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

**Liangjie Guo**
China University of Geosciences
**Fasheng Miao** ( fsmiao@cug.edu.cn )
China University of Geosciences
**Fancheng Zhao**
China University of Geosciences
**Yiping Wu**
China University of Geosciences

**Research Article**

**Keywords:** data mining technology, Shuping Landslide, governing factors, the Three Gorges Reservoir, reservoir level, rainfall

**Posted Date:** March 11th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1432895/v1

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

**Liangjie Guo [1], Fasheng Miao [1,2,*], Fancheng Zhao [1], Yiping Wu [1]**

**Liangjie Guo [1]**

1. Faculty of Engineering, China University of Geosciences, Wuhan, China, 430074

First author. Email: guoliangjie@cug.edu.cn

**Fasheng Miao [1,2,*]**

1. Faculty of Engineering, China University of Geosciences, Wuhan, China, 430074;

2. Engineering Research Center of Rock-Soil Drilling & Excavation and Protection, Ministry of Education, Wuhan, China, 430074;

*Corresponding author. Tel.: +86 15527160870, Email: fsmiao@cug.edu.cn

**Fancheng Zhao [1]**

1. Faculty of Engineering, China University of Geosciences, Wuhan, China, 430074

Email: zhaofancheng@cug.edu.cn

**Yiping Wu [1]**

1. Faculty of Engineering, China University of Geosciences, Wuhan, China, 430074

Email: ypwu@cug.edu.cn

## Acknowledgments:

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.

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

**Liangjie Guo [1], Fasheng Miao [1,2,*], Fancheng Zhao [1], Yiping Wu [1]**

(1. Faculty of Engineering, China University of Geosciences, Wuhan, China, 430074;
2. Engineering Research Center of Rock-Soil Drilling & Excavation and Protection, Ministry of Education, Wuhan, China, 430074;

**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 was 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). 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 (Nicu 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 175m, 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). 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).

The causes of landslides are miscellaneous and varying from disaster triggering factors (Gariano et al. 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 et al. 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 (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; Miao et al. 2020).

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), prediction (Du et al. 2013; Chen et al. 2017; Zhang et al. 2020), modeling (Stumpf et al. 2011; Vorpahl et al. 2012), susceptibility (Hong et al. 2016; Chen et al. 2018; Zhao et al. 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, Tian 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. Lutfiye 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.

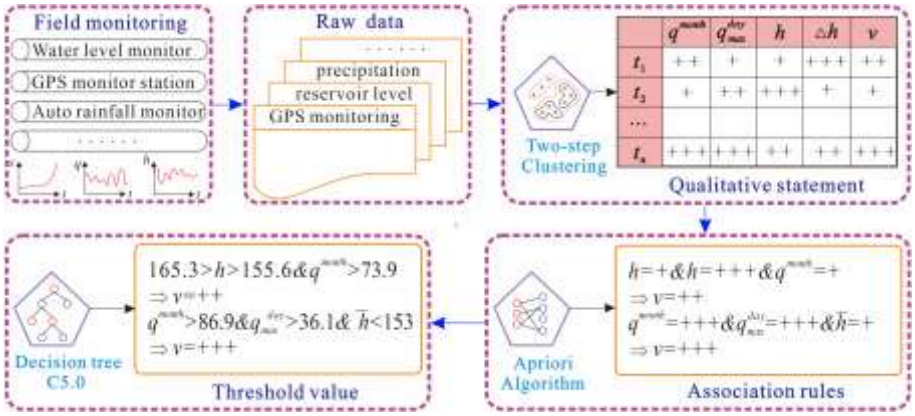


**Fig. 1** Data mining technology proposed in the paper

## 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 $C_{L'\rightarrow L-L'}$ is greater than a user-specified confidence threshold:

$$C_{L'\rightarrow(L-L')}=\frac{|T(L)|}{|T(L')|}\geq minconf \tag{1}$$

Then the association criterion is generated: $L'\rightarrow(L-L')$.

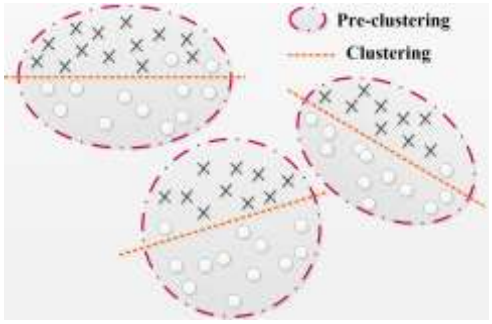


**Fig. 2** Sketch map of two-step clustering

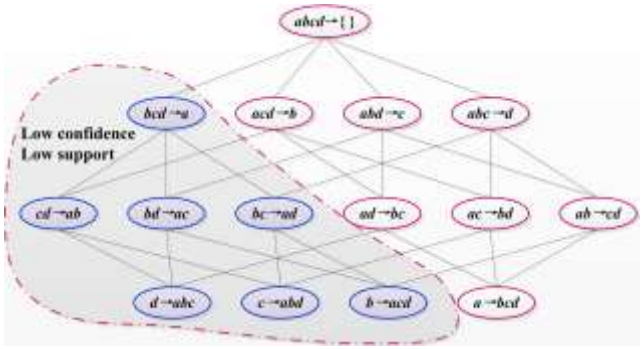


**Fig. 3** The sketch map of Apriori algorithm

## 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 term of classification and prediction, the nodes of the tree are sample attributes chosen for splitting while the branches are attribute values (Pandya et al., 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.

$$Ent(N) = -\sum_{i}^{t} p(C_i|N)\log_2 p(C_i|N) \tag{2}$$

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

$$Ent(N|T) = \sum_{j}^{k} \frac{|T_j|}{|N|} \times Ent(N) \tag{3}$$

The information gain can be expressed as:

$$Gains(N,T) = Ent(N) - Ent(N|T) \tag{4}$$

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

$$GainRatio = \frac{Gains(N,T)}{Ent(T)} \tag{5}$$

In addition, determining the optimal splitting point is also essential in decision tree growth. The core idea of the solution has the following steps: ①arrange the input variables in ascending order; ②define 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; ④calculate the chi-square value of each two adjacent groups; ⑤obtain the chi-square threshold based on the significance level and degrees of freedom, and merge the interval with the smallest cardinality value; ⑥repeat 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:

$$w_j(i) = \frac{1}{N}(i = 1,2,\mathrm{L}\ ,k;\ j = 1,2,\mathrm{L}\ ,N) \tag{6}$$

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:

$$w_j(i+1) = w_j(i) \times \beta(i) \tag{7}$$

$$\beta(i) = \frac{e_i}{1-e_i} \tag{8}$$

Among them, sample weight for misclassification remains the same:

$$w_j(i+1) = w_j(i) \tag{9}$$

Then, normalized:

$$w_j(i+1) = \frac{w_j(i+1)}{\sum_{j=1}^{N} w_j(i+1)} \tag{10}$$

⑤ Iterative steps ②~④ 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 $C_i(i=1,2,3,\ldots,k)$:

$$W_j(X) = -\log\left(\frac{e_i}{1-e_i}\right) \tag{11}$$

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 over-fitting 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 $55\times10^4$ $m^2$, with a total volume of $2.89\times10^7$ $m^3$.

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.

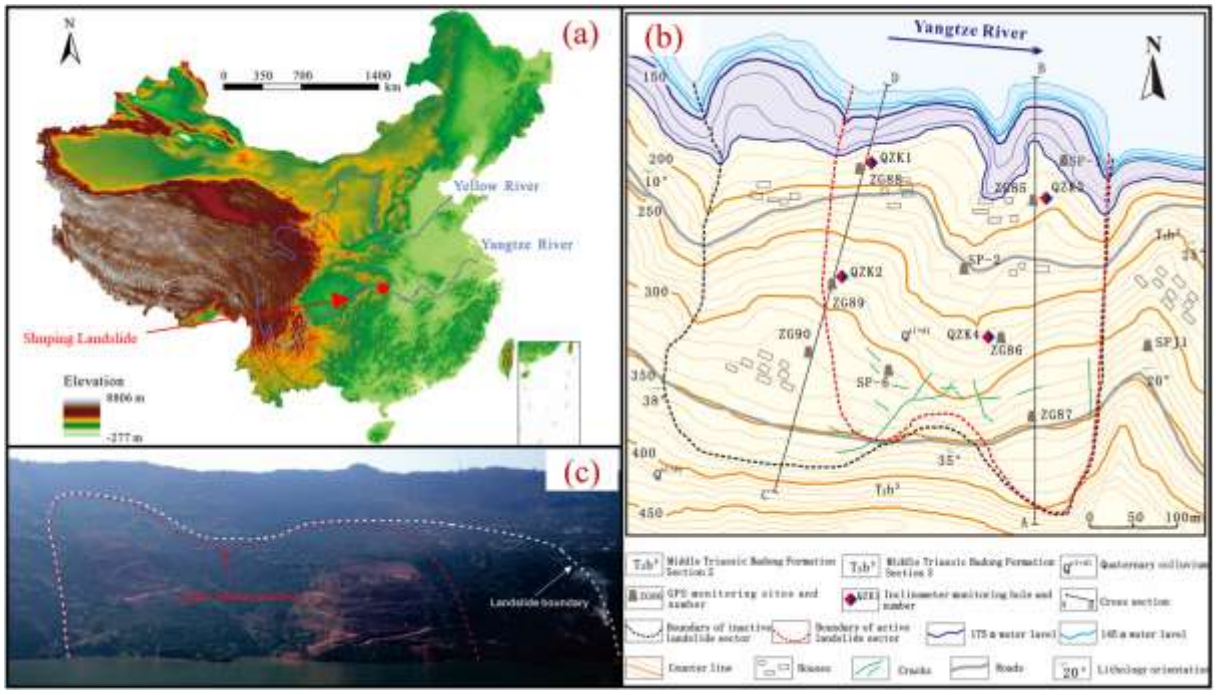


**Fig. 4** Location and plan of the Shuping landslide (a). Location; (b). Topographic map; (c). Overall view

### 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.

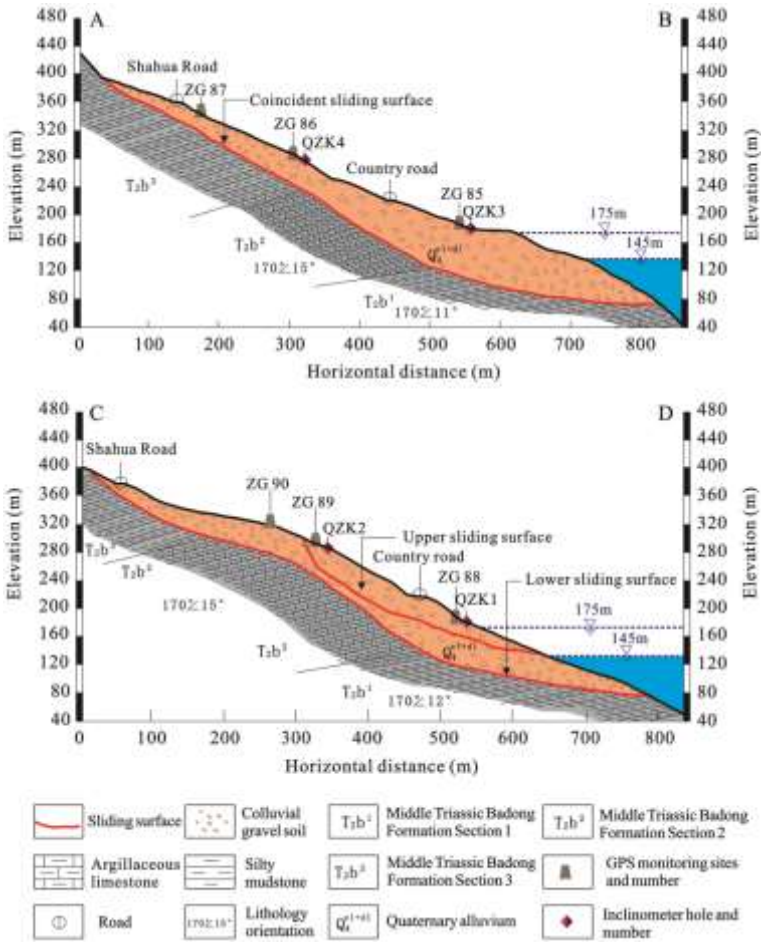


**Fig. 5** Schematic Geological Profile of the Shuping Landslide (A-B), (C-D)


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

### 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 m and 140 m. Then in 2007, the height of the reservoir level reached 155 m for the first time and fluctuated between 145 m 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 m 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 m 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 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.

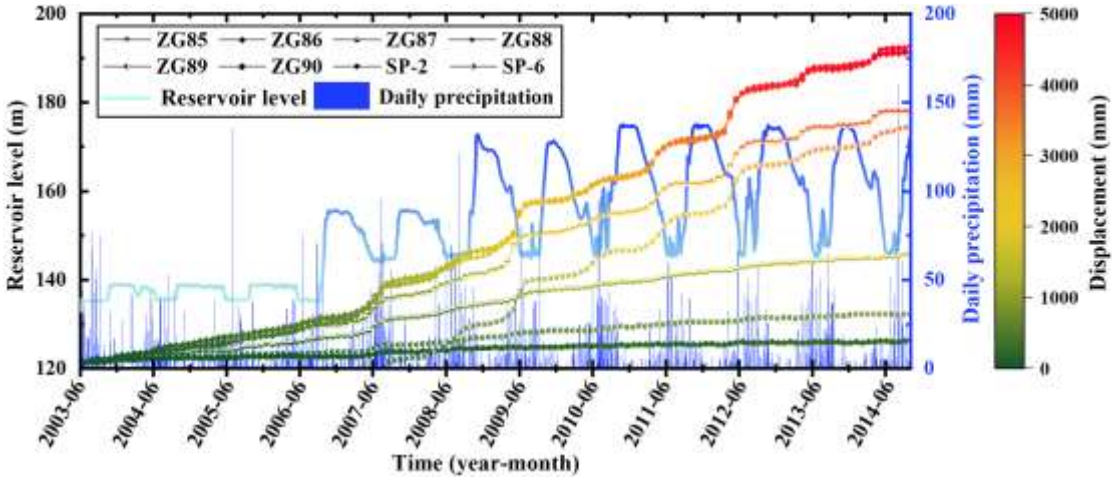


**Fig. 7** Monitoring data of Shuping Landslide

In this paper, we selected 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: $\hbar$, $\Delta h^d_m$, $\Delta h^r_m$, $\Delta h_1$, $\Delta h_2$. Additionally, the hydrological factors related to rainfall include: $q^e_c$, $q_1$, $q_2$, $q^d_m$, as shown in Table 1.

**Table 1** Hydrological factors of Shuping landslide

| Reservoir water level | Rainfall |
|---|---|
| Monthly average water level ($\hbar$) (m) | Monthly maximum continuous rainfall ($q^e_c$) (mm) |
| Monthly maximum daily drop of water level ($\Delta h^d_m$) (m/day) | Monthly cumulative rainfall ($q_1$) (mm) |
| Monthly maximum daily rise of water level ($\Delta h^r_m$) (m/day) | Bi-monthly cumulative rainfall ($q_2$) (mm) |
| Monthly variation of water level ($\Delta h_1$) (m/month) | Monthly maximum daily rainfall ($q^d_m$) (mm) |
| Bi-monthly variation of water level ($\Delta h_2$) (m/month) | |

# 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 2~10 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).

**Table 2** Clustering results based on the two-step clustering

| Category | factors | Clustering results | | Count |
|---|---|---|---|---|
| Reservoir water level | $\hbar$ | (160.14~174.74) | High-Water-Level | 46 |
| | | (144.21~158.47) | Medium-Water-Level | 53 |
| | | (135.23~138.95) | Low-Water-Level | 39 |
| | $\Delta h^r_m$ | (0.00~0.63) | Slowly-Rise | 106 |
| | | (0.73~1.87) | Medium-Rise | 22 |
| | | (2.15~3.69) | Sharply-Rise | 10 |
| | $\Delta h^d_m$ | (0.00~0.03) | Slowly-Drop | 104 |
| | | (0.49~1.70) | Sharply-Drop | 34 |
| | $\Delta h_1$ | (0~3.86) | Slowly-Variation | 89 |
| | | (3.95~10.45) | Medium-Variation | 35 |
| | | (11.28~18.25) | Sharply-Variation | 14 |
| | $\Delta h_2$ | (0.00~7.19) | Slowly-Variation | 85 |
| | | (7.78~25.56) | Sharply-Variation | 53 |
| Rainfall | $q^e_c$ | (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 |
| | $q^d_m$ | (1.30~25.60) | Light-Daily-Rainfall | 70 |
| | | (26.50~51.30) | Moderate-Daily-Rainfall | 44 |
| | | (51.70~95.80) | Medium-Daily-Rainfall | 21 |
| | | (121.10~160.70) | Heavy-Daily-Rainfall | 3 |

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.

**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 |

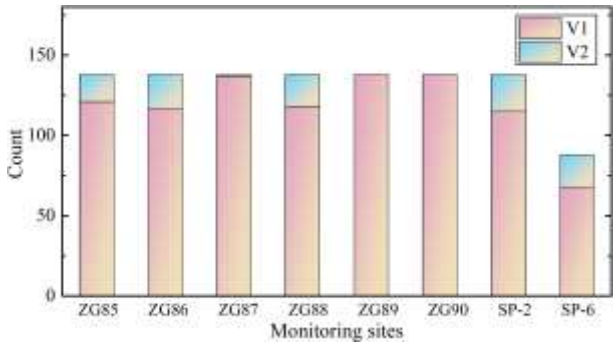


**Fig. 8** Count of clustering results of the monitoring sites

## 4.2 Association criteria mining

In the association criteria mining process, 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 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 2~5 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 6~10 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.

**Table 4** Association criteria of ZG85 monitoring point

| Rule ID | Rules | Support (%) | Confidence (%) |
|---|---|---|---|
| 1 | $\Delta h^r_m$ =Sharply-Rise & $\Delta h_1$=Sharply-Variation & $q^e_c$ =Light-Continuous-Rainfall → V2 | 0.72 | 100 |
| 2 | $\Delta h^d_m$ =Sharply-Drop & $\hbar$ =Medium-Water-Level & $q_2$=Light-Cumulative-Rainfall → V2 | 2.17 | 100 |
| 3 | $\Delta h^d_m$ =Sharply-Drop & $\Delta h^r_m$ =Medium-Rise & $\Delta h_1$=Slowly-Variation → V2 | 1.45 | 100 |
| 4 | $\Delta h^d_m$ =Sharply-Drop & $\Delta h^r_m$ =Medium-Rise & $\Delta h_2$=Slowly-Variation → V2 | 0.72 | 100 |
| 5 | $\Delta h^d_m$ =Sharply-Drop & $\Delta h_1$=Medium-Variation & $q_2$=Light-Cumulative-Rainfall → V2 | 1.45 | 100 |
| 6 | $\Delta h_1$=Sharply-Variation & $\Delta h^d_m$ =Sharply-Drop & $\Delta h^r_m$ =Medium-Rise → V2 | 0.72 | 100 |
| 7 | $\Delta h_1$=Sharply-Variation & $\Delta h^d_m$ =Sharply-Drop & $q^d_m$ =Medium-Daily-Rainfall → V2 | 1.45 | 100 |

| | | | |
|---|---|---|---|
| 8 | $\Delta h_1$=Sharply-Variation & $q^e_c$ =Light-Continuous-Rainfall & $q^d_m$ =Moderate-Daily-Rainfall → V2 | 1.45 | 100 |
| 9 | $\Delta h_2$=Sharply-Variation & $\Delta h^d_m$ =Sharply-Drop & $\Delta h_1$=Slowly-Variation → V2 | 1.45 | 100 |
| 10 | $\Delta h_2$=Sharply-Variation & $\Delta h^d_m$ =Sharply-Drop & $q_2$=Light-Cumulative-Rainfall → V2 | 0.72 | 100 |

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 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.

**Table 5** Association criteria of ZG86 monitoring point

| Rule ID | Rules | Support (%) | Confidence (%) |
|---|---|---|---|
| 1 | $q^e_c$ =Heavy-Continuous-Rainfall & $q_2$=Moderate-Cumulative-Rainfall & $\hbar$ =Medium-Water-Level → V2 | 0.72 | 100 |
| 2 | $q^e_c$ =Heavy-Continuous-Rainfall & $q^d_m$ =Moderate-Daily-Rainfall & $\hbar$ =Medium-Water-Level → V2 | 0.72 | 100 |
| 3 | $q^e_c$ =Heavy-Continuous-Rainfall & $q^d_m$ =Medium-Daily-Rainfall & $\Delta h^r_m$ =Slowly-Rise → V2 | 2.17 | 100 |
| 4 | $q^e_c$ =Heavy-Continuous-Rainfall & $q^d_m$ =Medium-Daily-Rainfall & $\Delta h^d_m$ =Slowly-Drop → V2 | 1.45 | 100 |
| 5 | $q^e_c$ =Heavy-Continuous-Rainfall & $q^d_m$ =Medium-Daily-Rainfall & $\Delta h_1$=Slowly-Variation → V2 | 1.45 | 100 |
| 6 | $q^e_c$ =Heavy-Continuous-Rainfall & $q^d_m$ =Medium-Daily-Rainfall & $\Delta h_2$=Slowly-Variation→ V2 | 1.45 | 100 |
| 7 | $q^e_c$ =Heavy-Continuous-Rainfall & $\Delta h_1$=Medium-Variation & $\Delta h^r_m$ =Slowly-Rise → V2 | 0.72 | 100 |
| 8 | $q_2$=Sharply-Cumulative-Rainfall & $\Delta h^r_m$ =Medium-Rise & $\Delta h_1$=Slowly-Variation → V2 | 0.72 | 100 |
| 9 | $q_2$=Sharply-Cumulative-Rainfall & $\Delta h^r_m$ =Medium-Rise & $\Delta h_2$=Slowly-Variation→ V2 | 0.72 | 100 |
| 10 | $q_2$=Sharply-Cumulative-Rainfall & $\Delta h_1$=Medium-Variation & $\Delta h^r_m$ =Slowly-Rise → V2 | 0.72 | 100 |

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 6** Association criteria of ZG88 monitoring point

| Rule ID | Rules | Support (%) | Confidence (%) |
|---|---|---|---|
| 1 | $\Delta h^r_m$ =Sharply-Rise & $\hbar$ =Medium-Water-Level & $q^e_c$ =Light-Continuous-Rainfall → V2 | 1.45 | 100 |
| 2 | $\Delta h^r_m$ =Sharply-Rise & $\Delta h_1$=Medium-Variation & $\hbar$ =Medium-Water-Level → V2 | 1.45 | 100 |
| 3 | $\Delta h^r_m$ =Sharply-Rise & $\Delta h_1$=Medium-Variation & $q^e_c$ =Moderate-Continuous-Rainfall → V2 | 0.72 | 100 |
| 4 | $\Delta h^r_m$ =Sharply-Rise & $\Delta h_1$=Medium-Variation & $q^d_m$ =Moderate-Daily-Rainfall → | 1.45 | 100 |

| | | | |
|---|---|---|---|
| | V2 | | |
| 5 | $\Delta h^r_m$ =Sharply-Rise & $\Delta h_1$=Sharply-Variation & $q^e_c$ =Light-Continuous-Rainfall → V2 | 0.72 | 100 |
| 6 | $\Delta h^d_m$ =Sharply-Drop & $\hbar$ =Medium-Water-Level & $q_2$=Light-Cumulative-Rainfall → V2 | 1.45 | 100 |
| 7 | $\Delta h^d_m$ =Sharply-Drop & $\Delta h^r_m$ =Sharply-Rise & $\Delta h_1$=Medium-Variation → V2 | 1.45 | 100 |
| 8 | $\Delta h^d_m$ =Sharply-Drop & $\Delta h^r_m$ =Sharply-Rise & $q^e_c$ =Light-Continuous-Rainfall → V2 | 1.45 | 100 |
| 9 | $\Delta h^d_m$ =Sharply-Drop & $\Delta h_1$=Medium-Variation & $q_2$=Light-Cumulative-Rainfall → V2 | 0.72 | 100 |
| 10 | $\Delta h_2$=Sharply-Variation & $\Delta h^d_m$ =Sharply-Drop & $q_2$=Light-Cumulative-Rainfall → V2 | 1.45 | 100 |

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 | $q^e_c$ =Heavy-Continuous-Rainfall & $q^d_m$ =Medium-Daily-Rainfall & $\Delta h^r_m$ =Sharply-Rise → V2 | 0.72 | 100 |
| 2 | $q_1$=Heavy-Cumulative-Rainfall & $q_2$=Moderate-Cumulative-Rainfall & $\Delta h^r_m$ =Sharply-Rise → V2 | 2.17 | 100 |
| 3 | $q_1$=Heavy-Cumulative-Rainfall & $q_2$=Moderate-Cumulative-Rainfall & $\Delta h_1$=Sharply-Variation → V2 | 1.45 | 100 |
| 4 | $q_1$=Heavy-Cumulative-Rainfall & $q_2$=Moderate-Cumulative-Rainfall & $q^e_c$ =Light-Continuous-Rainfall → V2 | 0.72 | 100 |
| 5 | $q_1$=Heavy-Cumulative-Rainfall & $q^d_m$ =Moderate-Daily-Rainfall & $q^e_c$ =Light-Continuous-Rainfall → V2 | 0.72 | 100 |
| 6 | $q_2$=Heavy-Cumulative-Rainfall & $q^d_m$ =Medium-Daily-Rainfall & $\Delta h_1$=Sharply-Variation → V2 | 0.72 | 100 |
| 7 | $q_2$=Heavy-Cumulative-Rainfall & $q^d_m$ =Medium-Daily-Rainfall & $\Delta h^r_m$ =Sharply-Rise → V2 | 0.72 | 100 |
| 8 | $q_2$=Heavy-Cumulative-Rainfall & $\Delta h^r_m$ =Medium-Rise & $\Delta h_1$=Slowly-Variation → V2 | 0.72 | 100 |
| 9 | $q_2$=Heavy-Cumulative-Rainfall & $\Delta h^r_m$ =Medium-Rise & $\Delta h_2$=Slowly-Variation → V2 | 0.72 | 100 |
| 10 | $q_2$=Heavy-Cumulative-Rainfall & $\Delta h_1$=Medium-Variation & $\Delta h^r_m$ =Slowly-Rise → V2 | 0.72 | 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 m 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).

**Table 8** Association criteria of SP-6 monitoring point

| Rule ID | Rules | Support (%) | Confidence (%) |
|---|---|---|---|

| | | | |
|---|---|---|---|
| 1 | $\Delta h^r_m$ =Sharply-Rise & $\Delta h^d_m$ =Sharply-Drop & $q^e_c$ =Light-Continuous-Rainfall → V2 | 1.45 | 100 |
| 2 | $\Delta h^r_m$ =Sharply-Rise & $\Delta h^d_m$ =Sharply-Drop & $\Delta h_1$=Medium-Variation → V2 | 1.45 | 100 |
| 3 | $\Delta h^r_m$ =Sharply-Rise & $\Delta h_1$=Sharply-Variation & $q^e_c$ =Light-Continuous-Rainfall → V2 | 0.72 | 100 |
| 4 | $\Delta h^r_m$ =Sharply-Rise & $\Delta h_1$=Medium-Variation & $q^e_c$ =Moderate-Continuous-Rainfall → V2 | 0.72 | 100 |
| 5 | $\Delta h^d_m$ =Sharply-Drop & $\Delta h_2$=Sharply-Variation→ & $q_2$=Light-Cumulative-Rainfall→ V2 | 1.45 | 100 |
| 6 | $\Delta h^d_m$ =Sharply-Drop & $\hbar$ =Medium-Water-Level & $q_2$=Light-Cumulative-Rainfall→ V2 | 1.45 | 100 |
| 7 | $\Delta h^d_m$ =Sharply-Drop & $\Delta h_1$=Medium-Variation & $q_2$=Light-Cumulative-Rainfall→ V2 | 0.72 | 100 |
| 8 | $\Delta h^r_m$ =Sharply-Rise & $\hbar$ =Medium-Water-Level & $q^e_c$ =Light-Continuous-Rainfall → V2 | 1.45 | 100 |
| 9 | $\Delta h^r_m$ =Sharply-Rise & $\Delta h_1$=Medium-Variation & $\hbar$ =Medium-Water-Level → V2 | 1.45 | 100 |
| 10 | $\Delta h^r_m$ =Sharply-Rise & $\Delta h_1$=Medium-Variation & $q^d_m$ =Moderate-Daily-Rainfall → V2 | 1.45 | 100 |

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

In the decision tree C5.0 model, boosting technology and combining cross validation were used to prevent over fitting of the model, as shown in Fig. 9, which contained 6 hydrological factors ($\hbar$, $\Delta h^r_m$, $\Delta h_2$, $q^e_c$, $q_1$, $q^d_m$). 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 ($q^d_m$ <95.8 mm).

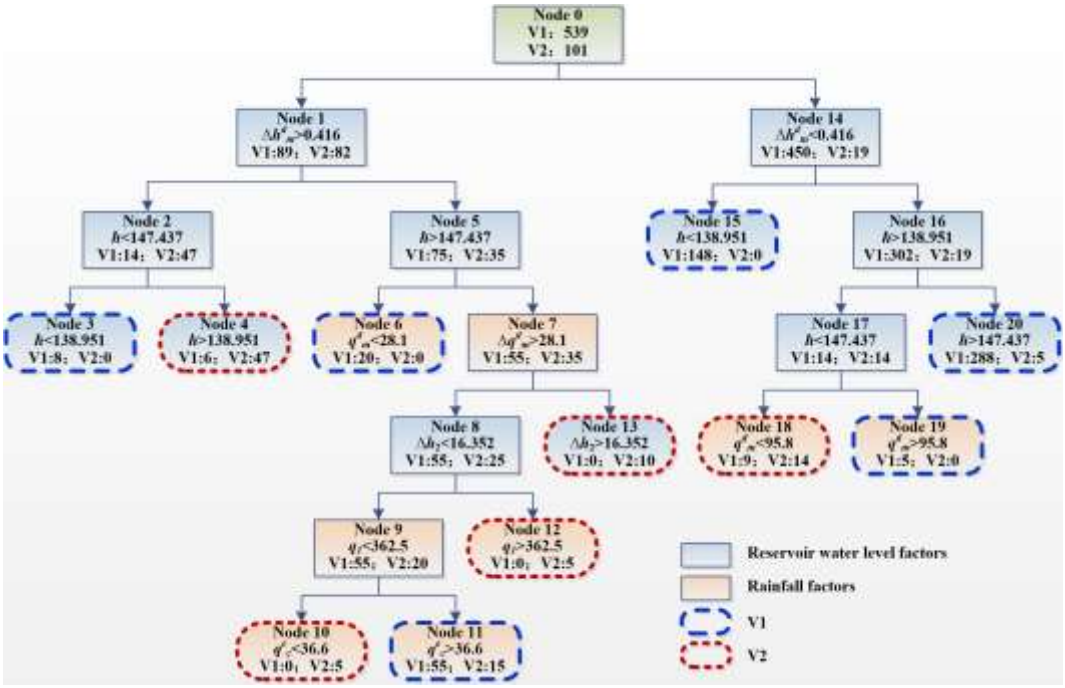


**Fig. 9** Decision Tree C5.0 model of ZG85, ZG86, ZG88, SP-2, SP-6 (accuracy: 94.53%)

**Table 9** Threshold values of the hydrological factors

| Criterion | Hydrological factors | | | | | | Mode | Count | Confidence |
|---|---|---|---|---|---|---|---|---|---|
| Criterion 1 | $\Delta h^d_m$>0.416 | $\hbar$>147.437 | $q^d_m$ >28.1 | $\Delta h_2$ <16.352 | $q_1$<362.5 | $q^e_c$<36.6 | V2 | 5 | 1.000 |
| Criterion 2 | | | | | $q_1$>362.5 | | V2 | 5 | 1.000 |
| Criterion 3 | | | | $\Delta h_2$>16.352 | | | V2 | 10 | 1.000 |
| Criterion 4 | | $\hbar$ <147.437 | $\hbar$ >138.951 | | | | V2 | 47 | 0.887 |
| Criterion 5 | $\Delta h^d_m$<0.416 | $\hbar$ >138.951 | $\hbar$ <147.437 | $q^d_m$ <95.8 | | | V2 | 14 | 0.609 |

## 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.

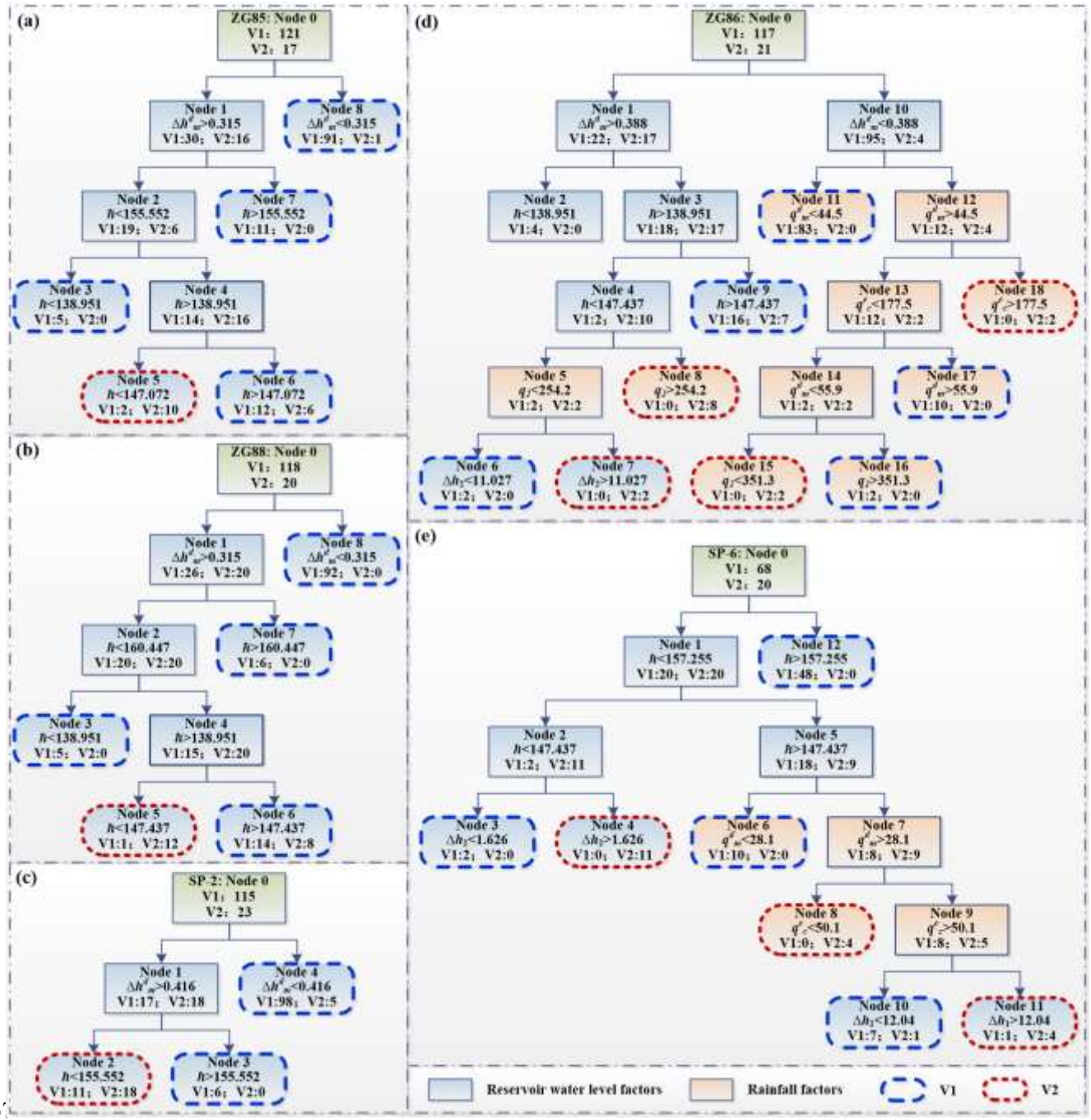


**Fig. 10** Decision Tree C5.0 models (a) ZG85 (accuracy: 93.48%); (b) ZG88 (accuracy: 93.48%); (c) SP-2 (accuracy: 88.41%); (d) ZG86 (accuracy: 94.93%); (e) SP-6 (accuracy: 97.73%)

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 $\Delta h^r_m$ and $\hbar$, and the contribution degrees of the two factors reached 0.58 and 0.42 respectively. When the reservoir water level was between 138.951 m 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 $\Delta h^r_m$ and $\hbar$, 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 ($\Delta h^d_m$, $q^d_m$, $\hbar$, $q^e_c$, $q_2$, $\Delta h_2$), 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 ($\Delta h^d_m$, $\hbar$, $\Delta h_2$), while the total contribution of the three rainfall related factors was 0.46 ($q^d_m$, $q^e_c$, $q_2$). When the reservoir water level was 138.951~147.437 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.

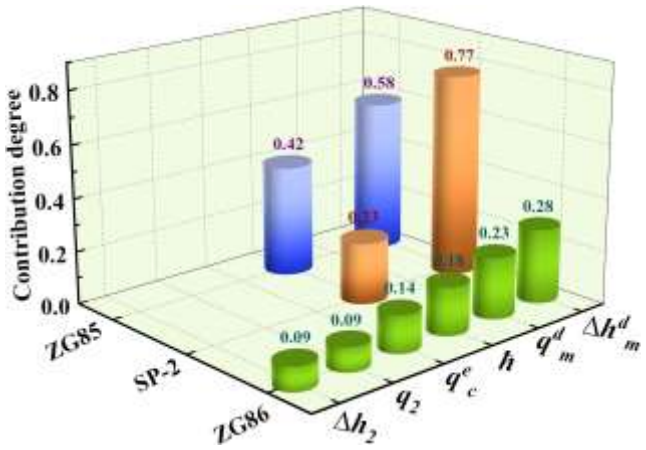


**Fig. 11** Importance degree of each governing factor

In this paper, when establishing the decision tree model, nine influencing factors related to reservoir water level and rainfall were selected ($\hbar$, $\Delta h^d_m$, $\Delta h^r_m$, $\Delta h_1$, $\Delta h_2$, $q^e_c$, $q_1$, $q_2$, $q^d_m$). 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.

## Acknowledgments

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 (No.ZRZY2020KJ12). The authors thank the colleagues in our laboratory for their constructive comments and assistance.

## Conflict of interest

The authors declared that they have no conflicts of interest to this work.

## 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 (p 207–216)

Althuwaynee, 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. Applied Sciences, 11(2).

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.

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. The Science of the Total Environment, 644(dec.10), 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-Science Reviews, 227-252.

Guo, W., Zuo, X., Yu, J., et al. (2019). Method for mid-long-term prediction of landslides movements based on optimized apriori algorithm. Applied Sciences, 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(Apr.15), 105-118.

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. Reliability Engineering and System Safety, 205.

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 Geosciences, 13(1), 1226-1244.

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

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

Miao, F., Wu, Y., Li, L., et al. (2020). Triggering factors and threshold analysis of baishuihe landslide based on the data mining methods. Natural Hazards, 1-20.

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.

Milevski, I., Slavoljub D., Zorn, M.. (2019). Statistical and expert-based landslide susceptibility modeling on a national scale applied to north macedonia. Open Geosciences, 11.

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

Nicu, I., Asndulesei, A.. (2018). Gis-based evaluation of diagnostic areas in landslide susceptibility analysis of bahluie river basin (moldavian plateau, ne romania). are neolithic sites in danger?. Geomorphology, S0169555X18301715.

Pandya, R., Pandya, J. (2015). C5.0 algorithm to improved decision tree with feature selection and reduced error pruning. International Journal of Computer Applications, 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. Climate and Development, 1-16.

Petley, D., Hearn, G., Hart, A., et al. (2007). Trends in landslide occurrence in Nepal. Nat Hazards, 42: 23– 532 44.DOI: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 – 539 236.DOI: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 Transactions on Geoscience and Remote Sensing, PP (99), 1-19.

Stumpf, A., Kerle, N.. (2011). Object-oriented mapping of landslides using random forests. Remote Sensing of Environment, 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. Engineering Geology, 295, 106411.

Tan, F., Hu, X., He, C., et al. (2017). Identifying the main control factors for different deformation stages of landslide. Geotechnical & Geological Engineering (2), 1-14.

Tang, HM., Wasowski., Juang, CH (2019). Geohazards in the three gorges reservoir area, china – lessons learned from decades of research. Engineering Geology, 261, 105267-105267.

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

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. Bulletin of Engineering Geology and the Environment, 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. Environmental Earth Sciences, 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. Natural Hazards (1), 1-31.

Zhang, Y., Tang, J., Liao, R., et al. (2020). Application of an enhanced bp neural network model with water cycle algorithm on landslide prediction. Stochastic Environmental Research and Risk Assessment (61–65), 1-19.
Zhang, Y., Zhang, Z., Xue, S., et al. (2020b). Stability analysis of a typical landslide mass in the Three Gorges Reservoir under varying reservoir water levels. Environmental Earth Sciences, 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. Applied Sciences, 10(1), 16.
Zhou, C., Yin K., Cao, Y., et al. (2016). Application of time series analysis and pso–svm model in predicting the bazimen landslide in the three gorges reservoir, china. Engineering Geology, 204, 108-120.