EVALUATING CLASSIFICATION EFFECTIVENESS OF SEQUENTIAL MINIMAL OPTIMIZATION (SMO) ALGORITHM ON CHEMICAL PARAMETIZATION OF GRANITOIDS

Olalekan S. Akinola[1] & Adetoun C. Afolabi[2]
Computer Science Department,University of Ibadan, Nigeria
[1]solom202@yahoo.co.uk, [2]tounafolabi@yahoo.com

ABSTRACT
Geologists have been facing the problem of identification and correct classification of rock types since time immemorial. In this study, we employed Sequential Minimal Optimization (SMO) data mining algorithm on the classification of granitoids such as granites. Granitoid data were obtained from previous works of thesis, projects and journals. The data were then subjected to data mining activities using the SMO. Briefly, results obtained from the study shows that SMO algorithm is suitable and reliable for classifying granitoid rock types based on the chemical composition up to 97% accuracy.

Keywords: *Granitoids, Rocks, Sequential Minimal optimization, Data Mining, Classification*.

1. INTRODUCTION

Identification and classification of rock types is a task that many geologists (especially petrologists) have been unravelling for over a century now. Granitoids include rocks of a wide range of composition which are; Granites, Granodiorites, Sodic granites, Charnockites, Quartz monzonites, Syenites, and Pegmatites. The granites having received a large attention by many workers are classified quite variedly into i) A type ii) S type iii) I type iv) M type v) E type vi) C type vii) G type. These divisions are based on the variations observed in the chemical properties of the granites [1].

Data mining techniques detect patterns in large volumes of data and use these patterns to predict future instances in similar data. It is currently being used in a wide range of profiling practices such as marketing, surveillance, fraud detection, and scientific discovery. This study seeks to achieve the classification of rock types within the granitoid system based on chemical composition using Sequential Minimal Optimization (SMO) data mining techniques.

The rest of this paper is divided as follows: In section 2, we present the lierature review and in Section 3, the methodology used for this study was presented. Discussion of our results is presented in Section 4 while Section 5 concludes the paper.

2. RELATED WORKS

Data mining is the term used to describe the process of extracting value from a database. According to Moxon [2]: "Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large volumes of data, using pattern recognition technologies as well as statistical and mathematical techniques. It is a "knowledge discovery process of extracting previously unknown, actionable information from very large databases." It can also be defined as "The nontrivial extraction of implicit, previously unknown, and potentially useful information from data" [3]. A later definition slightly expanded on this, defining data mining as the application of various algorithms for finding patterns or relationship in a data set [4].

Data mining models can be categorized according to the tasks they perform which are Classification, Prediction, Clustering, and Association Rules. The most common action in data mining is classification. It recognizes patterns that describe the group to which an item belongs. It does this by examining existing items that have already been classified and inferring a set of rules. Similar to classification is clustering. The major difference being that no groups are being predefined.

Prediction is the construction and use of a model to assess the class of an unlabeled object or to assess the value or range of values that a given object is likely to have. The next application is forecasting. This is different from predictions because it estimates the future value of continuous variables based on patterns within the data.

There are a number of data mining techniques with different levels of analysis and they are largely categorized as either descriptive or predictive. Descriptive data mining intends to summarize data and to highlight the interesting properties of the data, while predictive data mining aims at building models to forecast future behaviour [5]. They are as follows:

- **Artificial Neural Networks:** Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- **Genetic Algorithms**: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.
- **Decision Trees**: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.
- **Nearest Neighbor Method/*k*- Nearest Neighbor**: In pattern recognition, the *k*-nearest neighbor algorithm (*k*-NN) is a method for classifying objects based on closest training examples in the feature space.
- **Rule Induction**: The extraction of useful *if-then* rules from data based on statistical significance.
- **Data Visualization**: The visual interpretation of complex relationships in multidimensional data.

Not much has been done in the application of data mining techniques to classification of rock types. However, Steckhan and Roman [6] were able to apply ANN in the investigation of petrographical composition of reservoirs and wells and in identifying different rock types therein before petrophysical properties of such wells can be accurately evaluated. This is because the variety of different rock-building minerals was too large to be handled by a standard multi-mineral model, and also the responses of conventional logging measurements have not been calibrated to most of the minerals present in the basement.

Considering the practice of rock engineering, an attempt has been made to implement Artificial Neural Network (ANN) for the concept and method of rock mass blastability classification. A set of rock mass blastability data has been used for neural network training and testing. The model classification technique of ANN has been used to classify the rock mass blastability rank. The better ANN model for rock mass blastability classification is given and described after the ANN training and optimizing. The given ANN model is effective and of good stability and adaptability when used to give the result of rock mass blastability classification. It is concluded from this study that the ANN-based rock mass blastability classification can be developed well by proper training and learning algorithms based on a comprehensive data set, and the importance of every influencing factor can be directly got from its different net-weight [7].

Sequential Minimal Optimization (SMO) is a new algorithm for training Support Vector Machines (SVMs). Training a support vector machine requires the solution of a very large quadratic programming (QP) optimization problem. SMO breaks this large QP problem into a series of smallest possible QP problems. These small QP problems are solved analytically, which avoids using a time-consuming numerical QP optimization as an inner loop. The amount of memory required for SMO is linear in the training set size, which allows SMO to handle very large training sets. Because matrix computation is avoided, SMO scales somewhere between linear and quadratic in the training set size for various test problems, while the standard chunking SVM algorithm scales somewhere between linear and cubic in the training set size. SMO's computation time is dominated by SVM evaluation, hence SMO is fastest for linear SVMs and sparse data sets [8].

It is discovered that there exists a lot of variance within the granitoid system and that a single magma melt of heterogeneous composition has the ability to produce many varied rock types. What is more interesting is that during the process, batches of magma melts of largely uniform composition further produce rocks of varied composition (although slightly related) which could be quite tasking to identify and classify [1, 9, 10, 11, 12, 13]. Hence, the need to deploy SMO data mining techniques to alleviate the classification problem.

SMO has not been applied to any rock classification task just because it is a newly developed and evolving algorithm. We therefore evaluate the performance of this algorithm on the classification of granitoid rock system.

3. METHODOLOGY

Figure 1 represents the model for carrying out data mining task on a massive, historical data.

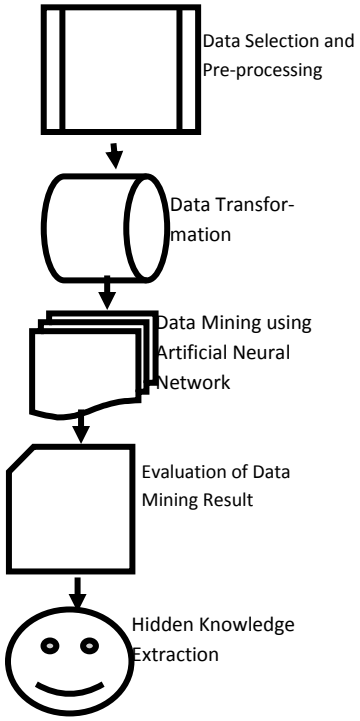

Figure 1: Data Mining Model

- **Data Collection**

The data used for this research was gathered from previous works of thesis, projects and journals. 191 data was used in all. 66% (126) of which was used for training and 34% (65) was used for testing. Out of the 65 data for testing, 8 was ignored because they could not be classified, while the remaining 57 was useful.

The dataset is composed of the chemical properties of a typical granitoid and it has sixty-seven (67) attributes, their types and description. Part of attributes are shown in Table 1.

Table 1: Part Description of Attributes of the Granitoid Dataset

	Attribute	Type	Description
1.	SiO_2	Numeric	Silica
2	TiO	Nominal	Titanium Oxide
3.	Al_2O_3	Numeric	Alumina

- **Data Transformation**

This is also referred to as data consolidation. Data is transformed into the format appropriate for the proposed data mining technique. The data files were saved in Comma Separated Value (CSV) file format. The dataset was normalized to reduce scalability on the data.

- **Stages of Data Mining**

The data mining stage was divided into three phases. At each phase, the algorithms were all experimented using the granitoid datasets. The testing method adopted for this research was percentage split that train on a percentage of the dataset, cross validate on it and test on the remaining percentage.

- **Training Phase**

At this phase, 126 (66%) of the granitoid dataset was used to build the model using all the classifiers.

- **Validation Phase**

This is one of the ways of determining the performance of a classifier. In this case, a number of folds (n) is specified, the dataset is randomly reordered and then split into n folds of equal size. In each iteration, one fold is used for testing and the other n-1 folds are used for training the classifier. The test results were collected and averaged over all folds. This gives the cross-validation estimate of the accuracy. The folds can be purely random or slightly modified to create the same class distributions in each fold as in the complete dataset. In this study, the validation phase was carried out using ten folds of the training sets.

- **Testing Phase**

The models were tested with the remaining 65 (34%) data, which was randomly selected from the dataset.

3.1 Sequential Minimal Optimization (SMO) Algorithm

Sequential Minimal Optimization (**SMO**) is an algorithm for efficiently solving the optimization problem which arises during the training of support vector machines. It was invented by John Platt in 1998 at Microsoft Research. SMO is widely used for training support vector machines and is implemented by the popular libsvm tool [14] and [15]. The publication of the SMO algorithm in 1998 has generated a lot of excitement in the SVM community, as previously available methods for SVM training were much more complex and required expensive third-party Quadratic Programming (QP) solvers [16].

Consider a binary classification problem with a dataset (x_1, y_1), ..., (x_n, y_n), where x_i is an input vector and $y_i \in \{-1, +1\}$ is a binary label corresponding to it. A soft-margin support vector machine is trained by solving a quadratic programming problem, which is expressed in the dual form as follows:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j K(x_i, x_j) \alpha_i \alpha_j,$$

subject to:

$$0 \leq \alpha_i \leq C, \quad \text{for } i = 1, 2, \ldots, n,$$
$$\sum_{i=1}^{n} y_i \alpha_i = 0$$

where C is an SVM hyperparameter and $K(x_i, x_j)$ is the kernel function, both supplied by the user; and the variables α_i are Lagrange multipliers.

SMO is an iterative algorithm for solving the optimization problem described above. SMO breaks this problem into a series of smallest possible sub-problems, which are then solved analytically. Because of the linear equality constraint involving the Lagrange multipliers α_i, the smallest possible problem involves two such multipliers. Then, for any two multipliers α_1 and α_2, the constraints are reduced to:

$$0 \leq \alpha_1, \alpha_2 \leq C,$$
$$y_1 \alpha_1 + y_2 \alpha_2 = k$$

and this reduced problem can be solved analytically.

The algorithm proceeds as follows:

1. Find a Lagrange multiplier α_1 that violates the Karush–Kuhn–Tucker (KKT) conditions [8] for the optimization problem.
2. Pick a second multiplier α_2 and optimize the pair (α_1, α_2).
3. Repeat steps 1 and 2 until convergence.

When all the Lagrange multipliers satisfy the KKT conditions (within a user-defined tolerance), the problem has been solved. Although this algorithm is guaranteed to converge, heuristics are used to choose the pair of multipliers so as to accelerate the rate of convergence.

3.2 Software Tool

The research was conducted using an open source software known as WEKA version 3.7.4. WEKA (Waikato Environment for Knowledge Analysis) is a collection of many data mining and machine learning algorithms. It was developed by a team of researchers at the University of Waikato in New Zealand and certified by IEEE.

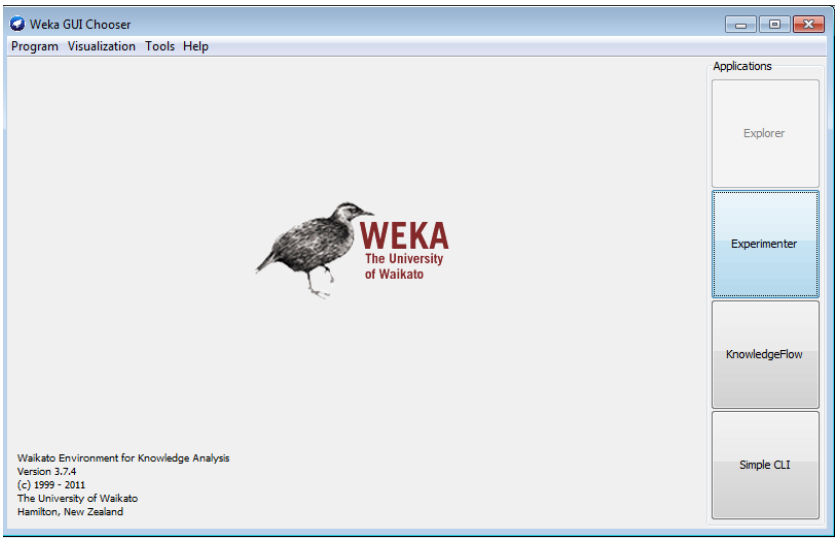

Figure 2: WEKA 3.7.4 GUI Chooser Interface

For the experiment, the Explorer application was used. Figure 3 displays the Explorer application window.

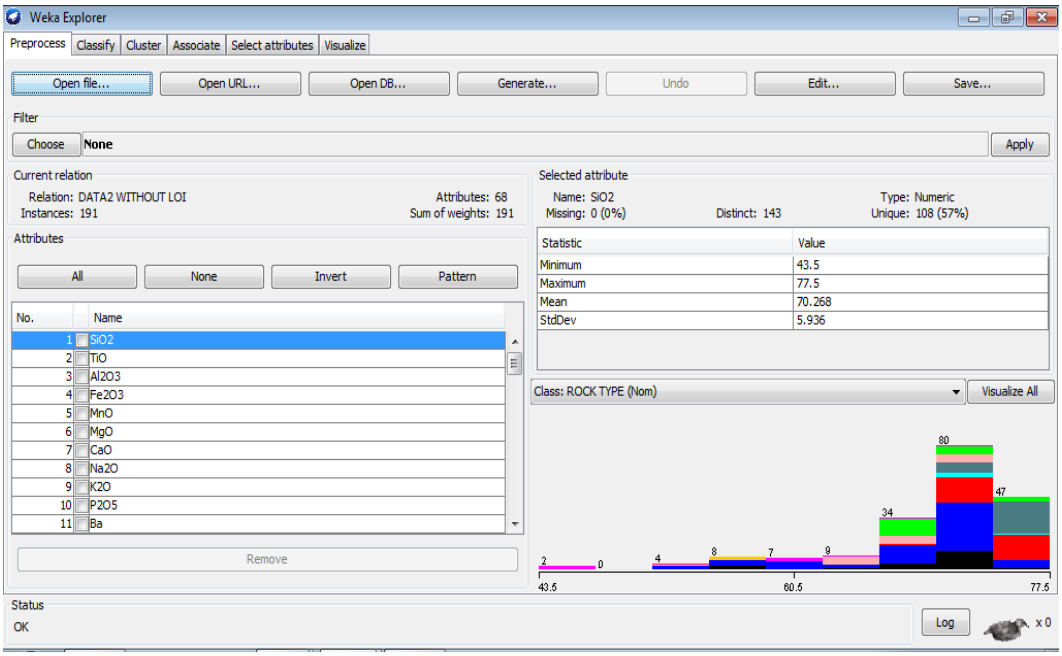

Figure 3: WEKA 3.7.4 Explorer Application Window

4. RESULTS

At the initial stage of the experiment, Expectation Maximization algorithm was used to get the clusters distribution for the granitoid dataset. Figure 4 shows part of the granitoid clusters. At the pre-processing stage of this experiment, the clusters generated from the granitoid dataset presented in Figure 4 shows that there is a high variability in patterns of the granitoid dataset. Therefore, there exists a need to employ appropriate tools and techniques that could be used to classify granite rocks with minimal error margins.

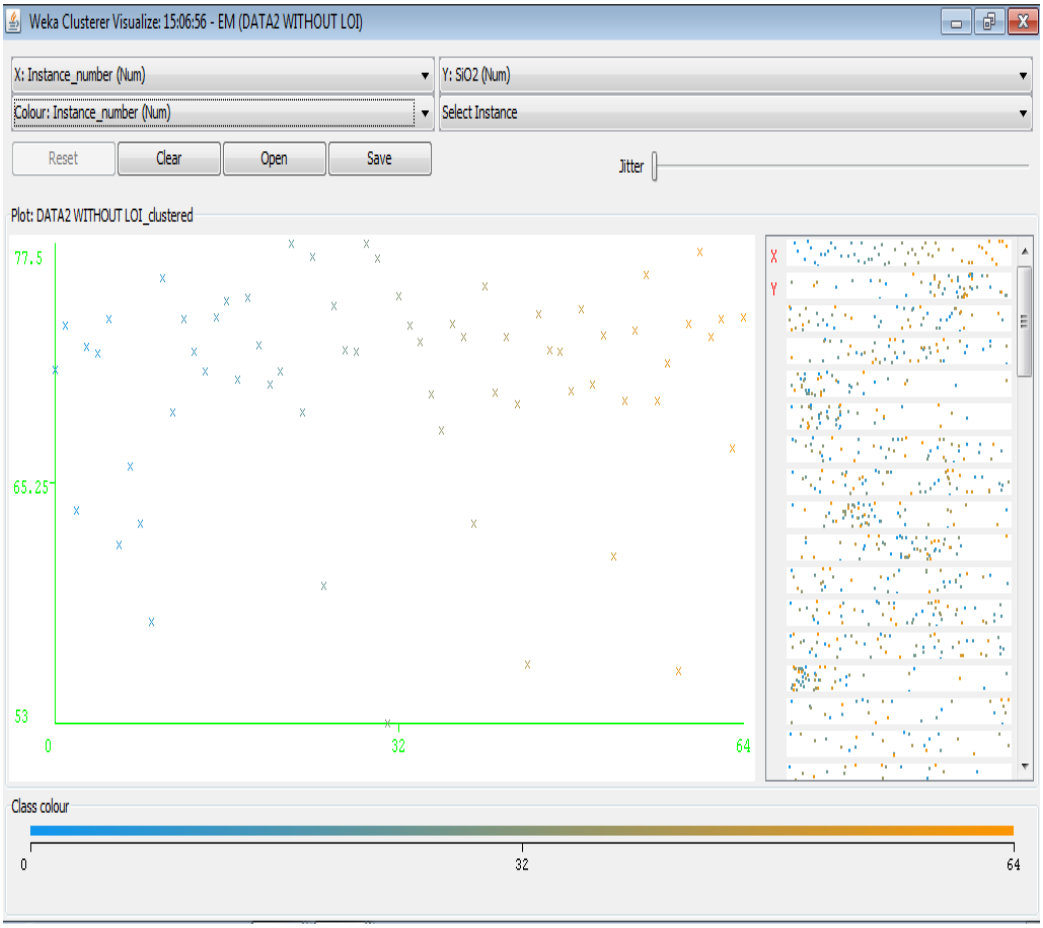

Figure 4: Granitoid Clusters

4.1 SMO Output for Granitoid Dataset Model

The results of performance of the algorithms are presented in Tables 5a and 5b while Figure 6 displays the confusion matrix.

Table 5a: Results of Modelling the Granitoid Dataset Using SMO Algorithm

Performance Measure	**Instances**	**SMO**
Time to Build Model (secs)		5.9secs
Percentage Classification	Correct	94.7368%
	Incorrect	5.2632%
Kappa Statistics		0.9328
MAE		0.1891
RMSE		0.2939
RAE (%)		94.9462%
RRSE (%)		92.7819%
Total No. Of Instances		57
Ignored Class Unknown Instances		8

Table 5b: More Results of Modelling of Granitoid Dataset Using SMO Algorithm

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0.053	0.905	1	0.95	0.902	Granite
	1	0	1	1	1	0.99	A-Type
	1	0	1	1	1	1	Sub-Alkali Granite
	1	0	1	1	1	1	Leucogranite
	1	0	1	1	1	1	Granodiorite
	1	0.02	0.857	1	0.923	0.992	S-Type
	0	0	0	0	0	1	Syenite
	0.5	0	1	0.5	0.667	0.906	Charnockite
Weighted Average	0.947	0.02	0.936	0.947	0.934	0.958	

Prediction

	a	b	c	d	e	f	g	h	<--- classified as
Actual	19	0	0	0	0	0	0	0	| a = GRANITE
	0	13	0	0	0	0	0	0	| b = A-TYPE
	0	0	1	0	0	0	0	0	|c = SUB- ALKALI GRANITE
	0	0	0	7	0	0	0	0	| d = LEUCOGRANITE
	0	0	0	0	6	0	0	0	| e = GRANODIORITE
	0	0	0	0	0	6	0	0	| f = S-TYPE
	0	0	0	0	0	1	0	0	| g = SYENITE
	2	0	0	0	0	0	0	2	| h = CHARNOCKITES

Figure 6: Confusion Matrix of Granitoid Dataset for SMO Algorithm

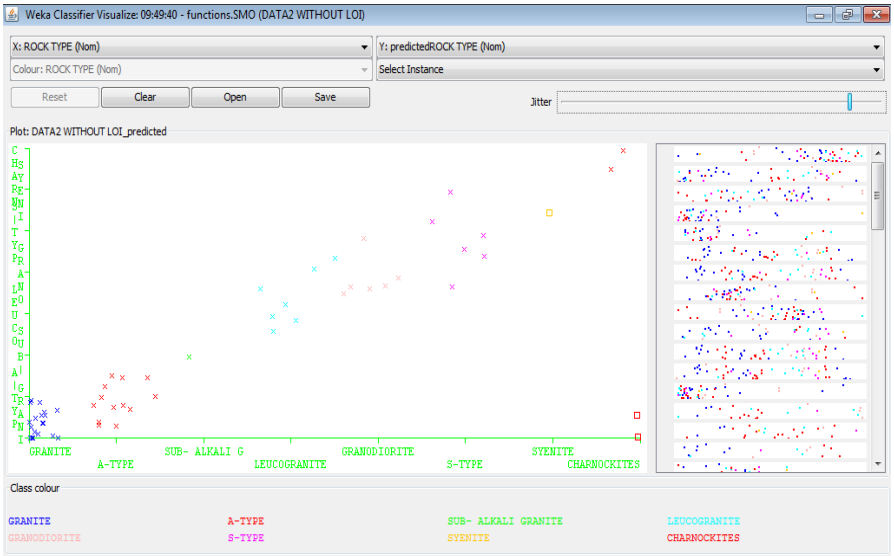

Figure 7: Chart Showing SMO Classifier Errors

4.2 Discussion of Results

Important results from this study shows that the time taken to build the classification model with SMO was 5.9 seconds and the percentage correct classification was 94.7%. This shows that Recall or True Positive Rate (*TP*), which is the proportion of positive cases that were correctly identified was very high. On the other hand, the False Positive Rate (FP), which is the proportion of negative cases that were incorrectly classified as positive was very small. TP is also regarded as the Precision, P.

The Kappa Statistics (KS) is a chance-corrected measure of agreement between the classifications and the true classes. It is calculated by taking the agreement expected by chance away from the observed agreement and dividing by the maximum possible agreement. A value greater than 0 means that the classifier is doing better than chance. 1.0 signifies complete agreement. Results from this study shows that the KS was 0.93. A high KS was thus obtained in the study.

ROC graphs (Relative Operating Characteristics) are another way to examine the performance of classifiers [17](Sweets, 1988). A ROC graph is a plot with the false positive rate on the *X* axis and the true positive rate on the *Y* axis. The point (0,1) is the perfect classifier: it classifies all positive cases and negative cases correctly. It is (0,1) because the false positive rate is 0 (none), and the true positive rate is 1 (all). The point (0,0) represents a classifier that predicts all cases to be negative, while the point (1,1) corresponds to a classifier that predicts every case to be positive. Point (1,0) is the classifier that is incorrect for all classifications. It has been suggested that the area beneath an ROC curve can be used as a measure of accuracy in many applications (Swets, 1988). The weighted average ROC area obtained in this study 0.958. This also shows the high performance measure for SMO classifier.

A confusion matrix is a plot used to evaluate the performance of a classifier during supervised learning. It is a matrix plot of the predicted versus the actual classes of the gene expression data. It can also be explained as an array showing relationships between true and predicted classes. Entries on the diagonal of the matrix count the correct calls. Entries off the diagonal count the misclassifications. Entries on the diagonal of the matrix obtained in this study shows that they were very high and that 'outliers' (misclassifications) were very small.

The Absolute Error (AE) is the magnitude of the difference between the exact value and the approximation. The absolute error of the measurement shows how large the error actually is, while the relative error of the measurement shows how large the error is in relation to the correct value. The relative error (RE) is the absolute error divided by the magnitude of the exact value. The mean squared error (MSE) of an estimator is one of many ways to quantify the difference between values implied by an estimator and the true values of the quantity being estimated.

MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. MSE measures the average of the squares of the errors. The error is the amount by which the value implied by the estimator differs from the quantity to be estimated. The difference occurs because of randomness or because the estimator does not account for information that could produce a more accurate estimate.

The mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The mean absolute error is given by

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i|\,.$$

As the name suggests, the mean absolute error is an average of the absolute errors $e_i = f_i - y_i$, where f_i is the prediction and y_i the true value.

MAE obtained in this study was 0.189 while the Root Mean Squared Error (RMSE) was 0.29. These values were very infinitesimal to undermine the high performance classification effectiveness of SMO.

The result of this research implies that geologists can have about 97% reliance on the SMO algorithm to classify rock types. Furthermore, SMO which is designed for SVMs can be employed when classifying datasets with many attributes as found in the granitoid dataset. The result showed that SMO algorithm is suitable and reliable for classifying granitoid rock types based on the chemical composition.

This result is in line with the claim of the originator of SMO, Platt John in 1988 [8] and Francis, *et al.* [18] that SMO is fastest for linear SVMs and sparse data sets. On real world sparse data sets, SMO can be more than 1000 times faster than the chunking algorithm.

5. CONCLUSION

The classification effectiveness of Sequential Minimal Optimization (SMO) data mining algorithm is demonstrated in this study. The results obtained from the study indicates that SMO can be effectively used for classifying granitoid rock types.

6. REFERENCES

[1] Russel, B. H. (2004). The application of multivariate statistics and neural networks to the prediction of reservoir parameters using seismic attributes A Thesis in the Department of Geology and Geophysics, Calgary, Alberta, September, 2004, http://www.crewes.org/ResearchLinks/GraduateTheses/2004/Russell-PhD-2004.pdf Downloaded in May 2012.

[2] Moxon B. (1996). Defining Data Mining, *DBMS Journal*, Vol. 9, Issue 9, August 1996.

[3] Frawley, W., Piatetsky-Shapiro, G. and Matheus, C., (1992). Knowledge Discovery in Databases – An Overview, pp. 11 – 30.

[4] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., (1996). Advances in Knowledge Discovery and Data Mining, *Menlo Park, CA: AAAI Press.*

[5] Han, J. and Kamber, M. (2006). Data Mining: Concepts and Techniques, San Francisco: Morgan Kaufmann.

[6] Steckhan J. and Roman S. (2010). Introduction of a Rock Typing Methodology in Crystalline Basement Reservoirs (Yemen)*, Search and Discovery Article #40524 April 26, 2010, *Adapted from extended abstract prepared for presentation at *AAPG International Conference and Exhibition*, Rio de Janeiro, Brazil, November 15-18, 2009, OMV Exploration & Production GmbH, Vienna, Austria.

[7] Jiang H., Xu W. and Xie S. (2000). Artificial Neural Network Method of Rock Mass Blastability Classification, Research Institute of Geotechnical Engineering, Hohai University, Nanjing, Jiangshu, P.R.China 210098.

[8] Platt, J. (1998), Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.43.4376&rep=rep1&type=pdf visited in May 2012.

[9] Pearce, J. A., (1982). Trace Element Characteristics Of Lavas From Destructive Plate Bounderies. In Thorpe, R. S. (ed.), Andesites: Orogenic Andesites And Related Rocks. *Wiley, Chichester*, pp. 525 – 548.

[10] Mullen, E. D., (1983). Mn/TiO2/P2O5: A minor element discriminant for basaltic rocks of oceanic environments and its implications for petrogenesis. Earth Planet. Sci. Lett., 62, pp 53 – 62.

[11] Pearce, J. A. and Cann, J. R., (1973). Tectonic setting of basic volcanic rocks determined using trace element analysis. Earth Planet. Sci. Lett., 19, pp 290 – 300.

[12] Winter, J. D., (2001). An Introduction to Igneous and Metamorphic Petrology. Prentice-Hall Inc. Upper Saddle River, New Jersey 07458. pp 697.

[13] Ragland, P. C., (1989). Basic Analytical Petrology. Oxford University Press. New York.

[14] Chih-Chung C. and Chih-Jen L. (2012). *LIBSVM: a Library for Support Vector Machines*, *http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf* visited in May 2012.

[15] Luca Z. (2006). Parallel Software for Training Large Scale Support Vector Machines on Multiprocessor Systems, *Journal of Machine Learning Research* 7 (2006) 1467–1492.

[16] Rifkin, R. (2002), "Everything Old is New Again: a Fresh Look at Historical Approaches in Machine Learning", *Ph.D. thesis*: 18, http://dspace.mit.edu/handle/1721.1/17549 visited in May 2012.

[17] Swets, J. A., (1988). Measuring the accuracy of diagnostic systems, *Science* 3 June 1988.

[18] Francis R. B., Gert R. G. L. and Michael I. J. (2004). Fast Kernel Learning using Sequential Minimal Optimization, Report No. UCB/CSD-04-1307, February, 2004, Computer Science Division (EECS), University of California, Berkeley, California 94720. Vol. 240, no. 4857, pp. 1285-1293, *DOI:* 10.1126/science.3287615