Evaluation of Logistic Regression and Support Vector Machine Approaches for XRF Based Particle Sorting for a Copper Ore

Yang Xu[a, *], Bern Klein[a], Genzhuang Li[a], Bhushan Gopaluni[b]

[a]Norman B. Keevil Institute of Mining Engineering, University of British Columbia, 6350 Stores Road, Vancouver, BC, Canada V6T 1Z4

[b]Chemical and Biological Engineering, University of British Columbia, 2360 East Mall, Vancouver, BC Canada V6T 1Z3

ABSTRACT

The study is aimed at particle sorting at the Copper Mountain Mine using XRF. Possible applications include the rejection of barren material from mill feed, the rejection of pebbles in the SABC circuit or the recovery of valuable material from low grade stockpiles. It is recognized that XRF is a surface measurement that can detect copper but depending on several operational conditions, such as the orientation of the particle or mineral texture, the sensor spot may not see the copper. However, XRF also provides information about the concentrations of a range of elements in minerals that are associated with copper mineralization which can improve sorting. The study described herein is aimed at improving XRF sensor-based sorting by the introduction of logistics regression (LR)- and support vector machine (SVM)-based machine learning approaches. To solve the collinearity and dimensionality issues in the input variables, the authors propose a combined approach of principal component analysis (PCA) and stepwise regression to extract the significant features. The combined PCA and stepwise regression approach is novel and has shown to be very effective for dimensionality reduction of the XRF spectrum data. By applying the ROC and AUC), the LR and SVM models are compared. Results showed that the LR model with the AUC of 0.847 outperforms the SVM with kernel functions with respect to classification accuracy; especially for data sets with a small number of features. The improved classification accuracy should benefit the economic performance of the particle sorting system.

KEYWORDS

Sensor-based ore sorting, receiver operator curve, logistic regression, support vector machine, principal component analysis

ACKNOWLEDGEMENTS

The support of Copper Mountain Ltd and Mitacs (IT10533) for this study is greatly appreciated.

1. Introduction

1.1 Sensor Based Particle Sorting

During the past 20 years, significant research and development programs have studied the application of sensor based sorting for improving productivity in the mining industry [1]. Bamber demonstrated 20% energy

*Corresponding Author.
Email address: xuyangon123@gmail.com (Y, Xu), bklein@mining.ubc.ca (B. Klein), genzhuang.li@ubc.ca(G,Li) bhushan.gopaluni@ubc.ca (B, Gopaluni)

reduction and cost savings through pre-concentration using particle sorting across Ni/Cu mines in the Sudbury basin. The benefits were mainly attributed to reduced transport and comminution requirements with additional benefits of increased and more consistent grades to the process plant resulting in higher metal recoveries [2][3]. Other studies have demonstrated the sorting of pebbles in the semi-autogenous-ball mill-crusher (SABC) circuit to reject barren rock in the circulating load thereby reducing energy requirements and improving throughputs [4][5]. Another application involves the sorting of low grade stockpiles and waste dumps to recover additional metal [6][7].

The various types of sensors currently available for SBS have been reviewed elsewhere [8]. The present paper is specifically focused on the use of X-ray fluorescence (XRF) for particle sorting. This technique is based on the spectral analysis of secondary X-rays emitted from the surface of the sample due to irradiation by primary X-rays. The physical principle of XRF is based on the excitation of electrons in the atoms on the particle surface. When these electrons subsequently fall from the higher energy shell into the vacancies that they left behind, they emit X-rays that are characteristic of the specific element and its abundance, thereby enabling the quantification of the metal content for scanned particles [9][10].

An example of an XRF spectrum is presented in Figure 1, where the distribution of each detected element is shown according to its XRF energy, and the proportion of each element corresponds to its number of counts.

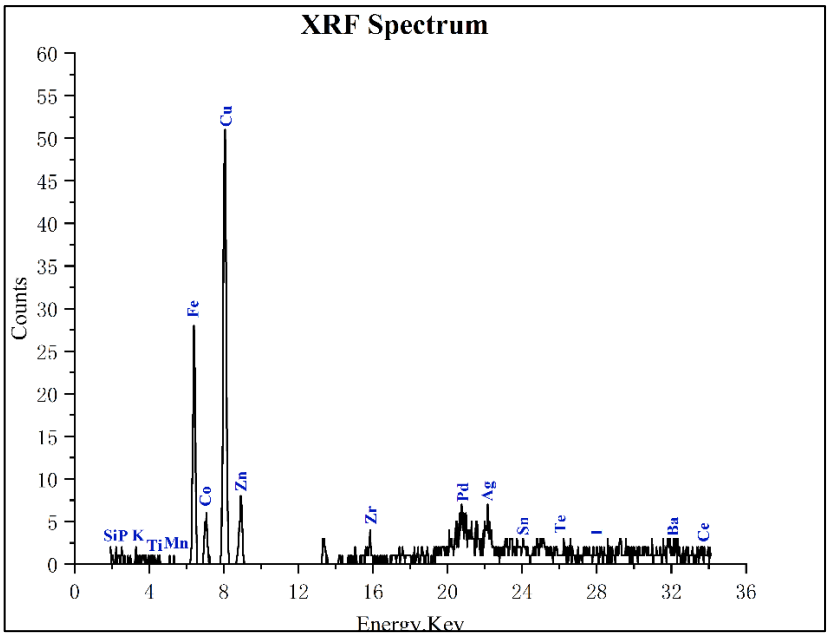

Figure 1: A typical XRF spectrum of an ore sample

Although studies have shown that SBS can significantly improve the productivity of mining operations, it is still regarded as a novel technology and, hence, there is insufficient confidence to secure financial support for greenfield projects that rely on SBS [11]. One challenge with the technology is that of accurately translating the sensor signals for classification, such as into classes of ore and waste.

Currently, sorting companies have been using simple linear regression algorithms that convert a single sensor response to a metal grade. However, there are cases when a spectrum of responses can be used to improve the linear correlation (multivariable linear regression). This is particularly important to XRF particle

sorting which measures a surface response. However, the surface XRF readings may not accurately represent the metal content in the whole rock. Also depending on the mineralization texture or orientation of the particle, the sensor may not see the metal. However, the sensor will detect elements that are found in the minerals associated with metal e.g., Cu and therefore serve as proxies to indicate Cu content.

At present, testing to assess particle sorting involves the collection of a range of particles (typically 100 to 300), to demonstrate the ability of a sensor to classify as ore or waste according to a threshold metal content. This stage of testing is followed by extensive and expensive pilot-scale testing requiring large samples. There is therefore a need to develop improved sorting algorithms that can be obtained using small sample sets; the outcome is particularly important for early-stage assessments such as scoping studies or for mineral exploration projects.

1.2 Review of Machine Learning Approaches

There have been some attempts to apply machine learning (ML) techniques to develop sorting algorithms [12][13]. The main difference between statistical methods and ML is their application purposes. Statistical modelling is focused on seeking the relationships between independent variables and dependent variables. By contrast, ML concentrates on prediction by using specified learning algorithms to find underlying patterns in large amounts of complex data. The machine learning methods can be effective even when the data are gathered without a carefully controlled experimental design and in the presence of complicated nonlinear interactions [14].

To take an example, XRF measures a surface response, which may not accurately represent the metal content of the whole rock. This issue cannot be resolved by conventional statistical methods (such as simple linear or multivariable linear regression), but potentially can be handled by the application of advanced machine learning techniques. There is also an opportunity to obtain more accurate correlations and improved decision-making with the machine learning approach [15].

In general, more than one model can be trained for a given machine learning application scenario. The model selection depends on the amount and quality of training data, the form of the relationship between the input and output variables, and computational constraints (iteration time and memory). For a specific application, it is important to review and compare the strengths and weaknesses of each approach to select the best approach.

Table 1 summarizes the advantages and limitations of different supervised machine learning algorithms. The LR, SVM, DT, RF and ANN are all commonly used as supervised learning models for classification problems. Gomez-Flores [16] presented four machine learning models including MLR, KNN, DT and RF to predict grade and recovery in flotation. The predicted performance of these four models were compared by use of mean absolute error and R2. The results showed that RF performed the best of the four models. However, RF is intrinsically suited for multiclass problems, while LR and SVM are more suitable for binary classification [17]. Also, RF performs well on large data sets because it is based on the theory of decision trees, however, it is highly prone to overfitting for data sets [18].

Table 1: Summary of the advantages and limitations of different supervised machine learning algorithms[19]

Algorithms	Applications	Advantages	Limitations
Decision Tree	investment decision	visual analysis; can be used in both classification and regression problems; fast calculation speed	overfitting; difficult to handle missing values and outliers
Random Forest	multi-classification	large size dataset; high accuracy	overfitting; unsuitable for small size dataset
KNN (k-nearest neighbour)	pattern recognition; multi-classification	online training; insensitive to outliers	not suitable for large-size databases; K value should be carefully set
SVM (Support vector machine)	text classification; imagine recognition; binary classification	small size dataset; high dimensional; strong generalization ability	sensitive to kernel function and outliers
Logistic Regression	binary classification	easy to understand and achieve; less calculation cost	underfitting; low Accuracy
ANN (Artificial neural networks)	voice recognition; computer vision	high accuracy; strong robustness	a large number of parameters; black box; long training time

DT models can be constructed relatively fast compared to other methods of classification. However, DT can only illustrate which predictors are essential for building the trees and does not provide information about the association path. Also, it is much more suited for large data sets, is more sensitive to missing values and outliers and is prone to overfitting [20][21].

ANN cannot indicate either the size or the direction of the relationship between input and output. This is due to the intrinsically multi-layered process and, in particular, the hidden layer between the input and output, where the function applies weights to the inputs before directing them to the output via an activation function [22].

LR is one of the most important nonlinear machine learning modelling methods which is used for the analysis and classification of categorical variables and has been widely used in many fields such as medicine, geology, biology, health care, and finance [23][24][25][26]. As the core of LR is a sigmoid function to transform linear regression into the logit function, it is more suitable for handling binary classification problems [27]. Specifically, LR gives a measure of both predictor relevance and the direction of the association, such that it generally provides a more informative output. However, the drawbacks are that it can underfit and have low accuracy.

KNN classification is a well-known decision rule that is widely used in pattern classification and recognition [28]. KNN is a typical nonparametric method which predicts the class of an object according to the class of its k nearest neighbour. The disadvantages of KNN are the computation of accurate distances as well as how to set K value [29].

The SVM algorithm is a comparatively effective machine learning method developed by Vapnik and Cortes [30] with the ability to analyze data and recognize patterns. Due to the advantages of the ability for

global optimization and a superior capacity to solve the overfitting problem, SVM has become the primary choice for binary and multi-class classification. Certain types of SVMs have been developed and applied successfully to text categorization, handwriting recognition, gene-function prediction, remote sensing classification and other applications [31][32]. It commonly uses small datasets but is sensitive to kernel functions as well as outliers. In addition, the performance of SVM largely depends on the selection of the kernel functions.

To summarize, different supervised machine learning algorithms were compared (Table 1). Considering that ore sorting is a typical binary classification scenario (where the feed to the sorter is classified as ore or waste), LR and SVM were considered more applicable for ore sorting applications. LR is capable of describing the relationship between the probability of a binary response variable and a set of corresponding explanatory variables. SVM has advantages because of its global optimized ability and a superior capacity to solve overfitting. Therefore, in this study, logistic regression and support vector machine were selected to develop the classification model for ore pre-concentration scenario.

2. Materials and Methodology

2.1 XRF Testing

Samples were collected from the Copper Mountain open pit mine in British Columbia, Canada. Copper Mountain mines an alkalic copper-gold porphyry ore with typical reserve grades of 0.25% Cu and 0.11 g/t Au. Mineralization occurs in veins, fracture fillings and dissemination within the volcanic rock and intrusive rocks. Copper minerals occur as chalcopyrite, bornite and chalcocite and the main gangue minerals associated with this mineralization are pyrite, magnetite and calcite.

The Copper Mountain mine processing plant was designed for a throughput of about 35,000 t/d to produce copper concentrate using SABC comminution and flotation. The present study supports the assessment of particle sorting for either pebble rejection in the SABC circuit or to recover value from low grade stockpiles.

A 400 kg sample of cone crusher product (plant feed material) was collected and screened into the size fractions as shown in Figure 2.

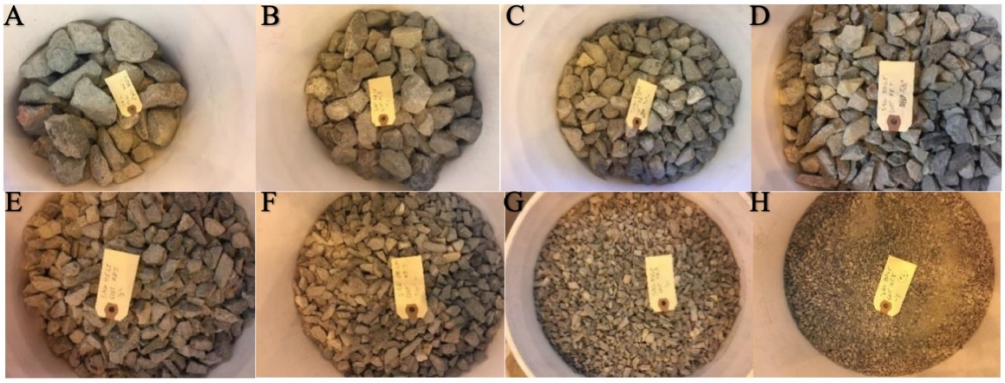

Figure 2: Screened size fractions of Copper Mountain mine plant feed. A is +37.5 mm, B is -37.5+25 mm, C is -25+19 mm, D is -19+16 mm, E is -16+12.5 mm, F is -12.5+9.5mm, G is -9.5+4.5 mm, H is -4.5 mm

The test procedure used is shown schematically in Figure 3. The three hundred particles in the -53+19mm size fraction were scanned using a handheld XRF unit (Olympus Alpha 6000) equipped with the Innov-X system to generate a spectrum for the scanned area. Each rock was scanned on four sides using an exposure time of 30s for each shot. After XRF scanning, the result was exported to a computer, and the grade of each element was obtained by averaging the grades from the four shots (sides). All rocks were then pulverized to -0.074mm and analyzed by XRF to determine the chemical compositions. The accuracy of the XRF on pulverized samples was validated on selected rocks using ICP, which was in good agreement. The main elements' ranges of concentration are presented in Table 2.

Table 2: Main elements' ranges of concentration based on chemical analysis

Element	Co (ppm)	Cr (ppm)	Cu (ppm)	Fe (%)	Mn (ppm)	P (ppm)	Pb (ppm)	Sr (ppm)	Ti (%)
Lowest	1	1	1	0.01	5	10	2	1	0.01
Highest	92	124	4540	6.17	420	2033	13	182	0.46

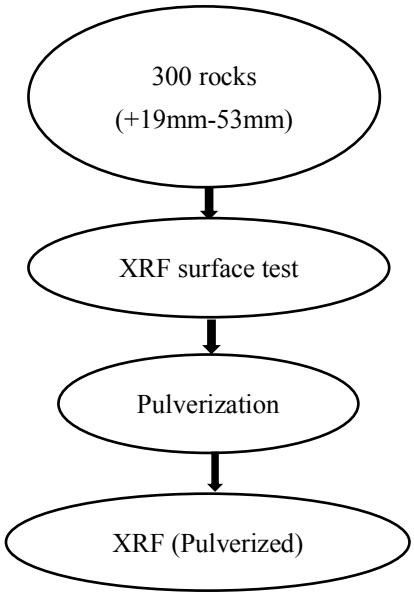

Figure 3: Test Procedure of samples collected from Copper Mountain

2.2 XRF Data Screening

The XRF sensor generates peaks for detectable elements. In some cases, peaks for more than one element may overlap. Therefore, the identity of elements from the XRF spectrum needs to be confirmed. For the present case study, the XRF responses were compared to periodic charts to identify the elements associated with each peak. The results were then compared to multi-element ICP to confirm the presence and significant concentrations of the identified elements. The following elements were found to be detectable by the XRF scan: iron (Fe), copper (Cu), zinc (Zn), arsenic (As), barium (Ba), cobalt (Co), lead (Pb), chromium (Cr), manganese (Mn), tantalum (Ta), titanium (Ti), and zirconium (Zr).

2.3 Data Processing and Modelling

In particle sorting systems, sensor data is used to predict the quality/grade of each particle. The classification is advised by a calibrated model (algorithm) that, based on a specific threshold, triggers and actuators divert rocks as either product or waste [33]. In the case of copper, the XRF spectrum can be used to identify proxy elements for copper and for training the classification model. The following steps summarize the procedures used for XRF spectral data processing and modelling.

1. The XRF data were screened using the ICP data to identify detectable elements. The screened data is a 300×13 matrix, containing 300 rock samples and 13 elements for each rock, which is considered typical of the small size of the dataset.
2. The input variables (XRF measurements of the 13 elements) were normalized to scale from -1 to 1 by applying a z-score calculation.
3. The training dataset was then subjected to stepwise regression combined with PCA to reduce the size of the input variables and extract the latent features to generate a new set of input variables.
4. The LR and SVM models were then trained with the data collected, and 10-fold cross-validation was performed by using RStudio Workbench which is an integrated development environment for R and Python, with a console and tools for plotting, history, debugging and workspace management. When finished, the cross-validation error can be computed based on the averaged estimate of out-of-sample accuracy from the 10 iterations. Based on cross-validation scores, the best model can be identified.
5. The LR and SVM models were then tested with the validation dataset, and the predictive performance of each model was evaluated using a confusion matrix, ROC, and AUC.

2.3.1 Multicollinearity Detection

In regression analysis, multicollinearity refers to the case where several independent variables are correlated. It is essential to implement a multicollinearity test before modelling. The VIF is an index that determines the variance of a regression coefficient is inflated resulting from the multicollinearity. In general, input variables with a VIF greater than 5 are considered subject to multicollinearity and should be removed [34].

2.3.2 Stepwise Regression Integrated with Principal Component Analysis

The principal component analysis is a statistical dimensionality reduction method that transforms a large dataset of interrelated variables into a smaller dataset of uncorrelated variables, namely the principal components (PCs). The PCs can be expressed as linear combinations of the original variables and retains the maximum information from the original data [35]. The selection of the number of PCs is an important step. In the present study, a PCA approach integrated with stepwise regression was selected for data pre-handling. Stepwise regression (or sequential replacement) is a combination of forward and backward selections starting with no predictors, then sequentially adding the most contributive predictors.

2.3.3 Logistic Regression Modelling

The LR is a particular case of a generalized linear regression model and is expressed as Equation (1), where P represents the probability of ore detection occurrence and x_1, x_2, …refer to the influential elemental variables; β_0, β_1, … β_m are the parameters or coefficients which could be established by the maximum likelihood estimation method. m is the number of selected variables; Logit(P) means a logit transformation of P (x) by the natural log of the odds (being defined as the ratio of occurrence probability to non-occurrence probability) [36]. In the present context of ore or waste prediction, x represents the XRF sensor measurements of detected elements and for each rock, an 'ore' or 'waste' label is assigned based on the predicted occurrence (P).

$$\text{Logit(P)} = \text{Ln}\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m \qquad (1)$$

2.3.4 Support Vector Machine Modelling

The basic SVM is a linear classifier that takes a set of input data and makes a prediction. As it is a binary classifier, there are two possible classes of output. These can be regarded as two clusters of samples in feature space, where the SVM searches for a hyperplane that stays as far as possible from marginal samples of each category. The basic idea is to map a low-dimensional input space (x) onto a higher-dimensional feature domain (X) via a nonlinear mapping function. With the assistance of SVM, the classification problem can be reformulated as a linear problem, thereby enabling the ordinary SVM to perform linear classification in the new feature space.

A critical step in SVM classification is the selection of an appropriate kernel for a particular application. Different kernel functions generate different SVMs and may result in different performances, and different applications need different kernels to achieve reliable results. Moreover, it is crucial to select the appropriate kernel before modelling because the incorrect transformation will lead to a poor result. Hence, the method for choosing a suitable kernel function is very important. At present, the most common selection method is to depend on prior knowledge. However, the best choice of an appropriate kernel for a given problem remains an open research issue [37]. Currently, the following three kernel functions are typically used in the SVM model: the linear kernel, the polynomial kernel, and the radial basis function (RBF).

Three kernel functions, namely linear, polynomial and RBF, are typically used in an SVM model. In this paper, RBF was used due to its localized and finite responses along the entire x-axis. The RBF expression is shown as equation (2), where $||x_i - x_j||^2$ is recognized as the squared Euclidean distance between the two feature vectors γ (gamma) defines how far the influence of each training sample reaches. The higher the gamma value, the more precisely the model tries to fit the training data set.

$$\text{k}(x_i, x_j) = \exp\left(-\gamma||x_i - x_j||^2\right) \qquad (2)$$

Cost (C), also called a regularization parameter, is the other significant parameter that is to determine an RBF-kernel function. It is the penalty parameter of the error term. If C is higher, the optimization will choose a hyperplane with a smaller margin, and the rate of misclassification will be lower. Commonly, the selection range of C and γ are from 0.001 to 100. In this paper, parameters C and γ are selected using a grid search algorithm via the 'tune' function in the 'e1071' package in R studios software.

3. Results and Discussion

3.1 Data Pre-processing and Modelling

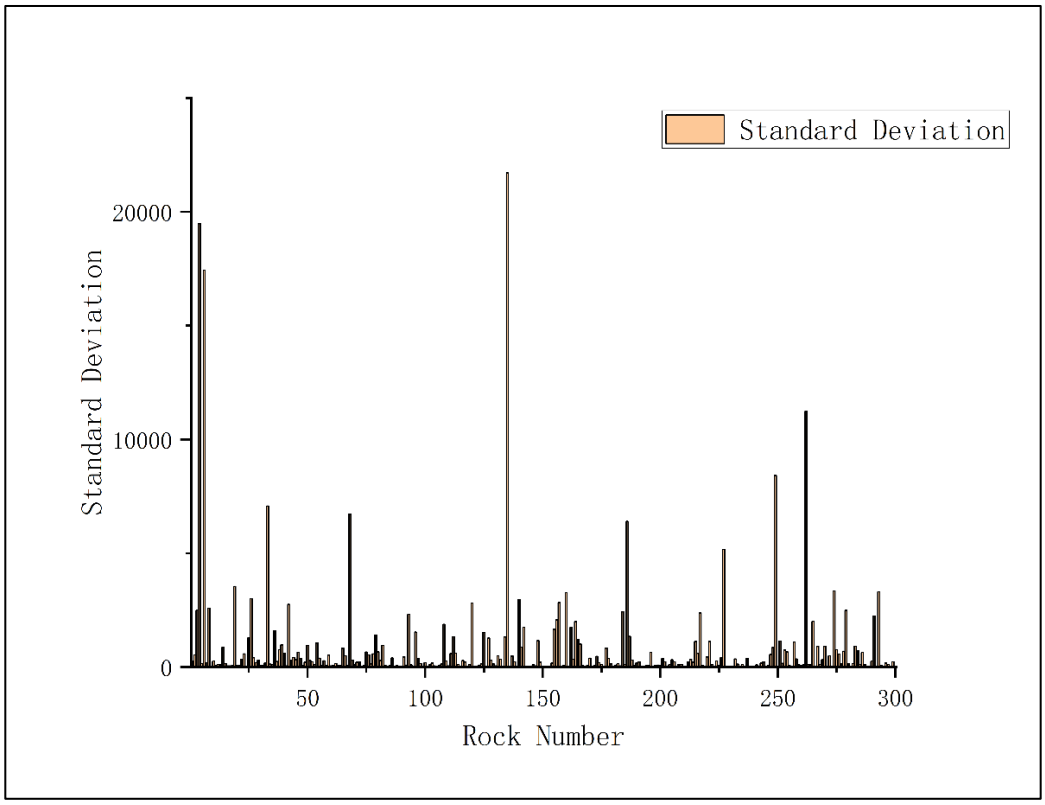

Figure 4: Standard deviations of grade measurements for each rock

The standard deviations of grade measurements for each rock were presented in Figure 4. This demonstrates that the average readings cannot be used to estimate the grade due to occasionally significant deviations. The main reason for this is attributed to the 'heterogeneity error' associated with the XRF surface testing process [38]. To explain, heterogeneous rocks are usually of sedimentary origin and consist of relatively stronger and weaker rocks alternately with varying thicknesses. Complex geological formations such as turbidites, flysch and molasses are typical examples of such rocks [39]. Thus, the content of minerals composition which is distributed on the surfaces of each rock may be dramatically different.

Figure 5 shows the Cu grade correlation between XRF readings on both surfaces and pulverized samples to ICP readings. The pulverized XRF readings show a perfect correlation with the ICP Cu grades with R^2=0.99, however, there is an apparent discrepancy between Cu grades determined by XRF on particle surfaces versus that on pulverized samples with R^2=0.43. An appropriate model, by correlating the XRF surface measurement to ideal classification, is therefore necessary for guaranteeing an accurate classification result.

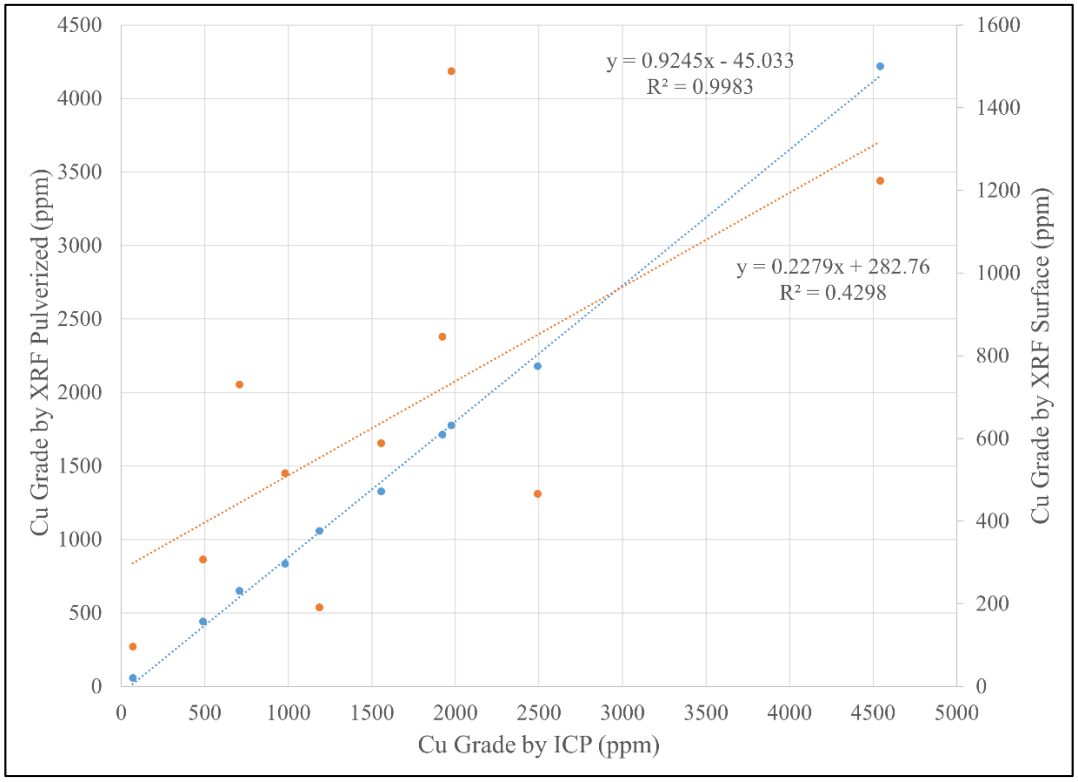

Figure 5: Comparison of Cu grades by XRF to ICP grades

3.1.1 Multicollinearity Detection

In this research, the VIF test was applied to all independent variables via the RStudio software. The VIF test scores of the samples are presented in Table 3. For the degree of correlation, L represents the low and H represents the high correlation. Here, the VIF scores for the elements Co and Fe are seen to exceed 5; In general, to eliminate the multicollinearity, there are two types of approaches generally applied. To create a set of new features or removed the features with higher VIF. It does not recommend removing the features because this inevitably causes information loss, which in turn may affect the prediction accuracy of the model. Hence, an approach for training the classification models without multicollinearity needs to be developed using feature extraction methods, for example, PCA.

Table 3: VIF indices of various elements in the Copper Mountain samples

Element	Cu	Co	Pb	Zn	Fe	As	Ba	Cr	Mn	Ta	Ti	Zr
VIF	1.5	**5.0**	1.5	1.9	**5.2**	1.4	1.3	1.4	2.5	1.1	1.7	1.7
Correlation (L/H)	L	**H**	L	L	**H**	L	L	L	L	L	L	L

3.1.2 Stepwise Regression Integrated with Principal Component Analysis

PCA is a linear method and allows sufficiently fast iteration to meet the needs of particle sorting. Four PCs, shown as Equations (3) ~(6), were selected by using the approaches that are based on the PCs selection

criteria mentioned in the section of materials and methodologies. The PC1~PC4 will be treated as new input variables for ore classification modelling.

$$PC1 = 0.46368Co + 0.32533Zn + 0.45884Fe + 0.34593Mn + 0.32135Ti \quad (3)$$

$$PC2 = 0.50382Cu + 0.32882Zn + 0.35184As - 0.52971Cr - 0.37473Ti \quad (4)$$

$$PC3 = 0.58884Pb - 0.39522Mn + 0.45899Zr \quad (5)$$

$$PC4 = 0.43668Zn + 0.42948Ta + 0.49227Zr \quad (6)$$

However, there is no guarantee that PCA integrated with stepwise regression will perform well for all types of data structures. Other approaches include KPCA, T-SNE, or AE, which are non-linear dimensional reduction approaches that should be considered in future work.

3.1.3 Logistic Regression Modelling for Ore Classification

The LR results trained by the input PCs are presented in Table 4. The coefficients of each independent variable are indicated, and variables with a P-value of < 0.05 are considered statistically significant. Here, PC4 has the lowest p-value and is significant for classifying a rock as ore. Finally, the LR results obtained with the stepwise regression on the input PCs are presented in Table 5.

Table 4: Logistic regression summary with PCA inputs

Input Variables	Estimated Intercept	P-value (Significance)
PC1	-3.79×10^{-5}	0.15
PC2	1.08×10^{-3}	0.09
PC3	-5.99×10^{-4}	0.33
PC4	-8.18×10^{-3}	3.49×10^{-6}

Table 5: The results of combined PCA and logistic regression optimized by stepwise selection

Input Variables	Estimated Intercept	P-value (Significance)
PC2	1.44×10^{-3}	0.01
PC4	-6.65×10^{-3}	9.17×10^{-7}

3.1.4 Support Vector Machine Modelling for Ore Classification

Figure 6 shows the tuning results of cost and gamma parameters for the Copper Mountain samples. C =2 and γ = 1 were finally selected to ensure the balanced predicted performance and robustness of the SVM model

The results of the RBF approach were compared to that of the linear and polynomial functions, as shown in Figure 7 and Table 6. Here, it can be seen that the RBF kernel function gives the best classification, with an accuracy of 83%. However, the difference in classification performance between the polynomial and RBF kernels is small. The results in Figure 7 indicate that the data structure is linearly impartible, and this explains why the linear kernel does not perform well here, even though the linear kernel is a particular form of RBF.

Meanwhile, the number of parameters in the polynomial kernel is much larger than in the RBF, thus resulting in a more complicated model.

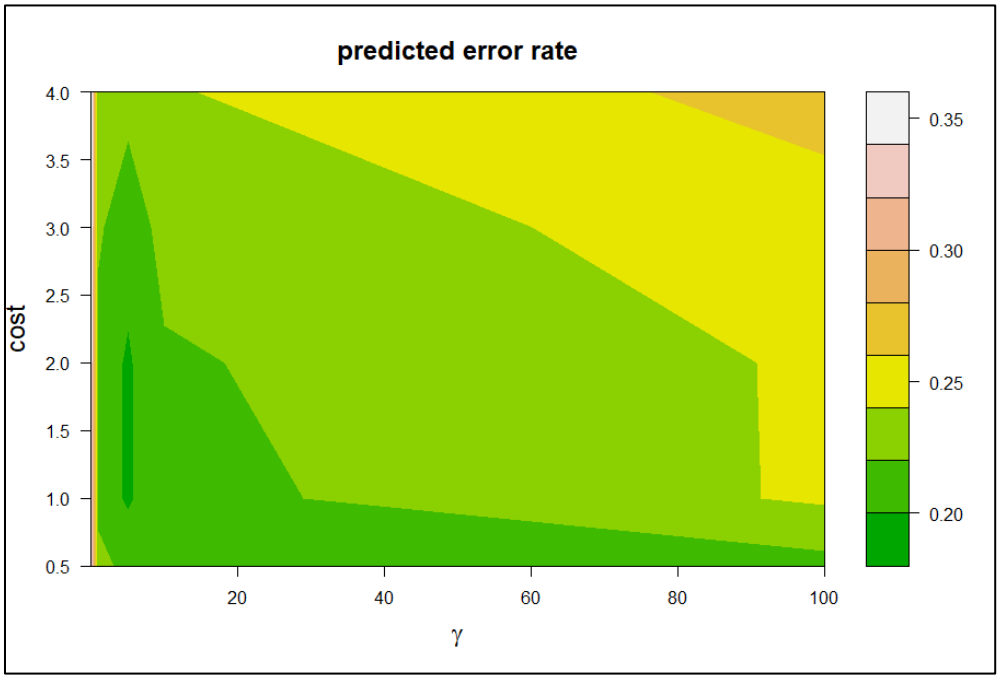

Figure 6: Tuning result of the SVM parameters for the Copper Mountain samples

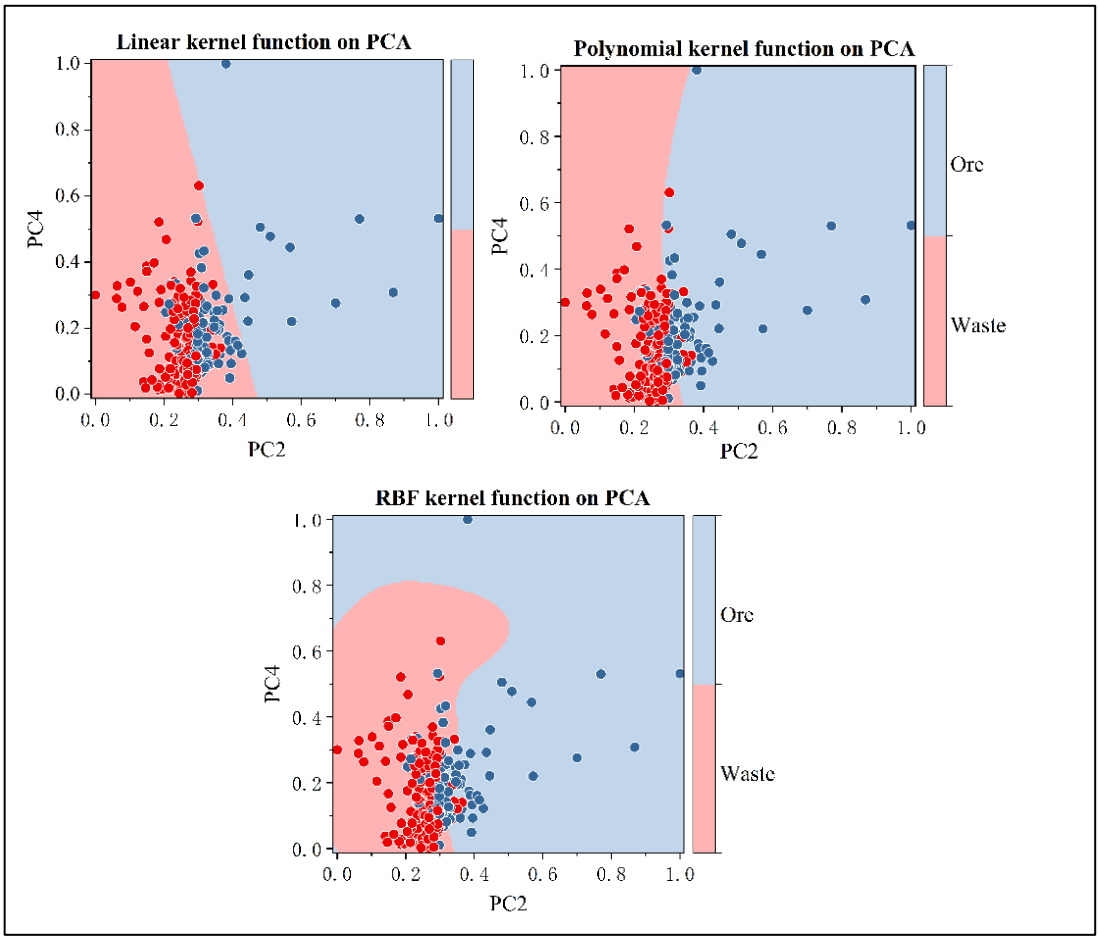

Figure 7: SVM classifier with various kernels on the PCA input dataset

Table 6: The SVM classification performance with various kernel functions

Kernel Function	Number of Support Vectors	Accuracy Rate
Linear	85,85	0.73
Polynomial	66,65	0.82
RBF	67,66	0.83

3.2 Performance Evaluation of Ore Classification Models

The ROC and AUC are used to assess the model classification performance at various threshold settings. The AUC-ROC performance of all the models are shown in Figure 8. Results showed that the PCA-LR approach obtains the highest AUC of 0.847, indicating logistic regression performs better in ore/waste classification for the Copper Mountain ore sample.

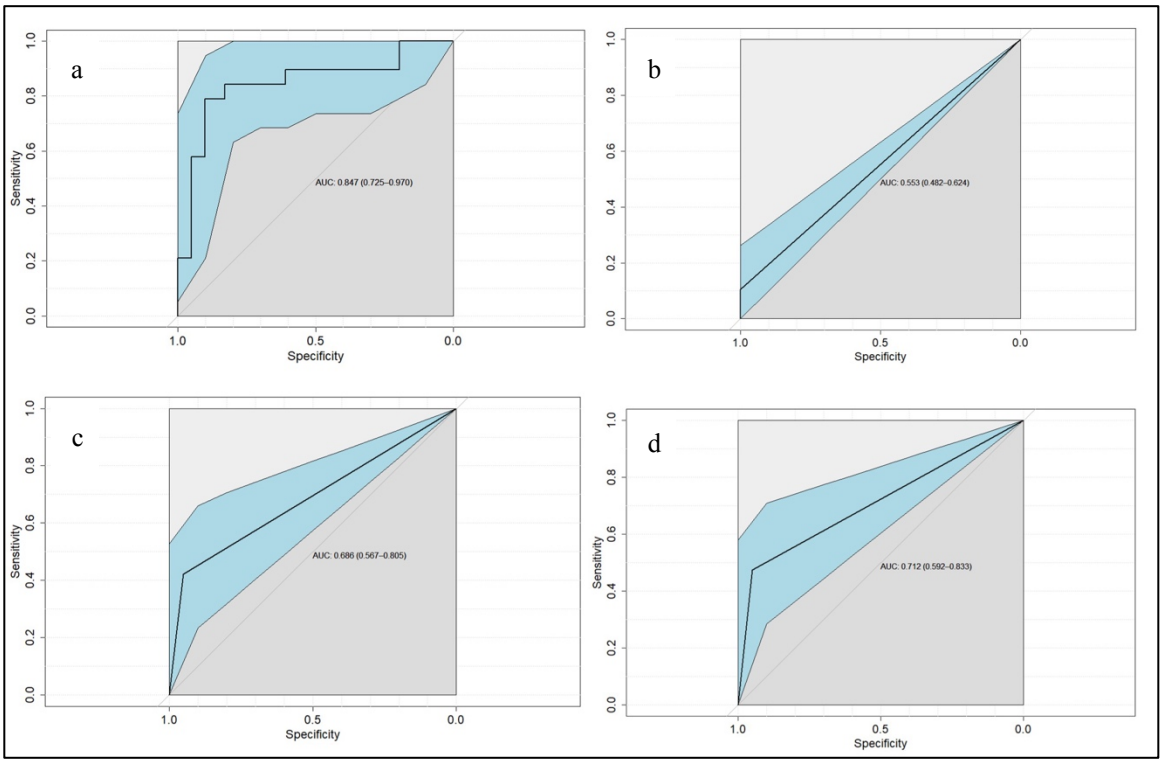

Figure 8: The ROC-AUC performance of models. (a) PCA-LR with AUC= 0.847 (b) PCA-SVM-polynomial kernel with AUC= 0.553 (c) PCA-SVM-linear kernel with AUC= 0.686 (d) PCA-SVM-RBF kernel with AUC=0.712

The ROC curves for these four models are displayed in Figure 9 where the blue curve represents the LR model; the black represents the linear SVM model; the red represents the RBF SVM model and the green represents the polynomial SVM model. In the context of sensor-based ore sorting, the most significant feature of a classification model is its sensitivity or TPR. This is because it might be accepted that a proportion of the waste samples is misclassified as ore class, but it must be ensured that as many valid samples as possible are classified into the ore group. In other words, a model with high sensitivity (or high TPR) is selected preferentially. For intuitive presentation, Table 7 below summarized the index parameters of ROC of each model. The LR and linear SVM showed the same accuracy as 0.8, however, the LR model with a TPR of 0.95 should be preferentially selected for ore sorting in this research.

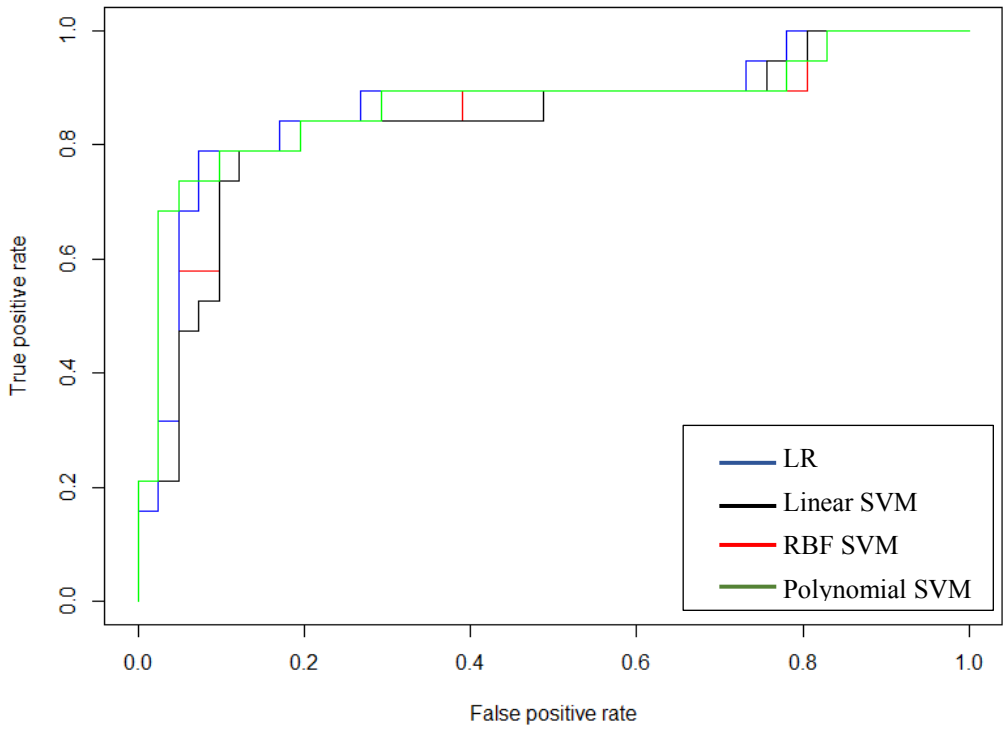

Figure 9: Comparison of the ROC curves between different models on the test dataset

Table 7: The parameters of ROC of each model

Model	Sensitivity	Specificity	Accuracy
LR	0.95	0.47	0.80
Linear SVM	0.79	0.82	0.80
RBF SVM	0.78	0.80	0.78
Poly SVM	0.71	0.95	0.72

To discuss, generally, SVM with linear kernel function has a similar algorithm structure to the LR and therefore their classification performance could also be similar. However, when the dataset is small, SVM would usually perform worse compared to LR, especially when the input variables are relatively less than the training dataset size. This is because the data points near the decision boundary may not be true representations of the actual decision boundary, so that will lead to a false maximum margin classifier boundary. For this study, only two input variables were identified after applying feature extraction by PCA and stepwise regression, while the number of training samples is 300 and is much larger compared to the number of the input features. Therefore, logistic regression was found to have a better classification performance compared to the support vector machine approach.

In addition, fewer parameters are used in LR modelling than in the SVM model which would likely allow faster iteration. From a mathematical perspective, LR is strictly convex with a smoother loss function, whereas SVMs are only convex, therefore LR is faster to process the sensor data and is considered more applicable for particle sorting. Besides, as the SVM is based on kernel functions, most implementations explicitly store this as an N×N matrix of distances between the training points to avoid computing entries over and over again. Therefore, in general, the LR performs faster with respect to computational time than the SVM with RBF and Polynomial kernel functions. However, the difference in computational time between LR and SVM is not very evident due to the bench-scale dataset with only three hundred samples.

4. Conclusions

The present study evaluated logistic regression and support vector machine approaches for XRF SBS. Copper ore samples from Copper Mountain in British Columbia, Canada, were scanned using XRF to obtain the spectral data for model development. PCA integrated with stepwise regression was selected for the data pre-processing and feature selection. By applying feature selection, information redundancy and noise in high dimension features are avoided and therefore better classification accuracy could be provided.

The classification performances of the LR and SVM algorithms are compared. The results indicated that the LR outperforms the SVM with different kernel functions, especially on small-sized data. In addition, polynomial, linear, and RBF kernels were compared to identify the most appropriate kernel tricks for the SVM classifier. The RBF kernel was shown to perform better than the other kernel tricks. Commonly, imagine having a $m \times n$ dataset, in which m represents the number of input features and n represents the number of samples, if $m>n$, then a linear kernel would be considered to use because, after mapping to a higher dimension with RBF kernel tricks, the data would be more complicated to separate and susceptible to overfitting. Instead, if $m<n$, RBF would be the priority choice.

To conclude, by applying machine learning techniques such as LR, SVM and PCA, there is an opportunity to improve the XRF based ore sorting performance. However, more ore samples with different mineralogy and size fractions should be tested to validate the robustness and generality of the sorting model.

Nomenclature

Nomenclature	Abbreviation
Sensor-Based Sorting	SBS
X-ray Fluorescence	XRF
Semi-Autogenous Grinding	SAG
Principal Component Analysis	PCA
Logistic Regression	LR
Support Vector Machine	SVM

Radial Basis Function	RBF
T-distributed Stochastic Neighbor Embedding	T-SNE
Kernel Principal Component Analysis	KPCA
Autoencoder Networks	AE
Receiver Operating Characteristic Curve	ROC
True Positive Rate	TPR
False Positive Rate	FPR
Area Under the ROC Curve	AUC
Variance Inflation Factors	VIF
Constitutional Heterogeneity	CH
Decision Tree	DT
Random Forest	RF
Artificial Neural Networks	ANN
K-nearest Neighbor	KNN
Multivariable Linear Regression	MLR

References

[1] Wotruba, H. (2006). Sensor Sorting Technology-is the Minerals Industry Missing a Chance? Plenary lecture. XXIII International Mineral Processing Congress, Istanbul.

[2] Bamber, A., et al. (2008). "Integrated mining, processing and waste disposal systems for reduced energy and operating costs at Xstrata Nickel's Sudbury operations." Mining Technology 117(3): 142-153.

[3] Bamber, A. S. (2008). Integrated mining, pre-concentration and waste disposal systems for the increased sustainability of hard rock metal mining, University of British Columbia.

[4] Dalm, M. (2011). "Applicability of near-infrared spectroscopy for sensor based sorting of mill pebbles from the Los Bronces copper mine, Chile."

[5] Seerane, K. and G. Rech (2011). Investigation of sorting technology to remove hard pebbles and recover copper bearing rocks from an autogenous milling circuit. 6th Southern African Base Metals Conference. The South African Institute of Mining and Metallurgy.

[6] Klein, B., Bamber, A.S., (2019), Mineral Sorting, Mineral Processing and Extractive Metallurgy Handbook, Society for Mining, Metallurgy and Exploration.

[7] Mazhary, A., Klein, B. (2015). Heterogeneity of Low-Grade Ores and Amenability to Sensor-Based Sorting, Proceedings of the Annual Meeting of the Canadian Mineral Processors, Ottawa, Canada

[8] Klein, B., Bamber, A.S., (2019), Mineral Sorting, Mineral Processing and Extractive Metallurgy Handbook, Society for Mining, Metallurgy and Exploration.

[9] Nienhaus, K., Pretz, T., Wotruba, H. (2014). Sensor Technologies: Impulses for the Raw Materials Industry, Shaker Verlag, Aachen, pp 88-100

[10] Fickling, R.S. (2011). 6th Southern African Base Metals Conf., Phalaborwa.

[11] Zhu, Z., Lin, C., Zhang, X., Wang, K., Xie, J., Wei, S. (2018). Evaluation of geological risk and hydrocarbon favourability using a logistic regression model with a case study, Marine and Petroleum Geology 92: 65-77.

[12] He X, Xu S. Process neural networks: Theory and applications. Springer Science & Business Media; 2010 Jul 5.

[13] Li, M. (2018). A neural network for electromagnetic based ore sorting. COMPEL - The international journal for computation and mathematics in electrical and electronic engineering 37(2): 691-703.

[14] Bzdok, D., Altman, N. & Krzywinski, M. Statistics versus machine learning. Nat Methods 15, 233-234 (2018). https://doi.org/10.1038/nmeth.4642

[15] Russell, S.J., Norvig, P. (2009). Artificial Intelligence: A Modern Approach (AIMA), Prentice Hall, 3rd Edition.

[16] Gomez-Flores, Allan, et al. "Prediction of grade and recovery in flotation from physicochemical and operational aspects using machine learning models." Minerals Engineering 183 (2022): 107627.

[17] Shah, Kanish, et al. "A comparative analysis of logistic regression, random forest and KNN models for the text classification." Augmented Human Research 5.1 (2020): 1-16.

[18] Grömping, Ulrike. "Variable importance assessment in regression: linear regression versus random forest." The American Statistician 63.4 (2009): 308-319.

[19] Uddin, S., Khan, A., Hossain, M. et al. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 19, 281 (2019). https://doi.org/10.1186/s12911-019-1004-8

[20] Priyam, Anuja, et al. "Comparative analysis of decision tree classification algorithms." International Journal of current engineering and technology 3.2 (2013): 334-337.

[21] Friedl, Mark A., and Carla E. Brodley. "Decision tree classification of land cover from remotely sensed data." Remote sensing of environment 61.3 (1997): 399-409.

[22] Dreiseitl, Stephan, and Lucila Ohno-Machado. "Logistic regression and artificial neural network classification models: a methodology review." Journal of biomedical informatics 35.5-6 (2002): 352-359.

[23] Bender, Ralf, and Ulrich Grouven. "Ordinal logistic regression in medical research." Journal of the Royal College of physicians of London 31.5 (1997): 546.

[24] Lee, S. A. R. O. "Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data." International Journal of remote sensing 26.7 (2005): 1477-1491.

[25] Sartor, Maureen A., George D. Leikauf, and Mario Medvedovic. "LRpath: a logistic regression approach for identifying enriched biological groups in gene expression data." Bioinformatics 25.2 (2009): 211-217.

[26] Shipe, Maren E., et al. "Developing prediction models for clinical use using logistic regression: an overview." Journal of thoracic disease 11.Suppl 4 (2019): S574.

[27] Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression. John Wiley & Sons; 2013 Apr 1.

[28] Wu, Yingquan, Krassimir Ianakiev, and Venu Govindaraju. "Improved k-nearest neighbor classification." Pattern recognition 35.10 (2002): 2311-2318.

[29] Taunk, Kashvi, et al. "A brief review of nearest neighbor algorithm for learning and classification." 2019 International Conference on Intelligent Computing and Control Systems (ICCS). IEEE, 2019.

[30] Vapnik, V., Golowich, S., Smola, A. Support vector method for function approximation. (1997). Regression estimation and signal processing. In: Advances in Neural Information Processing Systems, vol. 9, no. 9,. p. 281-7.

[31] Joachims, T., 1998. Text categorization with support vector machines: learning with many relevant features. In: Proceedings of European Conference on Machine Learning, Springer-Verlag, Berlin, pp. 137-142

[32] C. Huang, L.S. Davis, J.R.G. Townshend An assessment of support vector machines for land cover classification International Journal of Remote Sensing, 23 (2002), pp. 725-749

[33] Libin T., Klein B. (2015). Correlation and regression analysis in the X-ray fluorescence sorting of a low-grade copper ore. SAG Conference. Vancouver: SAG Conference.

[34] Yoo, W., Mayberry, R., Bae, S., Singh, K., He, Qinghua, P.H., Lillard, J.W. (2014). A Study of Effects of Multicollinearity in Multivariable Analysis. Int. J. Appl. Sci. Technol. 4, 9-19.

[35] Shang, X., Li, X., Morales-Esteban, A., Chen, G. (2017). Improving micro seismic event and quarry blast classification using artificial neural networks based on principal component analysis. Soil Dyn. Earthq. Eng. 99, 142-149. doi: 10.1016/j.soildyn.2017.05.008

[36] Li, N., Jimenez, R. (2017). A logistic regression classifier for long-term probabilistic prediction of rock burst hazard." Nat. Hazards 90(1): 197-215.

[37] Tsang, I.W., Kwok, J.T., Cheung, P.M. (2005). Core vector machines: fast SVMs training on very large data sets. J. Mach. Learn. Res. 6: 271-363.

[38] Li, G., Klein, B., Sun, C., & Kou, J. (2021). Lab-scale error analysis on X-ray fluorecence sensing for bulk ore sorting. Minerals Engineering, 164, 106812.

[39] Liakas, Spyridon, Catherine O'Sullivan, and Charalampos Saroglou. "Influence of heterogeneity on rock strength and stiffness using discrete element method and parallel bond model." Journal of Rock Mechanics and Geotechnical Engineering 9.4 (2017): 575-584.