CERN Summer Student Report
9 October 2025
ryxie24@connect.hku.hk

Calibration of a Generic Event-Level Anomaly Detection Trigger (GELATO) for the ATLAS Experiment

Author: Ruiyang Xie
Supervisors: Sagar Addepalli, Julia Lynne Gonski, Dylan Sheldon Rankin
CERN, Switzerland

Keywords: ATLAS, Anomaly Detection, Trigger, Machine Learning, VAE-GAN, Calibration, GELATO, Neural Network

---

Abstract: The Large Hadron Collider (LHC) produces approximately 40 million proton–proton collisions per second, far exceeding the capacity of the detectors to record and store every event. To cope with this challenge, a multi-tiered trigger system is employed to select potentially interesting events for further analysis. At the ATLAS experiment, the Level-1 (L1) trigger provides the first, real-time decision layer. However, as its algorithms are primarily optimized for Standard Model (SM) signatures, it may miss events exhibiting unexpected or unconventional features — a significant limitation in the search for physics beyond the SM. To overcome this, we propose a novel unsupervised machine-learning-based approach: the Generic Event-Level Anomaly Detection Trigger (GELATO). Unlike conventional triggers, GELATO exploits unsupervised learning to reduce reliance on physics-specific assumptions, thereby enhancing sensitivity to unknown phenomena. In this study, we present the calibration of GELATO by evaluating anomaly scores and trigger efficiencies across six SM Monte Carlo datasets, under software thresholds corresponding to additional trigger rates of 500 Hz, 1 kHz, and extended scenarios of 2–5 kHz. These results establish a foundation for deploying a robust, data-driven anomaly trigger, offering a promising pathway to increase the discovery potential of ATLAS in searches for new physics.

---

1 Introduction

1.1 Motivation and Context

The ATLAS Level-1 (L1) trigger system is the first stage of the data acquisition pipeline, responsible for reducing the event rate from the 40 MHz LHC bunch crossing frequency to a manageable level of $\mathcal{O}(100\ \text{kHz})$. While traditional cut-based triggers are effective for well-defined and expected physics signatures, they are inherently limited when facing rare, unexpected, or poorly modeled phenomena. This limitation is particularly critical in searches for beyond-the-Standard-Model (BSM) physics, where the characteristics of potential signals are often unknown.
To address this challenge, a Generic Event-Level Anomaly Detection Trigger (GELATO) based on a Variational Autoencoder - Generative Adversarial Network (VAE–GAN) architecture has been proposed. This unsupervised, data-driven approach assigns an anomaly score to each event, enabling the identification of unusual topologies without prior assumptions about their underlying physics models. Such a system not only improves sensitivity to unknown signatures but also provides broader coverage for rare Standard Model processes that might mimic BSM phenomena.

1.2 Scope of This Work

This work focuses on the initial calibration of the GELATO. The calibration looks at the efficiency curves against different features at various well-defined thresholds corresponding to different intake rates, aiming to ensure that the trigger operates stably and predictably under high-luminosity conditions. The main tasks carried out include:

- Computing anomaly scores for six signal Monte Carlo (MC) datasets using GELATO.
- Evaluating trigger efficiencies across multiple physics processes at 500 Hz and 1 kHz added unique rates, as well as some hypothetical thresholds at 2-5kHz rates.

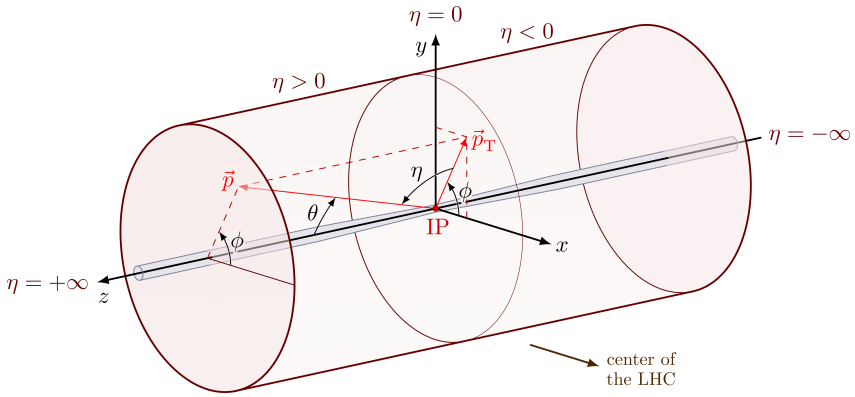

Figure 1: LHC Coordinate System (A segment of LHC)

- Establishing reproducible workflows and scripts as a foundation to develop a systematic calibration framework.

The results presented here lay a foundation for subsequent development and refinement.

2 Methodology

This study uses dedicated Monte Carlo (MC) datasets to calibrate the GELATO. Analyses focus on producing anomaly score profiles and trigger efficiency curves at multiple output-rate thresholds.

2.1 Monte Carlo Datasets

Six MC processes representing a range of signal-like events are used. All datasets are stored in `HDF5` (`.h5`) format. Each event is described by a 44-dimensional feature vector comprising:

- **Jets (jFexSR):** Six jets, each with p_T, η, and ϕ (18 features);
- **Taus (eFex):** Four taus, each with p_T, η, and ϕ (12 features);
- **Muons:** Four muons, each with p_T, η, and ϕ (12 features);
- **MET:** Missing transverse energy p_T and ϕ (2 features).

The sizes of the datasets are summarized in the table below.

Sample	Events	Size (GB)
Topo_2A_jz1	$\sim 5.0 \times 10^5$	~ 1.8 GB
Topo_2A_jz2	$\sim 4.8 \times 10^5$	~ 1.7 GB
Topo_2A_jz4	$\sim 3.5 \times 10^5$	~ 1.5 GB
Topo_2A_ttbar_1lep	$\sim 6.0 \times 10^5$	~ 2.0 GB
Topo_2A_ttbar_2lep	$\sim 5.5 \times 10^5$	~ 1.9 GB
Topo_2A_ZZ4lep	$\sim 1.0 \times 10^5$	~ 0.5 GB

Table 1: Summary of MC datasets used in the anomaly detection study.

These samples cover a broad range of physics topologies:

- `jz1`, `jz2`, `jz4`: Inclusive QCD dijet events, spanning low-, medium-, and high-energy regions, respectively;
- `ttbar_1lep` and `ttbar_2lep`: Semi-leptonic and dileptonic top–antitop production, respectively;
- `ZZ4lep`: A clean $ZZ \to 4\ell$ benchmark, providing a Standard Model control sample for multi-lepton final states.

2.2 Analysis Workflow and Pipeline

All datasets were accessed from EOS storage and processed in the CERN `lxplus` environment. Scripts, intermediate outputs, and final results were synchronized with CERNBox for backup and reproducibility. The calibration code and all anomaly score profiles and trigger efficiency plots are available on GitHub:

`https://github.com/ryxie-24/GELATO-Calibration`

The analysis pipeline consists of the following stages:

1. **Feature scaling:** Raw features are scaled using calibrated powers of two for each dimension, ensuring compatibility with the quantized model.

2. **Latent-space inference:** Scaled features are passed through the quantized VAE–GAN encoder to obtain the latent mean vector μ.

3. **Anomaly scoring:** Per-event anomaly scores are computed as the mean squared latent components:
$$\text{score}(x) = \sum_{j=1}^{d} \mu_j^2,$$
where $d = 3$ is the latent dimension.

4. **Data storage:** Scaled and unscaled features, along with per-event anomaly scores, are stored for further calibration and evaluation.

5. **Visualization and validation:** Scatter plots of anomaly scores versus jet p_T, η, ϕ, and missing transverse energy (MET) are produced. Efficiency-versus-kinematics curves are derived for thresholds corresponding to 500 Hz, 1 kHz, 2 kHz, 3 kHz, 4 kHz, and 5 kHz trigger output rates.

This integrated workflow ensures reproducibility and provides a foundation for systematic calibration studies and deployment tests.

2.3 Model Architecture

The GELATO model is a quantized Variational Autoencoder with Generative Adversarial Network regularization (VAE–GAN). It comprises three main components:

- **Encoder:** Accepts the 44-dimensional feature vector and processes it through two quantized dense layers with batch normalization (sizes 32 and 16), followed by quantized ReLU activations. The latent space has dimensionality $d = 3$. The encoder outputs the latent mean (z_μ), log-variance ($z_{\log \sigma^2}$), and sampled latent vector z via the reparameterization trick:
$$z = z_\mu + \exp\left(0.5 \cdot z_{\log \sigma^2}\right) \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I).$$

- **Decoder:** Mirrors the encoder, reconstructing the 44-dimensional feature vector from the 3-dimensional latent input through two ReLU layers (sizes 16 and 32) and a final linear layer.

- **Discriminator:** A binary classifier with hidden layers of sizes 8 and 2, followed by a sigmoid output, trained to differentiate real events from reconstructed ones.

2.4 Feature Scaling

Each input feature x_i is scaled as:
$$x_i^{\text{scaled}} = \frac{x_i}{2^{s_i}},$$
where s_i is the feature-specific scaling exponent determined during calibration. The complete list of features and their corresponding scaling exponents is shown in Table 2.

Category	Feature	s_i	Category	Feature	s_i
jFexSR jets (6)	jFexSR_jet0_pt (GeV)	7	Muons (4)	muon0_pt (GeV)	4
	jFexSR_jet0_eta	5		muon0_eta	4
	jFexSR_jet0_phi (rad)	5		muon0_phi (rad)	4
	jFexSR_jet1_pt (GeV)	7		muon1_pt (GeV)	2
	jFexSR_jet1_eta	5		muon1_eta	3
	jFexSR_jet1_phi (rad)	5		muon1_phi (rad)	3
	jFexSR_jet2_pt (GeV)	5		muon2_pt (GeV)	1
	jFexSR_jet2_eta	4		muon2_eta	1
	jFexSR_jet2_phi (rad)	4		muon2_phi (rad)	2
	jFexSR_jet3_pt (GeV)	5		muon3_pt (GeV)	-2
	jFexSR_jet3_eta	4		muon3_eta	-1
	jFexSR_jet3_phi (rad)	3		muon3_phi (rad)	-1
	jFexSR_jet4_pt (GeV)	4	MET	MET_pt (GeV)	6
	jFexSR_jet4_eta	3		MET_phi (rad)	5
	jFexSR_jet4_phi (rad)	3			
	jFexSR_jet5_pt (GeV)	3			
	jFexSR_jet5_eta	3			
	jFexSR_jet5_phi (rad)	2			
eFex taus (4)	eFex_tau0_pt (GeV)	6			
	eFex_tau0_eta	5			
	eFex_tau0_phi (rad)	5			
	eFex_tau1_pt (GeV)	5			
	eFex_tau1_eta	4			
	eFex_tau1_phi (rad)	5			
	eFex_tau2_pt (GeV)	4			
	eFex_tau2_eta	3			
	eFex_tau2_phi (rad)	4			
	eFex_tau3_pt (GeV)	3			
	eFex_tau3_eta	2			
	eFex_tau3_phi (rad)	3			

Table 2: Scaling exponents s_i for all 44 features.

3 Results and Findings

This section summarizes the main findings from the analysis of anomaly scores and efficiency profiles performed by GELATO on the MC signal datasets.

3.1 Anomaly Score vs. Jet p_T

The anomaly score profiles of all processes against all 44 kinematic features are studied. The details of all plots can be found in the GitHub repository. For most processes, the **anomaly score shows an approximately positive correlation with jet p_T**, independent of the jet rank. In contrast, the behavior of `jz4`, as shown in Figure 3, is more intricate: the anomaly score does not rise monotonically with increasing jet p_T. In the traditional Level-1 trigger configuration, `jz4` events are already accepted with almost 100% efficiency, showing that these topologies have been well captured by the standard trigger logic. This shows the anomaly score is governed not only by the magnitude of

individual jet p_T, but also by multi-jet correlations and the global event topology.

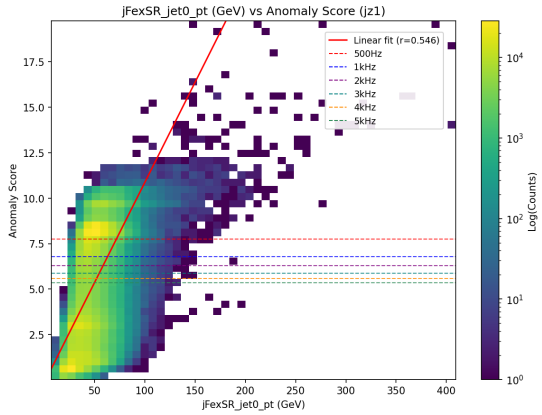

Figure 2: jz1 anomaly score vs first jet p_T (GeV)

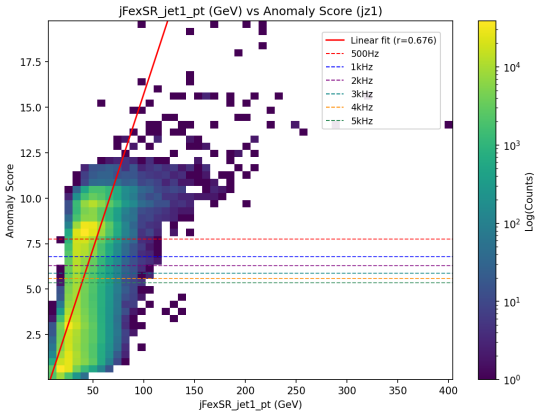

Figure 3: jz1 anomaly score vs second jet p_T (GeV)

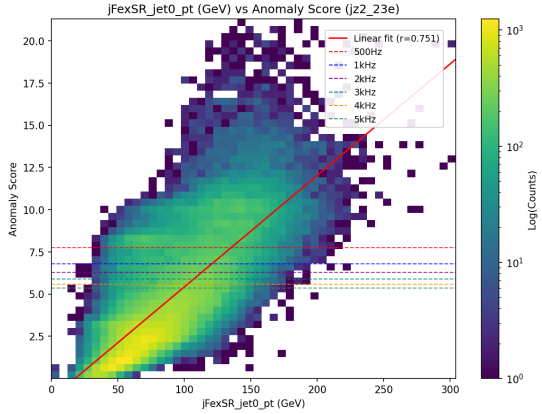

Figure 4: jz2 anomaly score vs first jet p_T (GeV)

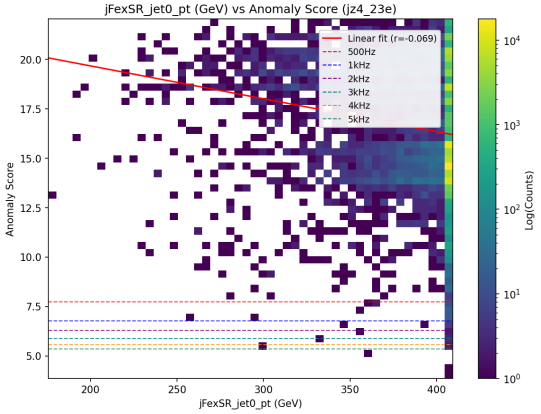

Figure 5: jz4 anomaly score vs first jet p_T (GeV)

3.2 Anomaly Score vs. Jet Multiplicity

Correlation between anomaly score and jet multiplicity of events of all processes are examined. For `jz1` (Figure 6), almost all events contain six jets, making it difficult to identify a clear correlation. However, events with fewer than six jets consistently show anomaly scores below 1, indicating that **low jet multiplicity rarely coincides with high anomaly scores.** For the other five processes, high anomaly score events tend to cluster at higher jet multiplicities, while the bulk of events with six jets remain at relatively low scores. This behavior suggests that events with multiple high-energy jets are more likely to display unusual correlations across kinematic variables interpreted as anomalous by GELATO.

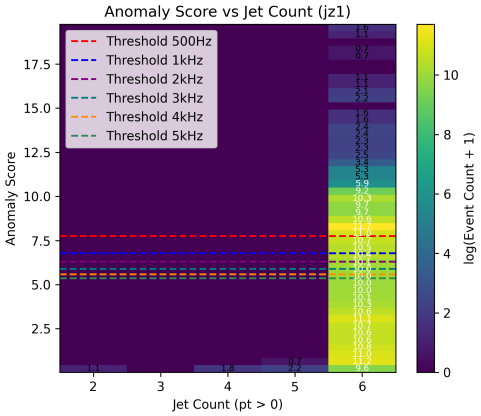

Figure 6: jz1 anomaly score vs jet multiplicity

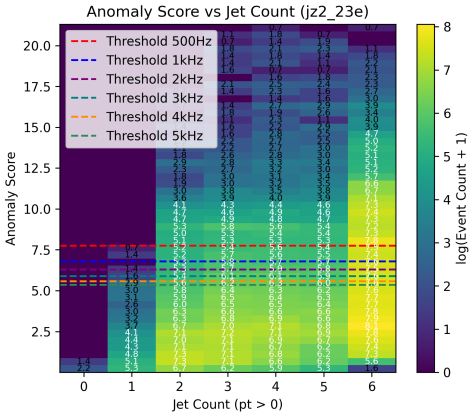

Figure 7: jz2 anomaly score vs jet multiplicity

3.3 Anomaly Score vs. Invariant Mass of the Two Leading Jets

Per-event invariant mass of two leading jets (m_{jj}) are calculated and their correlation with anomaly score are studied. **No universal trend between anomaly score and m_{jj} is observed across the processes.** In general, events above the anomaly score threshold tend to populate regions with relatively low m_{jj}, reflecting the intrinsic distribution of the dataset rather than a straightforward correlation between large dijet mass and high anomaly score.

3.4 Efficiency vs. Jet p_T

Efficiency curves by different thresholds are studied, providing a preliminary performance map across different regions of phase space. For conventional threshold-based triggers, efficiency curves typically exhibit an S-shaped profile: close to zero well below the threshold, approaching unity well above it, and a sharp transition in between. While the anomaly trigger is multivariate and its efficiency therefore depends on correlations among many features, the projected curves as a function of transverse momentum of different daughter particles still broadly follow a turn-on behavior. This agreement suggests that our GELATO is not only operating as intended, but also highlighting regions where more complex dependencies may appear. Interestingly, for the `ttbar_1lep`, `ttbar_2lep` and `ZZ4lep` processes, efficiency at very low values of some p_T starts at a finite value, dips, and then rises again in the classic S-shape. An example of this phenomena is illustrated in Figure 6. This suggests that **the anomaly trigger is sensitive to subtle multi-feature correlations in low-p_T regions, potentially identifying interesting topologies that could evade conventional triggers**. This capability is particularly promising, as it indicates the trigger's potential to uncover signatures of physics beyond the Standard Model in unconventional or low-energy regions of phase space.

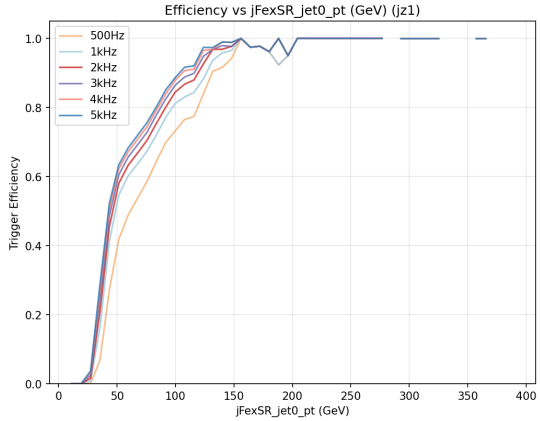

Figure 8: jz1 trigger efficiency vs first jet p_T (GeV)

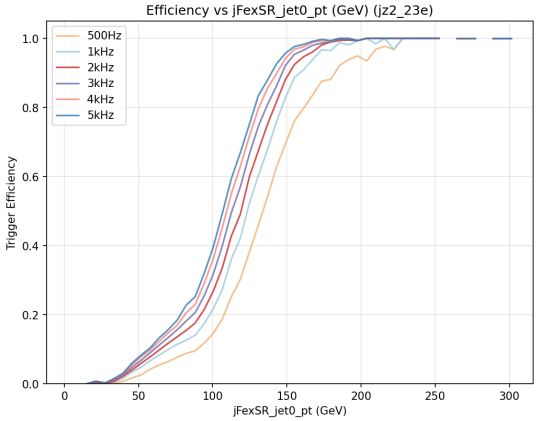

Figure 9: jz2 trigger efficiency vs first jet p_T (GeV)

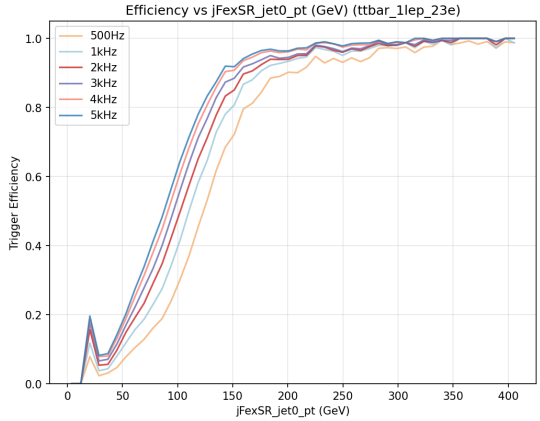

Figure 10: ttbar1lep trigger efficiency vs first jet p_T (GeV)

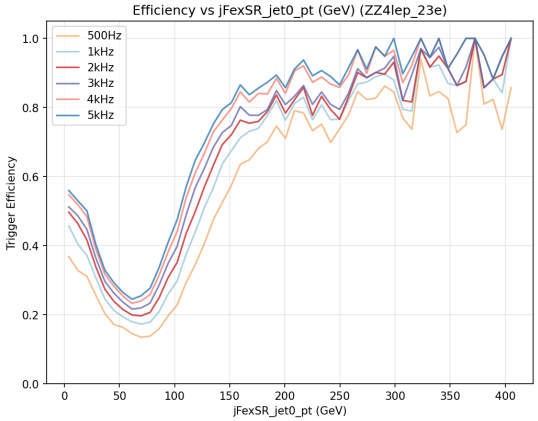

Figure 11: ZZ4lep trigger efficiency vs first jet p_T (GeV)

3.5 Efficiency vs. Endcap Muon p_T

The efficiency of the anomaly trigger as a function of the muon transverse momentum (p_T) provides insight into how the model utilizes lepton-related information. Figure 12-15 show the efficiency curves for the four reconstructed endcap ($\eta \geq 1.05$) muons in the `ZZ4lep` process. For the first three leading muon p_T, the efficiency curves remain relatively flat, showing only a mild dependence on p_T across the kinematic range. This indicates that the anomaly trigger does not rely strongly on the leading or subleading muon momenta in the event. However, a distinct behavior emerges for `muon3`. In the endcap region, the efficiency exhibits a pronounced turn-on structure, rapidly rising with increasing p_T. This pattern suggests that the model is effectively using information related to **muon multiplicity** — rather than just the hardest muon — as a discriminating feature. In other words, the presence and kinematics of the fourth muon play a significant role in determining the anomaly score, highlighting that the learned trigger is sensitive to final states with multiple leptons, a key characteristic of rare processes such as $ZZ \to 4\ell$.

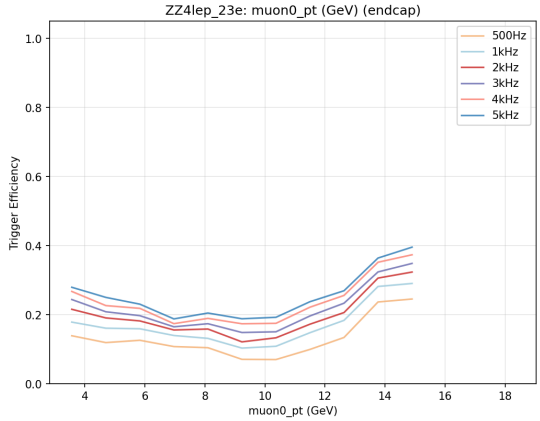

Figure 12: ZZ4lep trigger efficiency vs first muon p_T (GeV)

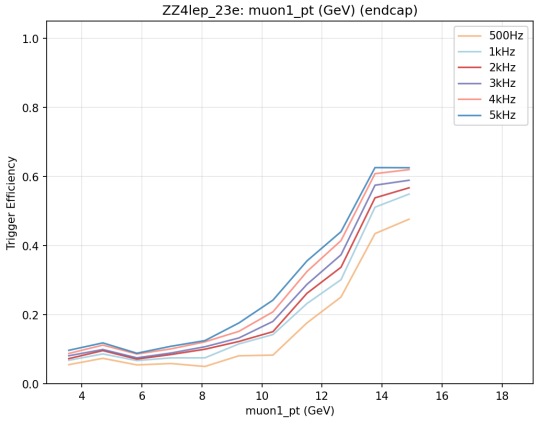

Figure 13: ZZ4lep trigger efficiency vs second muon p_T (GeV)

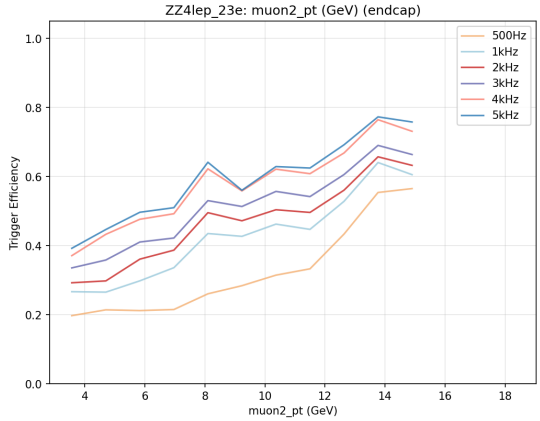

Figure 14: ZZ4lep trigger efficiency vs third muon p_T (GeV)

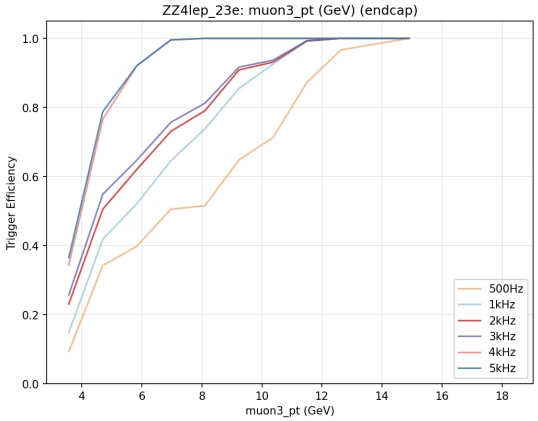

Figure 15: ZZ4lep trigger efficiency vs fourth muon p_T (GeV)

3.6 Summary of Findings

Overall, GELATO shows a broadly positive correlation between anomaly score and jet p_T, except in `jz4` where symmetric event topologies dominate. High anomaly scores cluster at higher jet multiplicities for most processes, while `jz1` events are concentrated in six-jet configurations, and `jz4` consistently shows high scores. No clear global correlation is found between anomaly score and dijet invariant mass, indicating that high anomaly scores are not limited to high-mass regions in our VAE–GAN model. Efficiency curves as a function of jet p_T display the anticipated turn-on behavior: low efficiency at small p_T, a rapid rise in the intermediate region, and a plateau at high p_T. Unlike traditional single-threshold triggers, the anomaly trigger is multivariate, so the turn-on is not a perfect S-shape and can exhibit process-dependent variations. Notably, the trigger demonstrates sensitivity to low-p_T events in `ttbar1lep`, `ttbar2lep`, and `ZZ4lep`, reflecting its ability to identify subtle correlations and potentially enhance sensitivity to unconventional or rare final states relevant for new physics searches. In addition, studies of muon kinematics reveal that for the `ZZ4lep` process, the anomaly trigger shows a pronounced efficiency turn-on with respect to the fourth muon's transverse momentum (p_T) in the endcap region ($\eta \geq 1.05$), whereas the leading three muons exhibit much weaker dependence. This behavior indicates that the model does not primarily rely on the leading muon p_T, but rather exploits muon multiplicity and subleading lepton information

to characterize anomalous topologies. Such behavior underscores the network's ability to leverage multi-object correlations beyond the reach of conventional triggers.

4 Challenges and Solutions

Several technical and analytical challenges were addressed during the study. This section summarizes the key issues and the strategies adopted to resolve them.

4.1 Large-scale Data Processing

The analysis involved large-scale Monte Carlo signal data, leading to significant storage and computational demands. To manage this, datasets were organized on EOS with CERNBox synchronization for reproducibility, and Python workflows were optimized with batch loading and vectorized operations to enhance efficiency and scalability.

4.2 Dataset Consistency

Inconsistent naming conventions across datasets and feature sets from previous work initially caused integration difficulties. All naming schemes were carefully standardized; users should exercise caution regarding feature names when utilizing the GitHub scripts.

5 Conclusion

This study computed anomaly scores for multiple Monte Carlo signal processes and systematically investigated their correlations with key kinematic variables, including jet transverse momentum, jet multiplicity, and dijet invariant mass. The analysis revealed process-dependent patterns in anomaly scores, providing a data-driven basis for interpreting trigger sensitivity and informing threshold selection for optimal Level-1 anomaly trigger performance in ATLAS.
Efficiency studies, in conjunction with detailed kinematic profiles and visualizations of anomaly scores, establish a rigorous foundation for future systematic calibration and scale-factor derivation. These results are expected to enhance the sensitivity of the ATLAS trigger system during Run 3 and may guide strategies for advanced anomaly detection in potential Phase-II upgrades.

6 References

1. ATLAS Run-3 Conditions. *arXiv:2305.16623*, `https://arxiv.org/pdf/2305.16623`.

2. ATLAS New Trigger R&D Twiki. `https://twiki.cern.ch/twiki/bin/view/Atlas/NewTriggerRandD`.

3. AD Trigger Training Repository. GitHub, `https://github.com/max-cohen54/AD_trigger_training`.

4. VAE-GAN Anomaly Trigger Concept. Ekaterina Govorkova et al., *arXiv:2108.03986*, `https://arxiv.org/abs/2108.03986`.