

Research Article

Using Artificial Intelligence Techniques to Improve the Prediction of Copper Recovery by Leaching

Victor Flores ¹, Brian Keith,¹ and Claudio Leiva ²

¹Department of Computing & Systems Engineering, Universidad Católica del Norte, 1270709 Antofagasta, Chile

²Department of Chemical Engineering, Universidad Católica del Norte, 1270709 Antofagasta, Chile

Correspondence should be addressed to Victor Flores; vflores@ucn.cl

Received 21 June 2019; Revised 19 November 2019; Accepted 3 January 2020; Published 7 February 2020

Guest Editor: Sidike Paheding

Copyright © 2020 Victor Flores et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copper mining activity is going through big changes due to increasing technological development in the area and the influence of industry 4.0. These changes, produced by technological context and more controls (e.g., environmental controls), are also becoming visible in Chilean mining. New regulations from the Chilean government and changes in the copper mining industry (such as a trend to underground mining) are fostering the search for better results in typical processes such as leaching. This paper describes an experience using artificial intelligence techniques, particularly random forest, to develop predictive models for copper recovery by leaching, using data from an enterprise present in northern Chile for more than 20 years. Two models, one of them with actual operational data and another one with data generated in a controlled environment (piling) are presented. Well-classified values of 98.90% for operational data and 98.72% for pile/piling data were obtained. The methodology devised for the study can be transferred to piling columns or piles with other characteristics, though the operation must focus on copper leaching. It can even be transferred to other leaching processes using another type of mineral, with proper adjustments.

1. Introduction

The Chilean mining industry, as in the whole world, is experimenting with big changes due to the rapid technological advance in the so-called industry 4.0 [1]. According to Pietrobelli et al. in [2], big mining companies typically tended to control their operations from remote centers located in multinational corporations, thus resulting in little local innovation and development. This way of operating helps the macroeconomy, but it makes difficult diversification, knowledge transfer, and regional innovation in the value chain [3]. Another factor producing changes in the above-mentioned trend is the significant fall of copper price since 2015, fostering both technological advances enabling companies to face production costs [4] and also greater regional innovation and development.

Chilean copper production represents 35% on a world basis [5, 6]. On a local basis, the copper production industry is the country's most profitable, providing almost 15% of Chilean GDP and representing 50% of exports [7, 8]. This Chilean predominant position in the copper industry is also

complemented with leadership in other mineral products, such as lithium. To keep this leadership in the world's mining activity, Chile must ensure mining profitability in the short term. A valid strategy for this may be investing in technology and innovation, together with mining industry diversification.

Recent papers [5–7] report a trend to technological diversification in the sector, even from mining suppliers. Furthermore, as stated in [6], a recent report from the Chilean government declares the objective of promoting the establishment of 250 local suppliers for the mining sector in 2035. This strategy is expected to create knowledge about business and technology appropriate for current challenges, both elements being directed to local mining development and exports as well. This would result in an income of about US\$10 thousand million.

For the aforementioned technological development and innovation, the Chilean mining industry is incorporating technology to develop intelligence system-type applications for supporting tasks such as copper recovery prediction. These systems are frequently based on artificial intelligence

computing models. Apart from representing a technological contribution, these models are becoming a great help for predicting or reducing production costs [9, 10], a very convenient fact for supporting modern technology characterized by a greater extraction complexity and increasing restrictions such as environmental ones [11].

In typical production processes such as leaching, predictive models have been satisfactorily used in the last decade to identify factors allowing production increase [9, 10]. There are several cases illustrating predictive model generation using artificial intelligence, specifically soft computing [12]. In particular, this paper fully describes the process for developing predictive models [13] to recover copper by leaching and the results were obtained at SCM Franke Company, from the KGHM International Group, present in Chilean mining exploitation since 2009.

Recently, research into the applicability of artificial intelligence techniques such as predictive model algorithms, for copper recovery prediction, has been conducted. In this context, comparative studies of which predictive model algorithms are the most appropriate according to the characteristics of the copper mining production process have been published. Thus, advantages of using support vector machine (SVM), random forest (RF), artificial neural networks (ANN), gradient boosted trees (GBT), or wavelet neural network (WNN) are frequently reported in the literature (such as [14]). For example, in [15], a predictive modeling using SVM for copper potential mapping in the Kerman copper bearing belt in the south of Iran is reported. In [16], a comparative analysis of ANN, WNN, and SVM models to mineral potential mapping for copper mineralization is presented. As a particular result of this work, the authors highlight that WNN exhibits excellent learning ability compared to the conventional ANN.

Also, in [17], SVM, ANN, and RF were used to conduct predictive modeling of mineral prospectively. For these algorithms, input data was obtained from GIS-based mineral prospectively mapping of the Tongling ore district (eastern China). As a conclusion from this work, authors highlight that the RF model outperformed the SVM and ANN models, giving a greater consistency and better predictive accuracy. Another example of comparative analysis of predictive models using GBT, ANNs, and RF is the work described in [16], where authors highlight that the RF models show the highest coefficients of determination (R^2) values and the lowest root-mean-square error (RMSE), and the highest residual prediction deviations (RPD) were obtained.

There are several papers that report that RF and GBDT perform the best (see Table 1 for a comparison among these methods); therefore, and based on the described information in the previous paragraphs, the use of RF can more appropriately lead to the achievement of the stated objective.

This paper describes the tasks done to generate predictive models for copper recovery in leaching piles with low-grade material, using data from actual pile operation and those produced in a controlled environment (pilot), using the same artificial intelligence technique (random forest technique) in both cases to develop predictive models.

TABLE 1: Accuracy ranking of different classification algorithms [32].

Algorithm	Accuracy ranking				
	Top 1	Top 2	Top 3	Top 4	Top 5
GBDT	42.25%	63.38%	67.61%	77.46%	83.10%
RF	21.13%	56.34%	77.46%	84.51%	92.96%
SVM	18.31%	30.99%	52.11%	64.79%	78.87%
ELM	15.49%	23.94%	30.99%	47.89%	56.34%
C4.5	12.68%	15.49%	28.17%	43.66%	56.34%
SRC	11.27%	22.54%	23.94%	33.80%	38.03%
LR	4.23%	11.27%	19.72%	26.76%	40.85%
AB	4.23%	7.04%	8.45%	19.72%	18.31%
KNN	2.82%	8.45%	12.68%	26.76%	36.62%
NB	2.82%	5.63%	7.04%	9.86%	19.72%
DL	1.41%	1.41%	2.82%	4.23%	4.23%

The remaining document is organized as follows: Section 2 describes the base concepts of the study and related work. Section 3 describes the experiment, the discretizing of the variables used in the model, data characteristics and how they were collected, work methodology, and the techniques used for analyzing results. Section 4 shows the results obtained for the two models, that is, operational data and piling data models. Section 5 deals with the discussion. Section 6 shows the conclusions of the paper. Finally, acknowledgments and bibliographical references are stated.

2. Concepts and Related Work

2.1. Leaching and Company Work. The copper leaching process involves tasks thoroughly identified by the industry, that is, irrigation beginning and maintenance, agglomerate condition evaluation, drainage distribution, pool solution inventory, PLS flow evaluation, and distribution and deposition of the material leached at the plant (harvest). These processes, due to the nature and variability of the input material, usually produce high levels of entropy and uncertainty (close to 20%) concerning copper recovery at the end of the harvest [9].

SCM Franke uses three industrial processes widely known in the industry of metallic copper production via hydrometallurgy. These processes are dynamic pile leaching, solvent extraction, and electrowinning [9]. The ultimate goal of these processes is to obtain the greatest copper production by saving resources and being the least possible aggressive to the environment (a kind of environmental trade-off). The leaching process has been shown to be one of the most convenient to achieve this environmental trade-off. The objective of this paper is predicting estimated copper recovery as accurately as possible at about 95% by dynamic pile leaching, using the least possible amount of leaching material and the best irrigation homogeneity.

2.2. Related Work and Predictive Models. The development of applications using predictive modeling to improve mineral

recovery estimation is prospectively becoming a central area of study in the mining industry [18–20].

Recent studies such as [3, 9, 10] reveal that one of the most critical tasks in prospective modeling is the selection of appropriate criteria and the application of sound innovating techniques to get the evidential characteristics of these criteria.

Traditionally, these criteria have been selected by different numerical methods, but in the last few decades, alternative techniques such as those from the artificial intelligence area have been applied for both criteria selection and the development of predictive models for mineral recovery [21]. In general, methods containing machine learning algorithms are being applied for building these predictive models.

In the literature, the methods referred to here have been grouped into two sets [21–23]: *knowledge-driven models* and *data-driven models*. Data-driven models are probabilistic models such as discriminant analysis or logistic regression [19, 24]. The algorithms of data-driven models, whose evidence of use is more often reported in the literature, are artificial neural networks (generally with backpropagation [25, 26]) (ANN) [19, 27, 28] and regression trees (RTs) [13, 24, 29] in sectors such as copper mining. Methods called support vector machines (SVM) and random forest (RF) [29] are sometimes used in this domain [13, 30, 31]. The common way of using the algorithms of the data-driven model group in concrete mining tasks such as studying copper recovery is using data themselves, while in knowledge-driven methods, an expert in mineral extraction via hydrometallurgy should be consulted for the job. As a whole, ANN, RT, or SVM models require enough amounts of records and parameters to achieve good quality in the models created as output.

The literature contains papers such as [32] that propose a comparison among the performance of predictive models. Table 1 shows that RF and GBDT perform the best, followed by SVM and ELM. Moreover, we observe that the interquartile ranges of RF, GBDT, and SVM are the smallest, showing that these three algorithms generally perform well, in terms of prediction accuracy, regardless of the datasets [32]. While ensemble and boosting methods have been reported to obtain good predictive performance in supervised learning, GBDT is generally less popular than RF. GBDT and RF show both best total average classification accuracy and best mean rank followed by SVM and ELM [32].

This study uses RF as a predictive model; it is a kind of predictive model based on decision trees. There are previous works as [33] that defined this kind of predictive models as “a type of predictive model that uses a decision tree to go from observations of an object (represented as the branches of a tree) to a certain conclusion about a target value of the object (represented by the tree leaves).”

Thus, the interest of using RF is twofold. First, data-driven model algorithms (like RF) are frequently used to predict values of the target variable influenced by other variables (predictor variables) in datasets [33, 34]. In this context, the RF model is adequate for generating a predictive model of the copper recovery by leaching (the target variable for this work), due to it providing a way to measure the influence

of each predictor variable on the target variable. And second, one of the main benefits of RF is that it can be used to determine the importance of variables in a regression or classification problem intuitively [35]. So, RF can be used to determine the importance of each predictive variable over the target variable.

Prediction is a highly interesting topic in machine learning, which is, in turn, one of the branches of artificial intelligence. As mentioned above, RF is based on decision trees (DT). DT have been widely used in areas such as medicine to yield a diagnosis since they are easy to interpret. Basically, DT is a hierarchical set of nodes (starting from a root node), where each node contains a decision based on the comparison of an attribute with a threshold value [36, 37]. DT-based learning goes from the observation of an object represented as branches of a tree to certain conclusions related to a target value of an object (represented by tree values) [36, 37].

Previous studies use artificial intelligence techniques for copper-related models. For example, in [8], a model based on fuzzy logic is reported to predict ground vibration and environmental impact due to blasts in the open-pit mine. For this model, the toolbox fuzzy logic of MATLAB was used. In [38], ANN was used to predict the copper ore flotation indices of separation efficiency within different operational conditions.

3. Materials and Methods

3.1. Experimental Description. Operational and piling data are available for attaining the objective set by SCM Franke company (environmental trade-off described in Section 2). The company keeps records of planning and copper recovery by heap leaching. These are called operational data (industrial operation). Work has also been done with data collected in a controlled environment. These data are known as piling data, which are the result of tests in leaching columns using strictly controlled measures on irrigation rates, acid concentration in irrigation solutions, and operational cycles.

For the specific case of this study, both operational and piling data were collected by two students in practice and Professor C. Leiva (students under the supervision of Professor C. Leiva, coauthor in this paper) all from the Chemical Department at the Universidad Católica del Norte, Chile. In a similar way to what worked in [9], the parameters of these data groups are fully described below:

- (i) Agglomerate is measured in mm, where 80% of the solids are below this value
- (ii) Irrigation rate (RL) (L hr/m^2) is the surface flow of sulfuric acid in the pile
- (iii) H+fed (gpl) is the volumetric flow of ILS (intermediate liquid solution) recirculating in the pile
- (iv) The height of a pile is defined by the production goals expected to be accomplished; that is, the piled fine copper tonnage with which the production to be obtained will be determined

- (v) Total Cu grade (%) is the total copper percentage existing in the pile in the n day of operation
- (vi) CO_3 grade (%) is the carbonate percentage existing in the pile in the n day of operation
- (vii) Leaching ratio (m^3/TMS) is defined by the amount of sulfuric acid with respect to the total material to be leached
- (viii) Days of operation refer to the days elapsing from the pile starting up to the end of leaching
- (ix) Soluble Cu (%) is the percentage of copper soluble in sulfuric acid present in the pile in the n day of operation
- (x) Class R (%) is the percentage of leached copper in the n day with respect to the soluble copper present in the pile in day 1 of operation

3.2. Operational Data. Operational data were collected during time periods called leaching cycles emerging after soil piling and the beginning of the irrigation process since day 1 to the last day of harvest. The leaching cycles in the company are planned to last 65–70 days. Operational data were obtained with a frequency of 4 hours during one year. Due to the conditions of the process and operational decisions, the irrigation of some piles or modules in service was stopped, a fact that could render incongruent results when modeling the system. For this reason and with the purpose of avoiding unnecessary “noise” in the system, along with storing poor data for the statistic model, the records of the nonirrigation periods were deleted from the database.

3.3. Piling Data. Piling (or pilot plant) was conducted in two agglomerate tanks of the same dimensions with a material whose granulometry was less than 13 mm in diameter. The mineral was put in contact (irrigated) with a solution of sulfuric acid and water and refined to form lumps of fine material; this was made in order to give the mineral a proper uniform size for the leaching stage and also help copper sulfidation via contact with acid solutions. The aforementioned conditions vary according to leaching cycles to obtain piling scenarios as close as possible to actual pile mineral exploitation. Piling data were obtained in the same way as explained for operational data.

3.4. Random Forest. As previously mentioned, random forest (RF) is a predictive model based on decision tree (DT). The RF supervised learning algorithm is based on the machine learning theory which belongs to the *ensemble methods* family [34]. These methods use supervised learning methodology over a set of labelled data (training set) to make predictions and produce a model which can be later used to classify nonlabelled data [39]. It uses supervised learning methodology to collect data from parameter values and threshold values, working on a set of training data [40]. The method combines the idea of bagging with the random selection of characteristics, so as to build decision trees using controlled variance [37].

The RF model is successfully used in classification and regression tasks, operating via the construction of multiple decision trees during training, with the purpose of discovering patterns existing in data. The method generates several trees as subsets by combining several automatic learning algorithms appropriately selected [33]. This method is a general technique of random decision trees that combines the idea of bagging with a random selection of characteristics, with the intention of building decision trees with controlled variance [34, 35].

RF is an ensemble method for classification and regression tasks, which operates through the construction of multiple decision trees during training [34]. Additionally, RF is useful for calculating the influence of predictive variables on the target and also for calculating the importance of each of these influences over the target. The calculation of this importance is made with a metric calculated according to impurity decrease in each node used for partitioning data. In case of a classification, the class determined corresponds to the mode of the classes provided by each tree. In case of a regression, it corresponds to the average prediction of individual trees. Random decision trees correct the DT tendency to overadjust to their training set [41].

3.5. Case Study. Using operational and piling data, a case study was conducted with a database of about 30,000 records. For each parameter above, discrete values of low, normal, and high were devised according to threshold values previously defined by SCM Franke, which are commonly used in copper leaching. In particular, this discretization considered data standard deviation (σ) defining low (low value of the variable), corresponding to values lower than a $-\sigma$; normal (normal value of the variable), corresponding to values at the interval $[-\sigma, \sigma]$; and values considered high (high value of the variable), that is, those greater than $+\sigma$.

3.6. Methodology. The methodology consists of 4 steps. The initial step to collect data of both operation and piling are considered a stage previous to the methodology described below since these data (mainly operational data) were collected during several years of operation. Parameter values were grouped in periods including days of operation while class (recovery) is described for each day of operation per each period. Figure 1 shows examples of what was described above. Figure 1(a) shows daily recovery in two consecutive periods of operation, while Figure 1(b) shows daily recovery in two consecutive periods, but with pilot plant (piling) data. In detail, the steps of our methodology are as follows:

- (1) **Data Preparation.** This stage included filtration tasks and data selection per leaching cycles. Plant data were obtained with a frequency of four hours in one year. Due to process conditions and operational decisions, the irrigation of some piles or modules in service was stopped during some periods, a fact that could render incongruent results when modeling the system. To ensure operational data congruence, records corresponding to irrigation suppression periods were deleted from the database; these records were being

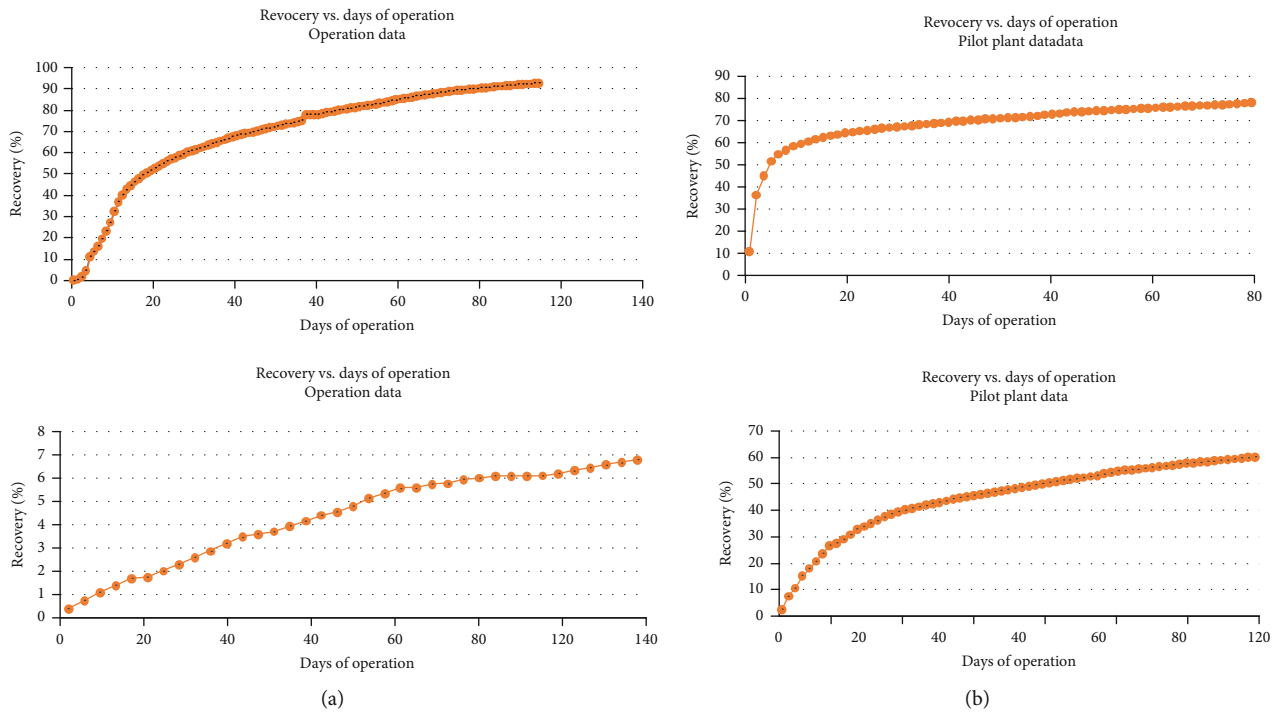


FIGURE 1: Examples of recovery periods. (a) shows daily recovery in two consecutive periods of operation, while (b) describes two consecutive periods with pilot plant (piling) data.

substituted by the leaching ratio. Leaching cycles with recovery values lower than 10% were also deleted after day 15 of the operation because this indicates an error in data acquisition

- (2) *Model Generation*. In this stage, data were collected and selected according to relevance in order to create a predictive copper recovery model on the conditions determined by the context of the study
- (3) *Model Visualization and Analysis*. In this stage, model results were visualized and analyzed to determine their validity. Evaluation consisted of checking the performance of the models obtained with RF for each dataset. To do this, values of certainty such as *accuracy*, *recall*, and *precision* were calculated and analyzed. The way these values of certainty were calculated and their importance for model quality are described below
- (4) *Result Analysis*. In this stage, the analysis is aimed at establishing if the results obtained are useful for the industry. This was done by analyzing aspects such as how optimal variable parametrization was or how well classified training set instances were (confusion matrix values)

To make the analysis in stage 3 above, a confusion matrix was considered. The confusion matrix facilitates the analysis necessary to determine an error in the classification, through a sample of error distribution in the different categories.

In this matrix, performance indicators [42] frequently used to evaluate classifier performance are described. They

are accuracy (Acc), recall (r), and precision (p). The way these indicators are calculated is described in Equations (1)–(3). The simplest indicator to evaluate a classifier performance is accuracy (Acc), corresponding to sample ratios correctly classified in the total number of examples of the dataset [33]. This indicator can be calculated on the basis of confusion matrix data according to Equation (1) (the dataset is supposed not to be empty). The other indicators, recall (r) and precision (p), are understood as relevance measures.

The p value is the ratio of true positives (a) among the elements predicted as positive ($a + b$). Conceptually, p value refers to the dispersion of the value set obtained from repeated measures of a quantity. Specifically, a high p value indicates low dispersion in measures. The r value is the ratio of true positives predicted among all the elements classified as negative.

$$\text{Acc} = \frac{a + b}{a + b + c + d}, \quad (1)$$

$$p = \frac{a}{a + b}, \quad (2)$$

$$r = \frac{a + d}{b + c}, \quad (3)$$

where a is the true positives, b is the false positives, c is the true negatives, and d is the false negatives.

4. Results

The problem described above was dealt with as a regression instance, looking for obtaining a copper recovery prediction numerically from data in each dataset (operational and piling). So, a model was obtained for both operational and piling data, the importance of associated variables being studied in both cases. To obtain the models, the free Rapid Miner Studio v 9.0 was used.

The strategy used in the model generation process was, first, preparing data according to task 1 of the methodology above. After the data preparation process (according to Section 3), a file with 1638 records for piling and another with 2001 records for operation were obtained (both files in CSV format). Previous studies such as [12, 34, 43] indicate that a minimum value of 1000 input cases for RF minimizes error in the classification and, at the same time, enables RF to make more stable predictions. So, both datasets are considered appropriate for generating the models.

In order to prepare the model evaluation and in a similar way to what is done in [34], a parameter tuning phase was performed. The models were evaluated using these parameters (40-fold crossvalidation 10 times) and averaging final results were taken. But the results of this validation were not good, for roundness. So, a method based on hold-out validation and similar to that performed in [34] was done as follows: for each dataset and using our defined optimal parameterization, one part of each dataset was taken to adjust the model and the rest of the sample for testing. In detail, to adjust the models, 70% of the total data in each dataset was used, leaving the remaining 30% for conducting the validation. The results and details of this are presented below.

4.1. Model Based on Random Forest Using Piling Data.

Table 2 summarizes the values obtained with RF in the parameter optimization process during training with the piling dataset. The parameters of interest for the optimal parametrization obtained in this model, that is, confidence (Con), number of trees (NoT), max depth (MDp), and accuracy (Acc), were used for interpreting results; these values are related to the confidence in a random tree model [43, 44].

Parameter Con is related to relative error, according to studies such as [1, 44]. Therefore, the values of $\text{Con} = \{25, 40, 55, 70, 85, 100\}$ were used for grouping the values of NoT, MDp, and Acc.

Figure 2 shows the values of Acc for each value of Con. Figure 2 also shows that all the graphs indicate a decreasing trend for parameter Acc, except for $\text{Con} = 40$. In this figure, the best mean value of Acc is for $\text{Con} = 100$, the following best values being for $\text{Con} = 40, 55$, and 85 . In all cases highlighted as the best, the average value of tree depth (MDp) is 8.5 . This may be interpreted as follows: the best combination of parameters is given when the mean tree depth of 8.5 is achieved; that is, this value represents the optimal depth in this classification.

On the basis of the piling data, the confusion matrix of this model was also obtained. In this optimization, 80% data were used for crossvalidation and 20% for validation [40] (Table 3).

TABLE 2: Mean values of NoT, MDp, and Acc (as a percentage) for each value of Con; optimal parametrization for each operational set.

Con	NoT	MDp	Acc (%)
25	62.5	8.5	83.4
40	62.5	8.5	83.5
55	62.5	8.5	83.6
70	62.5	12.7	83.5
85	62.5	8.5	83.6
100	62.5	8.5	83.8

Table 4 shows the importance of variables for this model. The most important variable is “agglomerate H dose,” followed by variable “RL.” In contrast, the least important variable is “soluble Cu.” Variables “operation day,” “H fed,” and “CO₃ grade” are over 10% of the value of importance, a fact that may be interpreted as their having a good predictive capacity for this model. This is not so for variable “Soluble Cu,” which does not exceed the threshold value of 10%.

4.2. RF-Based Model Using Operational Data. This section describes the results obtained with the operational data. Table 5 summarizes the statistical values obtained with RF in the parameter optimization process during training with the operational dataset. Like the model using piling data, parameters NoT, MDp, and Acc of optimal parametrization were used for interpreting results, grouped according to parameter Con. Figure 3 shows that all the graphs indicate a decreasing trend for parameter Acc. Also, all the mean values of Acc are quite close to one another (Table 5).

As can be seen in Figure 3, the best is when $\text{Con} = 25$. Other important aspects are, on the one hand, that the mean depth of trees increased (mean value = 12.7) as compared with the previous model (mean depth = 9.2). This indicates that a greater number of depth cases per each tree were classified, which is good for the model. On the other hand, the number of trees decreased (mean value = 23.4) as compared with the number of trees of the piling data model (mean value of the number of trees = 62.5). This may indicate that, as a whole, data were easier to group for the model algorithm.

Thus, on the basis of the abovementioned data and as shown in Figure 3, it may be stated that optimal parametrization for the operational data model is better than its equivalent with piling data.

Similar to the previous piling model, the confusion matrix for this model was also obtained, optimization procedure being the same as the previous model. Table 6 shows that all the values of recall (r) exceed 93%, the lowest being for the label high, thus coinciding with the previous model. Given this coincidence, the conditions for classifying records in this label should be improved to make future classifications better. The performance of the model is reliable, given the value $p = 98.90\%$ and the value of accuracy.

Table 7 shows the importance of variables for this model. The two most important variables here are the same as those of the piling data model (agglomerate H dose = 22.76% relative importance and RL = 18.86% relative importance). As

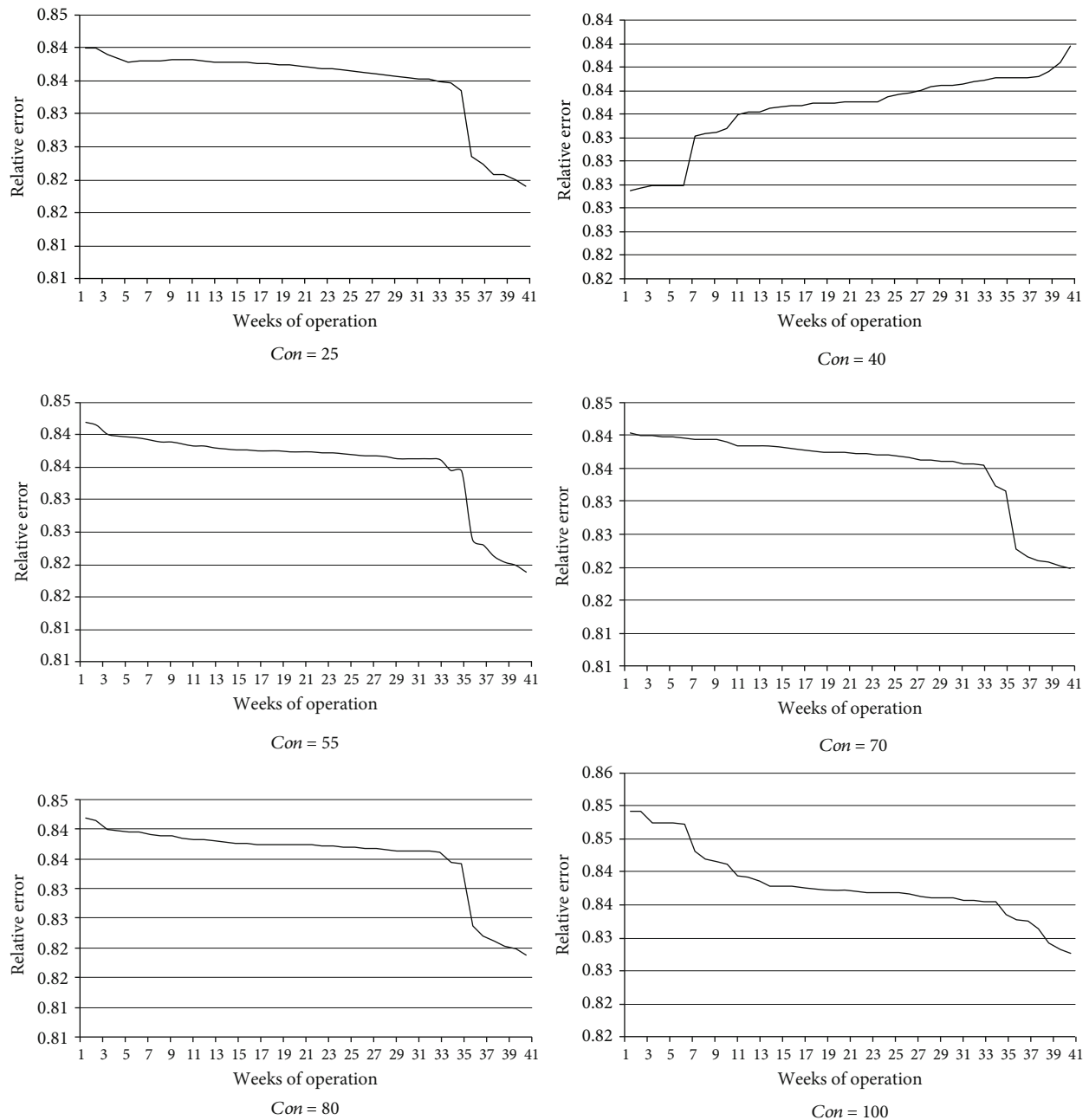


FIGURE 2: Evolution of parameter Acc for each of the values of parameter Con in optimization using piling data. The x values correspond to weeks of operations and y values correspond to accuracy values for each Con of relative error.

TABLE 3: Confusion matrix for operation with piling data. Value of accuracy = 98.72%.

	True low	True medium	True high	Class precision
Pred. low	1039	12	0	99.86%
Pred. medium	6	531	2	98.52%
Pred. high	0	1	47	97.92%
Class recall	99.43%	97.61%	95.92%	

can be seen in Table 7, the order or importance of variables is the same as shown in the previous model (Table 4), but the importance values are different. The least important variable in Table 7 is the same as in the previous model (Cu soluble). For this model, the percentage value of soluble Cu decreased in about 1%. This means that, although the order of importance of variables is maintained, the relative importance of the variables changes with respect to the previous model. Since this model was developed using operational data, it is prudent to consider that this order of importance is the most convenient. Figure 4 illustrates the contrast described above.

TABLE 4: Importance of variables for the piling data model.

Attribute	Importance	Relative importance (%)
Agglomerate H dose	13.74	19.31
RL	13.42	18.86
Total Cu grade	11.11	15.61
Day of operation	8.92	12.54
H fed	8.82	12.40
CO ₃ grade	8.39	11.79
Soluble Cu	6.75	9.49

TABLE 5: Mean values of NoT, MDp, and Acc (as a percentage) for each value of Con; optimal parametrization for each operational dataset.

Con	NoT	MDp	Acc (%)
25	23.0	13	91.9
40	23.6	12.7	91.6
55	23.6	12.7	91.6
70	23.6	12.7	91.6
85	22.9	12.7	91.6
100	24	12.7	91.6

Figure 5 summarizes the importance of variables according to RF models for each experiment. Particularly, the figure shows that variable H+fed (volumetric flow of ILS solution) is the most important, followed by variables RL, total Cu grade, and day of operation. The order of importance of the variables remains in both classifications; that is, reproductivity of the conditions of the leaching pile in a controlled environment (piling) is an accurate representation thoroughly describing the pile, and therefore, piling can be used to predict pile copper recovery, with a much lower cost and reliability in the predictive model resulting from piling.

5. Discussion

Artificial intelligence techniques, specifically soft computing, are being used in productive industry to generate predictive models that improve industrial activity [25]. Random forest (RF) was used in this study to predict copper recovery by leaching. Predictive models using RF have been recently published by the mining industry, showing good results such as those reported in [3, 12, 33], but these studies were directed to objectives different from copper recovery prediction.

In recent papers such as [9], artificial intelligence computing tools (particularly machine learning algorithms) have been reported, but no evidence of the use of RF has been found in the literature to predict copper recovery. However, these works have helped to identify and relate information that directly influences to improve the copper recovery process by leaching.

The study published in [3] highlights that machine learning algorithms, since they are artificial neural networks, regression trees, random forest, and support vector machines, make up powerful tools currently scarcely used

in the copper mining industry, though there should be a tendency to increasingly use these machine learning tools in the present mining industry.

In RF, each tree is developed on the basis of the bootstrap algorithm philosophy. This may mean that the classification obtained for each tree is precise, thus causing a positive impact on the models presented here. In addition, this philosophy of work has made it possible to use all datasets in the classification and generate the models. The model precision obtained in this study is similar in both cases. The model for both datasets shows that a wealth of information was used to interpret the influence of predictive variables on class. For example, the order of the variables of interest is similar in both models and the performance shown by variables Acc, p , and r enables concluding that both models have a good quality and could be used to predict copper recovery in new cases with a good reliability value.

The capacity to identify the importance of variables for the model using training data (piling) is similar to the one shown by the model using actual data (operation). This was an expected result since the leaching material was the same in both cases, but this result validates the applicability of the machine learning algorithm selected for generating the models.

On the basis of the above described information, the objective of environmental trade-off was accomplished because model performance is optimal, and in both cases, the greatest number of records was classified as normal, when the acid irrigation rate lies between 20 and 50 g/l (normal value).

6. Conclusions

Copper recovery prediction by hydrometallurgical methods and, particularly, leaching is usually made with the help of mathematical models, but soft computing techniques can help create complex computational models [45] that help in this prediction. Recently, an increase in using soft computing tools in the industry has been observed [9, 13, 39], but in this particular case, the literature does not contain many studies reporting the use of RF to generate a copper recovery prediction model.

This study resulted in the generation of two copper recovery prediction models using the leaching method. Actual data (operation) were used in one of the models, while the other model was generated with hive-simulated data which had the same characteristics as the material to be leached and the lixiviant. In both cases, the models achieved an excellent predictive quality, one of the cases reaching 100% prediction for the label high, the mean being higher than 95% precision. In this way, it excelled in what was posed in the objective of this study (described at the beginning of this document).

As recently published in [9], a comparison between a linear model and an artificial neural network (ANN) for predicting copper recovery is made. One of the conclusions of this study is that ANN exceeds the linear model in terms of precision, but as conclusion at the present work, the interpretation capacities of RF-generated models exceed those of

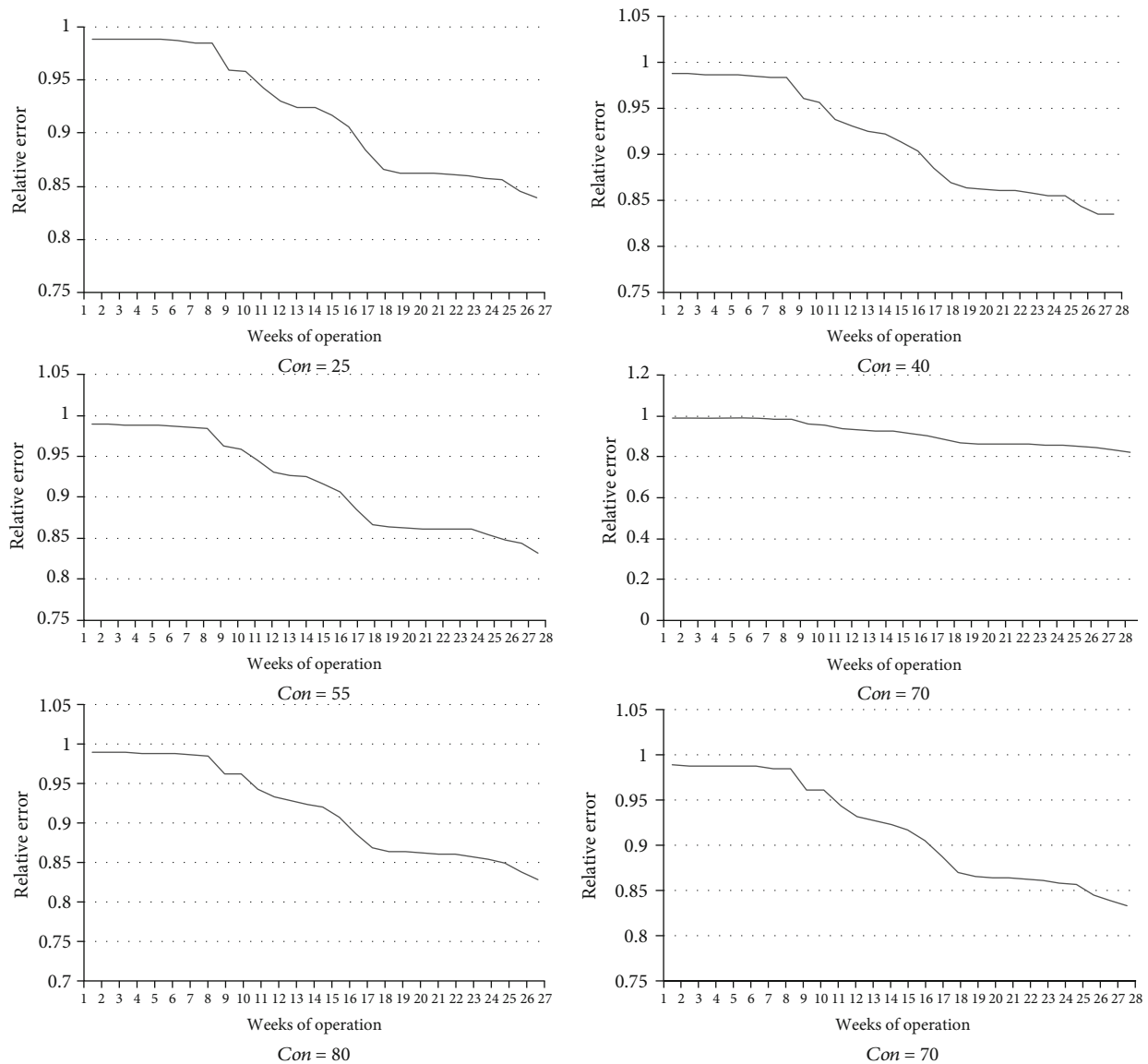


FIGURE 3: Evolution of parameter Acc for each value of parameter Con in optimization using operational data. The x values correspond to weeks of operations and y values correspond to accuracy values for each Con of relative error.

TABLE 6: Confusion matrix for the operation with operational data. Value of accuracy = 98.90%.

	True low	True medium	True high	Class precision
Pred. low	1487	7	0	99.53%
Pred. medium	10	434	5	96.58%
Pred. high	0	0	68	100.00%
Class recall	99.33%	98.38%	93.15%	

TABLE 7: Importance of variables for the operational data model.

Attribute	Importance	Relative importance (%)
Agglomerate H dose	23.02	22.76
RL	18.32	18.86
Total Cu grade	16.28	16.10
Day of operation	12.63	12.49
H fed	11.07	10.95
CO ₃ grade	11.01	10.89
Soluble Cu	8.79	8.69

ANN from the work previously mentioned, thus making it easier to arrive at conclusions.

This study helped make a comparison between two copper recovery prediction models in the same work context.

Adjustment precision measure indicates that the RF algorithm is highly useful for processes to predict future copper production.

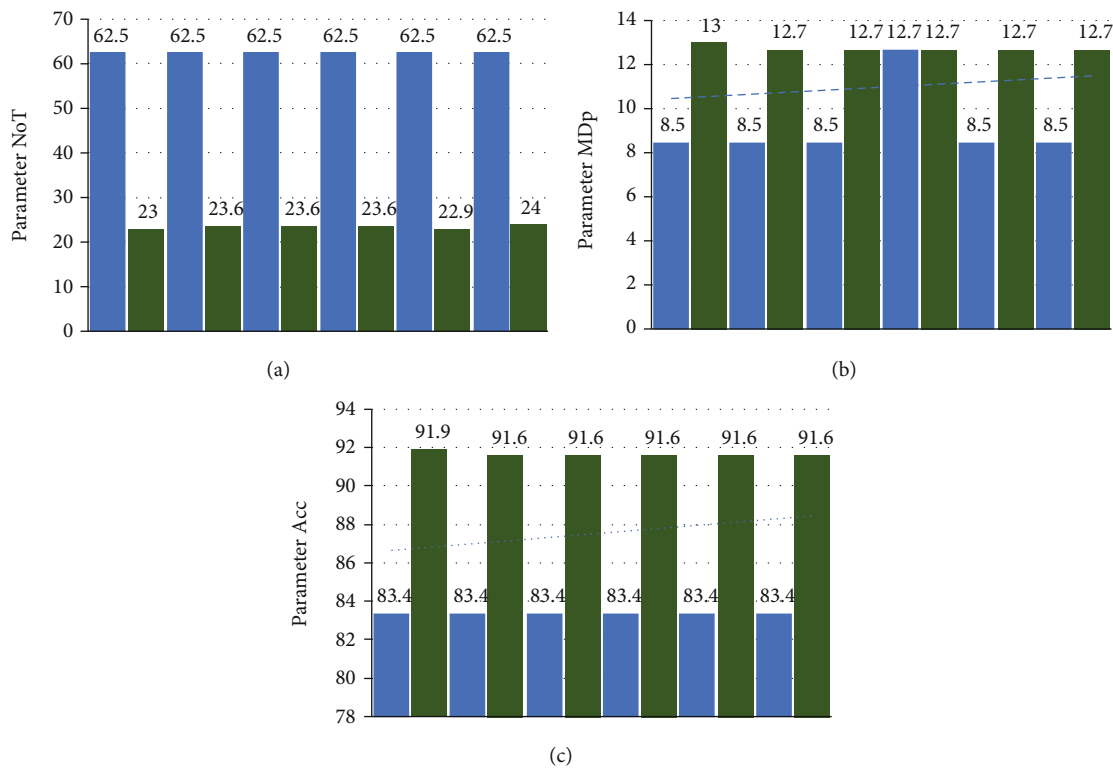


FIGURE 4: Comparison of parameter values for each value of parameter of interest for the optimal parametrization, obtained in each of the predictive models. (a) corresponds to parameter NoT, (b) to parameter MDP, and (c) to parameter Acc.

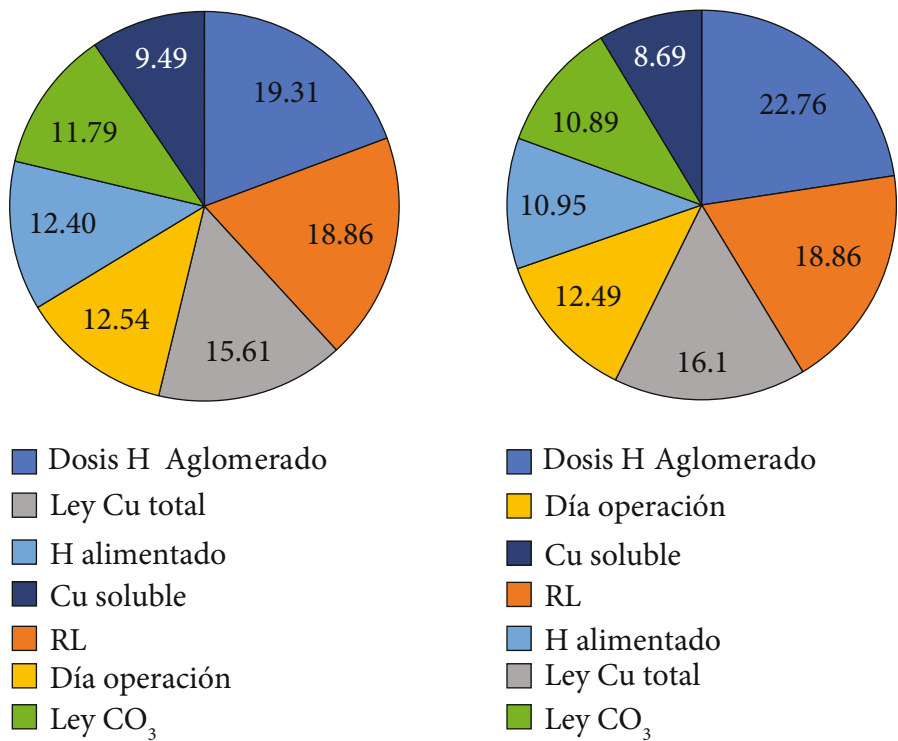


FIGURE 5: Importance of predictive variables for each RF-generated model.

In addition, experience was gained for defining and implementing the predictive model in the leaching domain on this specific work context. This experience may be used for other simulations of processes relative to the improvement of results to obtain copper at SCM Franke by means of soft computing techniques or other companies of the same industrial production sector.

What was said about model performance, the capacity to identify the influence of variables on class, and the capacity to interpret results, etc., is very important in the copper industry because it allows generating supporting tools for material exploitation planning, along with viewing, via indicators generated with this type of model, copper recovery results in the presence of a certain material. It also allows properly selecting both the most influential variables and the values of those variables to achieve the desired recovery. This may have a considerable impact on the intelligent exploitation of this mineral, considering the increasing demand and lack of this industrial activity.

To conduct this study, a methodology was proposed; results obtained by following the methodological steps devised show excellent quality and are replicable for other copper leaching piles to study the future performance of copper recovery using the prediction method. Also, this methodology can be transferred to other copper leaching processes, including the knowledge of this particular process to generate a predictive model. In this way, this study may indicate a future line of research.

Data Availability

The input data used to support the findings of this study could be available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors thank the SCM Franke company for the material support provided and their collaboration to this study, particularly their support for obtaining and storing data from the copper production process using the leaching method, as well as data from the pilot plant (column simulation) and their experience to select the parameters and criteria used for generating predictive models, and are thankful for the financial support provided by Universidad Católica del Norte.

References

- [1] A. C. Pereira and F. Romero, "A review of the meanings and the implications of the industry 4.0 concept," *Procedia Manufacturing*, vol. 13, pp. 1206–1214, 2017.
- [2] C. Pietrobelli, A. Marin, and J. Olivari, "Innovation in mining value chains: new evidence from Latin America," *Resources Policy*, vol. 58, pp. 1–10, 2018.
- [3] J. Katz and C. Pietrobelli, "Natural resource based growth, global value chains and domestic capabilities in the mining industry," *Resources Policy*, vol. 58, pp. 11–20, 2018.
- [4] L. Stubrin, "Innovation, learning and competence building in the mining industry. The case of knowledge intensive mining suppliers (KIMS) in Chile," *Resources Policy*, vol. 54, pp. 167–175, 2017.
- [5] J. De Gregorio and F. Labbé, "Copper, the real exchange rate and macroeconomic fluctuations in Chile," *Beyond the Curse: Policies to Harness the Power of Natural Resources*, pp. 203–233, 2011.
- [6] Y. Ghorbani and S. H. Kuan, "A review of sustainable development in the Chilean mining sector: past, present and future," *International Journal of Mining, Reclamation and Environment*, vol. 31, no. 2, pp. 137–165, 2017.
- [7] E. L. Esquenazi, B. K. Norambuena, Í. M. Bacigalupo, and M. G. Estay, "Evaluation of soil intervention values in mine tailings in northern Chile," *PeerJ*, vol. 6, p. e5879, 2018.
- [8] E. J. Lam, B. F. Keith, Í. L. Montofré, and M. E. Gálvez, "Copper uptake by *Adesmia atacamensis* in a mine tailing in an arid environment," *Air, Soil and Water Research*, vol. 11, 2018.
- [9] C. Leiva, V. Flores, F. Salgado, D. Poblete, and C. Acuña, "Applying soft computing for copper recovery in leaching process," *Scientific Programming*, vol. 2017, 6 pages, 2017.
- [10] C. A. Leiva, K. V. Arcos, D. P. Poblete, E. A. Serey, C. M. Torres, and Y. Ghorbani, "Design and evaluation of an expert system in a crushing plant," *Minerals*, vol. 8, no. 10, p. 469, 2018.
- [11] P. Meller and A. M. Simpasa, *Role of Copper in the Chilean & Zambian Economies: Main Economic and Policy Issues*, Global Development Network, GDN Working Paper Series, 2011.
- [12] L. A. Zadeh, "Fuzzy logic, neural networks, and soft computing," *Communications of the ACM*, vol. 37, no. 3, pp. 77–84, 1994.
- [13] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas, "Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines," *Ore Geology Reviews*, vol. 71, pp. 804–818, 2015.
- [14] S. Nawar and A. Mouazen, "Comparison between random forests, artificial neural networks and gradient boosted machines methods of on-line Vis-NIR spectroscopy measurements of soil total nitrogen and total carbon," *Sensors*, vol. 17, no. 10, p. 2428, 2017.
- [15] F. Zandiyyeh, M. R. Shayestefar, H. Ranjbar, and S. Saadat, "Prospectivity mapping of iron oxide-copper-gold (IOCG) deposits using support vector machine method in Feyzaabad area (east of Iran)," *Journal of Himalayan Earth Sciences*, vol. 49, no. 2, pp. 50–62, 2016.
- [16] B. S. Saljoughi and A. Hezarkhani, "A comparative analysis of artificial neural network (ANN), wavelet neural network (WNN), and support vector machine (SVM) data-driven models to mineral potential mapping for copper mineralizations in the Shahr-e-Babak region, Kerman, Iran," *Applied Geomatics*, vol. 10, no. 3, pp. 229–256, 2018.
- [17] T. Sun, F. Chen, L. Zhong, W. Liu, and Y. Wang, "GIS-based mineral prospectivity mapping using machine learning methods: a case study from Tongling ore district, eastern China," *Ore Geology Reviews*, vol. 109, pp. 26–49, 2019.

- [18] V. Rodriguez-Galiano, B. Ghimire, M. Chica-Olmo, and J. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93–104, 2012.
- [19] A. Porwal, E. Carranza, and M. Hale, "Knowledge-driven and data-driven fuzzy models for predictive mineral potential mapping," *Natural Resources Research*, vol. 12, no. 1, pp. 1–25, 2003.
- [20] M. Abedi, S. A. Torabi, G. H. Norouzi, M. Hamzeh, and G. R. Elyasi, "PROMETHEE II: a knowledge-driven method for copper exploration," *Computers & Geosciences*, vol. 46, pp. 255–263, 2012.
- [21] H. K. Haghighi, M. Rafie, and D. Moradkhani, "Modeling on transition of heavy metals from Ni–Cd zinc plant residue using artificial neural network," *Transactions of the Indian Institute of Metals*, vol. 68, no. 5, pp. 741–756, 2015.
- [22] R. Xu, *Improvements to random forest methodology*, [Ph.D. thesis], Iowa State University, Ames, IA, USA, 2013.
- [23] T. Kohonen, "An introduction to neural computing," *Neural Networks*, vol. 1, no. 1, pp. 3–16, 1988.
- [24] S. Visweswaran and G. Cooper, "Learning instance-specific predictive models," *Journal of Machine Learning Research*, vol. 11, pp. 3333–3369, 2010.
- [25] V. Flores and M. Correa, "Performance of predicting surface quality model using softcomputing, a comparative study of results," in *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pp. 233–242, Springer, 2017.
- [26] V. Skorpil and J. Stastny, *Neural networks and backpropagation algorithm*, Electronics, Bulgaria, Sozopol, 2006.
- [27] D. M. Skapura and J. A. Freeman, *Neural Networks, Algorithms, Applications and Programming Techniques*, Reading Addison Wesley, 1991.
- [28] M. Cilimkovic, *Neural networks and backpropagation algorithm*, Institute of Technology Blanchardstown, Blanchardstown Road North Dublin, 2015.
- [29] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [30] E. J. M. Carranza and A. G. Laborte, "Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines)," *Computers & Geosciences*, vol. 74, pp. 60–70, 2015.
- [31] T. K. Ho, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, pp. 278–282, Montreal, Quebec, Canada, Aug. 1995.
- [32] C. Zhang, C. Liu, X. Zhang, and G. Almpandis, "An up-to-date comparison of state-of-the-art classification algorithms," *Expert Systems with Applications*, vol. 82, pp. 128–150, 2017.
- [33] V. Flores and B. Keith, "Gradient boosted trees predictive models for surface roughness in high-speed milling in the steel and aluminum metalworking industry," *Complexity*, vol. 2019, Article ID 1536716, 15 pages, 2019.
- [34] L. Gutiérrez, V. Flores, B. Keith, and A. Quelopana, "Using the Belbin method and models for predicting the academic performance of engineering students," *Computer Applications in Engineering Education*, vol. 27, no. 2, pp. 500–509, 2019.
- [35] I. Barandiaran, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.
- [36] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest," in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 154–168, Springer, Berlin, Heidelberg, 2012.
- [37] R. Lior, "Data mining with decision trees: theory and applications," *World scientific*, vol. 81, pp. 11–50, 2014.
- [38] O. Salmani Nuri, E. Allahkarami, M. Irannajad, and A. Abdollahzadeh, "Estimation of selectivity index and separation efficiency of copper flotation process using ANN model," *Geosystem Engineering*, vol. 20, no. 1, pp. 41–50, 2017.
- [39] H. A. Nooruddin, F. Anifowose, and A. Abdulraheem, "Using soft computing techniques to predict corrected air permeability using Thomeer parameters, air porosity and grain density," *Computers & Geosciences*, vol. 64, pp. 72–80, 2014.
- [40] C. K. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, 1968.
- [41] P. M. Granitto, F. Gasperi, F. Biasioli, E. Trainotti, and C. Furlanello, "Modern data mining tools in descriptive sensory analysis: a case study with a random forest approach," *Food Quality and Preference*, vol. 18, no. 4, pp. 681–689, 2007.
- [42] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [43] S. Bhattacharyya, "Confidence in predictions from random tree ensembles," *Knowledge and Information Systems*, vol. 35, no. 2, pp. 391–410, 2013.
- [44] S. Wager, T. Hastie, and B. Efron, "Confidence intervals for random forests: the jackknife and the infinitesimal jackknife," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1625–1651, 2014.
- [45] M. Milivojevic, S. Stopic, B. Friedrich, B. Stojanovic, and D. Drndarevic, "Computer modeling of high-pressure leaching process of nickel laterite by design of experiments and neural networks," *International Journal of Minerals, Metallurgy, and Materials*, vol. 19, no. 7, pp. 584–594, 2012.