Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 59-32 (2025) 138–143

Improving Soft Sensor Reliability in the Mining Industry Using Incremental Learning

Rafael P. Mota, * **Saulo N. Matos,** **,*** **Jó Ueyama,** ** **Thomás Pinto,** ***,**** **Márcio F. Braga** *,†

* *Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração, Universidade Federal de Ouro Preto e Instituto Tecnológico Vale, Ouro Preto - MG (e-mail: rafael.mota@aluno.itv.org).*

** *Institute of Mathematical and Computer Sciences (ICMC), University of São Paulo, São Carlos - SP, Brazil (e-mail: {saulo.matos,joueyama.icmc}@usp.br).*

*** *Instituto Tecnológico Vale, Ouro Preto - MG (e-mail: thomas.pinto@itv.org)*

**** *Graduate Program in Electrical Engineering, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil*

† *Department of Electrical Engineering, Institute of Exact and Applied Sciences, Universidade Federal de Ouro Preto (UFOP), João Monlevade - MG, Brazil (e-mail: mfbraga@ufop.edu.br).*

Abstract: This work investigates the structure and potential of incremental learning methods to improve the mining industry by enhancing long-term soft sensor reliability, despite the frequent and dynamic changes in operating conditions, typical of this sector. Several incremental learning methods were evaluated using operational data from a case study involving mass flow rate estimation on a crushing circuit. Results showed that these models achieved better long-term accuracy; while requiring a computational cost adequate for industrial application, supporting their suitability for real-world deployment in mining environments.

1. INTRODUCTION

Soft sensors play a crucial role in industrial operations by providing real-time estimates of critical process variables, thereby enabling more efficient control and informed decision-making. In this context, machine learning (ML) techniques have emerged as powerful tools for the development of such sensors, allowing the construction of predictive models based on large volumes of historical and operational data.

Soft sensors are widely used in many industrial applications. In the mining industry, for example, they are used to estimate the mass flow rate of ore as it is transported on conveyor belts. This method is adopted because using physical sensors, like belt scales, can be impractical at certain points in the plant due to their measurement principles. To address this issue, various researchers have proposed soft sensor-based solutions.

For instance, Heinzl et al. (2021, 2022) have estimated mass flow using a linear regression model fed with power draw data and have validated the model in a virtual environment. Similarly, Väyrynen et al. (2013) have developed a soft sensor combining power transducers and belt geometry. Sobreira et al. (2023) have designed a soft sensor using current, torque, and motor speed to estimate ore mass flow, evaluating machine learning models such as REPTree and Random Forest. Pereira et al. (2024) have further implemented a soft sensor integrated into a control system for mineral processing, testing decision trees, multilayer perceptrons, and linear regression.

Despite the successful application of soft sensors in various industrial scenarios, these models face challenges when exposed to changing operating conditions, a common issue in mining processes (Parisi et al., 2019). This limitation arises because traditional machine learning models are typically trained only once using a static dataset available prior to deployment. As a result, the generated model reflects the data distribution of the original dataset and performs best under conditions that resemble that same distribution.

Therefore, in operations where the data distribution changes frequently (Hoi et al., 2021) or where new features are added over time (Bartz and Bartz-Beielstein, 2024), soft sensors tend to lose efficiency. A possible countermeasure is to manually retrain the model periodically using an updated dataset. However, this approach is often unfeasible in industrial contexts; since collecting, processing,

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2025.12.410

validating, and retraining models requires significant time and effort from the responsible engineer, resources that are often unavailable. Consequently, the soft sensor gradually loses effectiveness until it is ultimately deactivated.

Incremental learning (IL) emerges as a promising alternative to this problem. It is a paradigm that allows continuous improvement of machine learning models with low computational cost, based on the sequential processing of new data (Hoi et al., 2021). This allows the soft sensor to remain updated with the current operating conditions, avoiding sudden performance degradation. Furthermore, this approach makes the soft sensor more efficient and scalable for deployment in dynamic environments (Gama et al., 2014).

To mitigate model degradation, several studies have introduced incremental learning techniques in industrial settings. For example, Sun et al. (2024) have proposed a soft sensor with IL for hydrocyclone slurry concentration prediction. Li et al. (2024) have developed a particle size predictor for grinding circuits with an incremental random weights neural network and compact constraints. Zhang et al. (2021) have applied incremental learning to an NARX model for froth flotation characterization, balancing new learning and knowledge retention. Fan et al. (2024) have designed a fault recognition system for zinc flotation using Extended Shapelet Learning and incremental updates via stochastic gradient descent. Lastly, Kumar et al. (2020) have optimized short-term production decisions with a reinforcement learning policy updated incrementally through an ensemble Kalman filter, enhancing adaptability without full retraining. However, none of these works have addressed the model degradation problem in a mass flow rate measurement application, nor have discussed the IL model computational cost, which is a critical aspect of an industrial model deployment.

Hence, in this work, we fill this gap by expanding the study case presented in Pereira et al. (2024), in which the authors have implemented a soft sensor based on a linear regression model to estimate the mass flow rate in a crushing circuit. However, the authors have not addressed the long-term accuracy of the soft sensor, despite the usual changes in operating conditions of this circuit. We then assessed the accuracy and computational cost of several incremental learning methods using data collected several months after the implementation of the linear regression-based soft sensor, as alternatives to ensure long-term reliability of the soft sensor.

This paper is structured as follows: Section 2 provides an overview of the incremental learning framework. Section 3 details the case study under investigation. Section 4 describes the experimental setup; and Section 5 presents the results and their discussion. Conclusions and potential future work are provided in Section 6.

2. FUNDAMENTAL CONCEPTS

Incremental Learning is an advanced machine learning paradigm focused on continuously training models through the processing of streaming data, gradually accumulating knowledge from past experiences. Unlike traditional machine learning techniques, which assume a stationary distribution between training and testing data, IL is designed to handle non-stationary data by adapting to distributional changes over time (Shaheen et al., 2022).

For an IL application, it is important to examine the nature of streaming data. This refers to continuously generated data, which is typically unstructured, volatile, and subject to unpredictable changes (Bartz and Bartz-Beielstein, 2024). Due to the massive volume of such data and memory constraints, storing all input instances becomes impractical. As a result, only a subset of the data is retained using specific structures such as time windows. The most common approaches include damped window, landmark window, and sliding window models (Zubaroğlu and Atalay, 2021).

Among these, the sliding window technique proposes the use of a fixed-size window that shifts as new data arrives. In this way, each retraining of the prediction model is performed based on a different data subset. The window size is a crucial parameter for the model's update performance, as it directly affects the balance between recent and older data during the model's adjustment.

However, even with these strategies in place, a major challenge remains: dealing with changes in data distribution over time, commonly known as concept drift. This is a major cause of performance degradation in machine learning models over time (Sethi and Kantardzic, 2017). Concept drift detection techniques can be classified in two ways, depending on their reliance on labeled data. Implicit (unsupervised) detectors analyze the properties of unlabeled data features to signal distributional shifts. These methods are useful in scenarios where labeling is costly or impractical, although they are more prone to false alarms. Conversely, explicit (supervised) detectors rely on labeled data to monitor performance metrics, effectively identifying performance drops that may indicate relevant changes in the data stream (Sethi and Kantardzic, 2017).

For instance, ADaptive WINdowing (ADWIN) is an explicit detector that manages a variable-size window to store the most recent data from the data stream. It monitors the average of the data within this window, dynamically adjusting its size based on the observed rate of change; the window expands when no change is detected and shrinks when a change is identified. If the difference between the means of two sub-windows within the current window is statistically significant, the algorithm signals a concept drift, discards outdated data, and updates the model accordingly (Bartz and Bartz-Beielstein, 2024).

The general structure of incremental learning, illustrated in Figure 1, consists of a continuous cycle of updating the inference model based on data received over time. At each time step k, a new instance $s_k = (x_k, y_k)$ is observed, where x_k is a feature vector and y_k is its corresponding output. The vector x_k is passed to the current model, f_k, which produces a prediction $\hat{y}_k = f_k(\theta_k, x_k)$. In parallel, the instance s_k is stored in a buffer that maintains a limited history of recent samples, m_k, serving as the basis for model updates. After observing the true output value y_k, the prediction error is computed using a loss function $\ell(\hat{y}_k, y_k)$, resulting in a performance measure l, which is used to detect model degradation. Based on this performance indicator and the content of the buffer, an update function adjusts the model parameters θ, generating a new model f_{k+1}. Some techniques also incorporate a change detection mechanism, which analyzes the data and emits a signal δ whenever a significant shift in the data distribution is detected. This signal is used to adapt the data window size and enhance the model's update function.

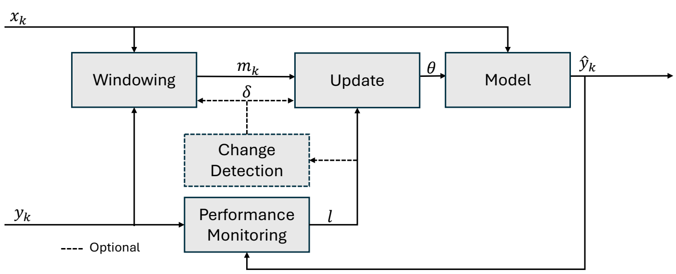


Fig. 1. General structure of incremental learning.

3. CASE STUDY

The circuit of the case study is depicted in Figure 2. It is a process of crushing and transporting materials from the mine to a copper processing plant operated by Vale S.A., a mining company in Brazil. In this system, the extracted material is first sent to a crusher, which reduces its particle size. The processed material is then stored in a silo, from which a feeder extracts it at a rate determined by its operating speed. The material is subsequently transferred onto a series of three conveyor belts that transport it to a yard, where an ore stockpile is formed.

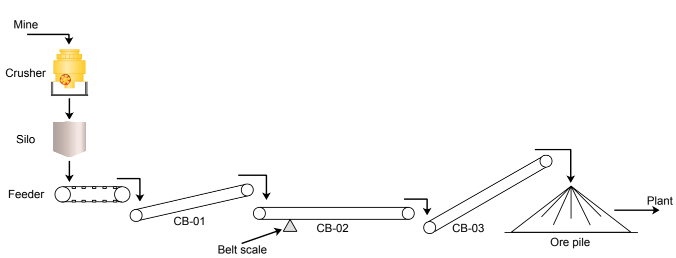


Fig. 2. Circuit diagram.

The soft sensor developed in Pereira et al. (2024) estimates the mass of the bulk material being transported on conveyor belt CB-02. It has been designed to anticipate access to the mass flow rate measurement, reducing dead time of the circuit's mass flow rate control system, which is implemented with a Proportional-Integral (PI) controller. When relying on the belt scale, the measurement would take over a minute due to the physical location of the belt scale — positioned approximately one kilometer away from the feeder — leading to poor control performance and high mass flow rate variability. After validation, the soft sensor predictions have replaced the belt scale readings as the controlled variable signal in the PI loop, resulting in improved controller effectiveness and circuit productivity.

The soft sensor uses two input variables: the average and standard deviation of the last 10 readings of current from CB-01. The soft sensor's structure was designed by evaluating correlations between the output and candidate input variables, and by testing various input configurations.

To assess how well the soft sensor would perform in future operational scenarios, we tested a linear regression model on three different datasets: (DS1) data from the period when the soft sensor presented in Pereira et al. (2024) was developed; (DS2) data collected six months after DS1; and (DS3) data collected 18 months after DS1. Each dataset consists of information from one day of operation, that is, 24 hours.

Figure 3 shows a comparison between the soft sensor estimates and belt scale measurements over a period of five hours for each of the three datasets. It clearly demonstrates a decline in the model's performance as the time gap between the data collection period and the model creation increases. Specifically, the DS3 comparison highlights a typical scenario of model degradation, where changes in operating conditions reduce model generalizability. During this period, the circuit mass flow rate setpoint was increased, resulting in higher values than those observed during the soft sensor development. As a result, the model becomes outdated and is unable to properly generalize to the new operating conditions.

Furthermore, the root mean squared error (RMSE) was employed to asses the soft sensor accuracy. For dataset DS1, the RMSE was 62.79 t/h, increasing to 79.42 t/h in dataset DS2, and reaching 231.99 t/h in DS3. This corresponds to an RMSE increase of 26.49% after six months and a degradation of 269.02% after 18 months, indicating a significant decline in the soft sensor performance over time.

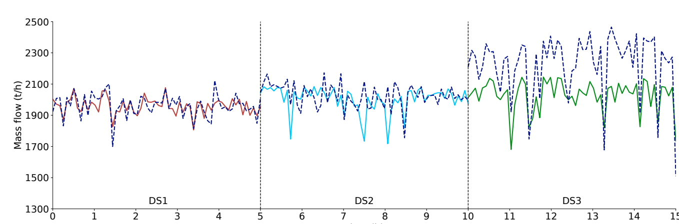


Fig. 3. Comparison between measurements of a belt scale (dashed line) and a soft sensor (solid lines) built with an LR model. For clarity, only 5 hours of data from each dataset are presented.

4. EXPERIMENTAL SETUP

4.1 Data collection and processing

In this work, the same datasets described in Section 3 were used, covering three different operational periods (from DS1, DS2, and DS3), with the aim of evaluating the robustness of the incremental learning models in real scenarios and with different levels of time lag. Operational data often exhibits errors or inconsistencies that can compromise the performance of machine learning models, if not addressed properly (Pereira et al., 2024).

First, an analysis was performed to remove data considered inconsistent or outside the operational range of the process. Then, to identify and remove outliers, the z-score method was applied, a statistical technique that standardizes the data to a distribution with zero mean and unit standard deviation. This method facilitates comparisons between datasets from different sources or with varying scales (Henderi et al., 2021).

In this study, a z-score threshold of 2 was adopted, meaning that values more than 2 standard deviations above or below the mean were considered outliers and removed. This threshold is commonly used because, in a normal distribution, approximately 95% of the data falls within 2 standard deviations from the mean, allowing for the exclusion of extreme values without significant loss of information (Navidi, 2021). After these preprocessing steps, the final dataset consisted of 180,446 valid instances (69.62% of the original dataset size), ready to be analyzed and used for model training.

The three datasets were converted into data streams to simulate how the models would perform in a real production environment. To achieve this, the datasets were concatenated sequentially, preserving their natural temporal order, in order to form a single data stream that reflects the evolution of the process over time. The models' performance was then evaluated through Progressive Validation. This method has the advantage of not requiring a separate validation set for testing, thus maximizing the use of available data. Each individual instance is processed in its arrival order. Before updating, the model generates a prediction, then based on the error between the prediction and actual value, the RMSE is updated incrementally, reflecting the model's performance over time.

4.2 Incremental learning models

Four incremental learning models were evaluated in this work: *i*) Incremental Linear Regression (I-LR); *ii*) Hoeffding Adaptive Tree Regressor (HATR); *iii*) Adaptive Random Forest (ARF); and *iv*) Incremental Multi-Layer Perceptron (I-MLP). The models' hyperparameters were empirically defined. These IL methods were selected to enable a direct comparison with traditional models previously evaluated by Sobreira et al. (2023).

Incremental Linear Regression (I-LR): This is the simplest model evaluated in this study, in which the output variable is expressed as a linear combination of the input variables. The model's coefficients are updated with each new sample using Stochastic Gradient Descent (SGD). Specifically, in this study, the Adam optimizer (Kingma and Ba, 2014) was adopted, a stochastic optimization method that requires only first-order gradients. It has low computational cost, and dynamically adjusts the learning rate for each parameter (Bishop and Bishop, 2023).

Hoeffding Adaptive Tree Regressor (HATR): This model is an extension of the Hoeffding Tree. It uses the initial samples from the data stream to select the best splitting feature at the root and continues adjusting the subsequent nodes based on the following samples. Predictions are made using either linear regression models at the leaves or sample averages. When a split no longer reflects the current data distribution, an alternative subtree is initiated at that node. HATR employs a concept drift detector to monitor the error within each subtree. When a change in the data stream is detected, a new subtree is constructed; if it outperforms the current one, it replaces it (Bartz and Bartz-Beielstein, 2024).

Adaptive Random Forest (ARF): An incremental version of the Random Forest, ARF combines Hoeffding Trees with online bagging using Poisson sampling to increase diversity among the trees. Randomness is further enhanced by limiting the number of attributes evaluated at each split. Each tree is equipped with a concept drift detector that, upon signaling a change, activates an alternate tree. The model also adopts weighted voting, where each tree's weight is proportional to its performance metric (Gomes et al., 2017). In this application, both ARF and HATR used the ADWIN method for drift detection, and linear regression with optimization via the Adam algorithm was employed as the base model at the leaves.

Incremental Multi-Layer Perceptron (I-MLP): A feedforward neural network composed of densely connected layers, capable of modeling nonlinear relationships between inputs and outputs. It is trained in a supervised manner using backpropagation, adjusting the weights to minimize the error. In the incremental mode (learning by pattern), the weights are updated after each instance based on the gradient of the individual error (Riedmiller, 1994). In this study, the I-MLP was configured with two hidden layers containing 20 and 10 neurons, respectively, using the Adam optimizer and a ReLU activation function.

5. RESULTS

Figure 4 depicts the evolution of the RMSE for each model, computed cumulatively over all instances observed up to each point in the data stream. The HATR model performed well initially, but exhibited progressive degradation over time, specially in the final third of the data stream, which corresponds to the most significant concept drift. This behavior may be due to its single-tree structure, which, when only partially updated, can become outdated as the data distribution gradually evolves, resulting in a steady increase in error. The I-LR and I-MLP models maintained stable and consistent performance throughout the entire data stream, demonstrating intermediate accuracy and robustness to changes in the data distribution. The ARF model, although it initially showed poor performance, progressively improved and ultimately achieved the lowest RMSE across the full data stream.

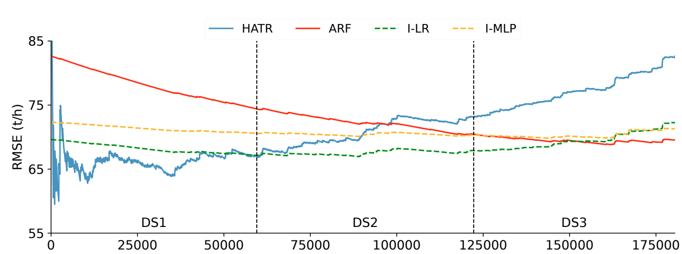


Fig. 4. Accuracy performance of incremental models.

Notably, the ARF model was the only one that exhibited a continuous improvement in the RMSE during the final third of the data stream, despite the significant changes on the data distribution. This consistent improvement can be attributed to the model robust architecture and adaptive mechanisms. It integrates predictions from multiple Hoeffding Trees, each trained on re-sampled data subsets, by averaging their outputs to yield the final prediction.

This ensemble voting strategy reduces the impact of errors from individual trees, leading to more stable and reliable predictions. Moreover, the inclusion of drift detectors enables rapid adaptation, with background trees are trained in parallel to promptly replace outdated ones upon concept drift detection.

In Table 1, the average computational cost observed for the evaluated incremental models is presented. The ARF and HATR exhibited highly variable memory usage and processing time across the samples, which is expected since updating these model involves changes in the size of their structure, thereby directly impacting both metrics. Consequently, they showed the highest average computational costs, with the ARF model presenting a notably higher value due to its structure being composed of multiple Hoeffding Trees. The I-MLP model demanded intermediate computational resources due to the backpropagation process, while the I-LR model stands out as the most lightweight in terms of computational cost, given its structural simplicity.

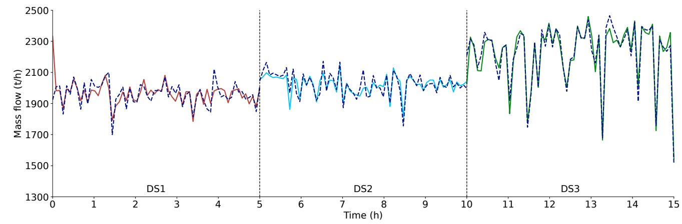
Table 1. Average performance of incremental models.

Metric	HATR	ARF	I-LR	I-MLP
Memory usage (KB)	51.52	1054.46	2.36	22.51
Processing time (ms)	0.55	12.24	0.06	0.61

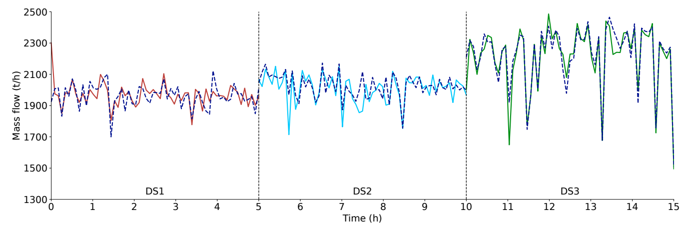
Overall, the results highlighted the capability of incremental learning models to achieve better long-term accuracy through structural updates. These models were able to reduce or maintain accuracy despite significant changes in the data distribution across the data stream, while presenting a computational performance suitable for a resource-constrained environment, such as an industrial Programmable Logic Controller (PLC). The exception was the HATR model, which demonstrated significant accuracy degradation over time.

To emphasize the effectiveness of incremental models over a traditional model, Figure 5 presents the performance of both the ARF and I-MLP models in estimating the mass flow rate, using the same data sample employed to evaluate the LR model in Section 3. The ARF model was selected for its superior predictive accuracy, and despite its higher computational cost, it remains suitable for industrial applications. Additionally, the I-MLP model was included for offering a well-balanced trade-off between accuracy and efficiency, along with greater stability, as it showed less performance degradation than the I-LR when processing data from DS3. The incremental models outperformed the traditional approach across all scenarios, demonstrating their ability to continuously adapt to evolving process dynamics, regardless of significant variations in operating conditions.

Figure 6 shows a comparison of RMSE performance for the ARF, I-MLP, and LR models across each dataset individually. While the LR model exhibited increasing degradation as the data distribution changed, both incremental models maintained stable and consistent performance. Moreover, Table 2 presents the percentage RMSE improvement of ARF and I-MLP compared to the LR model across all datasets. With the exception of the I-MLP model, which showed a slight degradation (indicated by the negative percentage RMSE) when processing data from DS1, all other scenarios demonstrated improvements in the accuracy metric. Notably, they achieved gains that exceeded 60% and 70% for the I-MLP and ARF, respectively, when processing DS3 — the dataset characterized by the most prominent concept drift in the datastrem.



(a) Soft sensor based on the ARF model.



(b) Soft sensor based on the I-MLP model.

Fig. 5. Comparison between measurements of a belt scale (dashed line) and soft sensor predictions (solid lines). For clarity, only 5 hours of data from each dataset are presented.

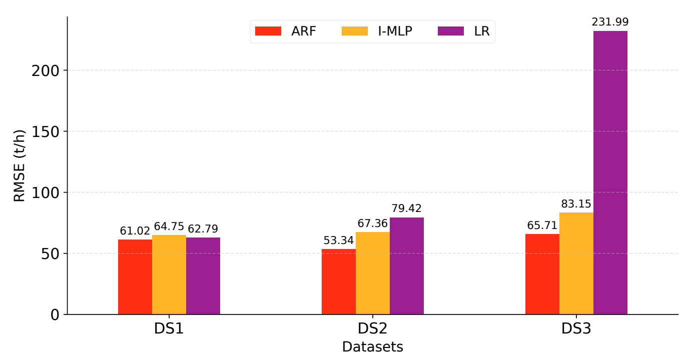


Fig. 6. Comparison of RMSE performance between the incremental models (ARF and I-MLP) and the traditional LR model.

Table 2. Percentual RMSE performance improvement of (ARF and I-MLP) over the traditional LR model.

Model	RMSE improvement (%)		
	DS1	DS2	DS3
ARF	2.82	32.84	71.67
I-MLP	−3.12	15.19	64.16

6. CONCLUSION

This work discussed the application of incremental learning in the mining industry, using the development of a soft sensor for estimating the mass flow rate in a crushing circuit as a case study. Several incremental learning models were evaluated in terms of both estimation accuracy and computational cost, using real operational data. While the ARF model stood out for its superior ability to adapt to changes in the data distribution, the I-MLP model offered the best trade-off between accuracy and computational efficiency, making it a cost-effective alternative. Moreover, most models showed the capacity to update incrementally and preserve performance, while requiring a computational effort compatible with industrial environments. These results highlight the potential of incremental learning models as effective solutions for maintaining the reliability of soft sensors in the dynamic environment of the mining industry.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Instituto Tecnológico Vale (ITV), the Universidade Federal de Ouro Preto (UFOP), and the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, processo APQ-05105-23).

REFERENCES

Bartz, E. and Bartz-Beielstein, T. (2024). *Online Machine Learning*. Springer Nature Singapore.

Bishop, C.M. and Bishop, H. (2023). *Deep Learning: Foundations and Concepts*. Springer Cham.

Fan, Y., Tang, Z., Luo, J., Xie, Y., Zhong, Y., and Gui, W. (2024). Extended shapelet learning-based discriminant dictionary for froth flotation fault recognition. *IEEE Sensors Journal*, 24(7), 10853–10864.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, 46(4), 1–37.

Gomes, H.M., Bifet, A., Read, J., Barddal, J.P., Enembreck, F., Pfharinger, B., Holmes, G., and Abdessalem, T. (2017). Adaptive random forests for evolving data stream classification. *Machine Learning*, 106, 1469–1495.

Heinzl, B., Hinterreiter, C., Roßbory, M., and Hinterdorfer, C. (2022). Challenges in mass flow estimation on conveyor belts in the mining industry: A case study. In *International Conference on Database and Expert Systems Applications*, 90–99. Springer.

Heinzl, B., Martinez-Gil, J., Himmelbauer, J., Roßbory, M., Hinterdorfer, C., and Hinterreiter, C. (2021). Indirect mass flow estimation based on power measurements of conveyor belts in mineral processing applications. In *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)*, 1–6. IEEE.

Henderi, H., Wahyuningsih, T., and Rahwanto, E. (2021). Comparison of min-max normalization and z-score normalization in the k-nearest neighbor (kNN) algorithm to test the accuracy of types of breast cancer. *International Journal of Informatics and Information Systems*, 4(1), 13–20.

Hoi, S.C., Sahoo, D., Lu, J., and Zhao, P. (2021). Online learning: A comprehensive survey. *Neurocomputing*, 459, 249–289.

Kingma, D.P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kumar, A., Dimitrakopoulos, R., and Maulen, M. (2020). Adaptive self-learning mechanisms for updating short-term production decisions in an industrial mining complex. *Journal of Intelligent Manufacturing*, 31, 1795–1811.

Li, Y., Xu, L., Li, J., Li, R., and Dai, W. (2024). Data-driven soft sensor modeling method for grinding particle size using IRVFLN with compact constraints. In *2024 7th International Symposium on Autonomous Systems (ISAS)*, 1–6. IEEE.

Navidi, W. (2021). *Principles of Statistics for Engineers and Scientists*. McGraw-Hill, 5th edition.

Parisi, G.I., Kemker, R., Part, J.L., Kanan, C., and Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks*, 113, 54–71.

Pereira, P., Pinto, T., Matos, S., Barbosa, H., Perez, J., and Pessin, G. (2024). Closing the loop: Enhancing industrial productivity through soft sensor. In *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 1–6. IEEE.

Riedmiller, M. (1994). Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms. *Computer Standards & Interfaces*, 16(3), 265–278.

Sethi, T.S. and Kantardzic, M. (2017). On the reliable detection of concept drift from streaming unlabeled data. *Expert Systems with Applications*, 82, 77–99.

Shaheen, K., Hanif, M.A., Hasan, O., and Shafique, M. (2022). Continual learning for real-world autonomous systems: Algorithms, challenges and frameworks. *Journal of Intelligent & Robotic Systems*, 105(1), 9.

Sobreira, S.G.A., Gomes, P.H., Filho, G.P.R., and Pessin, G. (2023). A data-driven soft sensor for mass flow estimation. *IEEE Transactions on Instrumentation and Measurement*, 72, 1–9.

Sun, K., Yang, C., Gao, C., Wu, X., and Zhao, J. (2024). Development of an online updating stochastic configuration network for the soft-sensing of the semi-autogenous ball mill crusher system. *IEEE Transactions on Instrumentation and Measurement*, 73, 1–11.

Väyrynen, T., Itävuo, P., Vilkko, M., Jaatinen, A., and Peltonen, M. (2013). Mass-flow estimation in mineral-processing applications. *IFAC Proceedings Volumes*, 46(16), 271–276.

Zhang, J., Tang, Z., Xie, Y., Ai, M., Zhang, G., and Gui, W. (2021). Data-driven adaptive modeling method for industrial processes and its application in flotation reagent control. *ISA transactions*, 108, 305–316.

Zubaroğlu, A. and Atalay, V. (2021). Data stream clustering: a review. *Artificial Intelligence Review*, 54(2), 1201–1236.