Autoencoder-based time series anomaly detection for ATLAS Liquid Argon calorimeter data quality monitoring

Vilius Čepaitis[1*] and Steven Schramm[1], on behalf of the ATLAS collaboration

1 University of Geneva, Switzerland

⋆ vilius.cepaitis@cern.ch

Abstract

The ATLAS experiment at the LHC employs comprehensive data quality monitoring procedures to ensure high-quality physics data. This contribution presents a long short-term memory autoencoder-based algorithm for detecting anomalies in ATLAS Liquid Argon calorimeter data, represented as multidimensional time series of statistical moments of energy cluster properties. Trained on good-quality data, the model identifies anomalous intervals. Validation is performed using a known short-term issue of noise bursts, and the potential for broader application to transient calorimeter issues is discussed.

1 Introduction

The Liquid Argon (LAr) calorimeter [1, 2] is a key component of the ATLAS [3] detector at the CERN LHC. It is designed to precisely measure electromagnetic shower energies, such as those originating from electrons and photons. Reliable data taking with many channels over decades necessitates identification of occasional issues. In these proceedings, a generic, proof-of-concept time series anomaly detection (AD) method for the LAr calorimeter data quality monitoring (DQM) is presented. While the LAr calorimeter has excellent data quality efficiency [4], there are several reasons for complementing the existing infrastructure with such a technique. First, existing algorithms focus on well-known detector issues, whereas an AD algorithm could act as an early warning system for rare, previously unseen problems. Furthermore, such a technique could serve in reducing the substantial human effort required to perform routine DQM checks.

2 Methodology

This section summarises the inputs to the autoencoder-based anomaly detection algorithm as well as the model's architecture and training setup. Additional details of the training procedure

can be found in Ref. [5].
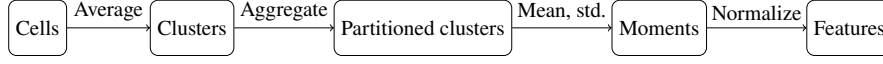
2.1 Liquid argon calorimeter inputs

Figure 1: A sketch of the data processing pipeline

A sketch of the data processing pipeline is shown in Figure 1. The whole calorimeter data is represented as a 16-dimensional time series, corresponding to different properties of the ensemble of calorimeter signals in each of the four electromagnetic detector partitions: EMBA, EMBC, EMECA, and EMECC.

The time series data is obtained in a staged approach. First, individual calorimeter signals (cells) are clustered into topo-clusters [6], a standard method used in ATLAS for hadronic object reconstruction. Two topo-cluster variables are used: the *Q*-factor and the timing, τ, calculated based on the pulse shape properties of the corresponding cells. These variables were chosen for their sensitivity to different detector issues. For each topo-cluster variable, two statistical moments (mean, standard deviation) are obtained. Finally, the topo-clusters are themselves aggregated over different physical calorimeter partitions, transforming the variable number of calorimeter signals into a fixed-size representation in the time series. As a final step of data preprocessing, each dimension is normalized to ensure it contributes a similar weight to the reconstruction loss, as discussed in the following section. The time series data represents an irregularly sampled event stream, with entries corresponding to events triggered by ATLAS during roughly one-minute intervals of data-taking.

2.2 Anomaly detection algorithm

An unsupervised autoencoder-based anomaly detection algorithm [7] is used to identify anomalous time periods. The autoencoder (AE) architecture is shown in Figure 2. The AE consists of an encoder f and a decoder g, which compress and reconstruct the input as well as possible. A long short-term memory (LSTM) [8] block is used for both the encoder and decoder architecture. The mean square error loss function is minimized between the input time series $\vec{X}$ and their auto-encoder replicas during training:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \|\vec{X}_i - g(f(\vec{X}_i))\|^2. \tag{1}$$

Several samples of ATLAS data were used for training the autoencoder, corresponding to both proton–proton and heavy–ion data, collected in different years. This combination was performed to obtain a more robust model as the detector, beam, and data collection conditions all changed. Only data certified as good for physics analyses by ATLAS was used in the training.

The algorithm exploits the fact that the detector generally operates well, only occasionally interrupted by brief periods of anomalous activity. The AE aims to reconstruct the bulk of the good data training distribution. LAr detector issues are generally rare, for instance the corresponding data loss during 2015–2018 data-taking was below 1% [4]. Therefore, such data will be more likely to be poorly reconstructed as an outlier, i.e., to have a higher reconstruction loss. This approach complements conventional DQM techniques within ATLAS, which are based on monitoring distributions in the form of histograms.
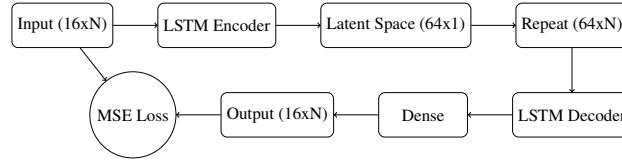
Figure 2: The LSTM autoencoder architecture.

3   Results

Several independent datasets were employed to evaluate the model performance. To validate the reconstruction of the AE and identify a threshold loss for flagging events as anomalous, a data-taking run heavily enriched in noise bursts was used.

The LAr calorimeter periodically suffers from bursts of coherent noise, which were found to intensify with increasing luminosity. Such bursts typically last for no longer than a millisecond, which is still a considerable period given the LHC bunch crossing rate of 25 ns. Figure 3a shows an example of a typical event with a total energy of 2 TeV recorded during a noise burst. A well-established algorithm to identify such noisy events already exists within ATLAS and was therefore used to provide ground-truth labels to benchmark the AE performance. Figure 3b shows that the AE reconstruction loss for the events identified as noisy by this algorithm is significantly higher compared to the rest. Based on this distribution, the threshold $L_{\mathrm{th}} = 1.3$ was chosen to classify events as anomalous. An anomalous time window is obtained by requiring several such consecutive events in order to reduce the risk of single rare events being flagged as detector issues, as is done for the reference algorithm.

Although the AE was primarily validated using noise burst-enriched data, its application remains more general. The AE complements the reference algorithm, occasionally identifying additional anomalous time windows. One such example is shown in Figure 4 for proton–proton collision data, likely corresponding to a temporary failure of a vacuum diffusion pump. The plot shows that the AE reconstruction loss passes the anomaly threshold $L > L_{\mathrm{th}}$ for a cluster of events and thus flags the issue, unlike the reference algorithm. Additional studies in Ref. [5] show that the model also performs well on cosmic or heavy–ion collision data.
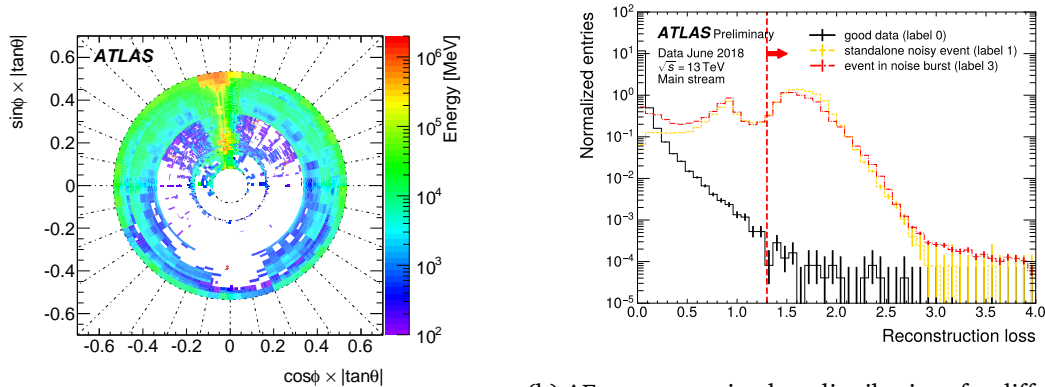
(a) A typical LAr calorimeter noise burst event, corresponding to an energy deposit of a few TeV, shown in the transverse plane. Figure from Ref. [9].

(b) AE reconstruction loss distributions for different events as categorized by reference noise burst algorithm. The chosen anomaly threshold $L_{\mathrm{th}} = 1.3$ separates good data from the bulk of the noisy events. Figure taken from Ref. [5].

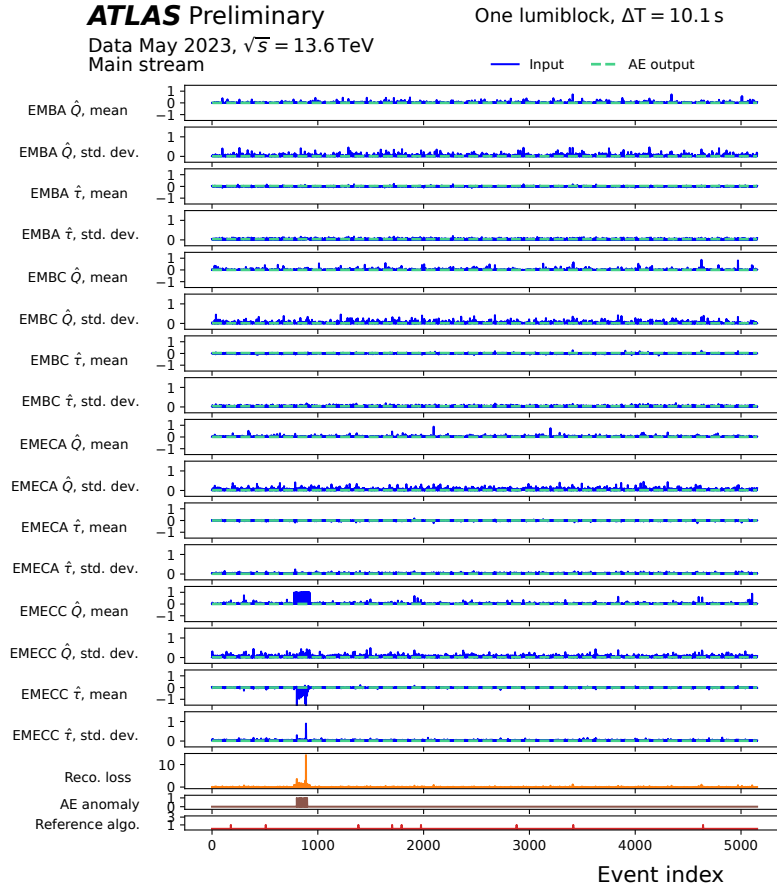Figure 3: Illustrations of LAr noise bursts and their reconstruction loss distributions.

Figure 4: Original (blue) and AE-reconstructed (cyan) time series for calorimeter partitions (top 16 rows). The bottom three rows show the reconstruction loss, L, events with $L > 1.3$, and noisy event labels from a reference algorithm. An anomalous window near event index 900 is flagged by the AE but not by the reference algorithm. Figure taken from Ref. [5].

4 Conclusion

A proof-of-principle time series anomaly detection method based on an LSTM autoencoder has been developed for the ATLAS Liquid Argon calorimeter. The algorithm is capable of identifying brief anomalous periods in the calorimeter time series, complementing existing data quality monitoring techniques. Validation studies show that the autoencoder successfully reconstructs good-quality data while identifying known anomalous events. Moreover, the method generalizes to previously unseen detector issues not flagged by existing algorithms. These results illustrate the potential of unsupervised anomaly detection as an early warning system for rare detector problems, paving the way for more robust and automated LAr calorimeter monitoring.

Acknowledgements

References

[1] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter: Technical Design Report*, ATLAS-TDR-2; CERN-LHCC-96-041, CERN, https://cds.cern.ch/record/331061 (1996).

[2] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter Phase-I Upgrade: Technical Design Report*, ATLAS-TDR-022; CERN-LHCC-2013-017, CERN, https://cds.cern.ch/record/331061 (2013).

[3] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3**, S08003 (2008), doi:10.1088/1748-0221/3/08/S08003.

[4] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15**, P04003 (2020), doi:10.1088/1748-0221/15/04/P04003, 1911.04632.

[5] ATLAS Collaboration, *Autoencoder-based time series anomaly detection for ATLAS Liquid Argon calorimeter data quality monitoring* (ATL-DAPR-PUB-2024-002) (2024), https://cds.cern.ch/record/2918639.

[6] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77**, 490 (2017), doi:10.1140/epjc/s10052-017-5004-5, 1603.02934.

[7] G. E. Hinton and R. R. Salakhutdinov, *Reducing the dimensionality of data with neural networks*, Science **313**(5786), 504–507 (2006), doi:10.1126/science.1127647.

[8] S. Hochreiter and J. Schmidhuber, *Long Short-Term Memory*, Neural Computation **9**(8), 1735 (1997), doi:10.1162/neco.1997.9.8.1735.

[9] ATLAS Collaboration, *Monitoring and data quality assessment of the ATLAS liquid argon calorimeter*, JINST **9**, P07024 (2014), doi:10.1088/1748-0221/9/07/P07024, 1405.3768.