

## Article

# A New Approach for Discontinuity Extraction Based on an Improved Naive Bayes Classifier

Guangyin Lu <sup>1,2,3</sup>, Xudong Zhu <sup>1,2,3</sup>, Bei Cao <sup>1,2,3,\*</sup>, Yani Li <sup>1,2,3</sup>, Chuanyi Tao <sup>1,2,3</sup> and Zicheng Yang <sup>1,2,3</sup>

<sup>1</sup> School of Geosciences and Info-Physics, Central South University, Changsha 410083, China; luguangyin@csu.edu.cn (G.L.); zhuxudong@csu.edu.cn (X.Z.); lyn219@csu.edu.cn (Y.L.); 215011082@csu.edu.cn (C.T.); 235012145@csu.edu.cn (Z.Y.)

<sup>2</sup> National Engineering Laboratory for High Speed Railway Construction, Changsha 410075, China

<sup>3</sup> Hunan Key Laboratory of Nonferrous Resources and Geological Hazards Exploration, Changsha 410083, China

\* Correspondence: beicao@csu.edu.cn

**Abstract:** An increasing number of methods are being used to extract rock discontinuities from 3D point cloud data of rock surfaces. In this paper, a new method for automatic extraction of rock discontinuity based on an improved Naive Bayes classifier is proposed. The method first uses principal component analysis to find the normal vectors of the points, and then generates a certain number of random point sets around the selected training points for training the classifier. The trained, improved Naive Bayes classifier is based on point normal vectors and is able to automatically remove noise points due to various reasons in conjunction with the knee point algorithm, realizing high-precision extraction of the discontinuity sets. Subsequently, the individual discontinuities are segmented using a hierarchical density-based spatial clustering method with noise application. Finally, the PCA algorithm is used to complete the orientation by plane fitting the individual discontinuities. The method was applied in two cases, Kingston and Colorado, and the reliability and advantages of the new method were verified by comparing the results with those of previous research, and the discussion and analysis determined the optimal values of the relevant parameters in the algorithm.

**Citation:** Lu, G.; Zhu, X.; Cao, B.; Li, Y.; Tao, C.; Yang, Z. A New Approach for Discontinuity Extraction Based on an Improved Naive Bayes Classifier. *Appl. Sci.* **2024**, *14*, 2050. <https://doi.org/10.3390/app14052050>

Academic Editor: Nikolaos Koukouras

Received: 31 January 2024

Revised: 22 February 2024

Accepted: 28 February 2024

Published: 29 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** rock mass; point cloud; discontinuity; automatic extraction; machine learning; naive Bayes classifier

## 1. Introduction

The presence of rock mass discontinuities, such as bedding planes, joints, faults, and other planes, plays a crucial role in the field of rock mechanics [1], altering the mechanical properties of rock masses, and consequently influencing the stability of rocky slopes [2,3]. The International Society for Rock Mechanics [4] proposed a set of quantitative parameters for characterizing rock discontinuities, including orientation, spacing, persistence, roughness, aperture, filling, wall strength, seepage, number of sets, and block size. Among these parameters, orientation, as a critical factor influencing the deformation of rock masses, holds extraordinary significance in the analysis of rock mass stability [5–9]. The traditional approach necessitates manual measurements of discontinuity orientation using a compass and inclinometer in accessible regions [10]. This technique is characterized by notable labor intensity and time consumption. Meanwhile, its application in regions featuring substantial topographical variations not only jeopardizes the safety of surveyors but also lacks assurance in ensuring the precision of data [11].

With the advancement of remote sensing technology, non-contact measurement methods such as photogrammetry [12] and LiDAR [13] have been introduced, facilitating rapid scanning of the rock surfaces within the study area. This generates high-precision point clouds composed of numerous three-dimensional coordinates, overcoming the

limitations associated with traditional methods [14,15]. Currently, a growing number of scholars are actively involved in extracting discontinuities from point clouds to gather comprehensive parameter information. The extraction process essentially involves segmenting points belonging to distinct discontinuities within the point cloud. The predominant methods for extracting discontinuities from point clouds can be classified into two distinct perspectives. Considering that rock mass discontinuities often exhibit a planar feature, one method involves linear or nonlinear fitting through geometric analysis and mathematical models to achieve plane extraction. Some researchers have employed plane detection algorithms, such as Random Sample Consensus (RANSAC) [16–19], 3D Hough Transform [20,21], and Principal Component Analysis (PCA) [22,23], to extract points belonging to the same plane within a point cloud, treating them as individual discontinuities. The Region Growing algorithm [24–26] relies on normal vectors and curvature to identify the nearest neighbors that share the same plane with the seed point, treating them as growth points, and carries out plane expansion to achieve the extraction of discontinuities. However, the method is very time-consuming. In the modified Region Growing algorithm proposed by Ge et al. [25], the homogeneity criterion is modified to increase the probability of potential growth points to become growth points, which improves the computational efficiency. By applying the Region Growing algorithm to the voxelized point cloud, some scholars [27,28] have greatly improved the efficiency of the extraction of the discontinuity. Aiming at the problems caused by the threshold selection of the traditional Region Growing algorithm, Yi et al. [26] improved on it and proposed a multi-rule Region Growing algorithm with higher accuracy. Singh et al. [29] found that the vertical and horizontal angle variations of points belonging to the same discontinuity set in a certain area appear as a sinusoidal signal, and proposed a new method to realize discontinuity set extraction by using the amplitude and phase of this signal. While such methods typically demonstrate effectiveness in simple scenarios, their application becomes time-consuming and vulnerable to interference from factors like noise when confronted with large and complex point clouds.

Given that points located in the same discontinuity in a point cloud often share similar characteristics, mainly including 3D coordinates, normal vectors, curvature, and color, another approach involves employing machine learning (ML) techniques for classification based on extracted features. For point cloud feature extraction, 3D coordinates and colors are directly available through laser scanning and photogrammetry [30]. For normal vectors and curvature, some studies have utilized methods such as triangulation [15,18,31], cube search [8], and other methods for extraction, but compared to the utilization of 3D methods to extract from a point set composed of points and neighboring points [32,33], the former methods may lose a certain amount of point cloud information. Some scholars have utilized normal vectors as features and applied the K-means algorithm [18,34] to cluster the point cloud, followed by the extraction of discontinuities from the clustered result. The Fuzzy C-means (FCM) algorithm, an improvement of K-means, calculates the probability that a sample point belongs to each cluster center, subsequently determining the category of sample points to achieve the automatic identification of discontinuities, and so exhibits reduced sensitivity to noise and superior performance in terms of classification results [35,36]. Nevertheless, both of the aforementioned methods necessitate a pre-determination of the cluster count and exhibit sensitivity to the initial selection of cluster centers. Variations in the choice of initial cluster centers frequently lead to discrepancies in the classification results. At this point, some research has introduced parameters such as Minimum Description Length (MDL) [37], and clustering validity indices [18,35], to provide references for determining the number of clusters. Optimization algorithms, including Particle Swarm Optimization (PSO) [38], Differential Evolution (DE) [39], and the Firefly Algorithm (FA) [40], have been introduced to optimize the update of cluster centers. Xu et al. [41] proposed a new fast fuzzy clustering (FFC) method by combining the FCM algorithm with peak density clustering (CFSFDP), overcoming the respective defects of the two algorithms. Riquelme et al. [32] integrated Density-Based Spatial Clustering of

Applications with Noise (DBSCAN), an unsupervised machine learning clustering algorithm based on density that is adept at handling irregular shapes, with Kernel Density Estimation (KDE) for the extraction of discontinuities from point clouds. Singh et al. [29] proposed five features for describing the points after filtering the noisy points in the point cloud using various methods, combined with K-Medoids and DBSCAN to complete the extraction of the discontinuity, which demonstrated strong adaptability in complex scenes. Chen et al. [42] proposed an adaptive DBSCAN (ADBSCAN) algorithm to address the challenge of determining DBSCAN parameters, which further improves the automation of discontinuity extraction. Distinct from unsupervised machine learning algorithms, Ge et al. [43] employed a well-trained artificial neural network model, trained by the manual selection of an ample training dataset, to classify point clouds, greatly enhancing the efficiency of discontinuity extraction. However, achieving a well-trained network involves repeated manual sample selection until the accuracy is satisfied, thereby imposing significant constraints on the level of automation and computational efficiency. To address the aforementioned issues, this paper proposes a semi-automated approach that efficiently generates a training sample set in a simple manner.

The Naive Bayes classifier [44], a supervised machine learning algorithm, is frequently utilized in document classification [45], medical diagnosis [46], slope stability prediction [47], and various other fields. However, there has been limited exploration of this algorithm in the field of rock mass discontinuity identification. This paper proposes a new method for discontinuity recognition based on an improved Naive Bayes classifier and specifically designed to address the needs of point cloud classification tasks. The structure of the article is as follows: Section 2 introduces the steps and principles of the method proposed in this paper through a case study in Colorado; in Section 3, the reliability of the method is validated through its application to the case study B, and a discussion on the selection of relevant parameters is presented; Section 4 compares the proposed method with previous approaches; Section 5 provides the conclusion.

## 2. Methodology

The methodology proposed in this study encompasses the following key steps, as delineated in Figure 1.

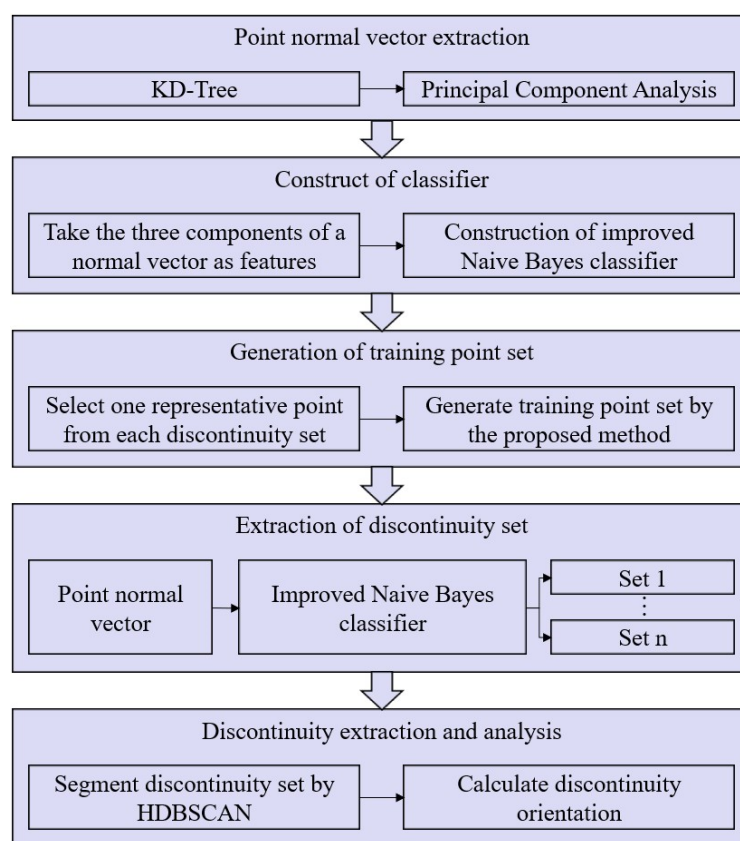
Step 1: Computation of point normal vector. The nearest neighbors of each point in the point cloud are quickly found by building a KD-Tree, and the normal vectors of the fitting plane for these points, obtained through Principal Component Analysis, will be used as the normal vectors of the points.

Step 2: Classifier construction. Three components of the point normal vector are used as features to construct an improved Naive Bayes classifier suitable for point cloud classification.

Step 3: Generation of training point sets. Based on the selection of one point per discontinuity set, the method proposed in this study is employed to generate training point sets.

Step 4: Extraction of discontinuity sets. The extraction of the discontinuity set is achieved by classifying the entire point cloud based on its normal vector using a trained, improved Naive Bayes classifier.

Step 5: Extraction and analysis of individual discontinuities. The Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm is utilized to extract individual discontinuities from the discontinuity sets. Principal component analysis for plane fitting is employed to calculate the dip direction and dip of individual discontinuities.



**Figure 1.** Flow chart of method.

### 2.1. Case Study A for Validation

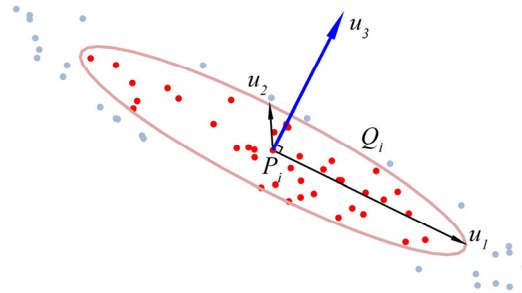
The proposed method in this paper is validated using data from a rock slope in Ouray, CO, USA (Figure 2). The dataset, obtained by Dr. John Kemeny in 2004 using the Optech ILRIS-3D laser scanner, consists of 1,024,521 points with a spacing of less than 2 cm. The data are publicly accessible in the Rockbench database [48]. Many scholars have extensively researched the use of this dataset for the identification of rock mass discontinuities from point clouds [49], providing a convenient basis for comparisons to validate the methods proposed in this study.



**Figure 2.** Actual scene of case study A in Ouray, CO, USA.

## 2.2. Extraction of Point Normal Vector

In this section, a KD-Tree is constructed for the rapid retrieval of the k-nearest neighbors for each point  $P_i$  within the three-dimensional point cloud, denoted as the point set  $Q_i$ . The normal vector of  $P_i$  is obtained through the PCA algorithm applied to the fitted plane of  $Q_i$  [50]. In the process of conducting Principal Component Analysis on  $Q_i$ , it is essential to initially calculate the covariance matrix. Subsequently, eigenvalue decomposition is applied to acquire both the eigenvalues and eigenvectors. The eigenvector corresponding to the smallest eigenvalue signifies the direction with the minimum variance and is commonly selected as the normal vector for  $P_i$  (Figure 3).

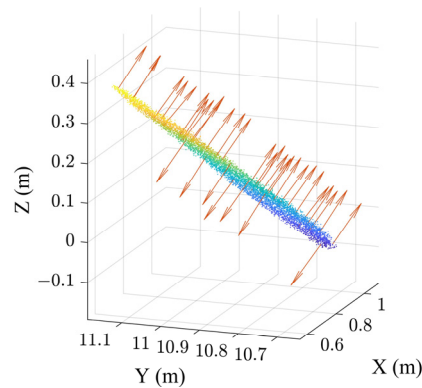


**Figure 3.** Schematic of point normal vector computation.  $u_1$ ,  $u_2$ , and  $u_3$  represent the three eigenvectors of the covariance matrix of  $Q_i$ , respectively.  $u_3$  is the eigenvector corresponding to the minimum eigenvalue. The red dots represent nearest neighbors, and the gray dots represent non-nearest neighbors.

Figure 4 demonstrates that the rough and uneven nature of the rock mass surface (caused by vegetation, among other reasons) may lead to opposite directions of the computed normal vectors for points belonging to the same discontinuity, potentially impacting the accuracy of point cloud classification. To address this issue, Equation (1) is employed to adjust the normal vector direction for each point:

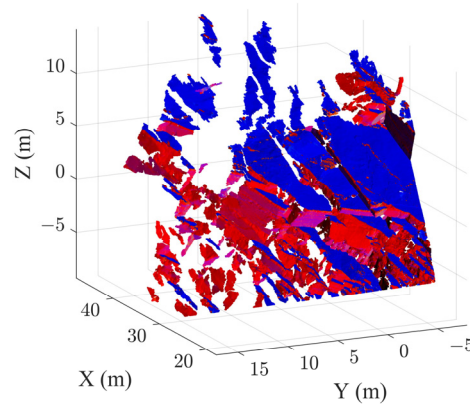
$$d = \frac{\vec{u} \times \vec{v}}{|\vec{v}|} \quad (1)$$

where,  $\vec{v}$  represents the selected direction, and  $d$  is the projection of the normal vector  $\vec{u}$  onto  $\vec{v}$ . The direction of the normal vector  $\vec{u}$  remains unchanged when  $d \geq 0$ , while the direction of the normal vector  $\vec{u}$  is reversed when  $d < 0$ .



**Figure 4.** Varying normal vectors for points on the same discontinuity.

Figure 5 depicts the color mapping of the point normal vector in the RGB color system applied to the point cloud. It can be seen that points on the same discontinuity set have similar colors.



**Figure 5.** Point cloud data of case A: assigning RGB colors based on normal vectors. One color per discontinuity set.

### 2.3. Extraction of Discontinuity Sets

#### 2.3.1. Construction of the Improved Naive Bayes Classifier

The Naive Bayes classifier, a widely employed supervised learning algorithm, is based on the fundamental Bayes' theorem, expressed as Equation (2) [51]:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (2)$$

where  $P(A)$  and  $P(B)$  represent the probability of events  $A$  and  $B$  occurring independently, respectively, with  $P(A) > 0$  and  $P(B) > 0$ .  $P(B|A)$  signifies the probability of event  $B$  occurring given the occurrence of event  $A$ , and  $P(A|B)$  denotes the probability of event  $A$  occurring given the occurrence of event  $B$ .

For each sample with  $d$  features  $X = \{x_1, x_2, \dots, x_d\}$  and all discontinuity set categories  $y_1, y_2, \dots, y_m$ , the Naive Bayes classifier is under the assumption that the  $d$  features are mutually independent. First of all, the probability  $P(y_j|X)$  ( $j = 1, \dots, m$ ) that the point to be classified belongs to category  $y_j$  is calculated by Naive Bayes theorem:

$$P(y_j|X) = \frac{P(y_j)P(X|y_j)}{P(X)} \quad (3)$$

where  $P(y_j)$  is the proportion of training samples belonging to category  $y_j$  in the training samples. The value of  $P(X)$  independent of the category is the evidence factor used for normalization, which can be obtained by the calculation of Equation (4):

$$P(X) = \sum_{j=1}^m P(y_j)P(X|y_j) \quad (4)$$

Due to the features of the sample  $X$  to be categorized being independent of each other,  $P(X|y_j)$  can be expressed as Equation (5):

$$P(X|y_j) = \prod_{i=1}^d P(x_i|y_j) \quad (5)$$

Then, substitute Equation (5) into Equation (3):

$$P(y_j|X) = \frac{P(y_j)}{P(X)} \prod_{i=1}^d P(x_i|y_j) \quad (6)$$

Considering that  $P(X)$  is independent of the category and its value will not affect the classification result, this paper introduces  $P_j$  to denote the probability that the sample  $X$  belongs to each category:

$$P_j = P(y_j) \prod_{i=1}^d P(x_i | y_j) \quad (7)$$

After calculating the probability that the sample  $X$  belongs to each category using Equation (7), the category with the highest probability is selected as the classification result:

$$h_{nb}(X) = \arg \max P_j = \arg \max P(y_j) \prod_{i=1}^d P(x_i | y_j) \quad (8)$$

The calculation method of  $P(x_i | y_j)$  in Equation (8) depends on the distribution of the features of the samples. In this paper, the components of the normal vectors of the points to be classified in the direction of the  $x$ ,  $y$ ,  $z$  coordinate axes are used as the three features used for classification. Due to its numerical changes being continuous and approximately normally distributed,  $P(x_i | y_j)$  in Equation (8) needs to be estimated using the following equation:

$$P(x_i | y_j) = \frac{1}{\sqrt{2\pi}\sigma_{y_j, x_i}} \exp \left( -\frac{(x_i - \mu_{y_j, x_i})^2}{2\sigma_{y_j, x_i}^2} \right) \quad (9)$$

where  $\mu_{y_j, x_i}$ ,  $\sigma_{y_j, x_i}$ , and  $\sigma_{y_j, x_i}^2$  is the mean, standard deviation, and variance of feature  $x_i$  of the samples belonging to the  $y_j$  in the training sample, respectively.

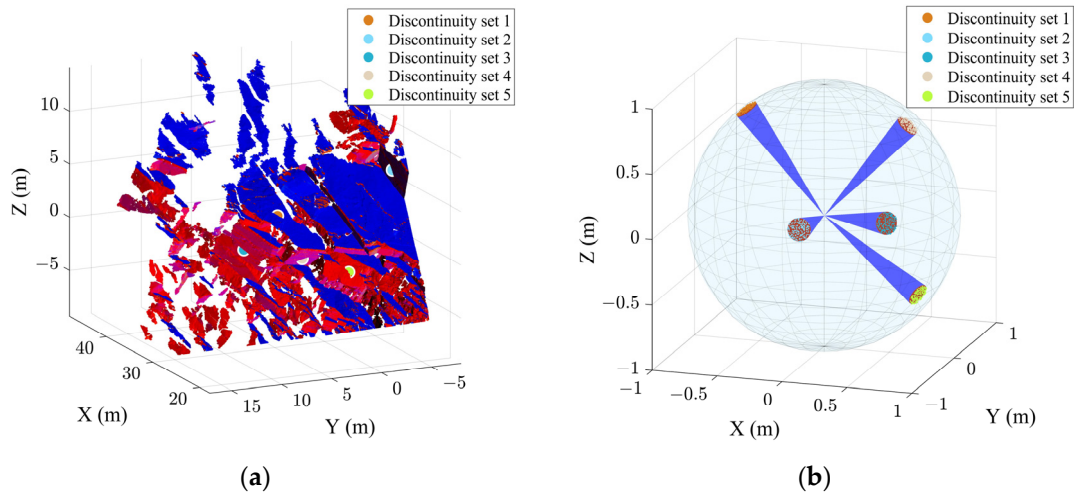
The above Naive Bayes classifier takes the category with the highest probability of belonging as the classification result. However, in the point cloud classification process, there are usually some noise points with a very small probability of belonging to any category that need to be removed. In this paper, the Naive Bayes classifier is improved on the basis of the distribution curve of logarithmic probability (LPR), i.e.,  $\ln(P_i)$ , arranged from small to large (logarithmic transformation prevents the multiplication of numbers less than 1 from trending toward zero, and the trend of the logarithmic probability distribution curve is more obvious, and the logarithmic probability is not subject to the restriction of 0–1 which can be an arbitrary positive or negative number). Then, the logarithmic probability threshold value is selected according to the distribution trend. The points with logarithmic probability less than this threshold are removed as noise points, which improves the performance of the Naive Bayes classifier in point cloud classification.

### 2.3.2. Generation of Training Set

To train the Naive Bayes classifier, a substantial number of training points are needed. Manual selection is not only time-consuming and labor-intensive but also has the potential to introduce noise, affecting the accuracy of point cloud classification. In response, this paper proposes a straightforward and efficient method for generating training point sets as follows:

Firstly, the number of discontinuity sets is determined according to the color of the point cloud (Figure 5). Subsequently, one point from each discontinuity set is selected to use its normal vector as the center, and  $N$  normal vectors are randomly generated as the training set in the range where the angle with it is less than the maximum angle  $maxAGL$ . Figure 6b shows the training set generated around the normal vectors of the points selected in Figure 6a with  $maxAGL = 5^\circ$  and  $N = 200$ .





**Figure 6.** (a) Training points selection; (b) training sets generation.

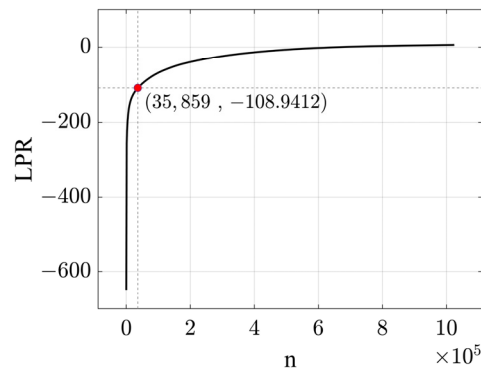
### 2.3.3. Extraction of Discontinuity Sets Based on the Improved Naive Bayes Classifier

The procedure for extracting discontinuity sets using the improved Naive Bayes classifier is outlined as follows:

Step 1: The training set is selected using the method proposed in Section 2.3.2 and the percentage  $P(y_j)$  of training points belonging to each category within the training sets is calculated.

Step 2: Based on the training sets, Equation (9) is utilized to calculate the conditional probability  $P(x_i|y_j)$  that the  $i$ th feature of the point  $X$  takes the value of  $x_i$ , provided that it belongs to each discontinuity set  $y_j$ .

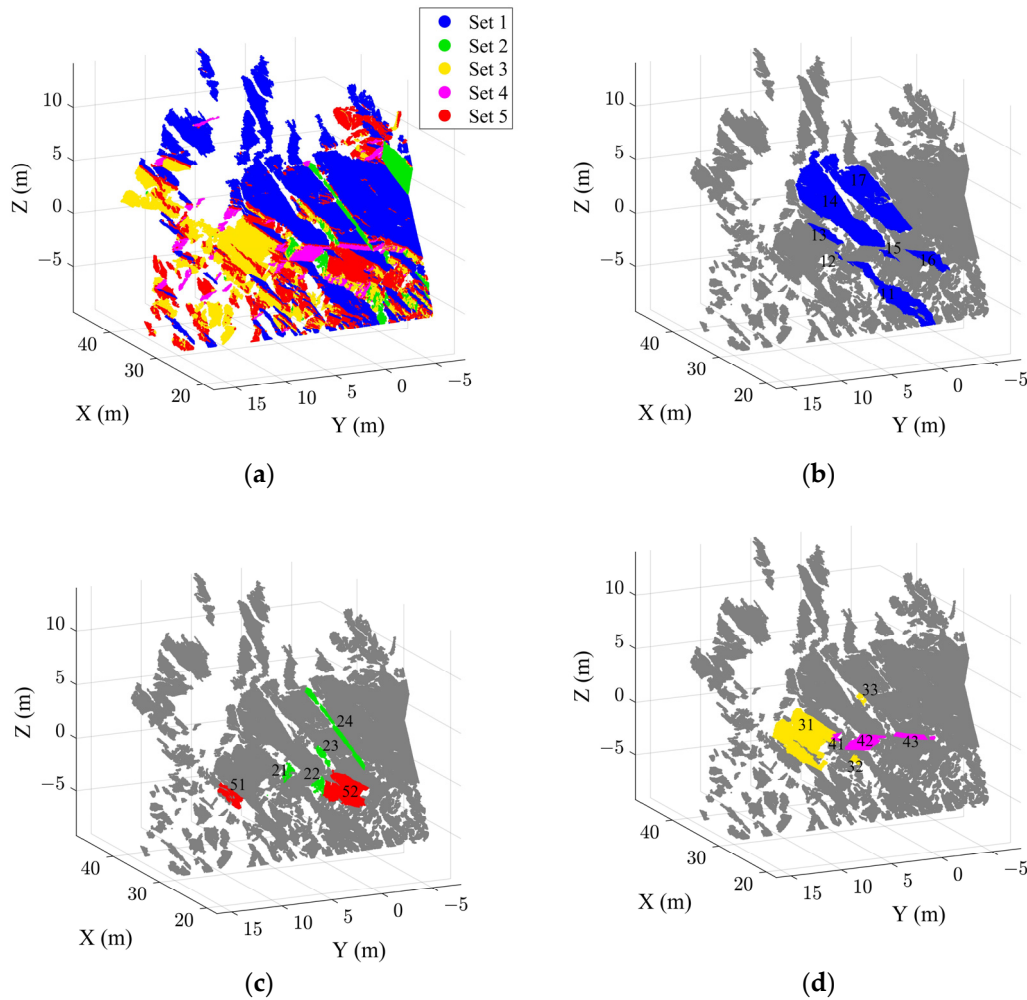
Step 3: The probability  $P_j$  that the sample  $X$  belongs to each category is calculated by substituting  $P(y_j)$  and  $P(x_i|y_j)$  derived from Step 1 and Step 2 into Equation (7). The distribution curve of logarithmic probability is plotted (Figure 7), from which it can be seen that the logarithmic probability of the points after the curve inflection point is higher and tends to be level. So, the logarithmic probability of the points at the curve inflection point is selected as the minimum logarithmic probability ( $minLPR$ ), and the points with logarithmic probability less than  $minLPR$  are removed as noise points. Substituting the probability  $P_j$  of the remaining points into Equation (8) yields  $h_{nb}(X)$  to complete the classification of the entire point cloud. At this point the procedure for extracting the discontinuity set using the improved Naive Bayes classifier terminates, where the points with the same category are those belonging to the same discontinuity set. So, the high accuracy extraction of the discontinuity set was achieved using the improved Naive Bayes classifier.



**Figure 7.** Logarithmic probability distribution curve of point classification. LPR: Logarithmic probability of point belonging to a category; n: point index.



The extraction results of discontinuity sets are shown in Figure 8.



**Figure 8.** Results of discontinuity sets extraction. (a) Assigned colors to each discontinuity set; (b–d) selected discontinuities for parameter calculations.

#### 2.4. Extraction of Individual Discontinuity Using the HDBSCAN Algorithm

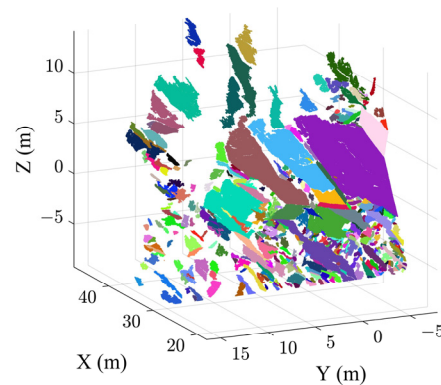
In this section, the extraction of individual discontinuity within each discontinuity set is accomplished through the utilization of clustering algorithms. Considering the inability to determine the number of individual discontinuities within each discontinuity set and the requirement of a predetermined number of clusters for K-means or FCM, it is deemed inapplicable. Currently, DBSCAN, a prominent density clustering algorithm, is widely used to extract individual discontinuities from discontinuity sets, necessitating input parameters: search radius ( $\epsilon$ ) and minimum number of points ( $m_{pts}$ ) to establish a density threshold for distinguishing discrete points. Additionally, variations in dataset density can affect the stability of this algorithm. Campello et al. [52] proposed the HDBSCAN algorithm, which combines hierarchical clustering with DBSCAN, providing enhanced stability and better clustering results for data with uneven density. This algorithm relies on a single key parameter, the minimum cluster size ( $minCluster$ ), and its performance is not sensitive to the specific value chosen for this parameter. The algorithm defines the mutual reachability distance between two points as follows:

$$d_{reach-k}(x_p, x_q) = \max\{core_k(x_p), core_k(x_q), d(x_p, x_q)\} \quad (10)$$

where  $core_k(x_p)$  and  $core_k(x_q)$  denote the distances from points  $x_p$  and  $x_q$  to their nearest  $k$  neighbors, while  $d(x_p, x_q)$  represents the distance between the points  $x_p$  and  $x_q$ . The

minimum spanning tree is generated based on the mutual reachability distance, and a hierarchical structure is constructed. Subsequently, the hierarchical structure is recursively split, removing noise clusters with sizes smaller than *minCluster*. This process accomplishes the compression of the hierarchical structure and the extraction of clusters.

In addition, the parameter *minDis* is used to eliminate extremely small clusters in the extraction results as noise, thereby further enhancing the accuracy of discontinuity extraction. For case A, after testing, parameters *minCluster* = 5 and *minDis* = 200 were chosen to successfully extract individual discontinuities (Figure 9):



**Figure 9.** Extraction results of individual discontinuity. One color per individual discontinuity.

## 2.5. Calculation of Rock Discontinuity Parameters

The PCA algorithm is employed for plane fitting on the extracted discontinuities to determine their normal vector  $\vec{u}$ . The components  $u_x$ ,  $u_y$ , and  $u_z$  of  $\vec{u}$  are then used in Equations (11) and (12) to calculate the dip direction ( $\alpha$ ) and dip ( $\beta$ ) of the discontinuity.

$$\alpha = \begin{cases} 90^\circ - \tan^{-1} \frac{u_y}{u_x} & u_x > 0 \\ 270^\circ - \tan^{-1} \frac{u_y}{u_x} & u_x < 0 \end{cases} \quad (11)$$

$$\beta = \cos^{-1} \left| \frac{u_z}{\sqrt{u_x^2 + u_y^2 + u_z^2}} \right| \quad (12)$$

In total, 19 discontinuities previously studied by Riquelme et al. [32] and Chen et al. [18], as shown in Figure 8b–d, were selected for orientation calculation to validate the effectiveness of the proposed approach in this paper. Table 1 presents the dip direction and dip results of the 19 discontinuities extracted using the approach proposed in this study, along with the methods introduced by Riquelme et al. [32] and Chen et al. [18]. It is observed that 94.7% of the discontinuities exhibit orientation differences within  $7^\circ$ , with an average difference remaining within  $2^\circ$ . The maximum deviation in the dip direction appears on Discontinuity 12, which is attributed to surface undulations.

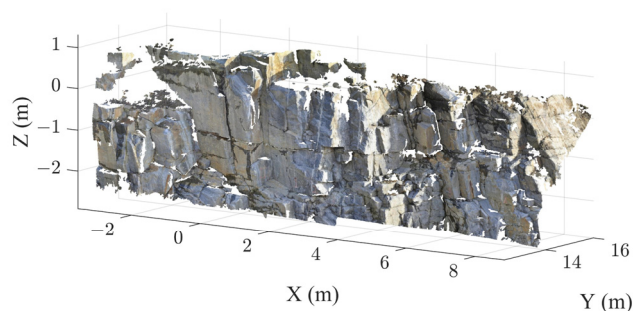
**Table 1.** Comparison of orientation results from different methods.

| Discontinuity ID  | Riquelme et al.[32] (°) | Chen et al.[18] (°) | New Method (°) | Riquelme et al.[32] (°) |              | Chen et al.[18] (°) |              |
|-------------------|-------------------------|---------------------|----------------|-------------------------|--------------|---------------------|--------------|
|                   |                         |                     |                | $\Delta DD $            | $\Delta DA $ | $\Delta DD $        | $\Delta DA $ |
| 11                | 246.24/39.02            | 244.62/38.38        | 246.25/39.03   | 0.01                    | 0.01         | 1.63                | 0.65         |
| 12                | 256.86/52.3             | 256.18/52.16        | 247.13/49.76   | 9.73                    | 2.54         | 9.05                | 2.4          |
| 13                | 70.26/35.8              | 251.04/36.17        | 250.23/35.79   | 0.03                    | 0.01         | 0.81                | 0.38         |
| 14                | 252.68/35.48            | 251.44/33.85        | 252.63/35.47   | 0.05                    | 0.01         | 1.19                | 1.62         |
| 15                | 249.74/35.91            | 250.82/36.83        | 249.33/35.55   | 0.41                    | 0.36         | 1.49                | 1.28         |
| 16                | 70.47/35.91             | 250.46/35.86        | 250.19/35.71   | 0.28                    | 0.2          | 0.27                | 0.15         |
| 17                | 255.12/32.82            | 253.19/33.46        | 254.90/30.87   | 0.22                    | 1.95         | 1.71                | 2.59         |
| 21                | 339.47/83.25            | 157.55/83.81        | 339.38/82.79   | 0.09                    | 0.46         | 1.83                | 1.02         |
| 22                | 166.33/76.58            | 166.31/78.73        | 166.64/77.31   | 0.31                    | 0.73         | 0.33                | 1.42         |
| 23                | 160.2/89.86             | 157.52/86.88        | 159.43/89.14   | 0.77                    | 0.72         | 1.91                | 2.26         |
| 24                | 173.55/76.85            | 353.07/77.82        | 173.77/77.83   | 0.22                    | 0.98         | 0.7                 | 0.01         |
| 31                | 136.59/82.58            | 314.73/80.04        | 135.33/81.71   | 1.26                    | 0.87         | 0.6                 | 1.67         |
| 32                | 131.23/82.67            | 136.52/89.85        | 137.43/87.91   | 6.2                     | 5.24         | 0.91                | 1.94         |
| 33                | 143.91/89.7             | 145.62/89.85        | 323.94/89.89   | 0.03                    | 0.19         | 1.68                | 0.04         |
| 41                | 97.55/63.22             | 285.98/59.84        | 99.33/62.05    | 1.78                    | 1.17         | 6.65                | 2.21         |
| 42                | 91.07/50.19             | 272.57/47.64        | 272.75/48.46   | 1.68                    | 1.73         | 0.18                | 0.82         |
| 43                | 96.64/47.97             | 277.31/49.31        | 96.72/47.79    | 0.08                    | 0.18         | 0.59                | 1.52         |
| 51                | 123.42/76.15            | 305.04/77.62        | 122.67/75.92   | 0.75                    | 0.23         | 2.37                | 1.7          |
| 52                | 105.75/69.94            | 109.29/76.61        | 106.65/69.93   | 0.9                     | 0.01         | 2.64                | 6.68         |
| Maximum deviation |                         |                     |                | 9.73                    | 5.24         | 9.05                | 6.68         |
| Average deviation |                         |                     |                | 1.31                    | 0.93         | 1.92                | 1.60         |

Note:  $\Delta|DD|$  —absolute value of dip direction difference;  $\Delta|DA|$  —absolute value of dip difference.

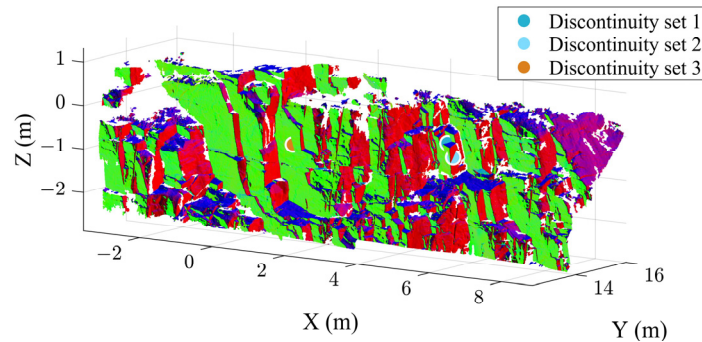
### 3. Case Study B

In this section, the practicality of the proposed method is validated using a point cloud case publicly available in the Rockbench repository [48] (Figure 10). The case is located along Highway 15 near Kingston, ON, Canada, and was acquired using a Leica HDS6000 laser scanner. It contains 2,167,515 data points along with their color information.

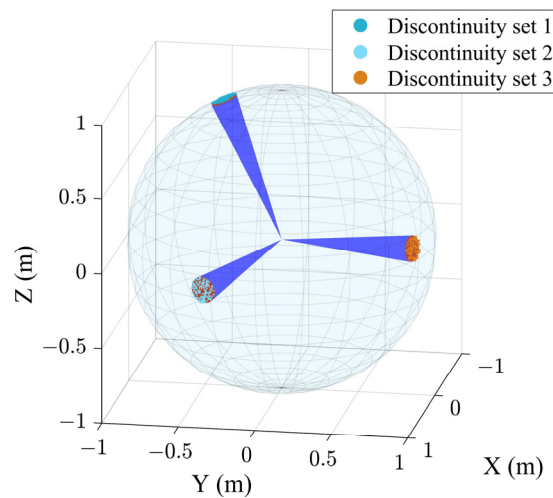
**Figure 10.** The point cloud data of case B.

### 3.1. Extraction Results of the Case B

According to the study by Ge et al. [43], in this section, the number of nearest neighbors is set to  $K = 40$  to calculate the point cloud normal vectors using the PCA algorithm. The colors assigned based on the normal vectors in Figure 11 mainly represent three sets of discontinuities in the rock mass (the artificial plane in the upper-right corner). Selecting three training points corresponding to three discontinuity sets shown in Figure 11 as centers, respectively, and generating 200 random normal vectors with angle less than  $\max\text{-AGL} = 5^\circ$  from each center forms the training set (Figure 12) for training the improved Naive Bayes classifier established in this study.

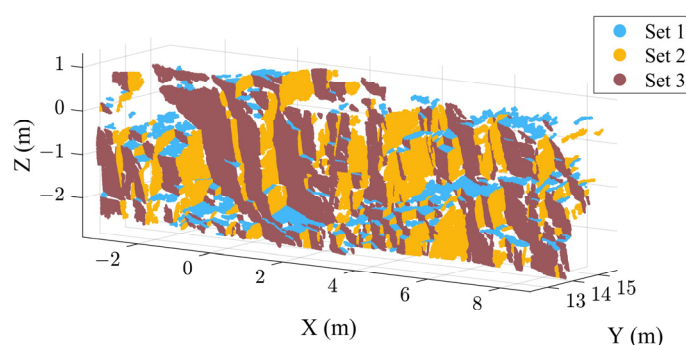


**Figure 11.** Training points selection of case B. One color per discontinuity set.

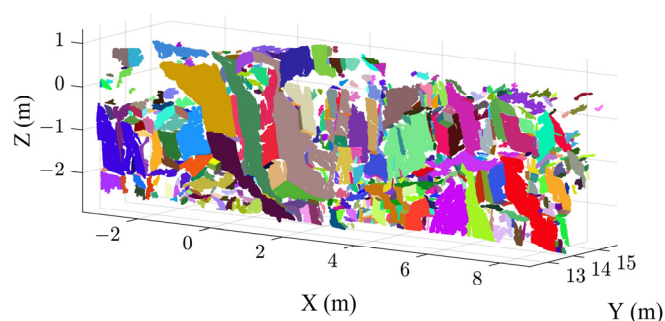


**Figure 12.** Training sets generation of case B.

Figure 13 illustrates the outcomes of discontinuity set extraction utilizing the improved Naive Bayes classifier trained with training sets applied to the entire point cloud, where each color denotes a distinct discontinuity set. It is evident that three discontinuity sets have been successfully extracted, and the majority of the artificial plane (the top-right corner of the point cloud), which does not belong to any discontinuity sets, has been effectively excluded by establishing a minimum logarithmic probability threshold  $\min\text{LPR} = -114.8513$ . The results of extracting individual discontinuities using the HDBSCAN algorithm under the  $\min\text{Cluster} = 10$  and  $\min\text{Dis} = 200$  are depicted in Figure 14. Herein, individual discontinuities are effectively extracted, with the removal of some excessively small discontinuities and noise points.



**Figure 13.** Extraction results of the discontinuity set for case B.



**Figure 14.** Results of extraction of individual discontinuity for case B.

After obtaining the normal vectors of each discontinuity using the PCA algorithm, and substituting them into Equations (11) and (12), the mean dip direction and dip of each discontinuity set are determined. By comparing these results with those obtained by Lato et al. [13] (Table 2), it is evident that Set 1 demonstrates a high level of agreement, with differences of  $0^\circ$  in dip direction and  $3^\circ$  in dip. Although Sets 2 and 3 exhibit slightly lower consistency, with differences of  $6^\circ$  and  $7^\circ$  in dip direction, and  $4^\circ$  in dip, respectively, all differences fall within an acceptable range, affirming the high reliability of the proposed method.

**Table 2.** Comparison of orientation results for case B.

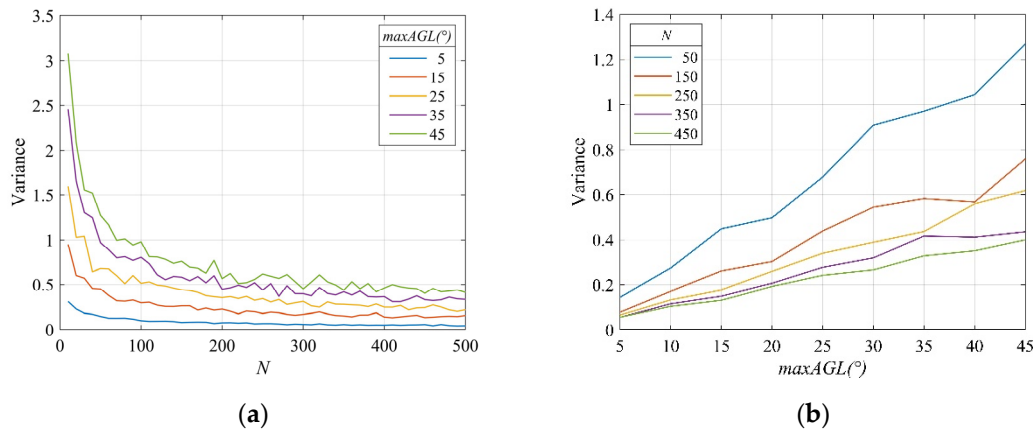
| Discontinuity Set | Number of Points | Number of Discontinuities | Lato ( $^\circ$ ) | Proposed Method ( $^\circ$ ) | Deviation ( $^\circ$ ) |
|-------------------|------------------|---------------------------|-------------------|------------------------------|------------------------|
| 1                 | 229,036          | 173                       | 194/34            | 194/31                       | 0/3                    |
| 2                 | 1,080,620        | 188                       | 29/76             | 35/80                        | 6/4                    |
| 3                 | 450,066          | 149                       | 309/90            | 136/86                       | 7/4                    |

### 3.2. Maximum Angle $maxAGL$ and Quantity $N$ for Training Sets Generation

Firstly, when generating the training set, the maximum value of the  $maxAGL$  should be less than half of the minimum angle between selected point normal vectors. This ensures that each training set is independent and non-overlapping with others. In case B, the minimum angle between selected point normal vectors is  $88.5^\circ$ . Therefore, the training sets should be generated within the range of angles less than  $44.25^\circ$ .

The generated training sets must exhibit uniform distribution within the chosen range to ensure ample stability and representativeness. To determine the optimal values for the  $maxAGL$  and  $N$ , this study iteratively generated 100 distinct training sets for each different combination of  $maxAGL$  and  $N$  around selected point normal vectors. The variance in angles between the central vector of the training set and the selected point normal vector served as a criterion for evaluating the stability of the generated training sets.

Figure 15 illustrates the variation in the stability of generated training sets with changes in the  $maxAGL$  and  $N$  over 100 repetitions.



**Figure 15.** Variation curves of stability for generated point sets: (a) different  $maxAGL$ ; (b) different  $N$ .

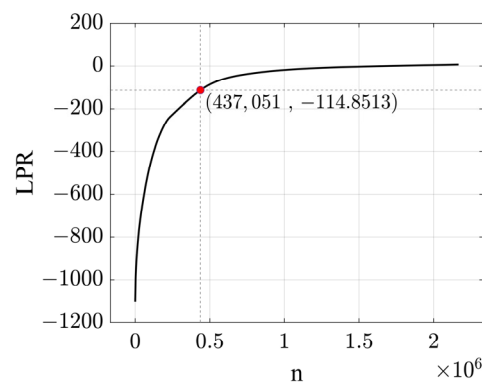
In Figure 15a, the stability of the generated training sets increases with the rise of quantity  $N$ , and the variation becomes relatively stable when  $N$  exceeds 200. However, there are still differences in stability among different  $maxAGL$ . In Figure 15b, under varying  $N$ , an increase in  $maxAGL$  corresponds to a decrease in stability. Moreover, higher values of quantity  $N$  result in less sensitivity of stability to  $maxAGL$ . Nevertheless, the impact of quantity  $N$  on stability diminishes beyond 150, consistent with the trend shown in Figure 15a.

In summary, the study configures the parameters for generating the training sets using the proposed method with a  $maxAGL$  set at  $5^\circ$  and an  $N$  set at 200, ensuring optimal performance of the algorithm.

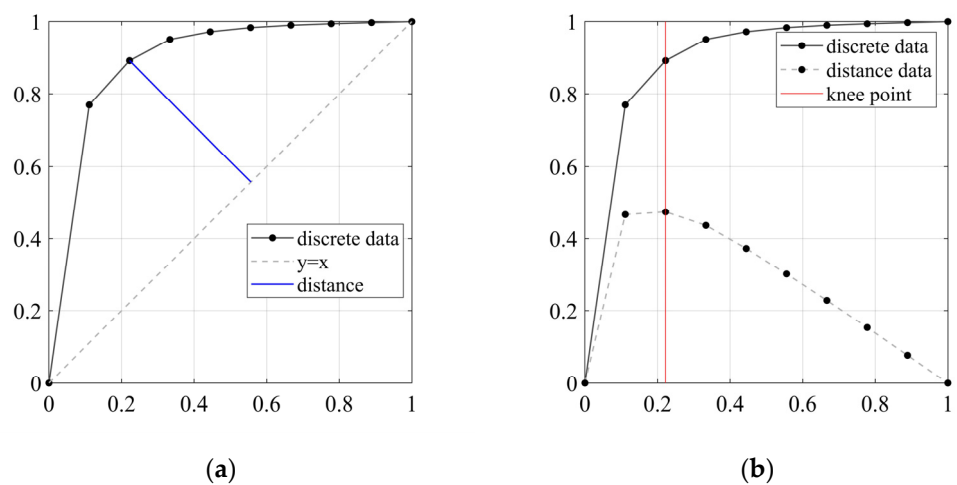
### 3.3. Minimum Logarithmic Probability $minLPR$

It is crucial to determine the  $minLPR$  for noise point removal when employing the proposed method for point cloud classification. Figure 16 illustrates the logarithmic probability distribution curve of the point cloud classification results for case B, with the  $minLPR$  set as  $-114.8513$ , determined through the knee point algorithm at the curve's inflection point. The inflection point is usually defined as the point of maximum curvature in the curve. However, the logarithmic probability distribution curve in Figure 16 is composed of many discrete points, and its inflection point is difficult to accurately obtain by simply taking the second-order derivatives of the curve. This paper therefore uses the "knee point" detection algorithm proposed by Satopaa et al. [53], which is suitable for discrete data. The basic principle of the algorithm is shown in Figure 17, where the discrete data are first normalized and plotted in an interval where the  $x$  and  $y$  coordinates are both 0–1 (this process preserves the morphology of the original data), and then the perpendicular distance from each discrete data point to the line  $y = x$  is found as shown in Figure 17a. The point with the largest perpendicular distance is taken as the knee point (Figure 17b). For more comprehensive information on the algorithm, refer to prior studies [53].





**Figure 16.** Logarithmic probability distribution for case B point cloud classification.



**Figure 17.** Principle of knee point algorithm: (a) perpendicular distance of the discrete point from the line  $y = x$ ; (b) selection of knee points.

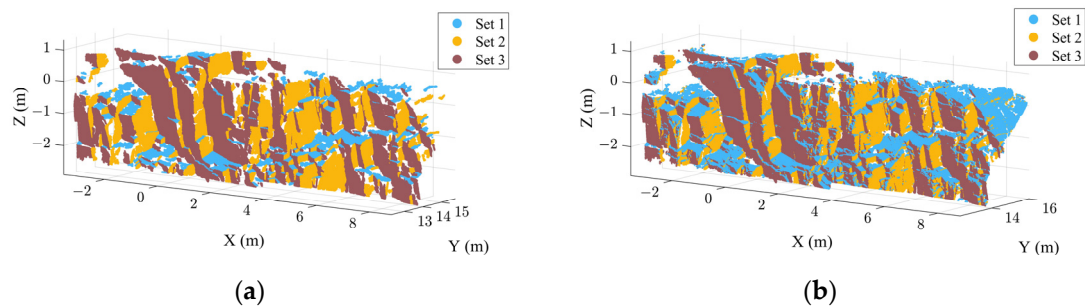
#### 4. Discussion

The point cloud dataset is extensive, and points representing discontinuities lack distinct characteristics, so it is usually necessary to extract certain features for the point cloud classification. Ge et al. [43] utilized four features—coordinates, normal vector, curvature, and density of points—as the basis input of the Artificial Neural Network (ANN) model, successfully achieving efficient point cloud classification. However, this method involves the manual selection of a sufficient number of training points, a process prone to errors that can adversely affect subsequent point cloud classification. To eliminate such effects as much as possible, it is generally necessary to repeatedly select training sets until better classification results are obtained, making it a time-consuming and labor-intensive process. The proposed method classifies point clouds solely based on the normal vectors of points, simplifying the extraction of point cloud features, and also demonstrating outstanding performance in discontinuity sets extraction. Employing a method that revolves around generating training sets around selected training points reduces the necessity for manual point selection, and decreases the likelihood of choosing erroneous points, thereby enhancing the efficiency and accuracy of acquiring training sets.

Within the three-dimensional point cloud acquired through scanning rock surfaces, the presence of noise points not associated with discontinuity, induced by factors like vegetation coverage, rock weathering, and human activities, adversely impacts subsequent calculations of discontinuity parameters. In the realm of prevailing methodologies for point cloud classification, mitigating the impact of noise points remains a formidable challenge. This study addresses the challenge by enhancing the Naive Bayes classifier through



the incorporation of a logarithmic probability threshold selected by the knee point algorithm, enabling more effective filtration of noise points. To gauge the denoising efficacy of the improved Naive Bayes classifier, this paper compares it with the Fuzzy C-Means (FCM) clustering algorithm applied to extract discontinuity sets from the point cloud of case B. As shown in Figure 18, the results of extracting discontinuity sets using the two methods are the same, but there are still slight differences. The FCM algorithm, in its indiscriminate classification of all points, fails to discern and eliminate noise points effectively. Notably, in the upper-right section of the point cloud, artificial rock discontinuities are inaccurately assigned to discontinuity set 1. In contrast, the method proposed in this paper identifies noise points by calculating the logarithmic probability of points on each discontinuity set, as illustrated in Figure 18a where the artificial rock plane has been almost completely removed. Furthermore, upon comparing Figure 18b, it becomes apparent that Figure 18a exhibits some conspicuous blank “voids”, which result from the removal of noise points situated in irregular locations, such as along the edges of the rock mass and in weathered or fragmented areas.



**Figure 18.** Extraction results of two methods: (a) proposed method; (b) FCM.

## 5. Conclusions

This paper introduces an innovative approach for the identification and extraction of rock discontinuities, leveraging an enhanced Naive Bayes classifier. The key conclusions can be summarized as follows:

In response to the arduous and time-intensive process of manually selecting training point sets, fraught with challenges such as randomness that could compromise classification accuracy, this paper introduces an innovative approach to generating training sets around the selected training points, streamlining the selection process and elevating the precision of the training sets. Tailored to meet the demands of point cloud classification tasks, the Naive Bayes classifier is enhanced to effectively filter out noise points during the point cloud classification process by incorporating a logarithmic probability threshold, thereby improving the accuracy of discontinuity set extraction. The HDBSCAN algorithm is employed for the swift extraction of individual discontinuities, while the PCA algorithm is employed to compute the normal vectors of these discontinuities, facilitating accurate positioning.

The proposed novel methodology has been applied to two cases located near Colorado and Kingston, respectively. Leveraging case A, the study explains the principles and procedures of the new method, validating its reliability through comparative analysis with previous research results. Case B is employed to discuss the optimal parameter selection, showcasing the reliable results obtained by the new method through comparisons with prior research outcomes and highlighting its advantages in the realm of noise reduction.

The innovative method proposed in this study for generating training point sets not only minimizes the manual selection of samples but also reduces the likelihood of errors in the training point set. In contrast to existing methods, the use of an improved Naive Bayes classifier for extracting discontinuity sets ensures the removal of noise points while preserving efficiency and reliability, leading to more trustworthy results.

Furthermore, the method proposed in this study operates under the assumption that the features used for classification are independent of each other, but many of the features may not be completely independent of each other, and thus subsequent research could further improve the Naive Bayes classifier in this direction to give it a better performance in terms of discontinuity extraction.

**Author Contributions:** conceptualization, G.L.; methodology, X.Z.; software, X.Z.; validation, Y.L. and Z.Y.; formal analysis, G.L.; investigation, X.Z. and B.C.; resources, G.L.; data curation, X.Z. and B.C.; writing—original draft preparation, X.Z.; writing—review and editing, X.Z. and B.C.; visualization, X.Z.; supervision, C.T.; project administration, X.Z.; funding acquisition, G.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Science and Technology Research and Development Project of China Railway Co., LTD (Grant No. 2022-Special-07-02), the National Natural Science Foundation of China (Grant No. 41974148), the Natural Resources Science and Technology Project of Hunan Province (Grant No.2022-01), and the Research Foundation of the Department of Natural Resources of Hunan Province (Grant No. 20230101DZ).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data and materials that support the findings of this study are available from the author, Xudong Zhu, upon reasonable request.

**Acknowledgments:** Thanks to John Kemeny for sharing the point clouds of case A. The raw data of case B was obtained from the Rockbench repository. The authors kindly appreciated M. Lato, J. Kemeny, R. M. Harrap, and G. Bevan for establishing the Rockbench repository. The authors' special appreciation goes to Editors and anonymous reviewers for valuable comments.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Slob, S. Automated Rock Mass Characterisation Using 3-D Terrestrial Laser Scanning. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 2010.
2. Ghosh, A.; Daemen, J.J. Fractal Characteristics of Rock Discontinuities. *Eng. Geol.* **1993**, *34*, 1–9. [https://doi.org/10.1016/0013-7952\(93\)90039-F](https://doi.org/10.1016/0013-7952(93)90039-F).
3. Zhang, F.; Damjanac, B.; Maxwell, S. Investigating Hydraulic Fracturing Complexity in Naturally Fractured Rock Masses Using Fully Coupled Multiscale Numerical Modeling. *Rock Mech. Rock Eng.* **2019**, *52*, 5137–5160. <https://doi.org/10.1007/s00603-019-01851-3>.
4. ISRM. Suggested Methods for the Quantitative Description of Discontinuities in Rock Masses. *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* **1978**, *15*, 319–368.
5. Bieniawski, Z.T. Engineering Rock Mass Classifications: A Complete Manual for Engineers and Geologists in Mining, Civil, and Petroleum Engineering; John Wiley & Sons: Hoboken, NJ, USA, 1989; ISBN 0-471-60172-1.
6. Sturzenegger, M.; Stead, D. Close-Range Terrestrial Digital Photogrammetry and Terrestrial Laser Scanning for Discontinuity Characterization on Rock Cuts. *Eng. Geol.* **2009**, *106*, 163–182. <https://doi.org/10.1016/j.enggeo.2009.03.004>.
7. Mah, J.; Samson, C.; McKinnon, S.D. 3D Laser Imaging for Joint Orientation Analysis. *Int. J. Rock Mech. Min. Sci.* **2011**, *48*, 932–941. <https://doi.org/10.1016/j.ijrmms.2011.04.010>.
8. Gigli, G.; Casagli, N. Semi-Automatic Extraction of Rock Mass Structural Data from High Resolution LIDAR Point Clouds. *Int. J. Rock Mech. Min. Sci.* **2011**, *48*, 187–198. <https://doi.org/10.1016/j.ijrmms.2010.11.009>.
9. Kong, D.; Wu, F.; Saroglou, C. Automatic Identification and Characterization of Discontinuities in Rock Masses from 3D Point Clouds. *Eng. Geol.* **2020**, *265*, 105442. <https://doi.org/10.1016/j.enggeo.2019.105442>.
10. Priest, S.D.; Hudson, J. Discontinuity Spacings in Rock. *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* **1976**, *13*, 135–148.
11. Gischig, V.; Amann, F.; Moore, J.; Loew, S.; Eisenbeiss, H.; Stempfhuber, W. Composite Rock Slope Kinematics at the Current Randa Instability, Switzerland, Based on Remote Sensing and Numerical Modeling. *Eng. Geol.* **2011**, *118*, 37–53. <https://doi.org/10.1016/j.enggeo.2010.11.006>.
12. Zhang, P.; Zhao, Q.; Tannant, D.D.; Ji, T.; Zhu, H. 3D Mapping of Discontinuity Traces Using Fusion of Point Cloud and Image Data. *Bull. Eng. Geol. Environ.* **2019**, *78*, 2789–2801. <https://doi.org/10.1007/s10064-018-1280-z>.
13. Lato, M.; Diederichs, M.S.; Hutchinson, D.J.; Harrap, R. Optimization of LiDAR Scanning and Processing for Automated Structural Evaluation of Discontinuities in Rockmasses. *Int. J. Rock Mech. Min. Sci.* **2009**, *46*, 194–199. <https://doi.org/10.1016/j.ijrmms.2008.04.007>.

14. Umili, G.; Ferrero, A.; Einstein, H. A New Method for Automatic Discontinuity Traces Sampling on Rock Mass 3D Model. *Comput. Geosci.* **2013**, *51*, 182–192. <https://doi.org/10.1016/j.cageo.2012.07.026>.
15. Li, X.; Chen, Z.; Chen, J.; Zhu, H. Automatic Characterization of Rock Mass Discontinuities Using 3D Point Clouds. *Eng. Geol.* **2019**, *259*, 105131. <https://doi.org/10.1016/j.enggeo.2019.05.008>.
16. Roncella, R.; Forlani, G. Extraction of Planar Patches from Point Clouds to Retrieve Dip and Dip Direction of Rock Discontinuities. In Proceedings of the ISPRS Workshop Laser Scanning, Enschede, The Netherlands, 12–15 September 2005; pp. 162–167.
17. Ferrero, A.M.; Forlani, G.; Roncella, R.; Voyat, H. Advanced Geostuctural Survey Methods Applied to Rock Mass Characterization. *Rock Mech. Rock Eng.* **2009**, *42*, 631–665. <https://doi.org/10.1007/s00603-008-0010-4>.
18. Chen, J.; Zhu, H.; Li, X. Automatic Extraction of Discontinuity Orientation from Rock Mass Surface 3D Point Cloud. *Comput. Geosci.* **2016**, *95*, 18–31. <https://doi.org/10.1016/j.cageo.2016.06.015>.
19. Chen, N.; Kemeny, J.; Jiang, Q.; Pan, Z. Automatic Extraction of Blocks from 3D Point Clouds of Fractured Rock. *Comput. Geosci.* **2017**, *109*, 149–161. <https://doi.org/10.1016/j.cageo.2017.08.013>.
20. Leng, X.; Xiao, J.; Wang, Y. A Multi-Scale Plane-Detection Method Based on the Hough Transform and Region Growing. *Photogramm. Rec.* **2016**, *31*, 166–192. <https://doi.org/10.1111/phor.12145>.
21. Yang, S.; Liu, S.; Zhang, N.; Li, G.; Zhang, J. A Fully Automatic-Image-Based Approach to Quantifying the Geological Strength Index of Underground Rock Mass. *Int. J. Rock Mech. Min. Sci.* **2021**, *140*, 104585. <https://doi.org/10.1016/j.ijrmms.2020.104585>.
22. Jaboyedoff, M.; Oppikofer, T.; Abellán, A.; Derron, M.-H.; Loye, A.; Metzger, R.; Pedrazzini, A. Use of LIDAR in Landslide Investigations: A Review. *Nat. Hazards* **2012**, *61*, 5–28. <https://doi.org/10.1007/s11069-010-9634-2>.
23. Gomes, R.K.; de Oliveira, L.P.; Gonzaga Jr, L.; Tognoli, F.M.; Veronez, M.R.; de Souza, M.K. An Algorithm for Automatic Detection and Orientation Estimation of Planar Structures in LiDAR-Scanned Outcrops. *Comput. Geosci.* **2016**, *90*, 170–178. <https://doi.org/10.1016/j.cageo.2016.02.011>.
24. Wang, X.; Zou, L.; Shen, X.; Ren, Y.; Qin, Y. A Region-Growing Approach for Automatic Outcrop Fracture Extraction from a Three-Dimensional Point Cloud. *Comput. Geosci.* **2017**, *99*, 100–106. <https://doi.org/10.1016/j.cageo.2016.11.002>.
25. Ge, Y.; Tang, H.; Xia, D.; Wang, L.; Zhao, B.; Teaway, J.W.; Chen, H.; Zhou, T. Automated Measurements of Discontinuity Geometric Properties from a 3D-Point Cloud Based on a Modified Region Growing Algorithm. *Eng. Geol.* **2018**, *242*, 44–54. <https://doi.org/10.1016/j.enggeo.2018.05.007>.
26. Yi, X.; Feng, W.; Wang, D.; Yang, R.; Hu, Y.; Zhou, Y. An Efficient Method for Extracting and Clustering Rock Mass Discontinuities from 3D Point Clouds. *Acta Geotech.* **2023**, *18*, 3485–3503.
27. Hu, L.; Xiao, J.; Wang, Y. Efficient and Automatic Plane Detection Approach for 3-D Rock Mass Point Clouds. *Multimed. Tools Appl.* **2020**, *79*, 839–864. <https://doi.org/10.1007/s11042-019-08189-6>.
28. Yu, D.; Xiao, J.; Wang, Y. High-Precision Plane Detection Method for Rock-Mass Point Clouds Based on Supervoxel. *Sensors* **2020**, *20*, 4209. <https://doi.org/10.3390/s20154209>.
29. Singh, S.K.; Banerjee, B.P.; Lato, M.J.; Sammut, C.; Raval, S. Automated Rock Mass Discontinuity Set Characterisation Using Amplitude and Phase Decomposition of Point Cloud Data. *Int. J. Rock Mech. Min. Sci.* **2022**, *152*, 105072. <https://doi.org/10.1016/j.ijrmms.2022.105072>.
30. Park, J.; Cho, Y.K. Point Cloud Information Modeling: Deep Learning-Based Automated Information Modeling Framework for Point Cloud Data. *J. Constr. Eng. Manag.* **2022**, *148*, 04021191. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002227](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002227).
31. Slob, S.; Hack, R.; Turner, A.K. An Approach to Automate Discontinuity Measurements of Rock Faces Using Laser Scanning Techniques. In Proceedings of the ISRM EUROCK, Madeira, Portugal, 25–28 November 2002; ISRM: London, UK, 2002; p. ISRM-EUROCK.
32. Riquelme, A.J.; Abellán, A.; Tomás, R.; Jaboyedoff, M. A New Approach for Semi-Automatic Rock Mass Joints Recognition from 3D Point Clouds. *Comput. Geosci.* **2014**, *68*, 38–52. <https://doi.org/10.1016/j.cageo.2014.03.014>.
33. Menegoni, N.; Giordan, D.; Perotti, C.; Tannant, D.D. Detection and Geometric Characterization of Rock Mass Discontinuities Using a 3D High-Resolution Digital Outcrop Model Generated from RPAS Imagery—Ormea Rock Slope, Italy. *Eng. Geol.* **2019**, *252*, 145–163. <https://doi.org/10.1016/j.enggeo.2019.02.028>.
34. Wu, X.; Wang, F.; Wang, M.; Zhang, X.; Wang, Q.; Zhang, S. A New Method for Automatic Extraction and Analysis of Discontinuities Based on TIN on Rock Mass Surfaces. *Remote Sens.* **2021**, *13*, 2894. <https://doi.org/10.3390/rs13152894>.
35. Van Knapen, B.; Slob, S. Identification and Characterisation of Rock Mass Discontinuity Sets Using 3D Laser Scanning. *Procedia Eng.* **2006**, *191*, 838–845.
36. Vöge, M.; Lato, M.J.; Diederichs, M.S. Automated Rockmass Discontinuity Mapping from 3-Dimensional Surface Data. *Eng. Geol.* **2013**, *164*, 155–162. <https://doi.org/10.1016/j.enggeo.2013.07.008>.
37. Olariu, M.I.; Ferguson, J.F.; Aiken, C.L.; Xu, X. Outcrop Fracture Characterization Using Terrestrial Laser Scanners: Deep-Water Jackfork Sandstone at Big Rock Quarry, Arkansas. *Geosphere* **2008**, *4*, 247–259. <https://doi.org/10.1130/GES00139.1>.
38. Li, Y.; Wang, Q.; Chen, J.; Xu, L.; Song, S. K-Means Algorithm Based on Particle Swarm Optimization for the Identification of Rock Discontinuity Sets. *Rock Mech. Rock Eng.* **2015**, *48*, 375–385. <https://doi.org/10.1007/s00603-014-0569-x>.
39. Cui, X.; Yan, E. A Clustering Algorithm Based on Differential Evolution for the Identification of Rock Discontinuity Sets. *Int. J. Rock Mech. Min. Sci.* **2020**, *126*, 104181. <https://doi.org/10.1016/j.ijrmms.2019.104181>.
40. Guo, J.; Liu, S.; Zhang, P.; Wu, L.; Zhou, W.; Yu, Y. Towards Semi-Automatic Rock Mass Discontinuity Orientation and Set Analysis from 3D Point Clouds. *Comput. Geosci.* **2017**, *103*, 164–172. <https://doi.org/10.1016/j.cageo.2017.03.017>.

41. Xu, W.; Zhang, Y.; Li, X.; Wang, X.; Ma, F.; Zhao, J.; Zhang, Y. Extraction and Statistics of Discontinuity Orientation and Trace Length from Typical Fractured Rock Mass: A Case Study of the Xinchang Underground Research Laboratory Site, China. *Eng. Geol.* **2020**, *269*, 105553. <https://doi.org/10.1016/j.enggeo.2020.105553>.
42. Chen, J.; Huang, H.; Zhou, M.; Chaiyasarn, K. Towards Semi-Automatic Discontinuity Characterization in Rock Tunnel Faces Using 3D Point Clouds. *Eng. Geol.* **2021**, *291*, 106232. <https://doi.org/10.1016/j.enggeo.2021.106232>.
43. Ge, Y.; Cao, B.; Tang, H. Rock Discontinuities Identification from 3D Point Clouds Using Artificial Neural Network. *Rock Mech. Rock Eng.* **2022**, *55*, 1705–1720. <https://doi.org/10.1007/s00603-021-02748-w>.
44. BAYES. An Essay towards Solving a Problem in the Doctrine of Chances. *Biometrika* **1958**, *45*, 296–315. <https://doi.org/10.1093/biomet/45.3-4.296>.
45. McCallum, A.; Nigam, K. A Comparison of Event Models for Naive Bayes Text Classification. In Proceedings of the AAAI-98 Workshop on Learning for Text Categorization, Madison, WI, USA, 26–27 June 1998; Volume 752, pp. 41–48.
46. Berchialla, P.; Foltran, F.; Gregori, D. Naïve Bayes Classifiers with Feature Selection to Predict Hospitalization and Complications Due to Objects Swallowing and Ingestion among European Children. *Saf. Sci.* **2013**, *51*, 1–5. <https://doi.org/10.1016/j.ssci.2012.05.021>.
47. Feng, X.; Li, S.; Yuan, C.; Zeng, P.; Sun, Y. Prediction of Slope Stability Using Naive Bayes Classifier. *KSCE J. Civ. Eng.* **2018**, *22*, 941–950. <https://doi.org/10.1007/s12205-018-1337-3>.
48. Lato, M.; Kemeny, J.; Harrap, R.; Bevan, G. Rock Bench: Establishing a Common Repository and Standards for Assessing Rock-mass Characteristics Using LiDAR and Photogrammetry. *Comput. Geosci.* **2013**, *50*, 106–114. <https://doi.org/10.1016/j.cageo.2012.06.014>.
49. Kemeny, J.; Turner, K.; Norton, B. LIDAR for Rock Mass Characterization: Hardware, Software, Accuracy and Best-Practices. In *Laser and Photogrammetric Methods for Rock Face Characterization*; ARMA: Golden, CO, USA, 2006; pp. 49–62.
50. Huang, C.M.; Tseng, Y.-H. Plane Fitting Methods of LIDAR Point Cloud. In Proceedings of the 29th Asian Conference on Remote Sensing 2008, ACRS 2008, Colombo, Sri Lanka, 10–14 November 2008; pp. 1925–1930.
51. Tsangaratos, P.; Ilia, I. Landslide Susceptibility Mapping Using a Modified Decision Tree Classifier in the Xanthi Perfection, Greece. *Landslides* **2016**, *13*, 305–320. <https://doi.org/10.1007/s10346-015-0565-6>.
52. Campello, R.J.; Moulavi, D.; Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Gold Coast, Australia, 14–17 April 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 160–172.
53. Satopaa, V.; Albrecht, J.; Irwin, D.; Raghavan, B. Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior. In Proceedings of the 2011 31st International Conference on Distributed Computing Systems Workshops, Minneapolis, MN, USA, 20–24 June 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 166–171.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.