------------------|-------------|----------------|
| 1       | $\Delta h_{m} = \text{Sharply-Rise} \& h = \text{Medium-Water-Level} \& q_{c}' = \text{Light-Continuous-Rainfall} \rightarrow V2$ | 1.45        | 100            |
| 2       | $\Delta h_{m} = \text{Sharply-Rise} \& \Delta h_{1} = \text{Medium-Variation} \& h = \text{Medium-Water-Level} \rightarrow V2$ | 1.45        | 100            |
| 3       | $\Delta h_{m} = \text{Sharply-Rise} \& \Delta h_{1} = \text{Medium-Variation} \& q_{c}' = \text{Moderate-Continuous-Rainfall} \rightarrow V2$ | 0.72        | 100            |
| 4       | $\Delta h_{m} = \text{Sharply-Rise} \& \Delta h_{1} = \text{Medium-Variation} \& q_{d}' = \text{Moderate-Daily-Rainfall} \rightarrow V2$ | 1.45        | 100            |
| 5       | $\Delta h_{m} = \text{Sharply-Rise} \& \Delta h_{1} = \text{Sharply-Variation} \& q_{c}' = \text{Light-Continuous-Rainfall} \rightarrow V2$ | 0.72        | 100            |
| 6       | $\Delta h_{m} = \text{Sharply-Drop} \& h = \text{Medium-Water-Level} \& q_{2} = \text{Light-Cumulative-Rainfall} \rightarrow V2$ | 1.45        | 100            |
| 7       | $\Delta h_{m} = \text{Sharply-Drop} \& \Delta h_{1} = \text{Medium-Variation} \rightarrow V2$ | 1.45        | 100            |
| 8       | $\Delta h_{m} = \text{Sharply-Drop} \& \Delta h_{1} = \text{Sharply-Rise} \& q_{c}' = \text{Light-Continuous-Rainfall} \rightarrow V2$ | 1.45        | 100            |
| 9       | $\Delta h_{m} = \text{Sharply-Drop} \& \Delta h_{1} = \text{Medium-Variation} \& q_{2} = \text{Light-Cumulative-Rainfall} \rightarrow V2$ | 0.72        | 100            |
| 10      | $\Delta h_{2} = \text{Sharply-Variation} \& \Delta h_{m} = \text{Sharply-Drop} \& q_{2} = \text{Light-Cumulative-Rainfall} \rightarrow V2$ | 1.45        | 100            |

Compared to ZG86, SP-6 was located at 310 m on the contour line, which was the closest to the rear of the landslide among all monitoring points. Table 8 showed the correlation criteria of SP-6. Precisely, Rules 1 ~ 4 indicated that if the monthly maximum continuous rainfall was between 1.25 and 2.39 m (Heavy-continuous-rainfall), and the monthly maximum daily rainfall was also in the interval of (0.52, 0.95), that was, medium-daily-rainfall, the landslide was prone to deform at a high rate (V2).

In general, longitudinal observation of the correlation criteria of ZG85, SP-2, ZG86 and SP-6 indicated that among the triggering factors of the toe, middle and rear of the landslide, the governing factors restricted each other, alternatively, the reservoir level factor was inversely proportional to the rainfall related factors. Specifically, the reservoir level factors played a leading role in the toe landslide. On the contrary, the rear landslide was jointly controlled by the reservoir water level and rainfall. In particular, the rainfall factor was the key factor that causes the deformation of the rear landslide.

4.3 Threshold values analysis

The 80% of the total data is defined as training samples to build the decision tree model, and the rest 20% data is set as the testing samples to check the accuracy of the model. In order to improve the generalization ability of the model and prevent the model from over fitting, this paper adopts the method of combining cross validation and boosting technology in the construction of decision tree C5.0 model. The number of tests for boosting is set to 10, the number of cross validation folds is set to 10, and the expected noise is set to 10%. As shown in Fig. 9, which contained 6 hydrological factors ($h, \Delta h_{m}', \Delta h_{2}, q_{c}', q_{1}, q_{d}'$). In this C5.0
model, the accuracy of total sample identification reached 94.53%. Generally, V2 mode should be paid enough attention. Therefore, nodes in V2 state were selected for analysis. In this model, there were 5 nodes in V2, as shown in Table 9. Among them, criterion 4 was only composed of reservoir water factor, while the remaining four criterion were composed of reservoir water and rainfall factor. It meant that when the water level was 138.951 ~ 147.437 m, once the daily drop of water level exceeded 0.416 m/d, the landslide will enter second stage (V2). This criterion was the most important one which accounted for 58.02% of all V2 state points. Therefore, during the period of low water level, 0.416 m/d can be identified as the threshold of daily drop of water level for the severe deformation of the landslide. Whereas, this did not mean that when the daily drop of water level was less than 0.416 m/d, the landslide will not have strong deformation. Criterion 5 indicated that under the same reservoir water level condition, the landslide might still appear V2 state as long as the rainfall meet certain conditions ($q_d < 95.8$ mm).
Table 9 Threshold values of the hydrological factors (Blue: reservoir water level factors; Orange: rainfall factors)

| Criterion | Hydrological factors | Mode | Count | Confidence |
|-----------|----------------------|------|-------|------------|
| Criterion 1 | \( \Delta h > 0.416 \) | \( h > 147.437 \) | \( q^m > 28.1 \) | \( \Delta h \) | \( q_r > 362.5 \) | V2 | 5 | 1.000 |
| Criterion 2 | \( \Delta h' > 0.416 \) | \( h < 147.437 \) | \( q^m > 16.352 \) | | \( q_r < 36.6 \) | V2 | 5 | 1.000 |
| Criterion 3 | | \( h > 138.951 \) | \( h < 147.437 \) | \( q^m < 95.8 \) | V2 | 10 | 1.000 |

5 Discussion

Significantly, Fig. 9 contained the data of 5 displacement monitoring points of Shuping landslide. Among them, monitoring points ZG85 and ZG88 were located at the toe landslide, SP-2 was located in the middle of the landslide, and ZG86 and SP-6 were located at the rear landslide. Although the accuracy of total sample identification reached 94.53%, the state identification of some nodes might be confusion due to the different governing factors of monitoring points at different part of the landslide. For instance, Node 18 displayed 9 points in V1 state and 14 points in V2 state. Therefore, it was necessary to establish some separate decision tree models for the monitoring points at different locations of the landslide to accurately identified the control factors of deformation.

The decision tree C5.0 model of each monitoring point was shown in Fig. 10. Obviously, the three monitoring points located in the middle and toe landslide were only controlled by the reservoir water factors (ZG85, ZG88, SP-2), while the two monitoring points located at the rear landslide were jointly controlled by the reservoir water and rainfall factors (ZG86, SP-6), which also made the decision tree model of these two monitoring points much more complex. In terms of the accuracy of the model, due to the inclusion of multiple factors related to reservoir level and rainfall, the accuracy of these two model was the highest, reaching 97.73% and 94.93% respectively. The two monitoring points at the toe landslide were only controlled by the reservoir water level, and the accuracy of the model was more than 0.9. It was worth noting that the decision tree model of the monitoring point SP-2 located in the middle landslide was the simplest, but the accuracy of the model was the lowest, only for 88.41%. This was because the middle landslide was in the transition zone from reservoir water control to rainfall control. On the one hand, the control effect of reservoir water related factors was gradually weakened. On the other hand, the control effect of rainfall related factors was gradually enhanced. In other words, the governing factors of landslide deformation in this area were fuzzy.

In order to quantitatively analyze the main control factors of monitoring points at different locations of the landslide, three monitoring points at different locations (toe landslide: ZG85, middle landslide: SP-2, and rear landslide: ZG86) were selected to count the factor contribution in the decision tree models, as shown in Fig. 11. Obviously, the daily drop of water level was the most important factor causing the deformation of Shuping landslide. ZG85 monitoring point at the toe landslide was only affected by \( \Delta h' \) and \( h \), and the contribution degrees of the two factors reached 0.58 and 0.42 respectively. When the reservoir water level was between 138.951 and 147.072 m, once the daily decline of reservoir water level exceeded 0.315 m/d, the toe landslide will enter the second stage (V2). Similarly, SP-2 monitoring point in the middle of the landslide was affected by \( \Delta h' \) and \( h \), and the contribution degrees of the two factors were 0.77 and 0.23 respectively. When the reservoir water level was lower than 155.552 m and the daily drop of water level exceeded 0.416 m/d, the middle of the landslide will enter the second stage (V2). The deformation of ZG86 at the rear landslide was controlled by six factors (\( \Delta h' \), \( q^m \), \( \bar{h} \), \( q_r \), \( q_2 \), \( \Delta h_2 \)), and the contribution degrees of each factor were 0.28, 0.23, 0.18, 0.14, 0.09 and 0.09 respectively. Among them, the total contribution of the three reservoir water level related factors was 0.55 (\( \Delta h' \), \( h \), \( \Delta h_2 \)), while the total contribution of the three rainfall related factors was 0.46 (\( q^m \), \( q_r \), \( q_2 \)). When the reservoir water level was between 138.951 and 147.437 m and the daily drop of water level exceeded 0.388 m/d, once the bi-monthly cumulative rainfall exceeded 254.2 mm, the rear landslide will enter the second stage (V2). Therefore, when the reservoir water level was at a low level, the daily drop of water level reaches 0.315 m/d, and the deformation of the toe landslide will be accelerated. When the daily drop of water level exceeds 0.416 m/d, accelerated deformation will occur in the middle landslide. The deformation of the rear landslide need to be warned in combination with various factors of rainfall and reservoir water level.
In this paper, when establishing the decision tree model, nine influencing factors related to reservoir water level and rainfall were selected ($h$, $\Delta h_1$, $\Delta h_2$, $q_e$, $q_1$, $q_2$, $q_d$). Nevertheless, it can be concluded from Fig. 7 that the selected 9 hydrological factors may have some time correlation. Therefore, in the established decision tree models, these two most unfavorable factors will not appear at the same time. Considering the results of data mining and decision tree models, Shuping landslide was prone to strong deformation from May to June every year. Therefore, the drop rate of water level need to be strictly controlled during this period. In case of heavy rainfall from May to June, it was necessary to monitor the deformation of landslide in real time and give early warning to prevent large-scale landslide disaster.
6 Conclusions

In this research, considering the spatial variability of governing factors in different parts of landslide, data mining algorithms were used to identify the governing factors and their thresholds of Shuping landslide. The following conclusions can be reached:

1. Under the joint influence of seasonal rainfall and periodic reservoir water level fluctuation, the displacement of Shuping landslide presented a "step-like" trend. Taking rainfall and reservoir water level as hydrological factors has clear physical significance.

2. Data mining results indicated that the governing factors of the toe, middle and rear landslide restricted each other. Specifically, the reservoir water level factor played a leading role in the toe landslide. On the contrary, the rear landslide was jointly controlled by the reservoir water level and rainfall.

3. The daily drop of water level was the most important factor causing the deformation of Shuping landslide. During the period of low water level, 0.416 m/d can be identified as the threshold of daily drop of water level for the severe deformation of the landslide.

4. Considering the results of data mining and decision tree models, Shuping landslide was prone to strong deformation from May to June every year. Consequently, it was necessary to monitor the deformation of landslide in real time and give early warning to prevent large-scale landslide disaster.

Acknowledgements This research was supported by the National Natural Science Foundation of China (42007267, 41977244), Science and Technology Project of Hubei Provincial Department of Natural Resources (ZRZY2022KJ07). The authors thank the colleagues in our laboratory for their constructive comments and assistance.

References

Agrawal R, Imieliński T, Swami A (1993) Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD international conference on management of data, pp 207–216

Alihuwaynee O, Aydda A, Hwang I et al (2021) Uncertainty reduction of unlabeled features in landslide inventory using machine learning t-SNE clustering and data mining apriori association rule algorithms. Appl Sci 11(2):556

Chen W, Pourghasemi H, Panahi M et al (2017) Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. Geomorphology. https://doi.org/10.1016/j.geomorph.2017.09.007

Chen W, Zhang S, Li R et al (2018) Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naive Bayes tree for landslide susceptibility modeling. Sci Total Environ 644:1006–1018

Du J, Yin K, Lacasse S (2013) Displacement prediction in colluvial landslides, three gorges reservoir, China. Landslides 10(2):203–218

Gariano S, Guzzetti F (2016) Landslides in a changing climate. Earth-Sci Rev 162:227–252

Gong W, Zhao C, Jiang C et al (2020) Stratigraphic uncertainty modeling with random field approach. Comput Geotech 125:103681

Gong W, Jiang C, Wasowski J (2021) Geohazards and human settlements: lessons learned from multiple relocation events in Badong, China-Engineering geologist’s perspective. Eng Geol 285:106051

Gong W, Tian S, Wang L et al (2022) Interval prediction of landslide displacement with dual-output least squares support vector machine and particle swarm optimization algorithms. Acta Geotech. https://doi.org/10.1007/s11440-022-01455-2

Guo W, Zuo X, Yu J et al (2019) Method for mid-long-term prediction of landslides movements based on optimized apriori algorithm. Appl Sci 9(18):3819

Hong H, Pourghasemi H, Pourtaghi Z (2016) Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models. Geomorphology 259:105–118

Huang F, Huang J, Jiang S et al (2017) Landslide displacement prediction based on multivariable chaotic model and extreme learning machine. Eng Geol 218:173–186

Jiang D, Wu B, Cheng Z et al (2021) Towards a probabilistic model for estimation of grounding accidents in fluctuating backwater zone of the Three Gorges Reservoir. Reliab Eng Syst Saf 205:107239

Juang CH, Dijkstra T, Wasowski J et al (2019) Loess geohazards research in China: advances and challenges for mega engineering projects. Eng Geol 251:1–10

Kusak L, Unel F, Alptekin A et al (2021) Apriori association rule and K-means clustering algorithms for interpretation of pre-event landslide areas and landslide inventory mapping. Open Geosci 13(1):1226–1244
Ma J, Tang H, Liu X et al (2017a) Establishment of a deformation forecasting model for a step-like landslide based on decision tree C5.0 and two-step cluster algorithms: a case study in the Three Gorges Reservoir area China. Landslides 14(3):1275–128

Ma J, Tang H, Hu X et al (2017b) Identification of causal factors for the majiagou landslide using modern data mining methods. Landslides 14(1):311–322

Ma J, Tang H, Liu X, Wen T, Zhang J, Tan Q, Fan Z (2018) Probabilistic forecasting of landslide displacement accounting for epistemic uncertainty: a case study in the Three Gorges Reservoir area China. Landslides 15(6):1145–1153

Ma J, Niu X, Tang H et al (2020) Displacement prediction of a complex landslide in the Three Gorges Reservoir Area (China) using a hybrid computational intelligence approach. Complexity. https://doi.org/10.1155/2020/2624547

Ma J, Wang Y, Niu X et al (2022) A comparative study of mutual information-based input variable selection strategies for the displacement prediction of seepage-driven landslides using optimized support vector regression. Stoch Env Res Risk Assess. https://doi.org/10.1007/s00477-022-02183-5

Miao F, Wu Y, Li L et al (2018a) Centrifuge model test on the retrogressive landslide subjected to reservoir water level fluctuation. Eng Geol 245:169–179

Miao F, Wu Y, Xie Y et al (2018b) Prediction of landslide displacement with step-like behavior based on multialgorithm optimization and a support vector regression model. Landslides 15(3):475–488

Miao F, Wu Y, Li L et al (2021) Triggering factors and threshold analysis of baishuihe landslide based on the data mining methods. Nat Hazards 105:2677–2696

Miao F, Wu Y, Ákos T et al (2022) Centrifugal model test on a riverine landslide in the Three Gorges Reservoir I induced by rainfall and water level fluctuation. Geosci Front 13(3):101378

Milevski I, Slavoljub D, Zorn M (2019) Statistical and expert-based landslide susceptibility modeling on a national scale applied to North Macedonia. Open Geosci 11:750–764

Moreiras S (2005) Climatic effect of ENSO associated with landslide occurrence in the Central Andes, Mendoza 530 Province, Argentina. Landslides 2:53–59. https://doi.org/10.1007/s10346-005-0046-4

Nici I, Asnuldeesi A (2018) GIS-based evaluation of diagnostic areas in landslide susceptibility analysis of Bahluiet River Basin (Moldavian Plateau, NE Romania). are Neolithic Sites in danger? Geomorphology 314:27–41

Pandya R, Pandya J (2015) C5.0 algorithm to improved decision tree with feature selection and reduced error pruning. Int J Comput Appl 117(16):18–21

Peng L, Xu D, Wang X (2018) Vulnerability of rural household livelihood to climate variability and adaptive strategies in landslide-threatened western mountainous regions of the Three Gorges Reservoir area. China. Clim Dev 11(6):469–484

Petley D, Hearn G, Hart A et al (2007) Trends in landslide occurrence in Nepal. Nat Hazards 42(23–532):44. https://doi.org/10.1007/s11069-006-9100-3

Sassa K, Nagai O, Solidum R et al (2010) An integrated model simulating the initiation and motion of earthquake 538 and rain induced rapid landslides and its application to the 2006 Leyte landslide. Landslides 7:219–236. https://doi.org/10.1007/s10346-010-0230-z

Shi W, Zhang M, Ke H et al (2020) Landslide recognition by deep convolutional neural network and change detection. IEEE Trans Geosci Remote Sens 59(6):4654–4672

Stumpf A, Kerle N (2011) Object-oriented mapping of landslides using random forests. Remote Sens Environ 115(10):2564–2577

Sun P, Wang H, Wang G et al (2021) Field model experiments and numerical analysis of rainfall-induced shallow loess landslides. Eng Geol 295:106411

Tan F, Hu X, He C et al (2017) Identifying the main control factors for different deformation stages of landslide. Geotech Geol Eng 2:1–14

Tang H, Wasowski J, Juang CH (2019) Geohazards in the Three Gorges Reservoir area, China: lessons learned from decades of research. Eng Geol 261:105267–105267

Vorpahl P, Eibenheer H, Mäker M et al (2012) How can statistical models help to determine driving factors of landslides? Ecol Model 239:27–39

Wang Y, Tang H, Huang J et al (2022) A comparative study of different machine learning methods for reservoir landslide displacement prediction. Eng Geol 298:106544

Wu S, Hu X, Zheng W et al (2021) Effects of reservoir water level fluctuations and rainfall on a landslide by two-way ANOVA and K-means clustering. Bull Eng Geol Env 80(7):5405–5421

Yao W, Li C, Zuo Q et al (2019) Spatiotemporal deformation characteristics and triggering factors of Baijiabao landslide in Three Gorges Reservoir region, china. Geomorphology 343:34–47

Zhang K, Wu X, Niu R et al (2017) The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area. China Environ Earth Sci 76(11):405

Zhang Y, Tang J, He Z et al (2020a) A novel displacement prediction method using gated recurrent unit model with time series analysis in the Erdaohe landslide. Nat Hazards 1:1–31

Zhang Y, Tang J, Liao R et al (2020b) Application of an enhanced BP neural network model with water cycle algorithm on landslide prediction. Stoch Env Res Risk Assess 61:65:1–19

Zhang Y, Zhang Z, Xue S et al (2020c) Stability analysis of a typical landslide mass in the Three Gorges Reservoir under varying reservoir water levels. Environ Earth Sci 79(1):1–14

Zhao X, Chen W (2019) GIS-based evaluation of landslide susceptibility models using certainty factors and functional trees-based ensemble techniques. Appl Sci 10(1):16

Zhou C, Yin K, Cao Y et al (2016) Application of time series analysis in the majiagou landslide using modern data mining methods. Nat Hazards 42(23–532):44. https://doi.org/10.1007/s11069-006-9100-3

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