An Acad Bras Cienc (2021) 93(4): e20201242 DOI 10.1590/0001-3765202120201242
Anais da Academia Brasileira de Ciências | *Annals of the Brazilian Academy of Sciences*
Printed ISSN 0001-3765 I Online ISSN 1678-2690
www.scielo.br/aabc | www.fb.com/aabcjournal

105 anos 1916 • 2021

ENGINEERING SCIENCES

Failure risk of brazilian tailings dams: a data mining approach

TATIANA B. SANTOS & RUDINEI M. OLIVEIRA

Abstract: This paper proposes the use of a hybrid method that combines Biased Random Key Genetic Algorithm (BRKGA) with a local search heuristic to separate Brazilian tailing dam data into groups. The goal was identifying dams similar to Fundão and B1 failed dams. The groups were created by solving the clustering problem by BRKGA. The clustering problem consists in separating a set of objects into groups such that members of each group are similar to each other. The data was composed by 427 dams, with the actual 425 dams of Brazilian Register of Tailing Dams and the two Brazilian failed dams from the last years. Computational experiments considering real data available are presented to demonstrate the efficacy of the proposed method producing feasible solutions. Thus, it is expected that the good results can be applied in the identification of tailings dams with risk potentials, assisting in the identification of these dams.

Key words: Tailing dams, clustering problem, biased random key genetic algorithm, mining.

1. INTRODUCTION

Extractive mineral industry is indispensable for the modern society and it is responsible by the production of essential materials for almost all human activities, like technology industry, civil construction and agriculture. The growing demand for ore and exploitation of low-content deposits have led to the increased production of tailings in the mineral processing plants.

Tailing dams consist of engineering structures built to tailing storage and they are the method mostly employed by the mineral industry. These structures have reached larger volumes and heights due to the growing production of these materials in the mineral processing plants due to the growing demand for ore and exploitation of low-content deposits. Mineral industry produces billions of tonnes of tailings each year. In year of 2010, the tailing production around the world was estimated in 14 billion of tonnes (Adiansyah et al. 2015).

Tailing dams are some of the largest structures built by geotechnical engineers. As larger is the tailing dam, as greater is the concern about its geotechnical management, since failure of these structures can lead to catastrophic consequences including human, environmental and economic loses. The failure risk disaster is a concern of the whole world and public concern about the risk and impacts of tailing dams growing since recent incidents occurred (Rico et al. 2008). Some of the last dam failures that caused serious consequences in the world occurred at Brumadinho (Brazil) (Thompson et al. 2020), Fundão (Brazil) (Martinho et al. 2016), Mount Polley (Canada) (Petticrew et al. 2015), Kayakiri (Japan) (Ishihara et al. 2015), Itabirito (Brazil) (CETEM 2016), Ajka (Hungary) (Ruyters et al. 2011), Karamken tailing plant (Russia) (Glotov et al. 2018), San Marcelino

Zambales (Philippines) (Piplinks 2007), Baia Mare (Romenia) (Soldán et al. 2001), Baia Borsa (Romenia) (Bird et al. 2008), Gällivare (Sweden) (Göransson et al. 2001) and Los Frailes (Spain) (López-Pamo et al. 1999).

In the last decade, the two major failure disasters of Brazilian tailing dams occurred in cities of Brumadinho and Fundão, all of them in Minas Gerais state. In 2019, at Córrego do Feijão iron mine in Brumadinho, the dam named B1 collapsed. The dam was built using upstream method and its height and crest length were 86 meters and 720 meters, respectively. The tailings were stored in an area of 249.5 thousand square meters and the volume was 11.7 million cubic meters. Besides the environmental damage, 259 people died in the disaster (Vale S.A. 2019).

In 2015, Fundão dam of Samarco's iron mine collapsed. More than 43 million of m^3 of iron ore tailings caused a large environmental damage, polluting 668 km of watercourses from Doce River Basin to the Atlantic Ocean. The dam was also built using upstream method and its height were 86 meters. The volume of waste and the extent of ecosystems affected have assume unprecedent proportions, involving the Brazilian Atlantic Forest (Carmo et al. 2017). Catastrophic failures of Fundão and B1 dams have thrust industrial scale mine tailing disasters onto to the global stage in unprecedented ways (Owen et al. 2020).

In order to define the Brazilian tailing dams similar to Fundão and B1 dams, research of Brazilian Register of Tailing Dams was carried out using data mining. Biased random-key genetic (BRKGA) with a local search algorithm was applied to the data with the goal of pattern recognition and clustering of the dams according to its similar characteristics. The knowledge of the dams with similar characteristics of Fundão and B1 dams become an important tool of prevention of future disasters.

BRKGA represents solutions in vectors of random keys. This vector is decoded into a real solution to the specific problem. This feature makes the method becomes independent of the problem. Thus, the only component that needs to be implemented is the decoder function. In this paper, a heuristic local search was also applied to speed up the convergence of the method, exploring new regions in the search space.

BRKGA method was applied to the Brazilian tailing dam data. For the separation of data in similar groups, the clustering problem was solved. Given a data that represent something in the real world, the clustering problem is defined as the process of separating a set of objects into groups such that members of a group are similar to each other (Berkhin 2006).

The paper is organized as follows. Section 2.1 and 2.3 presents the tailing dam data and the methodology research. Section 2.2 presents an overview of the hybrid method with BRKGA and local search heuristic. Section 3 presents the computational results and discussion. In Section 4, conclusions are mentioned.

2. MATERIALS AND METHODS

2.1. Dataset construction

The National Plan of Dam Security (NPDS) was established by the Brazilian government in year of 2010 with the objective of ensure dam safety standards, regulate and monitoring security actions, collect information to support the dam safety management by the government and foment a culture of dam safety and risk management. As part of NPDS, the Brazilian Register of Tailing Dams was created by the National Mineral Agency (ANM 2020).

The tailing dams with at least one of the characteristics presented below must to be inserted and registered in the NPDS:

a) dam height greater than or equal to fifteen meters (counted from the lowest point of the foundation to the crest);
b) total reservoir capacity greater than or equal to three million cubic meters;
c) reservoir containing hazardous waste;
d) medium or high potential damage associated (PDA).

Currently, 425 Brazilian tailing dams are inserted in the NPDS and registered in the Brazilian Register of Tailing Dams. This data is public and available on the National Mining Agency's Website (ANM 2019). Dam name, responsible company, geographical location, state, city, principal ore, actual height, actual volume, method of construction, risk category, associated potential damage and class are the information provided in the register for each tailing dam. Table I presents the variables used in the dataset construction for application of BRKGA. Only variables associated to the dam security were used, i.e., method of construction, risk category, associated potential damage, actual volume and actual height. The variable class was removed because it is a dependent variable obtained through the combination of risk category and associated potential damage. Values in a range of 1 to 3, 1 to 4 and 1 to 5 were attributed to the parameters of the ordinal qualitative variables. In case of quantitative variables, the original values were used.

A brief description of the database variables is presented below:

Method of construction: the most common construction methods of tailing dams are downstream, centerline, upstream and single dyke. The upstream method is the most economic method and the dikes are sequentially constructed above the tailing beach. In the downstream method the subsequent stages of dike construction are supported on top of the downstream slope of the previous section. Centerline method is an intermediate method to the previous methods. Figure 1 presents construction methods of tailing dams.

Risk category: is related to the aspects that can influence in the failure probability of a tailing dam. The tailing dams are classified

Table I. Variables and rating used in the dataset construction.

	Variable				
	Method of construction				
Description	**Undefined**	**Downstream**	**Centerline**	**Upstream or unknown**	**Single dyke**
Rating	5	4	3	2	1
	Risk category				
Description	High	Medium			Low
Rating	3	2			1
	Associated potential damage				
Description	High	Medium			Low
Rating	3	2			1
	Actual volume (millions of m^3)				
	Actual height (meters)				

as high, medium and low risk according to its technical characteristics, structural integrity, dam condition of conservation and dam safety plan (ANM 2020). The dams are classified as high, medium and low risk.

Associated potential damage: the tailing dams are classified as high, medium and low damage according to its failure consequences. The used criteria are the existence of downstream population with potential loss of human life, house or urban and community equipment; infrastructure or services; services of essential public equipment; protected areas defined in legislation; the nature of the stored tailing and dam volume (ANM 2020). The dams are classified as high, medium and low associated potential damage.

Actual volume: is related to the actual stored tailing quantity in millions of cubic meters. The variable is quantitative and the original values were used.

Actual height: is related to the actual dam height in cubic meters. The variable is quantitative and the original values were used.

Two Brazilian tailing dams which failed in the last decade were added to the dataset, beyond the 425 dams from Brazilian Register of Taling Dams. The added dams are B1 dam from Corrego do Feijão iron mine and Fundão dam from Samarco's mine. Table II presents

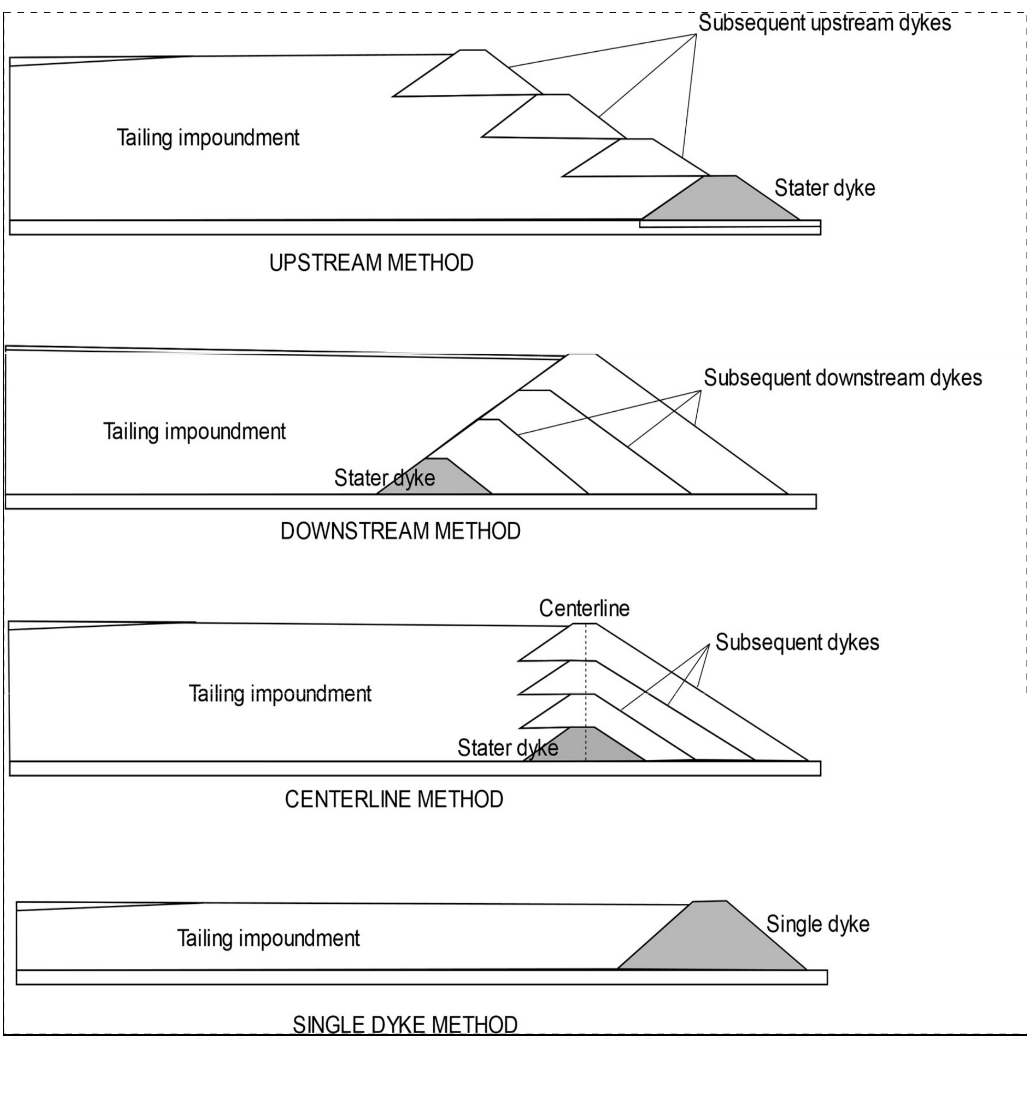

Figure 1. Construction methods of tailing dams.

the characteristics of these tailing dams in the moment of its failure occurrence.

The final dataset was composed by 427 dams (425 dams of Brazilian Register of Tailing Dams and the two failed dams from the last years). Table III presents the variable values of the dataset. These data have been standardized (Table IV).

2.2. Biased Random-key Genetic Algorithm and Local Search

The BRKGA was proposed by Gonçalves & Resende (2011) for solving combinatorial optimization problems and it is a variation of the Random-Key Genetic Algorithms (RKGA). This method has been used to solve sequencing in optimization problems (Bean 1994).

The BRKGA represents a solution with a vector of random-keys, which are real numbers in the interval [0, 1]. This vector (also called chromosome) is not considered as a solution of the problem. Therefore, it is necessary to decode the vector to a real problem solution. For each problem, we define a specific decoder. This is a deterministic algorithm that takes chromosome information and returns a solution to the problem. The fitness of each solution is also computed by the decoder.

The evolution process of the BRKGA is independent of the problem. A population of *p* random-key vectors evolves over a number of generations. In each generation, the population is sorted by the fitness. Then, a small group with the best p_e solutions in fitness values (elite group) are copied without modification to the population of the next generation. A number p_m of random-key vectors, randomly generated (mutants), are also introduced into this population. The remainder of the population (p - p_e - p_m solutions) is produced through the process of crossover, by combining an elite parent with a non-elite parent of the current population. The parameterized uniform crossover by is used in BRKGA (Spears & Jong 1991).

In this paper, BRKGA has been adapted to clustering the dams that are similar, solving the clustering problem. The method randomly creates an initial population of random-keys vectors. Each vector has *n* random-key, where *n* is the number of objects in the data.

The solution of the problem is accomplished by corresponding random-key values by decoder. In decoder, the interval [0, 1] is divided by *k* groups and the clusters are created with the objects that have the random-key in this interval. For example, a vector of random keys with 10 alleles randomly generated (Table V). After running the decoder, the vector of random keys is separated into two groups according with the similarity of objects. The objects between [0, 0.5] belong to cluster 1, and the objects between [0.51, 1] belong to cluster 2. Table VI shows the result of the decoding process considering two clusters, where the cluster 1 and cluster 2 are {(2,4,5,7,9), (1,3,6,8,10)}. The same idea applies to

Table II. **Characteristics of the Brazilian failed dams.**

ID	Dam	Method of construction	Risk category	Associated potential damage	Volume in the failure moment (m^3)	Height in the failure moment (m)
426	Fundão (Germano)	Upstream	Low	High	91866000	130
427	B1 (Córrego do Feijão)	Upstream	Low	High	11700000	86

the division into more clusters and modifying the interval values (Gonçalves & Resende 2011).

After decoding process, the fitness is calculated by the objective function (Babaki et al. 2014). The objective function (D) minimize the distance (d) between the elements of same cluster, divided by size of each cluster (C), as in Eq. 1.

$$D = \frac{\sum_{y_1, y_2 \in C} d^2\left(y_1, y_2\right)}{|C|} \tag{1}$$

Where, $y_1 \neq y_2$ belong to the cluster *C* and |*C*| is the cardinality of the cluster. Each pair of two points in *C* is included in the sum only once, without repetition. For example, the calculation of the distances of the objects in the Table VI will be: d(2,4), d(2,5), d(2,7), d(2,9), d(4,5), d(4,7), d(4,9), d(5,7), d(5,9) and d(7,9) in the cluster 1 and d(1,3), d(1,6), d(1,8), d(1,10), d(3,6), d(3,8), d(3,10), d(6,8), d(6,10) and d(8,10) in the cluster 2.

A local search was also implemented to improve the solution. It is applied to all offspring generated by BRKGA and is used in the decoded solution. The solution found by local search is not transferred to the vector of random keys. The diversity of the BRKGA is preserved because the local search solution is not transferred to the vector of random keys. It shifts the object of a cluster to the others. Then the best solution is stored. This heuristic is run while the fitness of the solution is improved. Figure 2 presents the algorithm that the proposed shift heuristic.

Table III. **First ten dams of the dataset.**

ID	Dam	Method of construction *x1*	Risk category *x2*	Associated potential damage *x3*	Actual volume (m^3) *x4*	Actual height (m) *x5*
1	0-1	1	1	2	27700000	22
2	103 (Cruz)	1	1	2	924000	10.30
3	111 (Índio)	1	1	2	48873	5
4	158 A-1	2	1	3	53380000	30
5	161 A-2	1	1	2	1018054	22
6	444 A-3	1	1	2	886241	14.70
7	81-1	1	1	2	75522	4
8	Pau D'Arco	1	1	2	52800	18
9	Barragem da cava	2	1	2	1755555	25
10	Mario Cruz	3	2	2	25366731	28

Table IV. First ten standardized tailing dams.

ID	Dam	Method of construction *z1*	Risk category *z2*	Associated potential damage *z3*	Actual volume (m^3) *z4*	Actual height (m) *z5*
1	0-1	0.2500	0.5000	1.0000	0.135802	0.069413
2	103 (Cruz)	0.2500	0.5000	1.0000	0.063580	0.002315
3	111 (Índio)	0.2500	0.5000	1.0000	0.030864	0.000122
4	158 A-1	0.5000	0.5000	1.5000	0.185185	0.133765
5	161 A-2	0.2500	0.5000	1.0000	0.135802	0.002551
6	444 A-3	0.2500	0.5000	1.0000	0.090741	0.002221
7	81-1	0.2500	0.5000	1.0000	0.024691	0.000189
8	Pau D'Arco	0.2500	0.5000	1.0000	0.111111	0.000132
9	Barragem da cava	0.5000	0.5000	1.0000	0.154321	0.004399
10	Mario Cruz	1.0000	1.0000	1.0000	0.172840	0.063566

Table V. Sample random key.

Index	1	2	3	4	5	6	7	8	9	10
Random keys	0.69	0.22	0.99	0.45	0.01	0.86	0.19	0.55	0.15	0.70

Table VI. Random keys after the decoder.

Index	2	4	5	7	9	1	3	6	8	10
Random keys	0.22	0.45	0.01	0.19	0.15	0.69	0.99	0.86	0.55	0.70
Clusters	1	1	1	1	1	2	2	2	2	2

Algorithm 1 Shift local search (S)

```
while (improve S ) do
    for (each object) do
        for (each cluster) do
            S′ ← S;
            S′ ← Move (an object to another clusters);
            if S′ better than S then
                S ← S′;
                STORE (the best solution S*);
        end for
    end for
end while
return (S*)
```

Figure 2. Algorithm.

2.3. Research methodology

The research methodology for defining the actual Brazilian dams with similar characteristics of the failed B1 and Fundão dams consisted of four main steps. The first step was the dataset organization. Characteristics of actual Brazilian tailing dams plus the two last failed dams were collected and organized. The second step consisted of the application of Biased random-key genetic (BRKGA) with a local search algorithm, aiming at clustering the data. The adequate number of clusters was defined by the Elbow Method (Thorndike 1953). Euclidean distance, Manhattan distance, Cosine similarity and Person similarity were used in order to test the most suitable metric to the dataset. In the third step, the best clustering was selected. This selection was made based on visual analysis of the results and clustering validation of the data. The validation was done using information of dams with high risk of failure provided by Brazilian Government Reports. Finally, the fourth step consisted of the definition of the actual Brazilian tailing dams with similar characteristics with B1 and Fundão failed dams. Figure 3 presents the flowchart of the methodology.

3. RESULTS AND DISCUSSION

This section presents the computational results, validation and discussion of BRKGA with Local Search Algorithm application. The algorithm application to the dataset were performed on a PC Intel Core i7, 2.9 GHz, 64 bits and 16 GB of RAM.

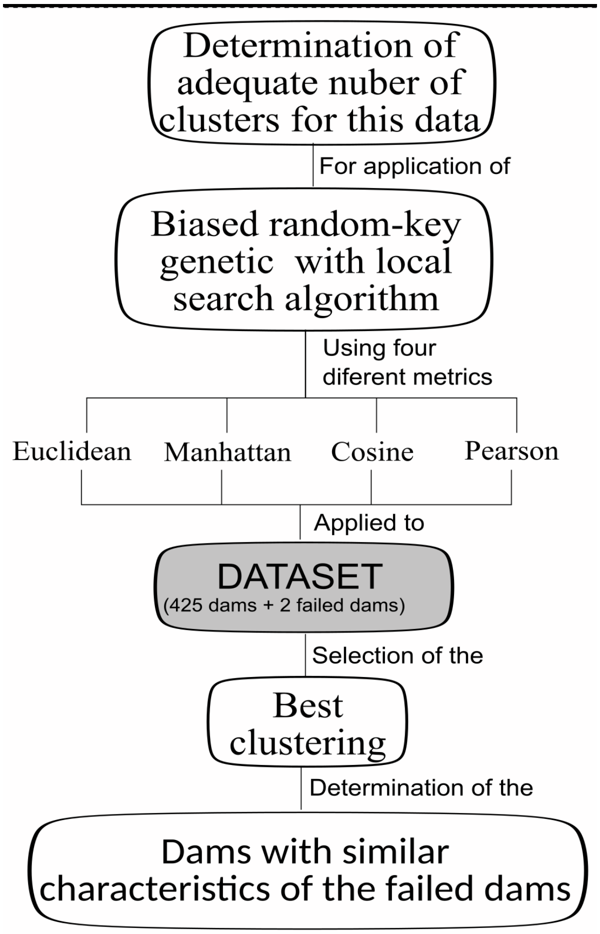

Figure 3. Methodology.

The parameters used for the BRKGA methods are presented in Table VII and they were obtained by Calibra software. Table VIII presents the computational results of the objective function minimization values for the tailing dam dataset obtained by the algorithm application with 2, 3, 4, 5, 6, 7, 8, 9 and 10 clusters. The objective function minimizes the distances between elements of the same clusters and all these solutions are feasible. The distances or dissimilarities were calculated using four metrics: Euclidean, Manhattan, Cosine and Pearson.

The adequate number of clusters for the studied dataset was defined by the Elbow Method (Thorndike 1953). The method consists of plotting the explained variation as a function of the number of clusters, and choosing the elbow of the curve as the number of clusters to use. For this dataset, the method chose four as an adequate number of clusters.

Table VII. BRKGA parameters used in the experiments.

Parameters BRKGA	Meaning	Value
P	Number of individuals in population	1000
Gen	Number of generations	1000
p_e	Size of the elite set in population	0.2
p_m	Number of mutants to be introduced in population at each generation	0.4
$\boldsymbol{\rho}_e$	Probability that an allele is inherited	0.6

With the goal of graphically presenting the clustering results, the dimensionality of the data was reduced from 5 to 2 variables using principal components technique. The technique consists of the construction of linear combinations of the original variables. The coefficients of the two first principal components (linear combinations) are the eigenvectors associated to the greater eigenvalues of the correlation matrix of the original standardized data. Eq. 2 and Eq. 3 presents the two first principal components (cp1 and cp2). They are capable of explaining 64.6% of the data variability.

$$cp1 = -0.43z1 + 0.21z2 - 0.38z3 - 0.04z4 + 0.19z5 \quad (2)$$

$$cp2 = -0.53z1 + 0.72z2 - 0.40z3 - 0.56z4 - 0.55z5 \quad (3)$$

Figure 4 presents the graphic results of clustering obtained through BRKGA with Local search.

Visually, the best clustering was obtained through the application of the BRKGA with Local Search using the Person measure.

Table VIII. Results with values of the objective function for the tailing dam dataset.

Clusters	Euclidean	Manhattan	Cosine	Pearson
2	105.612806	144.162819	9.471279	20.510861
3	84.608211	117.372262	5.918690	12.333456
4	74.439174	102.581991	4.237784	8.813047
5	63.586627	87.250852	3.783845	7.294127
6	54.872264	82.436485	3.052754	6.709703
6	51.707632	65.781028	2.931459	5.253705
8	47.912183	56.900554	2.681774	5.079224
9	42.319854	53.334356	2.160351	4.772565
10	40.945731	48.186336	2.116750	4.282325

Observing Figure 4, it possible to verify dams from a different cluster inserted inside an area of another cluster, *e.g.*, dams from Cluster 1 inserted inside the Cluster 3.

By means of Reports the Mining Agency from Brazilian Government, the tailing dams with risk of failure were identified in order to validate the clustering using Pearson measure. The Dams with risk are classified in Emergency Level 3 and Emergency Level 2. Dams classified as Level 3 are dams where the failure is imminent or is happening. Dams classified as Level 2 are dams where anomalies were not controlled by repair actions. Table IX presents these dams. It is expected that these dams have been clustered in the same group of the failed dams.

Figure 5 presents the tailing dams with emergency Level equal to 3 and equal to 2. All dams from Emergency Level 3 were clustered in the same cluster (Cluster 4) of Fundão (ID 426) and B1 (ID 427) failed tailing dams, see Figure 5(a). Fundão dam is more distant from the B1 dam and the dams with Emergency Level 3. This distance is related to this dam has greater height and volume than the others. Then, it presents characteristics slightly different. The dams with characteristics more similar to B1 (Corrego do Feijão Mine) are the dam identified by ID 195, ID 197 and 243. The dam with less similar characteristics to failed dams is the dam identified by the ID 55.

In case of dams with emergency level equal to 2, four dams were clustered in the same cluster (Cluster 4) of failed dams and two cluster were clustered in Custer 1, see Figure 5(b). The dams identified by ID 119, ID 191 and ID 196 presents characteristics more similar to the failed dams. The dam identified by ID 201 is far from these dams, then it presents less similar characteristics but it still belongs to Cluster 4. Of the dams belonging to Cluster 1, The dam identified by ID 150 is more similar to the dams from Cluster 3.

Considering all the dams used to validate the clustering, 80% of the dams with failure risk were clustered in the same cluster (Cluster 3) of the Fundão and B1 failed dams. The 20% of dams were classified in the same cluster (Cluster 1). This indicated that the Cluster 1 has more similar dams to Cluster 3, when compared to the other clusters. These results are sufficient to validate these clustering.

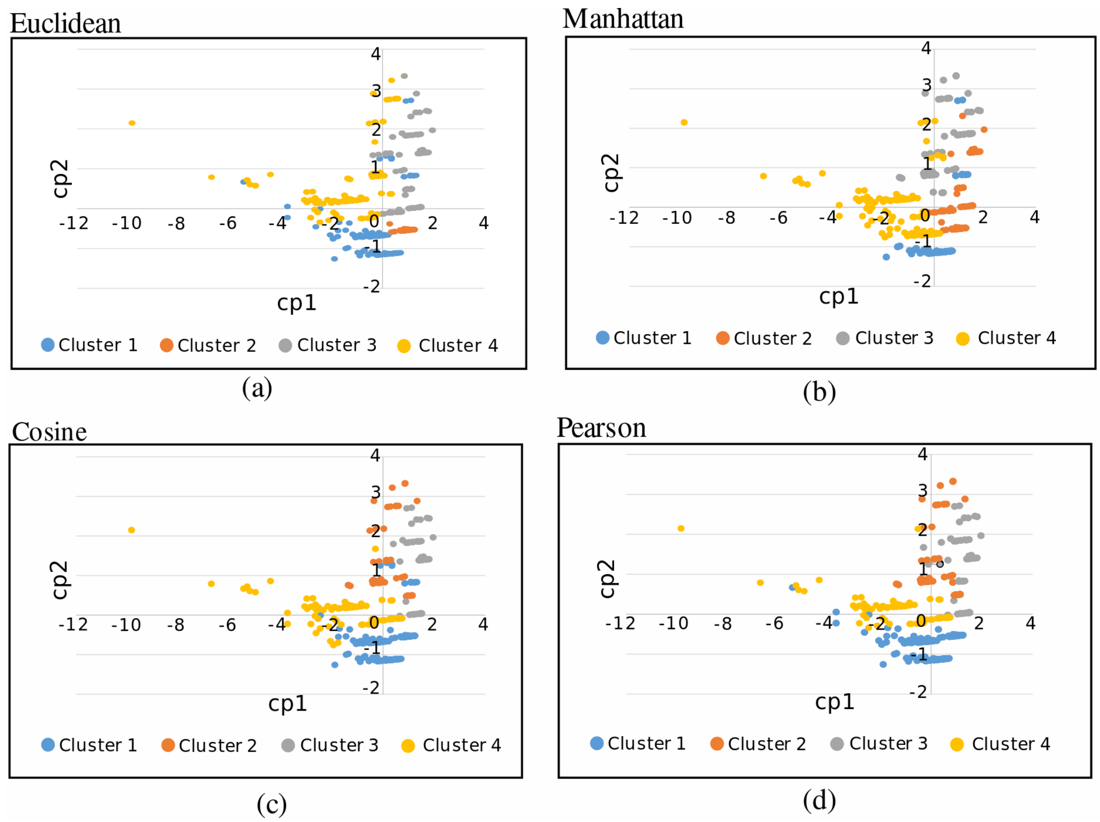

Figure 4. **Clustering results.**

Table IX. Brazilian Tailing Dams with high emergency level.

Emergency Level	Tailing dam	Company	City	ID	Source
Level 3	B3/B4	Minerações Brasileiras Reunidas Sa Filial: MBR Paraopeba	Nova Lima	55	(Neves et al. 2020)
	Forquilha I	Vale S A Filial: Vale Itabiritos	Ouro Preto	195	
	Forquilha III	Vale S A Filial: Vale Itabiritos	Ouro Preto	197	
	Sul superior	Vale S A Filial: Vale Minas Centrais	Barão de cocais	243	
Level 2	Barragem de Rejeitos	Arcelormittal Mineração Serra Azul S.a.	Itatiaiuçu	119	
	Capitão do mato	Vale S A Filial: Vale Vargem Grande	Nova Lima	150	
	Doutor	Vale S A Filial: Vale Mariana	Ouro Preto	191	
	Forquilha II	Vale S A Filial: Vale Itabiritos	Ouro Preto	196	
	Grupo	Vale S A Filial: Vale Itabiritos	Ouro Preto	201	
	Sul Inferior	Vale S A Filial: Vale Minas Centrais	Barão de Cocais	242	

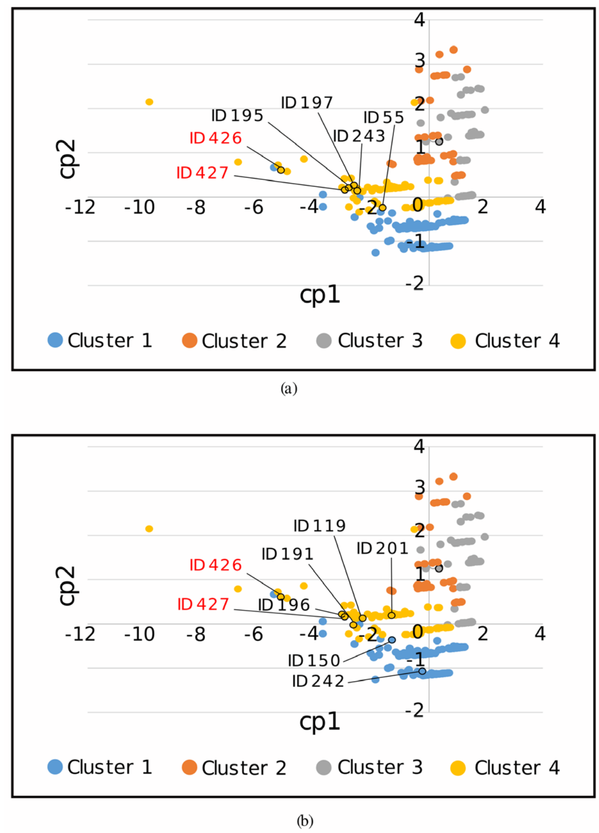

Figure 5. Dams with emergency Level equal to 3 and equal to 2.

Additional data of the Brazilian Dams data with the clusters that each dam belongs is available in the link bellow.

https://drive.google.com/file/d/1je9JFBkEkQr_tu5Sm6A5We4T04aOMbxA/view?usp=sharing

4. CONCLUSIONS

This paper presents a contribution to separate Brazilian tailing dam data in groups into groups and identifying dams similar to Fundão and B1 failed dams. The data used are from Brazilian Register of Tailing Dams with 425 dams and two failed dams from the last years.

The BRKGA method with a local search heuristic was proposed to create clusters with similar dams. The similar groups were obtained by solving the clustering problem. Euclidean distance, Manhattan distance, Cosine similarity and Person similarity were used in the experimental tests to verify the similarity between objects.

Computational results found feasible solutions for all metrics and created groups. Initially we tested and created 2, 3, 4, 5, 6, 7, 8, 9, 10 clusters and the Elbow method was used to define the adequate number of clusters for these data. The adequate number of clusters was equal to four. The better clustering was obtained using the Person measure. To validate the clusterization, dams with risk of failure were used. 80% of these dams were clustered in the same cluster of the Fundão and B1 failed dams.

Considering all data, the local search considerably improved the solutions by accelerating the convergence process. The good results obtained demonstrate that the hybrid heuristic BRKGA with local search can be used as a good alternative for identifying tailings dams with similar characteristics of failed dams and dams with failure risk.

Acknowledgments

The authors wish to thank Mining Department from Universidade Federal de Ouro Preto and Exat Siences Department from Universidade do Estado de Minas Gerais. Besides, the authors are grateful for the collaboration of the Research Group on Data Science in Engineering (CIDENG).

REFERENCES

ADIANSYAH JS, ROSANO M, VINK S & KEIR G. 2015. A Framework for a Sustainable Approach to Mine Tailings Management: Disposal Strategies. J Clean Prod 108: 1050-1062.

ANM - AGÊNCIA NACIONAL DE MINERAÇÃO. 2019. Cadastro Nacional de Barragens de Mineração. http://www.anm.gov.br/assuntos/barragens/cadastro-nacional-de-barragens-de-mineracao.

ANM - AGÊNCIA NACIONAL DE MINERAÇÃO. 2020. Cadastro Nacional de Barragens de Mineração. http://www.anm.gov.br/assuntos/barragens/relatorios-anuais-de-seguranca-da-barragens-de-mineracao-1/relatorio-anual-gsbm-2019-v-final.

BABAKI B, GUNS T & NIJSSEN S. 2014. Constrained Clustering Using Column Generation. In Int Conf AI and OR Tech in C Prog Opt Probl, p. 438-454.

BEAN JC. 1994. Genetic Algorithms and Random Keys for Sequencing and Optimization. ORSA J Comp 6: 107-220.

BERKHIM P. 2006. Survey of Clustering Data Mining Techniques. In: Grouping multidimensional data, Berlin: Springer, p. 25-75.

BIRD G, BREWER PA, MACKLIN MG, BALTEANU D, SERBAN M, DRIGA B & ZAHARIA S. 2008. River System Recovery Following the Novaţ-Roşu Tailings Dam Failure, Maramureş County, Romania. J Appl Geochem 23(12): 3498-3518.

CARMO FF ET AL. 2017. Fundão Tailings Dam Failures: The Environment Tragedy of the Largest Technological Disaster of Brazilian Mining in Global Context. Perspect Ecol Conserv 15(3): 145-151.

CETEM. 2016. Rompimento de Barragem de Rejeitos de Mineração de Ferro Em Itabirito (MG) Provoca Mortes. Http://Verbetes.Cetem.Gov.Br/Verbetes/ExibeVerbete.Aspx?Verid=209.

GLOTOV V, CHLACHULA J, GLOTOVA LP & LITTLE E. 2018. Causes and Environmental Impact of the Gold-Tailings Dam Failure at Karamken, the Russian Far East. Eng Geol 245: 236-247.

GONÇALVES JF & RESENDE MGC. 2011. Biased Random-Key Genetic Algorithms for Combinatorial Optimization. J Heuristics 17(5): 487-525.

GÖRANSSON T, BENCKERT A, LINDVALL M & RITZ R. 2001. Dam Failure at the Aitik Mine: Investigations, Conclusions and Measures Taken. In *Proceedings of Securing the Future: International Conference on Mining and the Environment*. Skelleftea, Sweden.

ISHIHARA K, KENNOSUKE U, YAMADA S, YASUDA S & YONEOKA T. 2015. Breach of a Tailings Dam in the 2011 Earthquake in Japan. Soil Dyn Earthq Eng 68: 3-22.

LÓPEZ-PAMO E, BARETTINO D, ANTÓN-PACHECO C, ORTIZ G, ARRÁNZ JC, GUMIEL JC, MARTÍNEZ-PLEDEL B, APARICIO M & MONTOUTO O. 1999. The Extent of the Aznalcollar Pyritic Sludge Spill and Its Effects on Soils. Sci Total Environ 242(1-3): 57-88.

MARTINHO MA, MENDES R, AMORIM FN, CIRANO M & DIAS JM. 2016. Fundão Dam Collapse: Oceanic Dispersion of River Doce after the Greatest Brazilian Environmental Accident. Mar Pollut Bull 112(1-2): 359-364.

NEVES LP, GONÇALVES JÚNIOR ES, SANTOS AAG, SOUSA GDB, SANTOS ACB, CRUZ CO, AMORIM MC, SANTOS JB & FREITAS AR. 2020. Report Semanal Barragens de Mineração 02 - 08/06/2020. http://www.anm.gov.br/assuntos/barragens/boletim-semanal-de-barragens-de-mineracao/boletim-semanal-2020-06-08-v5.

OWEN JR, KEMP D, LÈBRE E, SVOBODOVA K & MURILLO GP. 2020. Catastrophic Tailings Dam Failures and Disaster Risk Disclosure. IJDRR 42: 101361.

PETTICREW EL ET AL. 2015. The Impact of a Catastrophic Mine Tailings Impoundment Spill into One of North America's Largest Fjord Lakes: Quesnel Lake, British Columbia, Canada. Geophys Res Lett 42(9): 3347-3355.

PIPLINKS. 2007. Chronology of Tailings Dam Failures in the Philippines. http://www.piplinks.org/.

RICO M, BENITO G, SALGUEIRO AR, DÍEZ-HERRERO A & PEREIRA HG. 2008. Reported Tailings Dam Failures. A Review of the European Incidents in the Worldwide Context. J Hazard Mater 152(2): 846-852.

RUYTERS S, MERTENS J, VASSILIEVA E, DEHANDSHUTTER B, POFFIJN A & SMOLDERS E. 2011. The Red Mud Accident in Ajka (Hungary): Plant Toxicity and Trace Metal Bioavailability in Red Mud Contaminated Soil. Environ Sci Technol 45(4): 1616-1622.

SOLDÁN P, PAVONIČ M, BOUČEK J & KOKEŠ J. 2001. Baia Mare Accident - Brief Ecotoxicological Report of Czech Experts. Ecotoxicol Environ Saf 49(3): 255-261.

SPEARS VM & JONG KAD. 1991. On the Virtues of Parameterized Uniform Crossover. In Proceedings of the Fourth International Conference on Genetic Algorithms, p. 230-236.

THOMPSON F ET AL. 2020. Severe Impacts of the Brumadinho Dam Failure (Minas Gerais, Brazil) on the Water Quality of the Paraopeba River. Sci Total Environ 705: 135914.

THORNDIKE RL. 1953. "Who Belongs in a Family?" Psychometrica 18: 267-276.

VALE SA. 2019. "Brumadinho." http://Www.Vale.Com/Esg/En/Pages/Brumadinho.Aspx. 2019.

How to cite

SANTOS TB & OLIVEIRA RM. 2021. Failure risk of brazilian tailings dams: a data mining approach. An Acad Bras Cienc 93: e20201242. DOI 10.1590/0001-3765202120201242.

Manuscript received on August 7, 2020;
accepted for publication on October 10, 2020

TATIANA B. SANTOS[1]
https://orcid.org/0000-0001-5484-6675

RUDINEI M. OLIVEIRA[2]
https://orcid.org/0000-0001-9571-7423

[1]Programa de Pós Graduação em Engenharia Mineral, Universidade Federal de Ouro Preto, Departamento de Engenharia de Minas, Campus Universitário, s/n, Morro do Cruzeiro, 35400-000 Ouro Preto, MG, Brazil

[2]Departamento de Ciências Exatas, Universidade do Estado de Minas Gerais, Av. Brasília, 1304, Baú, 35930-314 João Monlevade, MG, Brazil

Correspondence to: **Tatiana B. Santos**
E-mail: tatiana.santos@ufop.edu.br

Author contributions

Tatiana Barreto dos Santos is responsible for the study conception, dataset construction, and Principal Component Analysis. Rudinei Martins de Oliveira is responsible for application of Biased Random Key Genetic Algorithm. Booth authors contributed with the paper writing.