PLC-Embedded Virtual Pressure Sensor with Continual Learning for Mining Tailings Pumping

Luís V. Souza[1], Wagner F. Timoteo[2], Guilherme Euler[3], Lucas M. Ribeiro[4], Eduardo S. Luz[1]

[1] Programa de Pós-Graduação em Ciência da Computação (PPGCC), Universidade Federal de Ouro Preto and Instituto Tecnológico Vale (ITV), Ouro Preto–MG, Brazil.
luis.vitorino@aluno.ufop.edu.br eduluz@ufop.edu.br
[2] Automation Maintenance Coordination, Vale S.A., Itabirito–MG, Brazil.
wagner.timoteo@vale.com
[3] Process Engineering Coordination, Vale S.A., Itabirito–MG, Brazil.
guilherme.euler@vale.com
[4] Programa de Pós-Graduação em Ciência da Computação (PPGCC), Universidade Federal de Ouro Preto (UFOP), Ouro Preto–MG, Brazil.
lucas.mr@aluno.ufop.edu.br

Abstract. Reliable pressure measurement in mining tailings pipelines is essential for safe and efficient operation, but physical sensors are prone to failures. This work proposes a virtual pressure sensor based on a compact LSTM network embedded in a PLC. Non-stationary data and hardware constraints introduce covariate shift and catastrophic forgetting in continuous learning. Three training strategies are compared: static dataset (A1), sequential learning without mitigation (A2), and learning with experience replay and targeted oversampling (A3). Results show that A2 suffers severe degradation ($R^2 = 0.48$), while A3 achieves the best performance (RMSE = 0.36 bar, $R^2 = 0.94$) without increasing inference cost. The model is validated in a SIMATIC PCS 7 environment, confirming real-time feasibility with 517 parameters.

CCS Concepts: • **Computing methodologies** → **Batch learning**; *Machine learning algorithms*.

Keywords: catastrophic forgetting, continual learning, LSTM, PLC, soft sensor, tailings pumping

1. INTRODUCTION

Tailings disposal is one of the greatest environmental challenges in mining. Tailings vary in composition and may be transported as slurry or dry material [Wills and Finch 2015], and inadequate management can contaminate water bodies with heavy metals and chemical reagents [Chalkley et al. 1989]. Slurry pumping through pipelines requires precise pressure control; significant variations can cause clogging, pipe ruptures, and severe economic losses. Given the long distances and remote areas involved, virtual sensors emerge as an innovative low-cost alternative to physical transmitters.

The studied system (Figure 1) consists of two parallel centrifugal pump trains driving slurry from tank TQ-01, with pressure monitoring at stations PMS-01 and PMS-02. The most critical point is PMS-01, at kilometer 2.47, in a high-elevation remote area. Failure of physical pressure transmitters there promotes *slack flow* — a regime in which the pipeline is no longer completely filled — increasing risks of clogging, wear, and unplanned shutdowns.

LSTM networks have proved efficient for virtual sensors in non-linear processes [Shen et al. 2024; Zhang et al. 2024; Zhou et al. 2022]. However, in fixed-setpoint systems, historical data concentrate in

This study was financed in part by CAPES (Finance Code 001), CNPq, Instituto Tecnológico Vale (ITV), and Universidade Federal de Ouro Preto (UFOP).

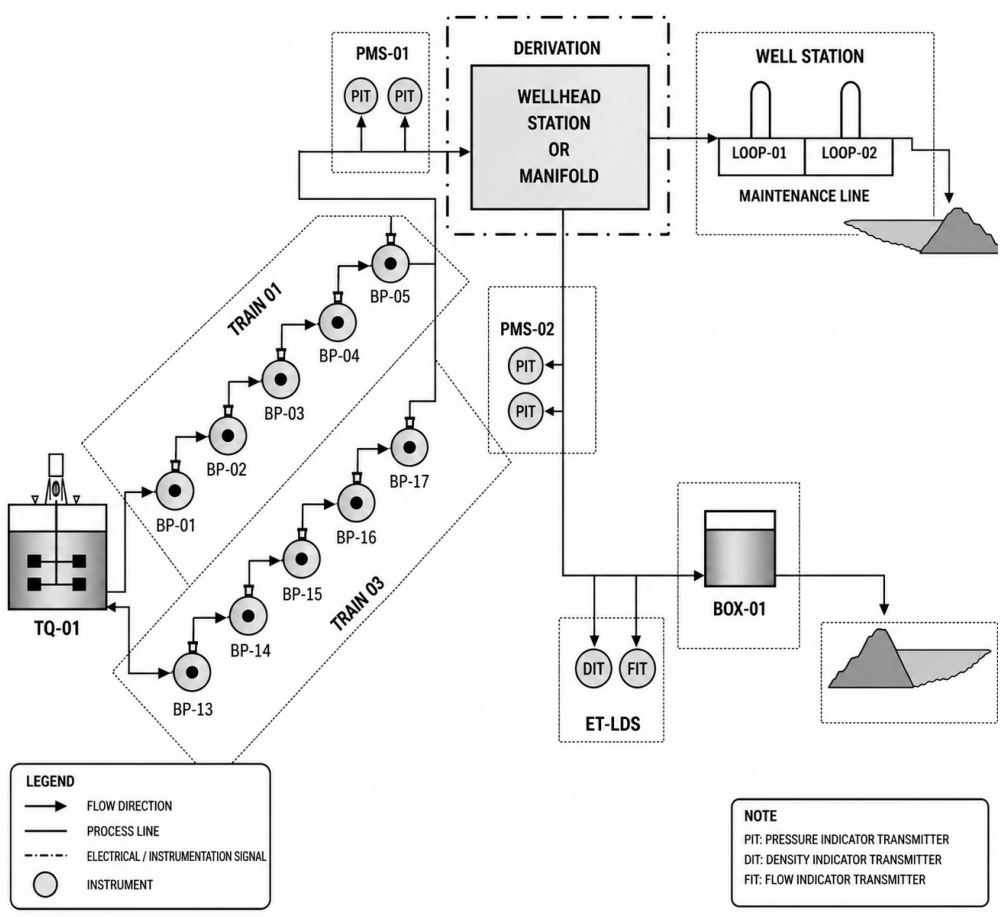

Fig. 1. Representation of the studied tailings pumping system. Source: Guilherme Euler (2024).

narrow pressure ranges, creating *covariate shift* [Zhang et al. 2024]. This is intensified by the resource constraints of industrial Siemens 410-5H PLCs: although a single virtual sensor occupies only a fraction of available memory, the PLC must simultaneously execute control loops, communication services, diagnostics, and other automation tasks, leaving only a limited computational budget for machine-learning models. When incremental learning is used to broaden operational coverage, a second challenge arises: *catastrophic forgetting*, whereby weight updates with new data degrade performance on previously learned distributions [McCloskey and Cohen 1989; Goodfellow et al. 2016], reflecting the plasticity–stability trade-off in continual learning [Parisi et al. 2019].

[de Souza Neto et al. 2025] investigate online and continual learning for temperature prediction on belt conveyors at the S11D mine using regularization-based LSTM variants, analyzing the plasticity–stability trade-off and catastrophic forgetting in a non-stationary industrial setting.

The present work differs in three key aspects: (i) the embedded PLC context makes EWC and sample-by-sample online learning infeasible due to scan-time and memory constraints; (ii) the sensor operates as an autonomous redundancy without access to real measurements in production; and (iii) training is *offline* in Python, with weights transferred to PCS 7 via data blocks, an approach incompatible with online learning but viable in the embedded industrial environment. Two hypotheses guide the investigation:

H1:. Catastrophic forgetting compromises generalization of compact LSTM networks under se-

quential updates with non-stationary distributions.

H2:. Experience replay and oversampling mitigate catastrophic forgetting in compact LSTM networks.

2. THEORETICAL BACKGROUND

2.1 Continual Learning, Paradigms, and Hardware Constraints

The pumping system produced eight datasets with substantially distinct pressure distributions over three months (Figure 2), reflecting setpoint changes, shutdowns, and pump replacements; no single dataset is representative of the full system behavior. Three paradigms address this problem on the plasticity–stability spectrum. *Online learning* updates the model sample-by-sample in real time, maximizing plasticity but offering no stability guarantees. *Incremental learning* trains in sequential batches without explicit knowledge preservation, risking catastrophic forgetting. *Continual learning* (CL) operates in sequential batches with explicit mechanisms to balance adaptation with retention [Parisi et al. 2019]. Formally, at each step t:

$$\theta_t = \arg\min_{\theta} \underbrace{\mathcal{L}(\theta;\, D_t)}_{\text{current adaptation}} + \lambda \underbrace{\Omega(\theta,\, \theta_{t-1})}_{\text{knowledge retention}} \tag{1}$$

where λ governs the plasticity–stability balance. In this work, CL operates in batch mode: the model is retrained offline in Python and weights are then transferred to the PLC — there is no update mechanism at inference time.

Two main families of mitigation strategies exist. *Regularization-based* methods such as Elastic Weight Consolidation (EWC) [Kirkpatrick et al. 2017] implement Ω via the Fisher information matrix, requiring storage quadratic in the number of parameters — infeasible for the Siemens 410-5H. [de Souza Neto et al. 2025] adopt EWC variants for belt-conveyor prediction on hardware without PLC constraints. *Replay-based* methods, specifically *experience replay* [Rolnick et al. 2019], reintroduce samples from past distributions into the current training batch, reducing the gradient's bias toward the new distribution without any additional inference-time cost. Random uniform sampling was preferred over reservoir sampling for its implementation simplicity within the offline Python retraining pipeline and lower bookkeeping overhead; reservoir sampling is identified as a direction for future work. The 30% replay fraction was chosen empirically: full retention would double the effective dataset size, while fractions below 20% showed insufficient protection against weight overwriting in preliminary experiments.

2.2 Catastrophic Forgetting and Covariate Shift in LSTMs

Catastrophic forgetting [McCloskey and Cohen 1989] is the tendency of a neural network to abruptly lose previously acquired knowledge when trained on new data. In the LSTM, this manifests through its internal gate dynamics. When retrained exclusively with D_t, the forget gate $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$ is adjusted to maximize performance on the current distribution, selectively suppressing the cell state $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$, which encoded patterns from $D_1, \ldots, D_{t-1}$. The resulting gradient $\nabla_\theta \mathcal{L}$, computed solely over D_t, carries no error signal from past distributions, directing θ away from the minima that ensured good prior performance [Goodfellow et al. 2016; de Souza Neto et al. 2025]: the stability side of the trade-off collapses, leaving the model highly plastic to recent data but unable to retain previously learned distributions.

In the studied system, the most recent training sets (Tr6–Tr8) concentrate on low-pressure ranges while the earliest sets (Tr1, Tr3) cover higher ranges (Figure 2). This shift in the input feature

distribution — a *covariate shift* [Zhang et al. 2024] — is the recurrent trigger of catastrophic forgetting: unconstrained sequential training causes model A2 to progressively specialize in low-pressure patterns, losing generalization to the higher-pressure ranges that dominate Test 3.

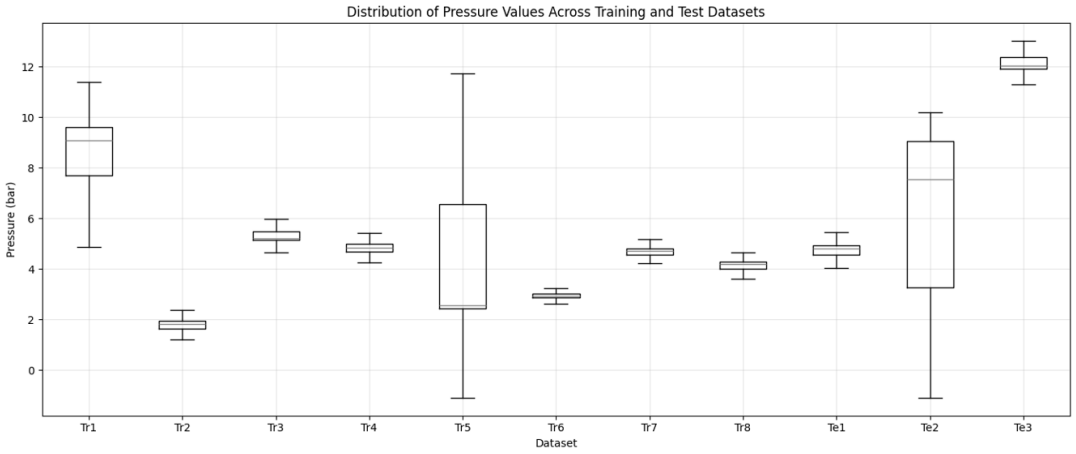

Fig. 2. Distribution of P_{out} across training (Tr1–Tr8) and test (Te1–Te3) sets. Source: the authors (2024).

3. MATERIALS AND METHODS

3.1 Dataset and Comparative Models

Data were collected at 1 Hz over three months from a real tailings pumping plant, comprising eight training and three test periods. Input variables are: slurry density ($\rho \in [0.95; 1.35]$ g/cm^3), discharge pressure ($h \in [0; 35]$ bar), volumetric flow rate ($Q \in [0; 4000]$ m^3/h), and upstream pressure ($P_{in} \in [-2; 20]$ bar). The target is $P_{out} \in [-2; 17]$ bar at PMS-01.

To justify the LSTM architecture, two reference models were evaluated under A1: an MLP (flattened W=20 window, two dense layers with 16 and 8 neurons) and a GRU (two recurrent layers with 8 and 2 units, Dropout 0.1). All models share the same hyperparameters (Adam, $lr = 10^{-3}$, batch 512, early stopping patience 10).

3.2 LSTM Architecture and Training Configuration

The LSTM was designed for the Siemens 410-5H PLC (SIMATIC PCS 7), totaling 517 trainable parameters: LSTM(8) → Dropout(0.1) → LSTM(2) → Dense(3) → Dense(1). Input tensors are (W=20, 4). The 20-sample window captures the 15–25 s hydraulic time constants of the PMS-01 section, estimated empirically from system step-response curves; windows below $W < 10$ caused loss of temporal context, while $W > 30$ yielded no additional gains in preliminary experiments. For embedded validation, the network was implemented in SCL functional blocks within PCS 7, with a supervisor block managing execution, intermediate outputs, and recurrent state updates.

All experiments ran in Python with TensorFlow/Keras. Variables are normalized to $[-1, +1]$ with fixed global limits compatible with PCS 7. Table I lists all training hyperparameters.

3.3 Training Approaches

Three main strategies were evaluated with the same architecture, plus two ablations. **A1 (static baseline)** trains exclusively on dataset 1. **A2 (no mitigation)** trains sequentially on all eight

Table I. Training hyperparameters common to all approaches.

Hyperparameter	**Value**
Optimizer	Adam
Learning rate	1×10^{-3}
Loss function	MSE
Batch size	512
Max. epochs / step	50
Early stopping (patience)	10 epochs (val_loss)
Validation fraction	20%
Random seed	42
Normalization	Global min-max $[-1, +1]$
Replay fraction	30% of previous sets
Oversampling criterion	$P_{out} < 6$ bar ($3\times$)

datasets reusing prior weights, illustrating the default behavior of incremental learning. **A3 (experience replay + oversampling)** at each step reincorporates 30% of stored sequences from previous sets (replay) and replicates all sequences with $P_{out} < 6$ bar by $3\times$ (oversampling): replay acts as a memory preservation mechanism, reducing the bias toward recent data and the destructive weight drift [Rolnick et al. 2019], while oversampling addresses under-representation of low-pressure regions. Ablations **A3a** (replay only) and **A3b** (oversampling only) isolate each component's contribution.

4. RESULTS AND DISCUSSION

4.1 Architecture Comparison

Table II compares models on Test 1 under A1. LSTM achieved the best performance (RMSE = 0.621 bar, $R^2 = 0.822$), confirming that long-term memory captures hydraulic temporal dependencies. GRU showed intermediate results ($R^2 = 0.616$), while MLP failed to generalize ($R^2 < 0$) without recurrent states.

Table II. Architecture comparison on Test 1 (strategy A1).

Model	**RMSE (bar)**	R^2
MLP	1.514	−0.058
GRU	0.929	0.616
LSTM	0.621	0.822

4.2 Continual Learning Evolution and Final Results

Figure 3 shows the validation RMSE over eight training steps. In A2, a rapid initial reduction (RMSE = 0.108 bar at step 2) is followed by growth to 0.328 bar at step 8, indicating progressive catastrophic forgetting: as recent low-pressure data dominate training, the model loses stability over earlier distributions. In A3, the error remains stable throughout (0.215 bar at the end), demonstrating that replay effectively preserves prior knowledge while adapting to new distributions.

Table III presents results on independent test sets. A3 achieved the best average performance (RMSE = 0.360 bar, $R^2 = 0.948$), confirming H2. On Test 1, A2 is competitive ($R^2 = 0.919$ vs. A3's $R^2 = 0.900$), which is expected: at that point the model has been exposed primarily to distributions similar to Test 1, so the absence of mitigation has not yet produced significant forgetting. The critical divergence emerges on Test 3, where A2 collapses to $R^2 = 0.483$ — well below even the static baseline A1 ($R^2 = 0.832$) — because the sequential training on recent low-pressure data (Tr6–Tr8) has irreversibly overwritten the weights that encoded higher-pressure behaviour. A3 maintains $R^2 = 0.953$

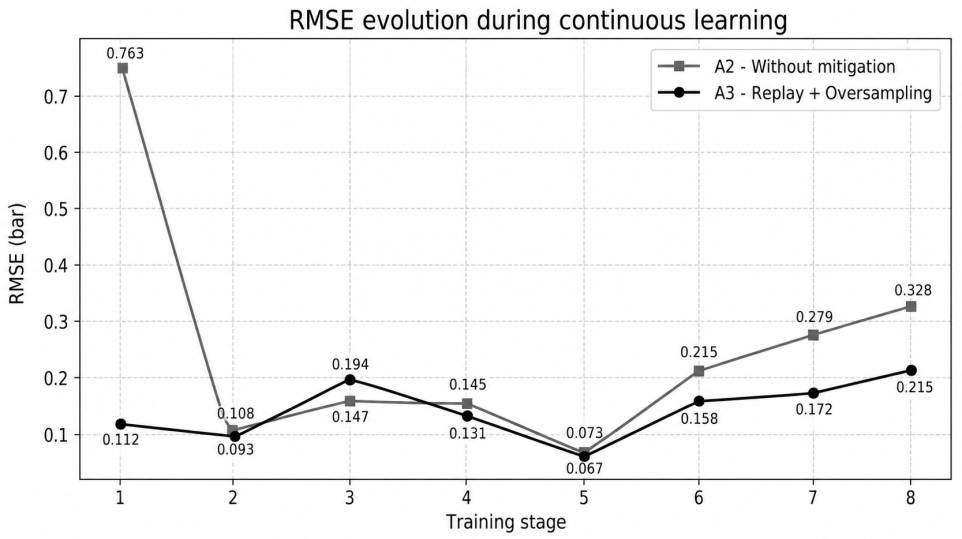

Fig. 3. Validation RMSE evolution during continual training (A2 vs. A3). Source: the authors (2024).

on Test 3, confirming H1 and H2: catastrophic forgetting is both a real risk and effectively mitigated by experience replay. The ablation study (Table IV) shows that replay (A3a) is the dominant component ($R^2 > 0.88$ across all tests), while oversampling alone (A3b) is insufficient ($R^2 = 0.531$ on Test 3) since without replay there is no protection against destructive weight overwriting. Oversampling refines statistical coverage of under-represented pressure regions but requires replay to produce consistent results.

Table III. Final results for A1, A2, and A3 across test sets.

Test	**Model**	**RMSE (bar)**	R^2
1	A1	0.6218	0.8215
1	A2	0.4179	0.9194
1	**A3**	**0.4652**	**0.9001**
2	A1	0.4798	0.9687
2	A2	0.5059	0.9652
2	**A3**	**0.2596**	**0.9908**
3	A1	0.6737	0.8316
3	A2	1.1800	0.4834
3	**A3**	**0.3558**	**0.9530**
Avg.	A1	0.5918	0.8739
Avg.	A2	0.7013	0.7893
Avg.	**A3**	**0.3602**	**0.9480**

Table IV. Ablation study of A3 components (values rounded to 2 d.p.; see Table III for full-precision results of A3).

Variant	**Test 1**		**Test 2**		**Test 3**	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
A3 (Replay + OS)	0.47	0.90	0.26	0.99	0.36	0.95
A3a (Replay only)	0.51	0.88	0.27	0.99	0.52	0.90
A3b (OS only)	0.89	0.63	0.74	0.93	1.13	0.53

4.3 Embedded Inference Validation on PCS 7

Approach A3 was implemented in SIMATIC PCS 7 with SCL functional blocks on an S7-410-5H CPU in PLCSIM. Weights trained in Python were exported to controller data blocks. Table V summarizes the computational metrics of the implementation.

Table V. Computational metrics of the embedded implementation (PCS 7).

Metric	**Value**
Controller	Siemens CPU 410-5H
Numeric precision	REAL IEEE 754 (32-bit)
Parameters	517
Scan time w/o LSTM	10 ms
Scan time w/ LSTM	12–15 ms
Inference overhead	2–5 ms
Working memory	409,914 bytes (0.02%)

The scan time without LSTM was measured with inference disabled, while the scan time with LSTM corresponds to the full model execution after the enable signal is activated. The inference added 2–5 ms to the scan cycle, maintaining ample operational margin for 1 Hz applications and using only 0.02% of the available CPU 410-5H memory.

Figure 4 shows real vs. predicted signals for all three tests. High adherence was observed across all scenarios ($R^2 > 0.91$; TEST1: 0.920, TEST2: 0.926, TEST3: 0.917), even under hydraulic transients and abrupt regime changes. A slight degradation relative to Python results is attributed to 32-bit precision limitations and PLC approximations of non-linear activation functions. Importantly, catastrophic forgetting mitigation was achieved without any increase in computational complexity.

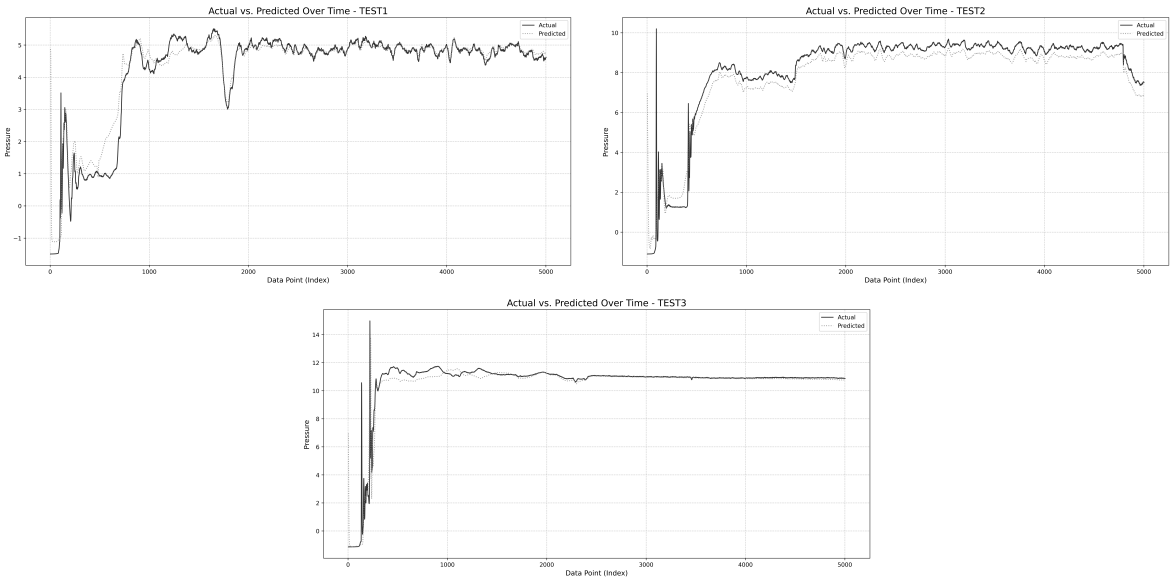

Fig. 4. Real vs. predicted signals on TEST1 (top-left), TEST2 (top-right), and TEST3 (bottom) — A3 model embedded in PCS 7. Source: the authors (2024).

5. CONCLUSION

This work presented a virtual pressure sensor based on a compact LSTM (517 parameters) embedded in a PLC for mining tailings pumping, addressing covariate shift and catastrophic forgetting via continual learning. Both hypotheses were confirmed: (H1) sequential training without mitigation (A2) yields $R^2 = 0.4834$ on Test 3, below static baseline A1 ($R^2 = 0.8316$), demonstrating that absent knowledge preservation is a critical limitation under non-stationary distributions — even though A2 remains competitive on Test 1 ($R^2 = 0.9194$), where the training distribution still resembles the test data; (H2) experience replay combined with oversampling (A3) achieves average RMSE = 0.3602 bar and $R^2 = 0.9480$, with replay confirmed as the dominant component ($R^2 > 0.88$ for A3a) and oversampling alone insufficient ($R^2 = 0.5300$ for A3b on Test 3). Compared to [de Souza Neto et al. 2025], which adopts online and continual learning on hardware without PLC constraints, this work demonstrates that batch continual learning with replay is competitive and viable where scan-time and memory constraints preclude EWC and sample-by-sample updates. The SIMATIC PCS 7 implementation confirmed $R^2 > 0.91$ in all scenarios with only 2–5 ms overhead and 0.02% memory occupancy.

Threats to validity include: (i) validation on a single industrial plant, which may limit generalization to other mining processes with different hydraulic characteristics; (ii) evaluation in PLCSIM rather than continuous live operation, although the implementation uses the same PLC architecture and data structures intended for deployment; and (iii) the 30% replay fraction was selected empirically and may require adjustment for different non-stationarity profiles and operational conditions. Future work includes: (i) reservoir sampling for principled buffer management; (ii) automatic concept drift detection for selective retraining; (iii) weight quantization for higher-capacity PLC architectures; (iv) adaptive replay fractions driven by distribution shift metrics; and (v) validation in continuous live operation at the industrial plant.

REFERENCES

Chalkley, M., Conard, B., Lakshmanan, V., and Wheeland, K. Tailings and Effluent Management. In *Proceedings of the International Symposium on Tailings and Effluent Management*. Halifax, Canada, 1989.

de Souza Neto, F. F., Silva, G. A. L., Silva, C. D., and Luz, E. J. S. Aprendizado Online e Contínuo para Previsão de Temperatura em Transportadores de Correia: Uma Abordagem Adaptativa com LSTM. In *Anais do Simpósio Brasileiro de Automação Inteligente (SBAI)*. Fortaleza, CE, Brasil, 2025.

Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. MIT Press, Cambridge, 2016.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., and Hadsell, R. Overcoming Catastrophic Forgetting in Neural Networks. *Proceedings of the National Academy of Sciences* 114 (13): 3521–3526, 2017.

McCloskey, M. and Cohen, N. J. *Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem*. Psychology of Learning and Motivation, vol. 24. Academic Press, New York, 1989.

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., and Wermter, S. Continual Lifelong Learning with Neural Networks: A Review. *Neural Networks* vol. 113, pp. 54–71, 2019.

Rolnick, D., Ahuja, A., Schwarz, J., Lillicrap, T. P., and Wayne, G. Experience Replay for Continual Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 32. Vancouver, Canada, 2019.

Shen, F., Wu, J., Ye, L., and Wang, B. Memory-Adaptive Supervised LSTM Networks for Deep Soft Sensor Development of Industrial Processes. *IEEE Sensors Journal* 24 (13): 21641–21654, 2024.

Wills, B. A. and Finch, J. *Wills' Mineral Processing Technology: An Introduction to the Practical Aspects of Ore Treatment and Mineral Recovery*. Butterworth-Heinemann, Oxford, 2015.

Zhang, M., Xu, B., Jie, J., Hou, B., and Zhou, L. A Novel Bidirectional Long Short-Term Memory Network with Weighted Attention Mechanism for Industrial Soft Sensor Development. *IEEE Sensors Journal* 24 (11): 18546–18555, 2024.

Zhou, J., Wang, X., Yang, C., and Xiong, W. A Novel Soft Sensor Modeling Approach Based on Difference-LSTM for Complex Industrial Process. *IEEE Transactions on Industrial Informatics* 18 (5): 2955–2964, 2022.