Classification of Volcanic Rocks based on Rough Set Theory

Shaaban M. Shaaban
Department of Electrical Engineering
Northern Border University, Arar, Saudi Arabia and
Department of Basic Science Engineering
Menofia University, Menofia, Egypt
shabaan27@gmail.com

Sameh Z. Tawfik
Department of Petrographic Studies
Research Division
Nuclear Materials Authority
Cairo, Egypt
samehzakaria@outlook.com

Abstract—**Classification of volcanic rocks is a fundamental task in the geologic studies. Volcanic rocks are igneous rocks that cooled rapidly above the surface of the Earth's crust. They are classified according to their oxide chemical content. Furthermore, volcanic rocks can also be classified numerically by statistical means. But these methods are mostly dependent on human expert decision making and have a high cost. In this paper, a novel approach in the classification of volcanic rocks is proposed. This method is based on the rough set mathematical theory. The continuous data of the information system are firstly discretized using the information loss method. Secondly, the discretized decision table is reduced and the decision rule sets are extracted. The results are consistent with previous methods and show that the proposed method reduces time and calculation costs.**

Keywords-decision rules; information loss-discretization; rough set; volcanic rocks

I. INTRODUCTION

One established area in geology is the research of volcanic rocks. Mineral composition in volcanic rocks is affected by the chemical magma structure and the physical-chemical conditions during crystallization. Different approaches are used for volcanic rock classification based on genetic, textural or chemical composition [1]. The genetic method classifies rocks according to their form. This method is only an initial solution, it does not say anything about mineralogy, rock chemistry, and cannot discriminate between basalt and andesite. Texture methodology depends on the shape, size, and structure of the grain of different rock minerals. This method has the same limitations as a genetic classification [2]. Chemical classification requires a complete chemical analysis of the rock. Volcanic rocks are categorized on the basis of their mineral or chemical composition. They are classified into basic rocks, acidic rocks, intermediate rocks, and various types of ultrabasic rocks [3]. On the basis of mineral composition, igneous rocks are classified into silicic, intermediates, mafic, and ultramafic rocks [4]. Several methods to develop classification architectures have been suggested including Artificial Neural Networks (ANNs) [5-7], decision trees [8], statistical techniques [9-10], and decision-making rules [11-12]. Rough Set (RS) theory is one of the most motivating areas of computational intelligent research, having become increasingly

popular with geologic applications. One of the main advantages of RS is that additional information on data such as probability distribution or grade participation is not needed [13], and it doesn't need mean and covariance matrices calculation. It takes time for ANNs to achieve acceptable accuracy, and the decision tree takes more time for computations because it depends on entropy, while such training in RS is not required.

In this paper, the collected features were firstly discretized through information loss technique. Then, RS was implemented as a feature reduction approach and rules extraction method was utilized on the discretized decision table to execute classification. The suggested technique finishes with a minimum of chemical composition that has a direct effect on the classification of volcanic rocks. The results indicate that the introduced method reduces significantly feature dimensions and increases classification accuracy while the results are consistent with previous approaches. The obtained results show that the proposed method reduces time and calculation cost.

II. ROUGH SET THEORY

To extract volcanic rocks information effectively, a large number of characteristic data must be objectively filtered out. When the best combination of characteristic parameters is achieved, it can be used to identify volcanic rocks precisely. After evaluating many non-linearity computational methods, no further data or previous knowledge were found to be needed for RS theory [14]. It can exclude individual or unimportant characteristics to effectively reduce decision systems with the same database classification ability [15]. The study of geological and volcanic rock information based on RS is a sort of new solution to the mainly geological high-dimensional complex NP (Nondeterministic Polynomial) problems.

A. Information System

We take the identification and extraction problems of mineral material, such as classification of the volcanic rocks, as a restricted method expressed as $S = (U, A, V, f)$, where, U is the non-empty finite set of samples called universe, A is non-empty finite set of parameters, $V = \bigcup_{a \in A} V_a$, V_a are the different values of attribute a, $f : U \times A \rightarrow V$ is the information function between U and A, $A = C \bigcup D$, $C \bigcap D = \varnothing$, where C

and *D* are the condition attributes and decision attributes respectively. The information system is called decision table.

B. Indiscernible Relation

Let $B \subseteq A$ define a binary relation $IND_A(B)$ on the universe U, $x_i, x_j \in IND_A(B)$. If any $a \in B$, $a(x_i) = a(x_j)$, then x_i and x_j are indiscernible and the equivalence relation R_B is given by:

$$IND_A(B) = \left\{ (x_i, x_j) \in U^2 \middle| \forall a \in B \; a(x_i) = a(x_j) \right\} \quad (1)$$

C. Attribute Reduction

One way to reduce dimensions is to keep only the attributes that preserve the relationship of indiscernibility, i.e. the accuracy of classification. The same set of equivalence classes are provided by the selected set of attributes which can be accessed with the entire attribute set. The other attributes are redundant and can be reduced without affecting the precision of classification. Typically there are many subsets of such attributes known as reducts, mathematically, $B \subseteq A$. The element *a* is redundant attribute in *B* if $IND(B) = IND(B - \{a\})$. The core is the set of all attributes of decision table, which cannot be removed from knowledge in the reduction process where $CORE(B) = \cap RED(B)$.

III. DISCRETIZATION OF DATA

The real value of attributes should be quantified data when using the RS theory in dealing with information systems. Discretization means dividing the continuous attribute into numerous sections, replacing each with a discrete value. There are different methods for discretization of data such as frequency algorithm, clustering method, the Naive scalar algorithm, etc. We will use the discretization method in this paper dependent on loss of information. The steps of the algorithm are:

Step 1: Let *S* be the universe set and *X* a feature set, *S* is ordered by ascending order according to *X* values. The result after sorting is $x_1, x_2, x_3, ..., x_n$.

Step 2: Construct the initial interval distribution $I_1, I_2, I_3, ..., I_n$ according to the equation $[x_1, \frac{x_1+x_2}{2}), [\frac{x_1+x_2}{2}, \frac{x_2+x_3}{2}), [\frac{x_2+x_3}{2}, \frac{x_3+x_4}{2}), ..., [\frac{x_{n-1}+x_n}{2}, x_n]$. Then merge into an interval some neighboring intervals with the same parameter value of classification.

Step 3: Evaluate the loss of information for each m neighboring intervals according to:

$$E(I_{p+1}, I_{p+2}, I_{p+3}, ..., I_{p+m}) = \sum_{i=1}^{m} \frac{|I_{p+i}|}{|I|} E(I_{p+i}) \quad (2)$$

$$E(I) = E(\bigcup_{i=1}^{m} I_{p+i}) = -\sum_{i=1}^{m} p(D_i, I) \log p(D_i, I) \quad (3)$$

$$Inf.\_Loss = E(I) - E(I_{p+1}, I_{p+2}, I_{p+3}, ..., I_{p+m}) \quad (4)$$

where $|I_{p+i}|$, $|I|$ are the number and the values of the *X* parameter for these samples in the interval I_{p+i}, *I* respectively, $E(I_{p+i})$ is the identification information entropy of the interval I_{p+i}, and $p(D_i, I)$ is the average of the number of samples which is equivalent to the classification attribute value D_i to the total number of interval samples *I* .

Step 4: Select one neighboring merger interval that has the least information loss and thus get a new interval.

Step 5: Go to Step 3 when the information loss from the present step is less than *k* times the last stage. Then obtain the discretized samples.

IV. VOLCANIC ROCK CLASSIFICATION BASED ON RS THEORY

A. Extraction and Assessment of Parameters

The presented volcanic rocks are parts of the Egyptian basement along the Red Sea coast and southern Sinai. In this research, 710 samples of the volcanic rocks were collected. According to the average chemical composition of the samples, the average chemical composition component was selected. Materials such as silicon dioxide (SiO_2) *A1*, titanium dioxide (TiO_2) *A2*, and water (H_2O) *A13* were selected as the condition attributes, while there are 6 decision classes, namely basalt, basaltic andesite, andesite, dacite, rhyodacites, and rhyolites. The used samples are given in Table I. As volcanic rock parameters have continuous values, the original information should be transferred to quantized data. After that, the values of the condition factors were divided into several intervals in accordance to the information loss, for instance the parameter's *A1* break points are 51.865, 54.785, 61.725 and 71.92. The discretized attribute values are shown in Table II. Table I is converted into Table III according to the algorithm for quantization based on information loss. The rock types as decision attributes are expressed as {1, 2, 3, 4, 5, and 6}, as shown in Table II.

B. Volcanic Rock Type Classification-Rules

Once the decision table is established, the parameters should be reduced by applying the proposed approach and the rules for correlating information between features and volcanic rock types should be obtained. Based on the approach of rough-sets, the most important variables can be filtered from the initial variables, with the goal being to obtain the optimal combination of parameters (parameter structural optimization). Without losing the essential information that has a direct and indirect relation to study objects, this filtration can reduce the dimensions of space and simplify the system. Based on the RS analysis, the core parameter is $\{A_1\}$ and the reduction attribute set is $\{A_1, A_3, A_{10}, A_{11}\}$. The extracted rules are shown in Table IV. Consider the third rule of Table IV as an instance to illustrate the meaning of this approach. When attribute values are $51.865 \leq A1 < 54.785$, $3.525 \leq A10$ and $A11 < 2.015$, the rock type is classified as 2 (basaltic andesite). "—" in the Table implies that the property value is indispensable.

TABLE I.   DOKHAN ROCK TYPE DECISION TABLE

No.	Condition Attributes													Decision Attribute
	A1	*A2*	*A3*	*A4*	*A5*	*A6*	*A7*	*A8*	*A9*	*A10*	*A11*	*A12*	*A13*	D
	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	FeO	FeO^t	MnO	MgO	CaO	Na_2O	K_2O	P_2O_5	H_2O	Rock type
1	49.94	0.67	14.48	3.95	4.92	8.48	0.15	9.64	8.48	3.07	0.64	0.08	3.98	**Basalts**
2	53.59	1.89	12.94	5.19	8.84	13.52	0.12	1.99	6.15	3.34	1.57	0.29	4.39	**Basalts**
3	61.97	0.63	16.65	4.08	1.97	5.65	0.08	3.63	4.68	2.96	2.26	0.20	1.24	**Dacites**
4	68.79	0.13	13.63	3.72	0.00	3.35	0.10	0.85	4.82	3.71	1.75	0.00	2.60	**Rhyodacites**
710	74.33	0.27	12.32	2.25	0	2.25	0.05	0.28	1.03	3.5	3.51	0.04	0.77	**Rhyolites**

TABLE II.   ATTRIBUTES DISCRETIZATION TABLE

Attribute			Discretization code
Condition Attributes	*A1*	SiO_2	0: $A1<51.865$ 1: $51.865 \le A1 < 54.785$ 2: $54.785 \le A1 < 61.725$ 3: $61.725 \le A1 < 71.92$ 4: $71.92 \le A1$
	A2	TiO_2	0: $A2<1.25$ 1: $1.25 \le A2$
	A3	Al_2O_3	0: $A3<15.53$ 1: $15.53 \le A3$
	A4	Fe_2O_3	0: $A4<10.55$ 1: $10.55 \le A3$
	A5	FeO	0: $A5<7.46$ 1: $7.46 \le A5$
	A6	FeO^t	0: $A6<6.586$ 1: $6.586 \le A6$
	A7	MnO	0: $A7<0.125$ 1: $0.125 \le A7$
	A8	MgO	0: $A8<5.753$ 1: $5.753 \le A8$
	A9	CaO	0: $A9<7.55$ 1: $7.55 \le A9$
	A10	Na_2O	0: $A10<3.525$ 1: $3.525 \le A10$
	A11	K_2O	0: $A11<2.015$ 1: $2.015 \le A11$
	A12	P_2O_5	0: $A12<0.756$ 1: $1.756 \le A12$
	A13	H_2O	0: $A13<1.886$ 1: $1.886 \le A13$
Decision	D	Rock type	1: Basalts 2: Basaltic Andesite 3: Andesites 4: Dacite 5: Rhyodacite 6: Rhyolites

V.   COMPARING RS AND LINEAR REGRESSION METHOD

The linear regression approach represents the linear relationship between independent and dependent parameters.

This approach presents the interaction as an equation that combines the condition attributes with the decision parameter. The decision variable is given by *Y* and the condition attributes by $x_1, x_2, x_3, ..., x_n$, where *n* gives the number of condition attributes [15]. The relationship among *Y* and $x_1, x_2, x_3, ..., x_n$, is estimated by:

$$Y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + ... a_n x_n \quad (5)$$

where, $a_0, a_1, a_2, ..., a_n$ represent the regression coefficients. In this article, an attempt was made to define the best chemical combination of minerals for rock type classification, so the dependent parameter is the volcanic rock type classification and the independent parameters are volcanic rock characteristics. Nevertheless, problems remain in the formulation of regression equations used to choose the best chemical combination, since it is technically challenging or even impractical to use all parameters and variables to construct the regression equation. Therefore, the stepwise approach is used to find the best arrangement of attributes. In this methodology, various parameters are utilized to build up the best linear relationship with the most significant estimation of R^2 by different condition variables. In this strategy, at first the magnitude of the correlation coefficient is determined between every independent and dependent variable. This is done to find which condition parameter can provide the greatest degree of correlation with the dependent parameter. This situation continues until the best second variable is established for the independent attributes. This procedure proceeds until the expansion of another condition parameter to the model negligibly affects R^2. Therefore the provided parameters are considered to be the most important defined parameters for volcanic rock classification in a linear regression equation that is developed in these steps. Table V shows the results of the regression equations. The results show that the RS model has the highest R^2. Therefore, in the RS method the accuracy and precision of the approximations are better.

TABLE III.   DISCRETIZED DECISION TABLE

No.	Condition Attributes													Decision attribute
	A1	*A2*	*A3*	*A4*	*A5*	*A6*	*A7*	*A8*	*A9*	*A10*	*A11*	*A12*	*A13*	D
	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	FeO	FeO^t	MnO	MgO	CaO	Na_2O	K_2O	P_2O_5	H_2O	Rock type
1	0	0	0	0	0	1	1	1	1	0	0	0	1	1
2	1	1	0	0	1	1	0	0	0	1	0	0	1	1
3	3	0	1	0	0	1	0	0	0	1	1	0	0	4
4	3	0	0	0	0	0	0	0	0	0	0	0	1	5
715	4	0	0	0	0	0	0	0	0	1	1	0	0	6

TABLE IV.   DOKHAN ROCK TYPE DECISION RULES

Rule No.	Condition attribute value				Decision attribute value	Certainty	Coverage
	A1	*A3*	*A10*	*A11*			
1	0	---	---	---	1	1	0.75
2	1	0	---	---	1	1	0.25
3	1	---	1	0	2	1	0.5
---	---	---	---	---	---	---	---

TABLE V.   STEPWISE REGRESSION EQUATIONS

Step	Parameters	Equation	R^2
1	SiO_2	Rock_type= -9.38+0.208*A1*	92.1%
2	K_2O	Rock_type= -9.561+0.207*A1*+0.074*A10*	94.4%
3	Na_2O	Rock_type= -9.642+0.208*A1*+0.081*A10*+0.022*A11*	95.6%
4	Al_2O_3	Rock_type= -9.645+0.208*A1*+0.082*A10*+0.021*A11*+0.003*A2*	96.7%
Rough set result		Rock_type= -9.741+0.213*A1*+0.0913*A10*+0.023*A11*+0.001*A3*	96.85%

VI.   CONCLUSIONS

This study suggested a methodology for classifying volcanic rocks based on rough set theory. According to rough set theory, the factors that influence the classification of volcanic rocks were examined through reduced attributes and the four principal variables which influence the classification of volcanic rocks were silicon dioxide (SiO_2), titanium aluminum oxide (Al_2O_3), sodium oxide (Na_2O) and potassium oxide (K_2O). The implementation of this approach to volcanic rocks has shown that the method is realistic, workable and gives guidelines for future projects. In addition, the obtained results demonstrate that the proposed method reduces time and calculation cost.

REFERENCES

[1] M. Meschede, "A method of discriminating between different types of midocean ridge basalts and continental tholeiites with the Nb-Zr-Y diagram", Chemical Geology, Vol. 56, No. 3-4, pp. 207-218, 1986

[2] M. J. Le Bas, A. L. Streckeisen, "The IUGS systematics of igneous rocks", Journal of the Geological Society, Vol. 148, No. 5, pp. 825-833, 1991

[3] B. R. Frost, C. D. Frost, "A geochemical classification for feldspathic igneous rocks", Journal of Petrology, Vol. 49, No. 11, pp. 1955-1969, 2008

[4] F. Tiecher, M. E. B. Gomes, D. C. C. Dal Molin, "Alkali-aggregate reaction: A study of the influence of the petrographic characteristics of volcanic rocks", Engineering, Technology & Applied Science Research, Vol. 8, No. 1, pp. 2399-2404, 2018

[5] Z. Wei, H. Hu, H. W. Zhou, A. Lau, "Characterizing rock facies using machine learning algorithm based on a convolutional neural network and data padding strategy", Pure and Applied Geophysics, Vol. 176, No. 8, pp. 3593-3605, 2019

[6] G. Cheng, J. Yang, Q. Huang, Y. Liu, "Rock image classification recognition based on probabilistic neural networks", Science Technology and Engineering, Vol. 13, pp. 9231-9235, 2013

[7] G. Cheng, W. Guo, P. Fan, "Study on rock image classification based on convolution neural network", Journal of Xi'an Shiyou University (Natural Science Edition), Vol. 32, No. 4, pp. 116-122, 2017

[8] Y. Pu, D. B. Apel, B. Lingga, "Rockburst prediction in kimberlite using decision tree with incomplete data", Journal of Sustainable Mining, Vol. 17, pp. 158-165, 2018

[9] Y. Pu, D. B. Apel, B. Lingga, "Regression analysis and neural network fitting of rock mass classification systems", Journal of Science and Engineering, Vol. 20, No. 59, pp. 354-368, 2018

[10] R. W. Le Maitre, "A new approach to the classification of igneous rocks using the basalt-andesite-dacite-rhyolite suite as an example", Contributions to Mineralogy and Petrology, Vol. 56, pp. 191-203, 1976

[11] T. Miranda, L. R. Sousa, A. T. Gomes, J. Tinoco, C. Ferreira, "Geomechanical characterization of volcanic rocks using empirical systems and data mining techniques", Journal of Rock Mechanics and Geotechnical Engineering, Vol. 10, No. 1, pp. 138-150, 2018

[12] C. A. Ozturk, E. Nasuf , "Strength classification of rock material based on textural properties", Tunnelling and Underground Space Technology, Vol. 37, pp. 45-54, 2013

[13] Z. Pawlak, Rough sets: Theoretical aspects of reasoning about data, Kluwer Academic Publishers, 1991

[14] S. M. Shaaban, "Application rough set theory and decision network as a new approach to simplify airline hubs network location", International Journal of Intelligent Engineering and Systems, Vol. 9, No. 2, pp. 1-7, 2016

[15] S. M. Shaaban, H. A. Nabwey, "Rehabilitation and reconstruction of asphalts pavement decision making based on rough set theory", Computational Science and its Applications–ICCSA 2012, Lecture Notes in Computer Science, Vol. 7334, pp. 316-330, Springer, 2012

[16] W. M. A. W. Ahmad, R. A. A. Rohim, N. H. Ismail, "Forecasting parameter estimates: A modeling approach using exponential and linear regression", Engineering, Technology & Applied Science Research, Vol. 8, No. 4, pp. 3162-3167, 2018