TRANSFER LEARNING FOR GENERALIZING A HYBRID AUTOENCODER-ISOLATION FOREST MODEL FOR TIME SERIES ANOMALY DETECTION IN ARRONAX CYCLOTRON OPERATIONAL DATA*

F. Basbous[†,1,2], F. Poirier[1,3], F. Haddad[1,2,3], D. Mateus[4]
IP ARRONAX, Saint-Herblain, France
[2]Nantes Université, Nantes, France
[3]Centre National de la Recherche Scientifique, Paris, France
[4]Nantes University, École Centrale Nantes, LS2N, UMR 6004, Nantes, France

Abstract

In the context of the operational monitoring of the ARRONAX C70XP cyclotron, our previous work addressed the limitations of the Isolation Forest (IF) algorithm in detecting local anomalies, particularly those occurring near the mean of normal data, due to its reliance on axis-parallel splits. To overcome this issue, we developed and validated a hybrid model combining an autoencoder and IF, using time series data from the proton beam intensity on target. This approach significantly improved the detection of both global and local anomalies, with no false alarms observed during evaluation. Building on these results, the present study investigates the use of transfer learning to generalize the hybrid model to other process variables originating from different subsystems, including the source, injector, and cyclotron core. Results suggest that the model can effectively label large volumes of multivariate operational data, supporting the development of a more scalable and integrated anomaly detection framework for the C70XP.

INTRODUCTION

Particle accelerators are safety-critical systems composed of complex and interconnected components and subsystems. Therefore, anomaly detection has become an important research area, as deviations from normal operational behavior can indicate potential faults and help prevent failures, irreversible damage, and costly repairs [1-4],

In our previous work [5], we proposed a hybrid Autoencoder-Isolation Forest (AE-IF) framework for anomaly detection in time-series operational data from the ARRONAX C70XP cyclotron. The approach relies on an autoencoder (AE) trained to learn a compact latent representation of normal operational behavior. The reconstruction errors produced by the AE are then used as input features for the Isolation Forest (IF) algorithm. Trained and validated on proton beam intensity on target time-series data, the model enables the detection of both global anomalies and subtle local anomalies occurring near the mean of normal data. The proposed approach was shown to outperform two alternative IF-based methods applied directly to the raw signal and to the PCA-transformed intensity space.

Despite these promising results, extending this approach to multiple process variables (PVs) remains challenging. The monitoring system involves numerous PVs originating from different subsystems, each exhibiting distinct dynamics and statistical distributions. Training a dedicated anomaly detection model from scratch for each variable requires significant computational effort and optimization time.

One of the techniques proposed to address the challenge of model adaptation is transfer learning (TL), which enables the reuse of knowledge learned from a source dataset when training models for related target datasets [6].

TL has demonstrated significant benefits in anomaly detection tasks, particularly in industrial time-series applications where labelled data are limited, enhancing detection robustness while reducing false positives [6,7]. It has also been successfully applied in Internet of Things monitoring and energy consumption analysis, where it improves detection accuracy while mitigating sensitivity to noise and overfitting [8,9]. For anomaly detection in detector monitoring systems, such as the Compact Muon Solenoid hadron calorimeter spatio-temporal data quality monitoring framework, TL has been shown to enable knowledge transfer across detector subsystems and improve model robustness [10]. However, to the best of our knowledge, the application of TL to anomaly detection in accelerator operational time-series data remains limited.

TL can be categorized according to the similarity of tasks and domains between the source and target problems. In general, TL is divided into inductive, transductive, and unsupervised paradigms [6,10]. Inductive TL is applied when the source and target tasks differ while labelled data are available for the target task. Transductive TL is used when the task remains the same but the source and target domains differ, typically with labelled data available only in the source domain. Unsupervised TL addresses scenarios where both the source and target tasks differ and the data are unlabeled.

Among these categories, transductive TL is particularly relevant to our study since the anomaly detection task remains unchanged while the data originate from different sensors. In this work, we propose a TL framework that reuses the latent representation learned from normal proton

*The Arronax cyclotron is supported by the CNRS, Inserm, INCa, Nantes Université, the Regional Council of Pays de la Loire, local authorities, the French government, and the European Union. This work has been supported in part by the French National Research Agency (ANR) "France 2030 investment plan" under the references I-SITE NExT (ANR-16-IDEX-0007), and Labex DHOLMEN, by financial support from the Pays de la Loire Region and by a grant from INCa-DGOS-INSERM-ITMO Cancer_18011 (SIRIC ILIAD).
†basbous@arronax-nantes.fr

beam intensity data to generalize anomaly detection across multiple PVs. By freezing the pretrained encoder and the latent representation and adapting only the output layer, the proposed approach reduces training complexity while maintaining robust detection of both global and local anomalies across several PVs despite class imbalance and variability in anomaly rates across experimental runs.

PROPOSED METHODOLOGY

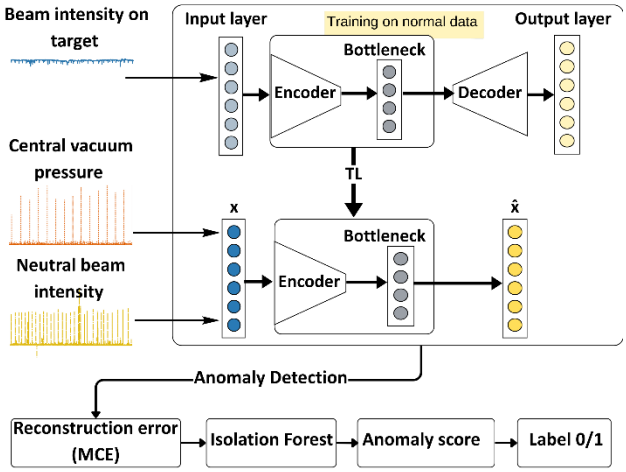

Figure 1: Illustration of the proposed TL framework. (a) The baseline AE is trained on normal beam intensity data. (b) TL is then applied by freezing the encoder and bottleneck and replacing the decoder with a single output layer trained for each PV. (c) Reconstruction errors (MCE) are computed and used as input to the IF for anomaly detection. Only two PVs are shown for illustration.

Baseline AE-IF Architecture

The baseline model corresponds to the hybrid AE-IF architecture introduced in our previous work [5]. The training set consists of N vectors $x_i \in \mathbb{R}^a$, denoted $X = \{x_i\}_{i=1}^N$, where each vector represents a standardized temporal window of dimension a. A fully connected AE, composed of an encoder and a decoder, is first trained exclusively on normal samples to learn a compact bottleneck capturing the intrinsic structure of nominal operational behavior (Fig. 1(a)). Rather than directly applying IF to the raw windows, anomaly detection is performed using reconstruction-error features. For each reconstructed vector $\hat{x}_i$, the Mean Cubic Error (MCE) with respect to the original input x_i is computed as defined in Eq. (1)

$$MCE(x_i, \hat{x}_i) = \frac{1}{a}\sum_{j=1}^{a}|x_{ij} - \hat{x}_{ij}|^3. \quad (1)$$

These reconstruction-error features are then used as input to the IF (Fig. 1(c)), which assigns anomaly scores based on the average path length in random trees [5]. Final anomaly decisions are obtained by thresholding the IF anomaly score based on the contamination parameter, set equal to the anomaly rate in the training subset of each PV.

Transfer Learning Strategy

To preserve the intrinsic structure of normal operational behavior learned from proton beam intensity data and to prevent catastrophic forgetting [11], the encoder and bottleneck layers of the pretrained model were frozen during transfer. The original decoder was replaced with a single dense output layer, as illustrated in Fig. 1(b). Within the proposed transductive TL framework, this output layer is trained separately for each target PV using its corresponding training subset, filtered according to normal data identified in the proton beam intensity on target training set.

To ensure stable adaptation of the new output layer, a small learning rate η is used to limit the magnitude of parameter updates during optimization. The weights of this layer, denoted W_{new} are updated at each iteration according to Eq. (2),

$$W_{new}^{t+1} = W_{new}^{t} - \eta \nabla_{W_{new}} L, \quad (2)$$

where L denotes the Mean Absolute Error (MAE) defined in Eq. (3).

$$L_{MAE} = \frac{1}{K}\sum_{i=1}^{K}|x_i - \hat{x}_i| \quad (3)$$

Here, x_i denotes the input sample, $\hat{x}_i$ its reconstruction, and K the total number of samples in the training batch. Anomaly detection is then performed using IF, retrained for each PV as in the baseline AE-IF framework.

Operational Definition of Anomalies

Despite differences in domain and scale across PVs, the definition of normal and anomalous behavior remains unchanged, and the AE-IF model after TL retains the same objective as the baseline, namely the detection of global and local anomalies. In the operational context of the C70XP cyclotron, global anomalies correspond to deviations from the normal operating range defined by the expert interval $[S_{low}, S_{high}]$. These anomalies occur when one or more observations fall outside this interval and may appear as sudden drops, abrupt rises, or stable sequences outside the defined interval. In contrast, local anomalies are more subtle. In this case, all observations remain within the interval $[S_{low}, S_{high}]$, while the internal variability within the time window is abnormally high and exceeds a predefined threshold α.

RESULTS AND DISCUSSION

Table 1: Validation Performance of TL Configurations Across PVs (Mean ± Standard Deviation)

PV	F1-score (Conf. 1)	F1-score (Conf. 2)
NBI	0.90 ± 0.05	**0.95 ± 0.03**
CVP	0.81 ± 0.19	**0.89 ± 0.01**
AC	0.88 ± 0.22	**0.99 ± 0.00**
PC	0.84 ± 0.12	**0.91 ± 0.02**
RF AC	0.65 ± 0.24	**0.89 ± 0.05**
IDV	0.44 ± 0.26	**0.78 ± 0.10**
IV	0.38 ± 0.27	**0.77 ± 0.08**
RF FP	0.69 ± 0.15	**0.75 ± 0.07**

Experimental Setup

To evaluate the performance of the proposed TL framework, 26 time-series datasets were used in this study, each corresponding to a run of approximately 7 days of continuous operation and including 8 representative PVs from the C70XP cyclotron. These PVs originate from multiple subsystems, including the source (arc current (AC), puller current (PC)), the cyclotron core (neutral beam intensity (NBI), central vacuum pressure (CVP), RF forward power (RF FP), RF anode current (RF AC)), and the injector (inflector voltage (IV), injection deflector voltage (IDV)). The signals were resampled at 1 Hz [1] and segmented into non-overlapping windows of length k=6, leading to an average of 74,000 sequences per run for each PV.

For model development, 60% of the runs (15 runs) were allocated to training and validation, while the remaining 40% (11 unseen runs) were reserved for final testing. The anomaly rate across the PV datasets ranges from approximately 0.1% to 16%, reflecting high to moderate class imbalance and variability across runs even for the same PV.

For each PV, all input sequences were standardized to zero mean and unit variance using a StandardScaler function, fitted on the normal samples of the training set. The same transformation was then applied to the validation and test sets. The model was trained using the Adam optimizer [12] for 30 epochs per PV, with η set to 10^{-5}.

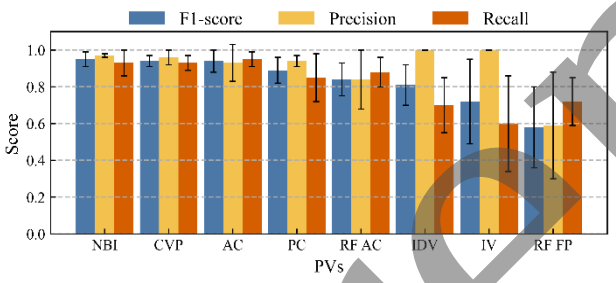

Figure 2: Detection performance of the proposed TL framework across PVs on unseen runs (mean ± std).

Validation Results

To assess whether setting the IF contamination parameter equal to the anomaly rate of the training subset remains appropriate under varying anomaly rates, a 10-fold cross-validation across runs [5] was performed. Two TL configurations were evaluated, one based on fine-tuning the decoder of the pretrained AE (Conf. 1) and the other corresponding to the proposed framework (Conf. 2), where the decoder is replaced with a single output layer.

The results summarized in Table 1 show that the proposed framework consistently achieves higher F1-scores across all PVs, with more significant differences for PVs affected by overfitting such as IV and IDV, where the fine-tuning configuration yields lower F1-scores (F1 < 0.5). Based on these results, the proposed framework was retained, and the IF contamination parameter set to the anomaly rate of the training subset remains effective across most PVs despite variations between runs, while RF FP exhibits the lowest F1-score, highlighting the limitations of a fixed contamination-based threshold for this PV.

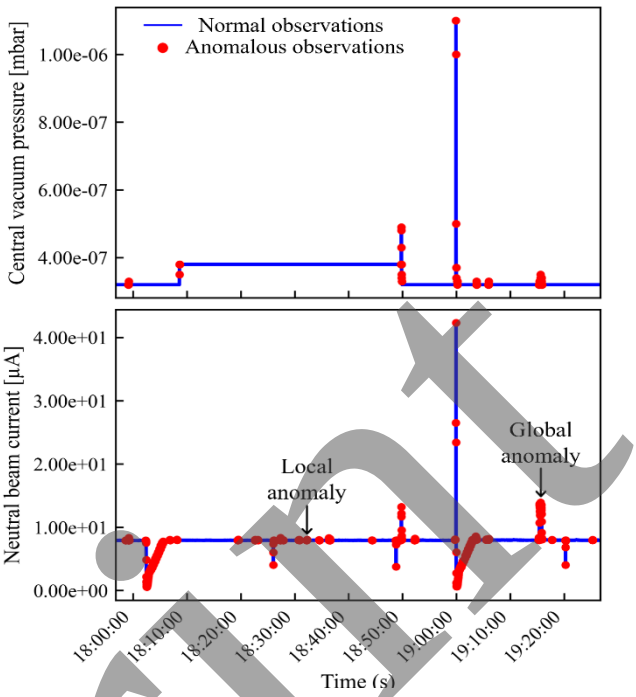

Figure 3: Example of global and local anomaly detection for CVP and NBI using the AE-IF model after TL.

Final Test Performance on Unseen Runs

The final test performance on completely unseen runs, summarized in Fig. 2, shows that the proposed framework achieves strong detection performance across several PVs, with mean F1-scores above 0.8. As shown in Fig. 3, the model detects both global and local anomalies while correctly classifies normal sequences. Among the studied PVs, RF FP achieved the lowest F1-score, consistent with the validation results. IV shows the second lowest performance, which appears to be unrelated to the decision threshold or to the model itself, but is instead attributable to data scaling. As defined in the Operational Definition of Anomalies section, a sequence is considered anomalous when it falls outside the normal operating range, even if it remains stable. However, the StandardScaler was fitted on the normal samples from the development training set, capturing the range $[S_{low},\ S_{high}]_{train}$. For some unseen runs, the run-specific range $[S_{low},\ S_{high}]_{test}$ may lie within this interval. As a result, stable sequences that are anomalous with respect to run-specific thresholds may still fall within the training range, leading to misclassification and lower recall. This scaling limitation also explains the noticeable variability in recall observed for IDV.

CONCLUSION

The proposed TL framework improves local and global anomaly detection across multiple PVs, reduces model complexity through the reuse of the encoder and its latent representation, and maintains robustness under varying anomaly rates. Limitations related to standardization motivate future work on more adaptive normalization strategies.

REFERENCES

[1] F. Poirier, D. Mateus, J. Rioult, and C. Lassalle, "First anomalies exploration from data mining and machine learning at the ARRONAX cyclotron C70XP", in *Proc. IPAC'23*, Venice, Italy, May 2023, pp. 2273-2276. doi:10.18429/JACoW-IPAC2023-TUPM036

[2] A. Ghribi et al., "Artificial intelligence for advancing particle accelerators", *Europhys. News*, vol. 56, no. 1, pp. 15–19, 2025. doi:10.1051/epn/2025106

[3] D. Leite, E. Andrade, D. Rativa, and A. M. A. Maciel, "Fault detection and diagnosis in Industry 4.0: A review on challenges and opportunities", *Sensors*, vol. 25, no. 1, p. 60, Dec. 2024. doi:10.3390/s25010060

[4] Y. Suetsugu, "Machine-learning-based pressure-anomaly detection system for SuperKEKB accelerator", *Phys. Rev. Accel. Beams*, vol. 27, no. 6, p. 063201, Jun. 2024 doi:10.1103/PhysRevAccelBeams.27.063201

[5] F. Basbous, F. Poirier, D. Mateus, and F. Haddad, "Hybrid autoencoder-isolation forest approach for time series anomaly detection in C70XP cyclotron operational data at ARRONAX", in *Proc. CYC2025*, Oct. 2025. doi:10.48550/arXiv.2603.20335

[6] S. J. Pan and Q. Yang, "A survey on transfer learning", *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010. doi:10.1109/TKDE.2009.191

7] J. Liang, H. Shui, R. Gupta, D. Upadhyay, and E. Darve, "Transfer learning for anomaly detection in rotating machinery using data-driven key order estimation", *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 13310–13326, 2025. doi:10.1109/TASE.2025.3552009

[8] Z. Khais Shahid, S. Saguna, C. Åhlund, and K. Mitra, "Anomaly detection using transfer learning for electricity consumption in school buildings: A case of northern Sweden", *Energy Build.*, vol. 346, p. 116129, Nov. 2025. doi:10.1016/j.enbuild.2025.116129

[9] M. Rezakhani, T. Seyfi, and F. Afghah, "A transfer learning framework for anomaly detection in multivariate IoT traffic data", in *Proc. IEEE ICC'25*, Jun. 2025, pp. 4975–4980. doi:10.1109/ICC52391.2025.11161334

[10] M. W. Asres et al., "Data quality monitoring for the hadron calorimeters using transfer learning for anomaly detection", *Sensors*, vol. 25, no. 11, p. 3475, Jan. 2025. doi:10.3390/s25113475

[11] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification", in *Proc. ACL'18*, Melbourne, Australia, Jul. 2018, pp. 328–339. doi:10.18653/v1/P18-1031

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", Jan. 2017, arXiv:1412.6980 doi:10.48550/arXiv.1412.6980