An LSTM Approach for SAG Mill Operational Relative-Hardness Prediction

Sebastian Avalos 1,* , Willy Kracht 2,3 and Julian M. Ortiz 1

1 The Robert M. Buchan Department of Mining, Queen’s University, Ottawa, ON K1G 5Z3, Canada; julian.ortiz@queensu.ca
2 Department of Mining Engineering, Universidad de Chile, Santiago 8370448, Chile; wkracht@uchile.cl
3 Advanced Mining Technology Center, AMTC, Universidad de Chile, Santiago 8370451, Chile
* Correspondence: sebastian.avalos@queensu.ca

Received: 1 August 2020; Accepted: 18 August 2020; Published: 20 August 2020

Abstract: Ore hardness plays a critical role in comminution circuits. Ore hardness is usually characterized at sample support in order to populate geometallurgical block models. However, the required attributes are not always available and suffer for lack of temporal resolution. We propose an operational relative-hardness definition and the use of real-time operational data to train a Long Short-Term Memory, a deep neural network architecture, to forecast the upcoming operational relative-hardness. We applied the proposed methodology on two SAG mill datasets, of one year period each. Results show accuracies above 80% on both SAG mills at a short upcoming period of times and around 1% of misclassifications between soft and hard characterization. The proposed application can be extended to any crushing and grinding equipment to forecast categorical attributes that are relevant to downstream processes.

Keywords: semi-autogenous grinding mill; operational hardness; energy consumption; mining; deep learning; long short-term memory

1. Introduction

In mining operations, the primary energy consumer is the comminution system, responsible for more than half of the entire mine consumption [1]. From all pieces of equipment that integrate the comminution circuit, the semi-autogenous grinding mill (SAG) is perhaps the most important in the system. With an aspect ratio of 2:1 (diameter to length), these mills combine impact, attrition and abrasion to reduce the ore size. SAG mills are located at the beginning of the comminution circuits, after a primary crushing stage. Although there are small SAG mills, their size usually ranges from 9.8 × 4.3 to 12.8 × 7.6 m, with a nominal energy demand of 8.2 and 26 MW, respectively [2], which make SAG mills the most relevant energy consumer within the concentrator. Modelling their consumption behaviour supports the operational control and energy demand-side management [3].

Most theoretical and empirical models [4–6] demand input feed characteristics, such as hardness, size distribution and inflow rate, SAG characteristics, such as sizing and product size distribution, and operational variables such as bearing pressure, water addition and grinding charge level. Although they are suitable to provide adequate design guidelines, they lack accurate in-situ inference since most assume steady-state and isolation from up and downstream processes. In response, model predictive control, SAG MPC [7], combines those methods with real-time operational information. However, expert knowledge is required to model the SAG mill dynamics properly.

From a geometallurgical perspective, the integration of new predictive methods that account for space and time relationships over real-time attributes has been defined as a fundamental challenge [8,9] in mining operations, particularly in an integrated system such as comminution.
In response, data-driven approaches have been proposed ranging from support vector machines [10] and gene expression programming [11] to hybrid models that combine genetic algorithms and neural networks [12] and recurrent neural networks [13]. As data-driven methods are sensitive to the context (available information) and representation (information workflow), the authors have studied the use of several machine learning and deep learning methods in modelling the SAG energy consumption behaviour based only on operational variables [14].

The energy consumed by a SAG mill is related to several factors such as expert operator decisions, charge volume, charge specific gravity and the hardness of the feed material. Knowing the output hardness material becomes relevant for the downstream stage in the primary grinding circuit. Ore hardness can be characterized at sample support by combining the logged geological properties and the result of standardized comminution tests. They can be used to predict the hardness of each block sent to the process. However, these attributes are not always available. In response, a qualitative characterization of the ore hardness processed at time $t$, relative to the operational hardness of the ore processed at time $t + 1$ can be done using only operational variables rather than a set of mineralogical characterizations. This qualitative characterization is referred and here used as operational relative-hardness (ORH).

We take advantage of previous works [14] by knowing that the Long Short-Term Memory (LSTM) [15] outperforms other machine learning and deep learning techniques on inferring the SAG mill energy consumption. Therefore, Section 2 presents the ORH and LSTM models, Section 3 establishes the SAG mill experimental framework, the results of which are presented in Section 4, and conclusions are drawn in Section 5.

2. Model

2.1. Operational Relative-Hardness Criteria

From the several operational parameters that can be captured and associated to SAG mill operations, we consider the energy consumption (EC) and feed tonnage (FT) to build our operational relative-hardness criteria.

Let us assume that data \{EC,FT\}$_t$ is collected over a period of time $T$ using a $\Delta t$ discretization. By considering the one-step forward time difference of energy consumption ($\Delta$EC$_t$ = EC$_{t+1}$ − EC$_t$) and feed tonnage ($\Delta$FT$_t$ = FT$_{t+1}$ − FT$_t$), a qualitative assessment of the operational relative-hardness can be done. For instance, if the energy consumption is increasing and the feed tonnage is constant, it can be interpreted as an increase in ore hardness relative to the previous period. Similarly, if the feed tonnage is constant and the energy decreases, a decrease in ore hardness relative to the previous period can be assumed. Particularly, when both $\Delta$EC$_t$ and $\Delta$FT$_t$ show the same behaviour, the SAG can be either processing ore with medium operational relative-hardness or being filled up or emptied. To avoid misclassification in this last case, the operational relative-hardness is labelled as undefined.

Table 1 summarizes the nine combinations of states and the associated operational relative-hardness. The qualitative labelling of $\Delta$EC$_t$ and $\Delta$FT$_t$ as increasing, constant or decreasing can be established based on their global distribution over the period $T$ as:

$$\begin{align*}
\Delta$EC$_t &= \begin{cases} 
\text{Increasing} & \text{if } \Delta$EC$_t > \lambda \cdot \sigma_{\Delta$EC} \\
\text{Constant} & \text{if } |\Delta$EC$_t| \leq \lambda \cdot \sigma_{\Delta$EC} \\
\text{Decreasing} & \text{if } \Delta$EC$_t < -\lambda \cdot \sigma_{\Delta$EC}
\end{cases} \\
\Delta$FT$_t &= \begin{cases} 
\text{Increasing} & \text{if } \Delta$FT$_t > \lambda \cdot \sigma_{\Delta$FT} \\
\text{Constant} & \text{if } |\Delta$FT$_t| \leq \lambda \cdot \sigma_{\Delta$FT} \\
\text{Decreasing} & \text{if } \Delta$FT$_t < -\lambda \cdot \sigma_{\Delta$FT}
\end{cases}
\end{align*}$$

(1)

where $\sigma_{\Delta$EC} and $\sigma_{\Delta$FT} represent the standard deviations over the period $T$ of EC and FT, respectively, and $\lambda$ is a scalar value that modulates the labelling distribution. Note that (i) a $\lambda$ value above 1.5 would make the entire definition meaningless since most values would remain as constant, and (ii) the $\lambda$ value definition is an external model parameter and can be guided either subjectively or via statistical meaning.
Table 1. Operational relative-hardness criteria based on one time-step difference of energy consumption and feed tonnage.

| Energy Consumption | Feed Tonnage | Operational Relative-Hardness |
|--------------------|-------------|-----------------------------|
| Constant           | Decreasing  | Hard                        |
| Increasing         | Constant    | Hard                        |
| Increasing         | Decreasing  | Hard                        |
| Decreasing         | Decreasing  | Undefined                   |
| Increasing         | Increasing  | Undefined                   |
| Constant           | Constant    | Undefined                   |
| Decreasing         | Constant    | Soft                        |
| Decreasing         | Increasing  | Soft                        |

2.2. Long Short-Term Memory

The Long Short-Term Memory (LSTM) [15] neural network architecture belongs to the family of recurrent neural networks in Deep Learning [16]. They are suitable to capture short and long term relationships in temporal datasets. Internally, LSTM applies several combinations of affine transformations, element-wise multiplications and non-linear transfer functions, for which the building blocks are:

- \( x_t \): input vector at time \( t \). Dimension \((m, 1)\).
- \( W_f, W_i, W_c, W_o \): weight matrices for \( x_t \). Dimensions \((n_H, m)\).
- \( h_t \): hidden state at time \( t \). Dimension \((m, 1)\).
- \( U_f, U_i, U_c, U_o \): weight matrices for \( h_{t-1} \). Dimensions \((n_H, m)\).
- \( b_f, b_i, b_c, b_o \): bias vectors. Dimensions \((n_H, 1)\).
- \( V \): weight matrix for \( h_t \) as output. Dimension \((K, m)\).
- \( c \): bias vector for output. Dimension \((K, 1)\).

where \( m \) is the number of variables as input, \( K \) is the number of output variables, and \( n_H \) is the number of hidden units. Let \( \tau \in \mathbb{N} \) be a temporal window. At each time \( t \in \{ 1, ..., \tau \} \), the LSTM receives the input \( x_t \), the previous hidden state \( h_{t-1} \) and previous memory cell \( c_{t-1} \). The forget gate \( f_t = \sigma( W_f x_t + U_f h_{t-1} + b_f ) \) is the permissive barrier of the information carried by \( x_t \). The input gate \( i_t = \sigma( W_i x_t + U_i h_{t-1} + b_i ) \) decides the relevance of the information carried by \( x_t \). Note that both \( f_t \) and \( i_t \) use sigmoid \( \sigma(x) = (1 + e^{-x})^{-1} \) as the activation function over a linear combination of \( x_t \) and \( h_{t-1} \).

By passing the combination of \( x_t \) and \( h_{t-1} \) through a Tanh function, a candidate memory cell \( \tilde{c}_t = \text{Tanh}( W_c x_t + U_c h_{t-1} + b_c ) \) is computed. The final memory cell \( c_t = f_t \odot \tilde{c}_{t-1} + i_t \odot \tilde{c}_t \) is computed as a sum of (i) what to forget from the past memory cell as an element-wise multiplication (\( \odot \)) between \( f_t \) and \( c_{t-1} \), and (ii) what to learn from the candidate memory cell as an element-wise multiplication (\( \odot \)) between \( i_t \) and \( \tilde{c}_t \).

Similar to \( i_t \) and \( f_t \) the output gate \( o_t = \sigma( W_o x_t + U_o h_{t-1} + b_o ) \) passes through a sigmoid function a linear combination between \( x_t \) and \( h_{t-1} \). It controls the information passing from the current memory cell \( c_t \) to the final hidden state \( h_t = \text{Tanh}( c_t ) \odot o_t \) as an element-wise multiplication between \( o_t \) and \( \text{Tanh}( c_t ) \). At the final step \( \tau \), the output is computed as \( y_\tau = ( V h_\tau + c ) \). When dealing with more than one categorical prediction \( (K > 1) \), as in the present work for ORH forecasting, a softmax function is applied over \( y_\tau \) to obtain the normalized probability distribution, and the category \( k \) has a probability of \( \hat{p}(k) = \frac{\exp(y_{\tau,k})}{\sum_{k=1}^{K} \exp(y_{\tau,k})} \).

An illustrative scheme of the internal connection at time step \( t \) inside an LSTM is shown in Figure 1 (left). The ORH prediction has three categories (hard, soft and undefined) and the probability is computed at the last unit, at time step \( \tau \), as shown in the unrolled LSTM in Figure 1 (right).
Figure 1. Schemes. Information flow inside Long Short-Term Memory (LSTM) (left) and unrolled LSTM where the output is computed at the last recurrence (right).

3. Experiment

3.1. Dataset

We used two datasets containing operational data for two independent SAG mills every half hour over a total time of 340 and 331 days, respectively. Each one of the SAG mills receives fresh feed and is connected in an open circuit configuration (SABC-B) where the pebble crusher product is sent to ball mills. At each time $t$, the dataset contains Feed tonnage (FT) (ton/h), Energy consumption (EC) (kWh), Bearing pressure (BPr) (psi) and Spindle speed (SSp) (rpm). They are split into two main subsets (a validation dataset is not considered since the optimum LSTM architecture to train is drawn from previous work [14]): training and testing (Table 2). This is an arbitrary division, and we seek to have a proportion of $\sim 50/50$, respectively.

Table 2. Summary statistics over training testing dataset on semi-autogenous grinding mill (SAG) mills.

| SAG Mill 1                  | Training | Testing | Dataset |
|-----------------------------|----------|---------|---------|
| Variable                    | Min      | Mean    | Max     | St Dev  | Count |
| Feed Tonnage (ton/h)        | 0        | 911     | 884     | 2111    | 1953  |
| Energy Consumption (kWh)    | 0        | 9927    | 8920    | 12,248  | 10,809 |
| Bearing Pressure (psi)      | 0        | 12.7    | 11.9    | 13.7    | 2.2   |
| Spindle Speed (rpm)         | 0        | 9.2     | 9.1     | 10.3    | 0.7   |

| SAG Mill 2                  | Training | Testing | Dataset |
|-----------------------------|----------|---------|---------|
| Variable                    | Min      | Mean    | Max     | St Dev  | Count |
| Feed Tonnage (ton/h)        | 0        | 2077    | 2073    | 3477    | 3452  |
| Energy Consumption (kWh)    | 0        | 16,709  | 17,445  | 19,688  | 19,533 |
| Bearing Pressure (psi)      | 0        | 13.8    | 14.8    | 18.3    | 3.5   |
| Spindle Speed (rpm)         | 0        | 9.1     | 8.9     | 10.0    | 9.9   |

As it can be seen in Table 2, the predictive methods are trained with the first 50% and tested with the upcoming 50%, without being fed with the previous 50% of historical data.

Note that the comminution properties of the ore, such as $a \times b$ or BWi, are not included in the datasets; therefore, the relationship between forecasted ORH and comminution properties is not explored in this work. The results herein presented, however, serve as a basis to examine such a relationship if those properties were known.

3.2. Assumptions

SAG mills are fundamental pieces in comminution circuits. As no information regarding downstream/upstream processes is available, recognizing bottlenecks in the dataset becomes subjective. We assume that SAG mills will potentially show changes from steady-state to under capacity and vice versa along with the dataset. Thus, stationarity of all operational variable distributions is
assumed throughout this work, including the ore grindability. It means that the entire dataset belongs to a
known and planned combination of ore characteristics (geometallurgical units). By doing so, we limit
the applicability of the present models beyond the temporal dataset without a proper training process.

As explained in the problem statement section, we make use of the temporal average over energy
consumption and feed tonnage as input for operational hardness prediction. Thus, we assume an
additivity property over those variables as their units are kWh and ton/h, respectively, over constant
temporal discretization so averaging adjacent data points is mathematically consistent.

In the operation from which the datasets were obtained, the SAG mill liners are replaced every
5–7 months. Since the datasets cover almost a year, we can ensure that the liners were replaced
in each SAG mill at least once during the tested period, which may alter the relationship between
energy consumption and other operational variables, inducing a discontinuity in the temporal plots.
However, since in this work the temporal window for ORH evaluation is eight hours, the local
discontinuity associated with liners replacement is not expected to affect the forecast at that time frame.
The ORH is related to what was happening in the corresponding mill within the last few hours, and
not to the mill behaviour prior to the last replacement of liners.

3.3. Problem Statement

The aim is to forecast the operational relative-hardness. To do so, we need to label the datasets
with the associated ORH category at data point. We know from Equation (1) that the ORH labelling
process requires as input (i) the one-step forward differences on energy consumption (ΔECt) and feed
 tonnage (ΔFTt), and (ii) a lambda (λ) value. In addition, we are interested in forecasting the ORH at
different time supports.

Since the information is collected every 30 min, the upcoming energy consumption ECt+1 and
feed tonnage FTt+1 at 0.5 h support are denoted simply as ECt+1 and FTt+1 in reference to ECt+1(0.5h)
and FTt+1(0.5h), respectively. An upcoming EC and FT at 1 h support, ECt+1(1h) and FTt+1(1h), are computed
by averaging the next two energy consumption, ECt+1 and ECt+2, and the two feed tonnage, FTt+1
and FTt+2. Similarly, by averaging the upcoming ECs and FTs, different supports can be computed.
Let s be the time support in hours, which represents the average over a temporal interval of a given
duration, then ECt+1(s) and FTt+1(s) are calculated as:

\[
EC_{t+1}^{(s)} = \frac{EC_{t+1} + \ldots + EC_{t+2s}}{2s} \quad \text{and} \quad FT_{t+1}^{(s)} = \frac{FT_{t+1} + \ldots + FT_{t+2s}}{2s} \quad (2)
\]

In this experiment, three different supports (sh) are considered: 0.5, 2 and 8 h.

Figure 2 illustrates the ORH criteria using a half-hour time support on SAG mill 1 dataset.
From the daily graph of ECt(0.5h) and FTt(0.5h) at the top, the graph of ΔECt(0.5h) and ΔFTt(0.5h)
are extracted and presented at the centre and bottom, respectively. Three different bands, corresponding to
λ: 0.5, 1.0 and 1.5, are shown. The values that are above the band are considered as increasing, the ones
below it are considered as decreasing and inside as undefined (relatively constant). The corresponding
categories for EC and FT are used to define the operational relative-hardness (as in Table 1). It can be
seen that, when λ increases, the proportions of hard and soft instances decrease. Since λ is an arbitrary
parameter, a sensitivity analysis is performed in the range [0.5, 1.5] to capture its influence on the
resulting LSTM accuracy to suitably learn to predict the ORH at the different time supports.

At each time t the input variables considered to predict ORHt+1(λk) are FTt, BPr, and SSp. To account
for trends, and since FT and SSp are operational decisions, the differences FTt+1−FTt and SSp_{t+1}−SSp_t
are also considered as inputs. Therefore, the dataset of predictors and output \{X, Y\} ∈ R^5 × R, at
each time support sh, has samples \{x_i, y_i\} ∈ \{X, Y\} made by \(x_i = \{\text{FT}_t, \text{BPr}_t, \text{SSp}_t, \text{FT}_{t+1}−\text{FT}_t,
\text{SSp}_{t+1}−\text{SSp}_t\}\) and \(y_i = \{\text{ORH}_t^{(λk)}\}\). We also tried several other combinations of input variables, but
all led to results with lower quality. A temporal window of the previous four hours (previous eight
consecutive data points) are used as input for training and testing the LSTM models.
3.4. Preprocessing Dataset

A preprocessing step is performed over the raw datasets to make them suitable for deep neural network training and inference processes. The aim is to make all input attributes fall into certain regions of the non-linear transfer functions via normalization and to be properly coded in categories via one-hot encoding. Thus, we normalize the entire raw dataset with the mean and standard deviation of the training dataset.

Let \( x^{(\text{var})}_t \) be one of the five input variables \( \text{var} \) at time \( t \), its normalized expression is computed as \( x^{(\text{var})}_t = \frac{\text{var}_t - m_{\text{var}}}{s_{\text{var}}} \), where \( m_{\text{var}} \) and \( s_{\text{var}} \) represent the mean and standard deviation of \( \text{var} \) in the training dataset. We normalize the first three attributes of \( x_t \), \( \text{FT}_t \), \( \text{BPr}_t \) and \( \text{SSp}_t \) while for last two attributes, the differences between the original values \( \text{FT}_{t+1} - \text{FT}_t \) and \( \text{SSp}_{t+1} - \text{SSp}_t \), are replaced by the differences between the normalized values of \( \text{FT} \) and \( \text{SSp} \).

The known operational relative-hardness at time \( t \) (\( y_t \)) is one-hot encoding such that soft, undefined and hard are encoded as \([1, 0, 0]\), \([0, 1, 0]\) and \([0, 0, 1]\), respectively.

3.5. Optimal LSTM Architecture

From the training dataset, sequence \{\( x_1, ..., x_\tau \)\} of length \( \tau \) are extracted to train the LSTM model in order to forecast the operational relative-hardness at next time step \( \tau + 1 \), at different time supports. The chosen length is four hours (\( \tau: 8 \)).

The external hyper-parameter to be optimized on any LSTM architecture is the number of hidden units, \( n_H \). Based on a previous work [14], the optimum number of hidden units was found and here used. They are displayed in Table 3.

Adam Optimizer is used to train the LSTM with hyper-parameters \( \epsilon = 1 \times e^{-8} \), \( \beta_1 = 0.9 \) and \( \beta_2 = 0.999 \) as recommended by [17].
4. Results

Directly from the datasets, the real operational relative-hardness ORH\(_R\) is calculated from Equation (1), varying \(\lambda\) in the set \((0.5, 0.6, ..., 1.4, 1.5)\) at each time \(t\) and for each time support. On the other hand, a probability vector with soft, undefined and hard ORH states is predicted. By taking the highest probability, the predicted ORH\(_P\) is obtained. Then, a confusion matrix, filled with the number of instances of pairs (RH\(_R\), RH\(_P\)), is built for each time support and each \(\lambda\) value. Table 4 summarizes and presents the cases of \(\lambda\): 0.5, 1.0 and 1.5, and supports 0.5, 2 and 8 h over the SAG mill 1, while the Table 5 summarizes the same results over the SAG mill 2.

### Table 4. SAG mill 1. Confusion matrices (number of instances) of operational relative-hardness (ORH) predictions using \(\lambda\): 0.5, 1.0 and 1.5 at 0.5, 2 and 8 h time supports.

| Time support | \(\lambda = 0.5\) | \(\lambda = 1.0\) | \(\lambda = 1.5\) |
|--------------|-------------------|-------------------|-------------------|
| 0.5 h Prediction | Soft | Und | Hard | Total | Soft | Und | Hard | Total | Soft | Und | Hard | Total |
| Real | 1515 | 591 | 60 | 1172 | 957 | 268 | 1 | 1199 |
| Und | 295 | 3179 | 390 | 4255 | 242 | 5255 | 137 | 5885 |
| Hard | 6 | 555 | 1606 | 2012 | 0 | 362 | 931 | 1069 |
| 0.5 h Prediction | Accurate | 6300 | Accurate | 7143 |
| 2.0 h Prediction | Soft | 1204 | 922 | 288 | 2165 | 406 | 1120 | 15 | 582 |
| Real | Und | 731 | 1659 | 904 | 3495 | 160 | 4793 | 102 | 7261 |
| Hard | 230 | 914 | 1301 | 2493 | 16 | 1348 | 193 | 310 |
| 2.0 h Prediction | Accurate | 4164 | Accurate | 5392 |
| 8.0 h Prediction | Soft | 1277 | 805 | 370 | 2105 | 683 | 873 | 33 | 1030 |
| Real | Und | 626 | 1457 | 1083 | 3076 | 324 | 4135 | 504 | 6061 |
| Hard | 202 | 814 | 1519 | 2972 | 23 | 1053 | 525 | 1062 |
| 8.0 h Prediction | Accurate | 4253 | Accurate | 5343 |

### Table 5. SAG mill 2. Confusion matrices (number of instances) of ORH predictions using \(\lambda\): 0.5, 1.0 and 1.5 at 0.5, 2 and 8 h time supports.

| Time support | \(\lambda = 0.5\) | \(\lambda = 1.0\) | \(\lambda = 1.5\) |
|--------------|-------------------|-------------------|-------------------|
| 0.5 h Prediction | Soft | Und | Hard | Total | Soft | Und | Hard | Total | Soft | Und | Hard | Total |
| Real | 1704 | 434 | 13 | 2039 | 1085 | 334 | 2 | 1266 |
| Und | 330 | 2718 | 416 | 3600 | 180 | 4485 | 360 | 5119 |
| Hard | 5 | 448 | 1882 | 2311 | 1 | 300 | 1203 | 1565 |
| 0.5 h Prediction | Accurate | 6304 | Accurate | 6773 |
| 2.0 h Prediction | Soft | 1026 | 1049 | 149 | 1614 | 676 | 786 | 47 | 1119 |
| Real | Und | 460 | 2224 | 720 | 4491 | 418 | 4178 | 395 | 6012 |
| Hard | 128 | 1218 | 976 | 1845 | 25 | 1066 | 395 | 819 |
| 2.0 h Prediction | Accurate | 4226 | Accurate | 5231 |
| 8.0 h Prediction | Soft | 917 | 1151 | 196 | 1368 | 789 | 735 | 29 | 1169 |
| Real | Und | 361 | 2052 | 896 | 4285 | 358 | 4118 | 353 | 6001 |
| Hard | 90 | 1082 | 1205 | 2297 | 22 | 1148 | 398 | 780 |
| 8.0 h Prediction | Accurate | 4174 | Accurate | 5305 |
The accuracy of the model prediction, ORH\(_p\), defined as the percentage of right predictions is computed as:

\[
\text{ORH}\text{Accuracy} = \frac{\#(\text{soft}_R, \text{soft}_P) + \#(\text{und}_R, \text{und}_P) + \#(\text{hard}_R, \text{hard}_P)}{\#\text{Total}} 
\]

(3)

and it represents the percentage of elements in the confusion matrix diagonal. The relative percentage of predictions of each class (rows) is shown in Table 6 for SAG mill 1 and in Table 7 for SAG mill 2.

As shown in Tables 6 and 7 at 0.5 h time support, the LSTM is able to predict with enough confidence the ORH regardless the value of \(\lambda\). Nevertheless, as \(\lambda\) increases, the number of instances of soft and hard ORH decreases improving the final accuracy since the higher the value of \(\lambda\), the more data points are classified as undefined. Particularly, for 0.5 h time support, increasing \(\lambda\) from 0.5 to 1.5 makes real undefined points increase from 4325 to 6577 (from 53.0% to 80.7%) in SAG mill 1 and from 3600 to 6469 (from 45.3% to 81.4%) in SAG mill 2. Therefore, increasing \(\lambda\) improves accuracy, but the price is resolution. On the other hand, the number of extreme cases (\(\text{soft}_R, \text{hard}_P\)) and (\(\text{hard}_R, \text{soft}_P\)) is close to zero. This is a great result, since predicting soft hardness when it is actually hard (or vice versa) may induce bad short term decisions on how to operate the SAG mill, along with other downstream decisions.

**Table 6.** SAG mill 1. Confusion matrices (percentage) of ORH prediction using \(\lambda\): 0.5, 1.0 and 1.5 at 0.5, 2 and 8 h time supports.

| \(\lambda\) (h) | 0.5 h Prediction | 0.5 h Prediction | 0.5 h Prediction | 0.5 h Prediction |
|----------------|------------------|------------------|------------------|------------------|
|                | Real             | Total            | Real             | Total            | Real             | Total            | Real             | Total            |
|                | Soft             | Und             | Hard            | Total            | Soft             | Und             | Hard            | Total            | Soft             | Und             | Hard            | Total            | Soft             | Und             | Hard            | Total            |
| 0.5            |                  |                  |                 |                 |                  |                  |                 |                 |                  |                  |                  |                 |                  |                 |                 |                  |                  |
| 1.0            |                  |                  |                 |                 |                  |                  |                 |                 |                  |                  |                  |                 |                  |                 |                 |                  |                  |
| 1.5            |                  |                  |                 |                 |                  |                  |                 |                 |                  |                  |                  |                 |                  |                 |                 |                  |                  |

**Table 7.** SAG mill 2. Confusion matrices (percentage) of ORH prediction using \(\lambda\): 0.5, 1.0 and 1.5 at 0.5, 2 and 8 h time supports.

| \(\lambda\) (h) | 0.5 h Prediction | 0.5 h Prediction | 0.5 h Prediction | 0.5 h Prediction |
|----------------|------------------|------------------|------------------|------------------|
|                | Real             | Total            | Real             | Total            | Real             | Total            | Real             | Total            | Real             | Total            | Real             | Total            | Real             | Total            | Real             | Total            |
|                | Soft             | Und             | Hard            | Total            | Soft             | Und             | Hard            | Total            | Soft             | Und             | Hard            | Total            | Soft             | Und             | Hard            | Total            |
| 0.5            |                  |                  |                 |                 |                  |                  |                 |                 |                  |                  |                  |                 |                  |                 |                 |                  |                  |
| 1.0            |                  |                  |                 |                 |                  |                  |                 |                 |                  |                  |                  |                 |                  |                 |                 |                  |                  |
| 1.5            |                  |                  |                 |                 |                  |                  |                 |                 |                  |                  |                  |                 |                  |                 |                 |                  |                  |

The percentage of extreme cases ((\(\text{soft}_R, \text{hard}_P\)) and (\(\text{hard}_R, \text{soft}_P\))) using \(\lambda\): 0.5 increases when moving from 0.5 to 8 h time support, on both SAG mills. However, they decrease to a value close to zero when increasing \(\lambda\) from 0.5 to 1.5, at all time supports. However, LSTM loses accuracy in terms of
predicting the relevant cases (soft\textsubscript{R}, soft\textsubscript{P}) and (hard\textsubscript{R}, hard\textsubscript{P}) as soon as the time support increases, on both SAG mills.

The accuracy graph (Figure 3) shows the $\lambda$ sensitivity at all time supports on both SAG mills. The lower accuracy is 51% and is achieved at 2 h time supports with $\lambda$: 0.5 on SAG mill 1. Its accuracy increases to 66% with $\lambda$: 1.0 and 81% with $\lambda$: 1.5. The best results are achieved at 0.5 h time support (same support as the original data) where 77%, 88% and 93% of accuracy are obtained with $\lambda$: 0.5, 1.0 and 1.5, respectively on SAG mill 1, and 79%, 85% and 90% of accuracy with $\lambda$: 0.5, 1.0 and 1.5 on SAG mill 2.

![Figure 3. Accuracy of operational relative-hardness prediction at different time support as function of lambda ($\lambda$) on both SAG mills.](image)

5. Conclusions

This work proposes the use of Long Short-Term Memory networks to forecast relative operational hardness in two SAG mills using operational data. We have presented the internal architecture of the deep networks, how to deal with raw operational datasets, and qualitative criteria to estimate the operational hardness of processing material inside the SAG mill based on the consumed energy, feed tonnage and a statistical distribution using a lambda value. Particularly, Long Short-Term Memory models have been trained to predict the operational relative-hardness based only on low-cost and fast acquiring operational information (feed tonnage, spindle speed and bearing pressure).

The LSTM network shows great results on predicting the relative operational hardness at 30 min time support. On SAG mill 1, using a lambda value of 0.5, the obtained accuracy was 77.3% while increasing the lambda to 1.5 led to an increase in accuracy of 93.1%. Similar results were found on the second SAG mill. As the time support increases to two and eight hours, the accuracy drops to around 52% using a lambda value of 0.5 and 78% with a lambda value of 1.5, on both SAG mills.

The inaccuracy of LSTM, when predicting extreme cases such as soft hardness when it is hard and vice-versa, is pretty low. Extreme misclassification is close to 1% at 0.5 h time support on both SAGs regardless of the lambda value. Although it increases to around 20% when increasing the time support using a lambda value of 0.5, it rapidly decreases to around 1% as lambda increases.

Lastly, the proposed application can be extended to any crushing and grinding equipment, under a similar context of real-data acquisition in order to forecast categorical attributes that are relevant to downstream processes.

**Author Contributions:** Conceptualization, S.A. and W.K.; methodology, S.A.; codes, S.A.; validation, S.A., W.K. and J.M.O.; formal analysis, S.A.; investigation, S.A.; resources, W.K.; data curation, S.A.; writing—original draft preparation, S.A.; visualization, S.A.; supervision, W.K. and J.M.O.; project administration, W.K.; funding acquisition, W.K. and J.M.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Natural Sciences and Engineering Council of Canada (NSERC) grant number RGPIN-2017-04280 and RGPAS-2017-507956, and by the Chilean National Commission for Scientific and Technological Research (CONICYT), through CONICYT/FIA Project AFB180004, and the CONICYT/FONDAP Project 15110019.

**Conflicts of Interest:** The authors declare no conflict of interest.
Abbreviations
The following abbreviations are used in this manuscript:

LSTM Long Short-Term Memory
ORH Operational relative-hardness
FT Feed tonnage
BPr Bearing pressure
SSp Spindle speed
SAG Semi-autogenous grinding

References
1. Cochilco. Actualización de Información sobre el Consumo de Energía asociado a la Minería del Cobre al año 2012; COCHILCO: Santiago, Chile, 2013.
2. Jones, S.M.; Fresko, M. Autogenous and semiautogenous mills 2010 update. In Proceedings of the Fifth International Conference on Autogenous and Semiautogenous Grinding Technology, Vancouver, BC, Canada, 25–28 September 2011.
3. Ortiz, J.M.; Kracht, W.; Pamparana, G.; Haas, J. Optimization of a SAG mill energy system: Integrating rock hardness, solar irradiation, climate change, and demand-side management. Math. Geosci. 2020, 52, 355–379. [CrossRef]
4. Jnr, W.V.; Morrell, S. The development of a dynamic model for autogenous and semi-autogenous grinding. Miner. Eng. 1995, 8, 1285–1297.
5. Morrell, S. A new autogenous and semi-autogenous mill model for scale-up, design and optimisation. Miner. Eng. 2004, 17, 437–445. [CrossRef]
6. Silva, M.; Casali, A. Modelling SAG milling power and specific energy consumption including the feed percentage of intermediate size particles. Miner. Eng. 2015, 70, 156–161. [CrossRef]
7. Salazar, J.L.; Valdés-González, H.; Vyhmeister, E.; Cubillos, F. Model predictive control of semiautogenous mills (sag). Miner. Eng. 2014, 64, 92–96. [CrossRef]
8. Ortiz, J.; Kracht, W.; Townley, B.; Lois, P.; Cardenas, E.; Miranda, R.; Alvarez, M. Workflows in geometallurgical prediction: Challenges and outlook. In Proceedings of the 17th Annual Conference of the International Association for Mathematical Geosciences IAMG, Freiberg, Germany, 5–13 September 2015.
9. Van den Boogaart, K.; Tolosana-Delgado, R. Predictive Geometallurgy: An Interdisciplinary Key Challenge for Mathematical Geosciences. In Handbook of Mathematical Geosciences; Springer: Berlin, Germany, 2018; pp. 673–686.
10. Curilem, M.; Acuña, G.; Cubillos, F.; Vyhmeister, E. Neural networks and support vector machine models applied to energy consumption optimization in semiautogeneous grinding. Chem. Eng. Trans. 2011, 25, 761–766.
11. Hoseinian, F.S.; Faradonbeh, R.S.; Abdollahzadeh, A.; Rezai, B.; Soltani-Mohammadi, S. Semi-autogenous mill power model development using gene expression programming. Powder Technol. 2017, 308, 61–69. [CrossRef]
12. Hoseinian, F.S.; Abdollahzadeh, A.; Rezai, B. Semi-autogenous mill power prediction by a hybrid neural genetic algorithm. J. Cent. South Univ. 2018, 25, 151–158. [CrossRef]
13. Inapakurthi, R.K.; Miriyala, S.S.; Mitra, K. Recurrent Neural Networks based Modelling of Industrial Grinding Operation. Chem. Eng. Sci. 2020, 219, 115585. [CrossRef]
14. Avalos, S.; Kracht, W.; Ortiz, J.M. Machine learning and deep learning methods in mining operations: A data-driven SAG mill energy consumption prediction application. Min. Metall. Explor. 2020, 37, 1–16. [CrossRef]
15. Hochreiter, S.; Schmidhuber, J. Long short-term memory. Neural Comput. 1997, 9, 1735–1780. [CrossRef] [PubMed]
16. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. Deep Learning; MIT Press: Cambridge, MA, USA, 2016; Volume 1.
17. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. arXiv 2014, arXiv:1412.6980.