PROCEEDINGS OF SCIENCE

Transformer Autoencoder Anomaly Detection for Enhanced Discovery in High-Energy Physics

Asrith Krishna Radhakrishnan,[a,*] **Maximiliano Sioli**[a] **and Lorenzo Rinaldi**[a]

[a] *Department of Physics and Astronomy, University of Bologna, Via Irnerio 46, Bologna, Italy*

E-mail: asrith.radhakrishna2@unibo.it

An unsupervised machine learning framework is developed for identifying anomalous events in proton–proton collisions at $\sqrt{s}$ = 13 TeV, trained on the ATLAS Run-2 open dataset collected during 2015–2016. The analysis utilize a model independent approach based on TAB-transformer-autoencoder trained to characterize Standard Model event topologies and identify deviations. The anomaly scores show clear separation between background and MC generated beyond-Standard-Model and rare benchmark processes. Clustering of high-scoring events reveals distinct kinematic structures consistent with signatures of rare and BSM dynamics. The result demonstrates the sensitivity of the Tab-Transformer-based anomaly detection framework to diverse rare and BSM signatures, enabling data-driven exploration beyond traditional search strategies.

The European Physical Society Conference on High Energy Physics (EPS-HEP 2025)
7–11 July 2025
Palais du Pharo, Marseille, France.

*Speaker

ISSN 1824-8039. Published by SISSA Medialab.

PoS(EPS-HEP2025)663

1. Introduction

Traditional LHC search studies for novel physics typically assume a specific signal hypothesis and optimize event selection or classifiers for that target [11, 12]. Although these approaches achieve high sensitivity for known topologies, they may overlook unexpected signatures. The absence of significant deviations from Standard Model (SM) predictions motivates complementary, model-independent approaches [10, 21]. Anomaly detection (AD) methods aim to identify rare events deviating from the learned SM background without a predefined signal hypothesis [11]. Fully unsupervised methods, trained exclusively on background data, are particularly appealing [10]. Autoencoders and related models learn the implicit SM density and flag high-reconstruction-error events as anomalies [13, 14]. Recent studies show these techniques can complement bump-hunt analyses [13, 14], e.g., ATLAS [13] and CMS [14] applied unsupervised AD to jet mass spectra, improving sensitivity to broad or non-resonant signals. Building on these advances, we develop a flexible, event-level autoencoder framework to capture complex correlations among reconstructed physics objects and robustly identify deviations from the learned SM manifold. To further enhance representational power and capture global dependencies among event features, we employ a Transformer-based architecture, offering improved modeling flexibility over conventional autoencoders.

2. Data and Preprocessing

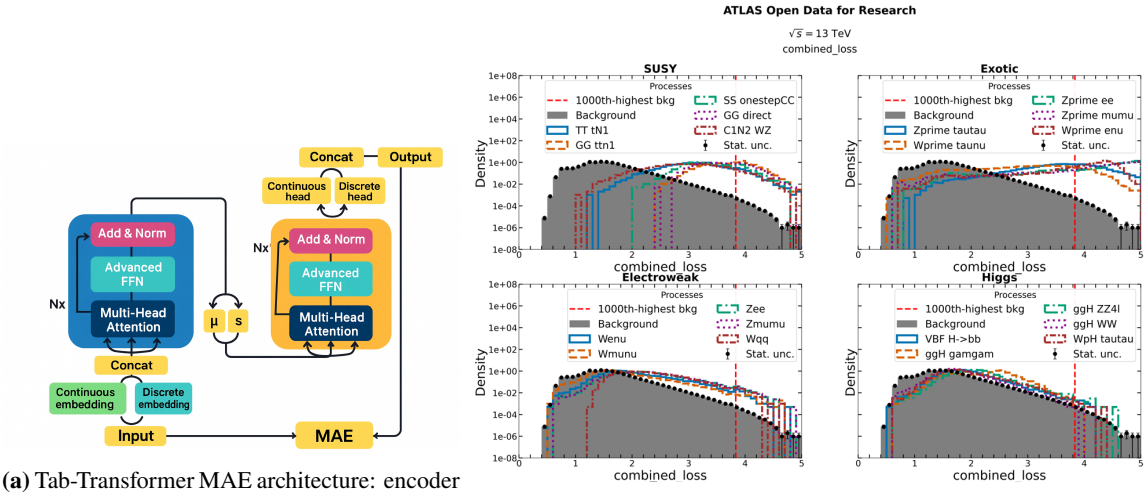
We use ATLAS Run-2 open data collected during 2015–2016 from proton–proton collisions at $\sqrt{s} = 13$ TeV [1], comprising $\sim 7 \times 10^9$ events (about 45 TB) with corresponding Monte Carlo (MC) simulated samples. Events are stored in the DAOD_PHYSLITE format, containing fully reconstructed objects such as jets, leptons, photons, and missing transverse energy. Data processing is performed using the ROOT `RDataFrame` interface [2], a high-level C++/Python framework providing a declarative API for efficient, parallelized event analysis.

`Dask.distributed` [5] scheduler in the HTCondor batch computing pool [4, 5]. This setup enables the concurrent execution of numerous RDataFrame tasks across the cluster, achieving scalable processing of millions of collision events. We extracted and filtered the physics observables such as jet and lepton kinematics, photon and `Egamma cluster` variables, candidates multiplicities, and missing transverse momentum (E_T^{miss}), and preprocessed in parallel. The resulting reduced ROOT files are subsequently merged and converted into tabular datasets for downstream machine-learning analysis. We employ the TMVA TF batch generator [7] to interface these datasets with TensorFlow-based models, ensuring optimized data throughput during training. This integrated `RDataFrame`, `Dask`, `SWAN`, and `HTCondor`, and provides a robust and scalable framework for handling the full ATLAS open-data volume, as well as the associated MC samples used for validation.

3. Model: Tab-Transformer Masked Autoencoder

We employ a masked autoencoder (MAE) based on the Tab-Transformer architecture [6], which will extend the Transformer framework [8] to heterogeneous tabular data. Event level features such as jet and lepton kinematics, multiplicities, and missing transverse energy are standardized, with

missing or invalid entries replaced by a missing token (large negative value) and tracked by a binary mask. Continuous features are linearly projected, while discrete variables are embedded through learnable categorical embeddings, allowing the model to capture complex correlations among physics observables. The overall architecture, including the encoder–decoder configuration and attention flow, is illustrated in 1a.



(a) Tab-Transformer MAE architecture: encoder embeds features with self-attention; decoder reconstructs masked inputs.

(b) Anomaly score distribution for SM background and BSM signals, with higher scores indicating anomalous events.

Figure 1: (a) Tab-Transformer masked autoencoder architecture. (b) Distribution of anomaly scores showing separation between background and signal events.

Regularization masking and fine-tuning are applied, with the model reconstructing missing values via multi-head attention (MHA) and feed-forward layers (FFN)

$$\mathbf{H}^{(l)} = \mathrm{BN}\big(\mathbf{H}^{(l-1)} + \mathrm{MHA}(\mathbf{H}^{(l-1)})\big), \quad \mathbf{Z}^{(l)} = \mathrm{BN}\big(\mathbf{H}^{(l)} + \mathrm{FFN}(\mathbf{H}^{(l)})\big). \tag{1}$$

The encoder output is compressed into a latent representation $\mathbf{z}$, which the decoder uses to reconstruct the complete event. Training optimizes a masked mean absolute error (MAE) reconstruction loss:

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{\sum_i m_i} \sum_i m_i |\hat{x}_i - x_i|, \tag{2}$$

where $\mathbf{m}$ is the validity mask and $\hat{\mathbf{x}}$ the reconstructed input. The anomalies are quantified in the residual space using a Mahalanobis distance:

$$D_M = \sqrt{(\mathbf{r} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{r} - \hat{\mathbf{x}})}, \quad \mathbf{r} = \mathbf{x} - \hat{\mathbf{x}}, \tag{3}$$

with $\boldsymbol{\Sigma}$ estimated via Ledoit–Wolf shrinkage. We compute a normalized score

$$z_{\mathrm{corr}} = \frac{D_M - \langle D_M \rangle_{\mathrm{bg}}}{\sigma_{\mathrm{bg}}^{0.9}}. \tag{4}$$

σ_{bg} the standard deviation of Mahalanobis distances from a fraction of Run2 background data, not used for training or testing, and the 0.9 exponent was tuned to boost sensitivity to deviations.The final anomaly score combines both the logarithm of D_M and its z-score, enhancing sensitivity to deviations from the Standard Model manifold.

Table 1: AUC values for the combined loss across different signal samples. Higher values indicate better separation from the Standard Model background.

Category	Signal	AUC	Signal	AUC	Signal	AUC
SUSY	TT tN1	0.9906	GGM HinoZh	0.9772	TT higgsino RPV	0.9961
SUSY	GG ttn1	0.9968	TT RPVdirectBL	0.9986	SS onestepCC	0.9960
SUSY	GG direct	0.9977	Stau Stau	0.8724	C1N2 WZ	0.9646
SUSY	SM N2C1p	0.7927	GGM N1N2C1	0.9915	–	–
Exotic	Zprime tautau	0.9960	Wprime qq	0.9967	Wprime taunu	0.9885
Exotic	Zprime ee	0.9998	Zprime bb	0.9971	Zprime mumu	0.9980
Exotic	Wprime enu	0.9987	DM A	0.9634	Wprime munu	0.9949
Exotic	Zprime tt	0.9968	LQd	0.9979	WpL tblep	0.9964
Exotic	WpL tbhad	0.9970	DM 4-top scalar	0.9899	–	–
Higgs	VBF H→bb	0.7942	ggH gamgam	0.8870	ggH ZZ4l	0.7766
Higgs	ttH dilep	0.9358	ttH gamgam	0.9680	VBFH gamgam	0.9162

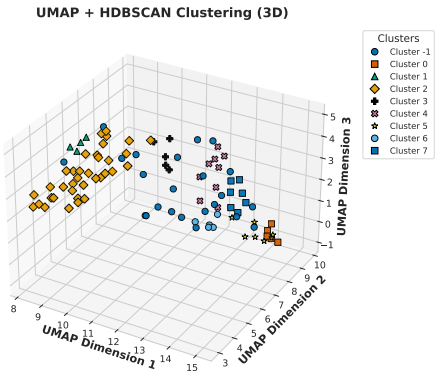
4. Benchmarking on BSM Signals

We benchmark the model on several simulated signal hypotheses available from ATLAS Open data, including representative super-symmetric scenarios (e.g. RPV gluino production), exotic resonances (e.g. heavy Z' bosons), non-standard Higgs decays, while electroweak and top samples are used to validate background reconstruction fidelity. Each signal sample is mixed with RUN 2 open data of 10 million events and passed through the trained model. We then produce anomaly score distribution (Figure. 1b) and receiver operating characteristic (ROC) curves for SM vs each signal Table. 1), using the anomaly score. In practice, we observe substantial discrimination for signals with distinct kinematics, with area-under-curve (AUC) values up to ~0.9 for clear signatures. These results are consistent with prior studies of unsupervised autoencoders in jet and event classification [10, 14, 21], validating the Tab-Transformer-based MAE as a promising model-agnostic search tool.

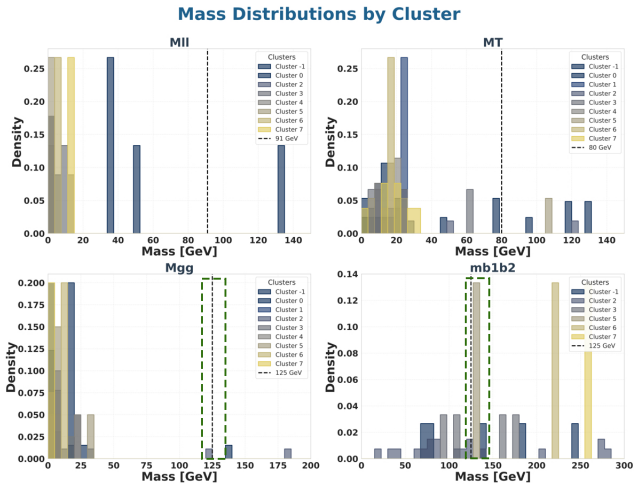
5. Clustering Analysis and Physics Interpretation

To interpret the anomalous events, we applied an unsupervised clustering workflow. The high-dimensional embeddings were first reduced to 2D/3D using UMAP [16], which preserves both local and global structure. We then performed HDBSCAN clustering [17] on the reduced space to automatically identify dense clusters and label noise points. For each cluster, we evaluated its statistical behaviour using the p-value studies in Table 2, and extracted the most discriminating variables using Cohen's d method [18]. These variables often correspond to physically meaningful event topologies. Figure 2a shows the 3D UMAP distribution with HDBSCAN cluster assignments. We additionally inspected mass-related variables: di-lepton, di-photon, di-$b$jet, and m_T, focusing on regions near the Higgs and W boson masses (Figure. 2b). Among the 150 most anomalous events selected from 10 million, we found a cluster (id: cluster 5) with 7 events where 3 events exhibited di-$b$jet masses close to the Higgs boson mass, and one event from Cluster 2 appeared near the Higgs

region in $m_{\gamma\gamma}$. These localized structures highlight specific regions of phase space (green box) that could motivate targeted follow-up studies. Cohen-d confirmed the variables driving cluster separation from the background.



(a) 3D UMAP embedding of events with HDBSCAN clustering. Each color represents a cluster, showing the structure of the data in low-dimensional space.



(b) Mass distributions of key physics observables for each cluster, highlighting differences in kinematics among event clusters.

Figure 2: Exploratory analysis of events using clustering and reconstructed features. (a) shows the 3D UMAP projection with HDBSCAN clusters. (b) presents the distributions of selected mass variables within each cluster.

Table 2: Cluster-level statistics using the Bonferroni global p-value.

Cluster	n_{ev}	S_{obs}	p_{local}	$p^{\text{Bonf}}_{\text{global}}$	**Cluster**	n_{ev}	S_{obs}	p_{local}	$p^{\text{Bonf}}_{\text{global}}$
2	36	166.55	0.0001	0.0008	5	7	31.53	0.0001	0.0008
4	9	40.48	0.0001	0.0008	6	5	23.49	0.0001	0.0008
7	7	32.48	0.0001	0.0008	3	5	23.33	0.0001	0.0008
0	5	23.06	0.0001	0.0008	1	4	17.98	0.0001	0.0008

6. Summary

We present an unsupervised anomaly detection pipeline for LHC data using a Tab-Transformer masked autoencoder, which learns SM event structure and identifies outliers without signal-specific training. Key findings include:

- The Tab-Transformer autoencoder handles mixed-feature inputs, producing rich embeddings [6].
- Mahalanobis-based anomaly scores flag signal events; and the combinations yields optimal ROC.
- Clustering anomalous events with UMAP + HDBSCAN uncovers distinct physics topologies, interpretable via Cohen's d [18].

Future work includes utilizing the full available Run−2 dataset, expanding to other datasets, and performing hyperparameter tuning to enhance sensitivity to rare and BSM events, further improving AI-assisted searches for unexpected LHC signals.

References

[1] ATLAS Collaboration, *ATLAS Open Data: Proton–Proton Collisions at* $\sqrt{s}$ = 13 *TeV (2015–2016)*, DOI:10.7483/OPENDATA.ATLAS.9HK7.P5SI.

[2] R. Brun and F. Rademakers, *ROOT – An Object-Oriented Data Analysis Framework*, *Nucl. Instrum. Meth. A* **389** (1997) 81–86.

[3] CERN SWAN Platform, https://swan.cern.ch.

[4] D. Thain, T. Tannenbaum, M. Livny, *Distributed Computing with Condor*, *Concurrency Comput. Pract. Exper.* **17** (2005) 323–356.

[5] M. Rocklin, *Dask: Parallel Computation with Blocked Algorithms and Task Scheduling*, *Proceedings of the 14th Python in Science Conference* (2015) 130–136.

[6] S. Huang et al., *TabTransformer: Tabular Data Modeling Using Contextual Embeddings*, AAAI (2021), arXiv:2012.06678.

[7] D. Niewenhuis et al., "RBatchGenerator_TensorFlow.py — example for loading batches from a ROOT dataset into a TensorFlow workflow," ROOT Reference Guide (v6-32), available at https://root.cern/doc/v632/RBatchGenerator__TensorFlow_8py.html, accessed 27 Nov 2025.

[8] A. Vaswani et al., *Attention Is All You Need*, arXiv:1706.03762 (2017).

[9] P. C. Mahalanobis, *On the Generalized Distance in Statistics*, *Proc. Natl. Inst. Sci. India* **2** (1936) 49–55.

[10] O. Cerri et al., *Variational Autoencoders for New Physics Mining at the LHC*, *JHEP* **05** (2019) 036.

[11] B. Nachman, *Anomaly Detection for New Physics*, *Rev. Phys.* **7** (2022) 100091.

[12] T. Heimel, G. Kasieczka, T. Plehn, J. Thompson, *QCD or What?*, *SciPost Phys.* **6** (2019) 030.

[13] ATLAS Collaboration, *Search for Anomalies with Autoencoders in LHC Run-2 Data*, ATLAS-CONF-2022-049, https://cds.cern.ch/record/2836746.

[14] CMS Collaboration, *Anomaly Detection for New Physics Searches in Di-Jet Events*, CMS-EXO-22-026 (2023), https://cds.cern.ch/record/2863945.

[15] CMS Collaboration, *Anomaly Detection for ECAL Data Quality Monitoring*, CMS-DP-2021-043 (2021), https://cds.cern.ch/record/2779069.

[16] L. McInnes, J. Healy, J. Melville, *UMAP: Uniform Manifold Approximation and Projection*, arXiv:1802.03426.

[17] R. J. G. B. Campello et al., *Density-Based Clustering Based on Hierarchical Density Estimates*, Springer (2013).

[18] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edition, Lawrence Erlbaum Associates (1988).

[19] L. van der Maaten, G. Hinton, *Visualizing Data Using t-SNE*, *J. Mach. Learn. Res.* **9** (2008) 2579–2605.

[20] R. Kansal et al., *Event Generation and Anomaly Detection with RMMs*, *JINST* **17** (2022) P07035.

[21] M. Romão, N. F. Castro, R. Pedro, *Machine Learning for Anomaly Detection in Particle Physics*, *Rev. Phys.* **7** (2022) 100091.

[22] A. De Simone, T. Jacques, *Rapidity–Mass Matrices for Event Classification Problems in HEP*, *Comput. Phys. Commun.* **254** (2020) 107258.