Anomaly Detection to Search for New Hadrons at BESIII

Yuyao Xu, Ke Li, Jingde Chen, Minggang Zhao

2026.4.15

Nankai

Introduction

Motivation:

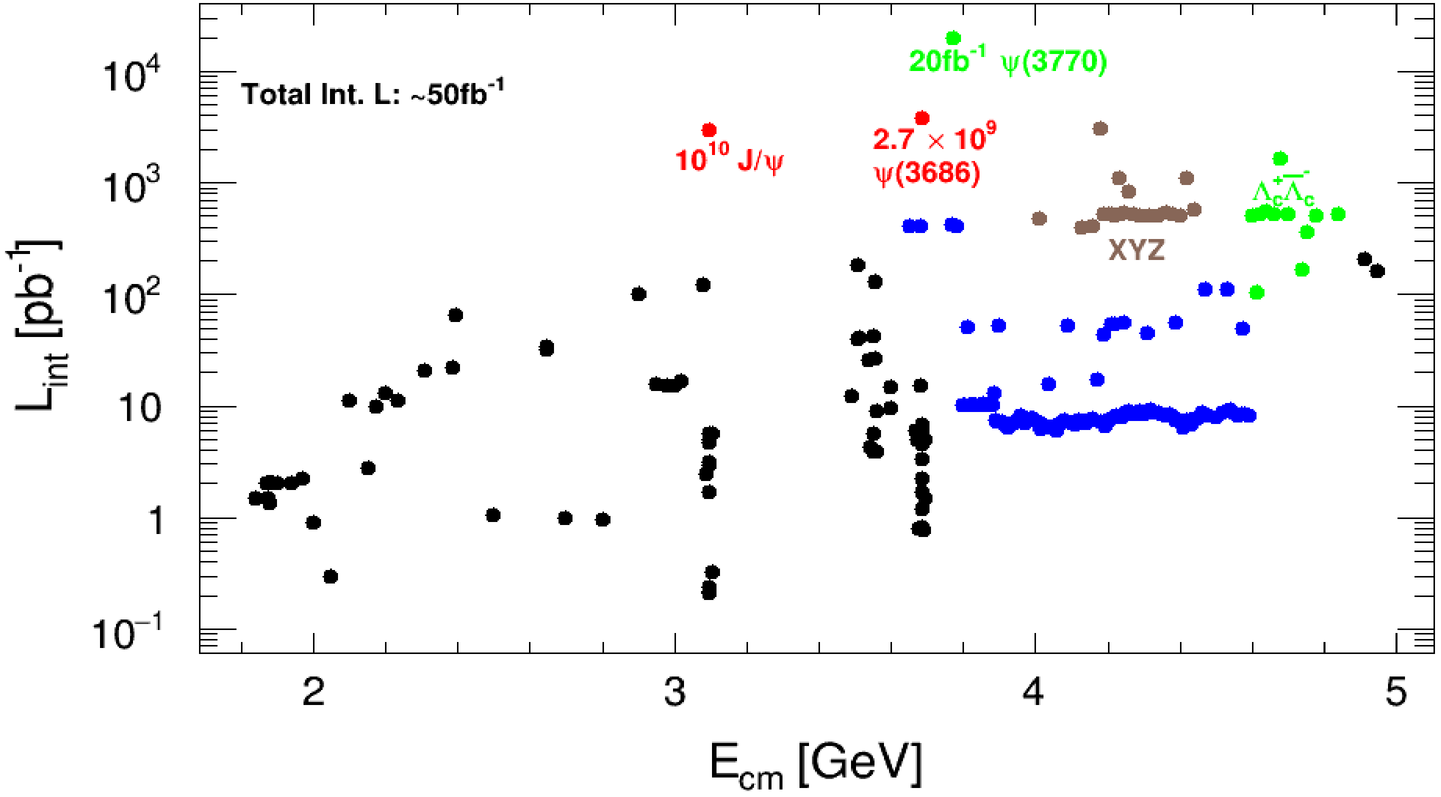
- To find unpredicted, rare signals in vast BESIII data.

Method:

- Learn normal patterns from data.
- Identify deviations as potential signals.

Traditional anomaly detection:

- Typically applied within a single dataset.
- Focus on identifying rare events.
- Limited generalization to new conditions.

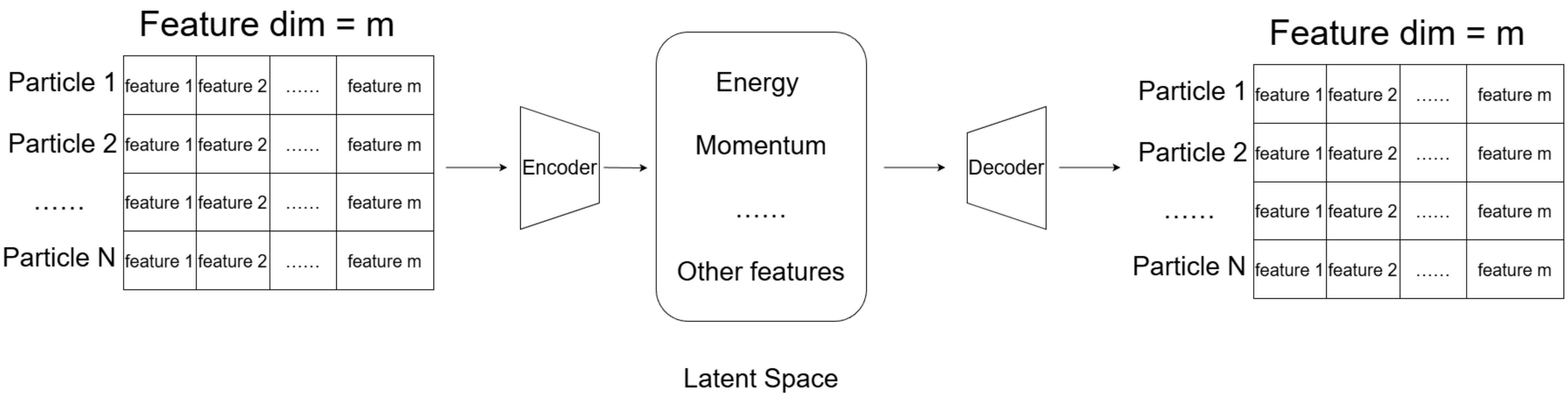


Introduction

Method	*Anomaly Score*	*Example Papers*
AE	***MSE***	*Autoencoders for Anomaly Detection in High Energy Physics; A Study of Unsupervised Anomaly Detection with Autoencoders*
Flow	$-\log p(x)$	*Anomaly Detection with Normalizing Flows; Generator-Based Inference for Anomaly Detection in Particle Physics*
CWoLa	*classifier output*	*Classification Without Labels (CWoLa) for High Energy Physics; TRANSIT: Transport-based Anomaly Detection in Resonant Searches*
Transformer	*learned score*	*DepthViT: Vision Transformer for Anomaly Detection; Transformer-based Anomaly Detection for High Energy Physics*
Latent Density	$-\log p(z)$	*Anomaly Detection with Tensor Networks in High Energy Physics*

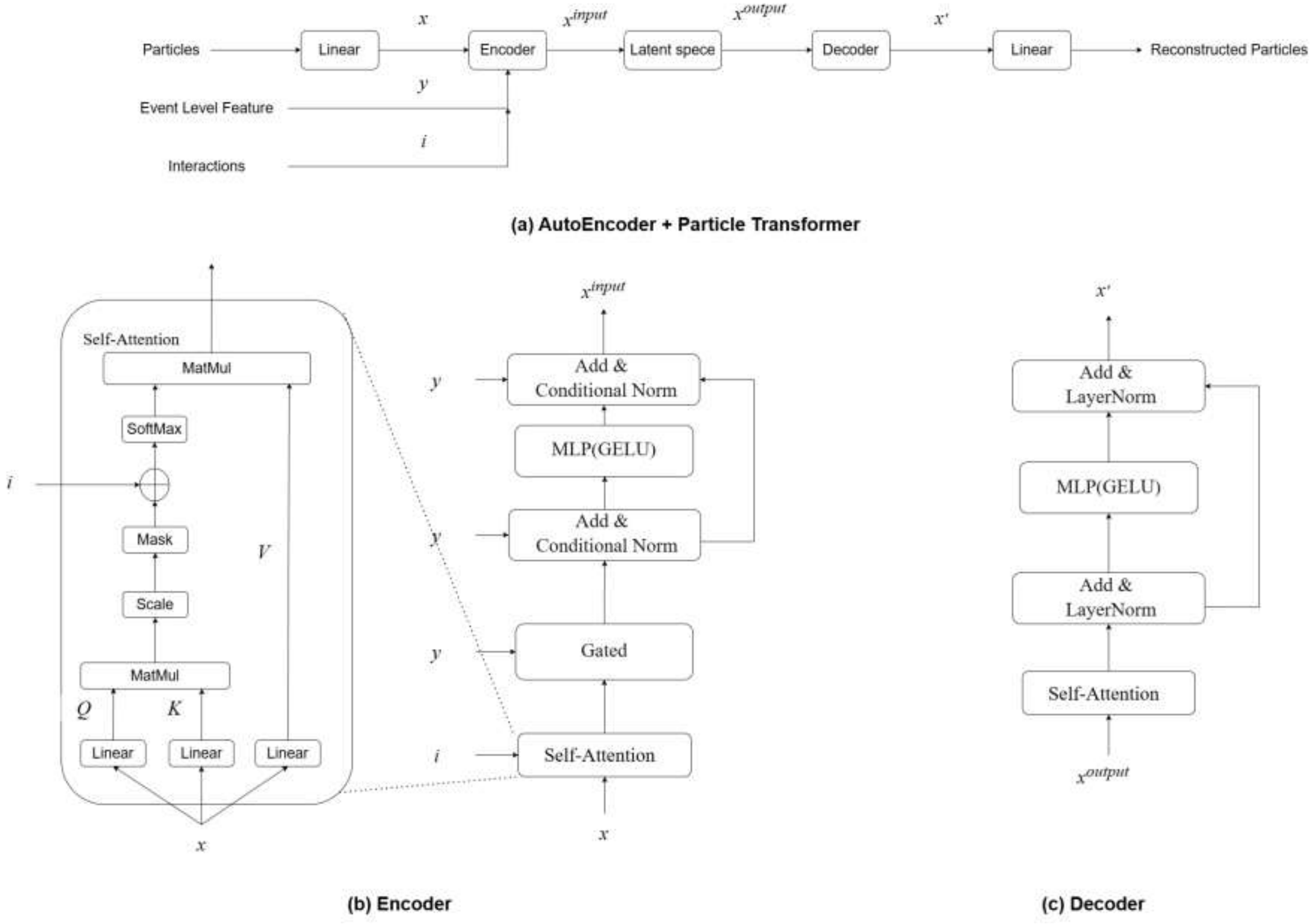
Limitation:
All methods operate within a single dataset.
Lack generalization across different datasets.

Method: AE



Feature dim = m
Particle 1
Particle 2
……
Particle N
feature 1
feature 2
……
feature m
Encoder
Energy
Momentum
……
Other features
Latent Space
Decoder
Feature dim = m
Particle 1
Particle 2
……
Particle N
feature 1
feature 2
……
feature m

Method: AE+Transformer



(a) AutoEncoder + Particle Transformer

(b) Encoder

(c) Decoder

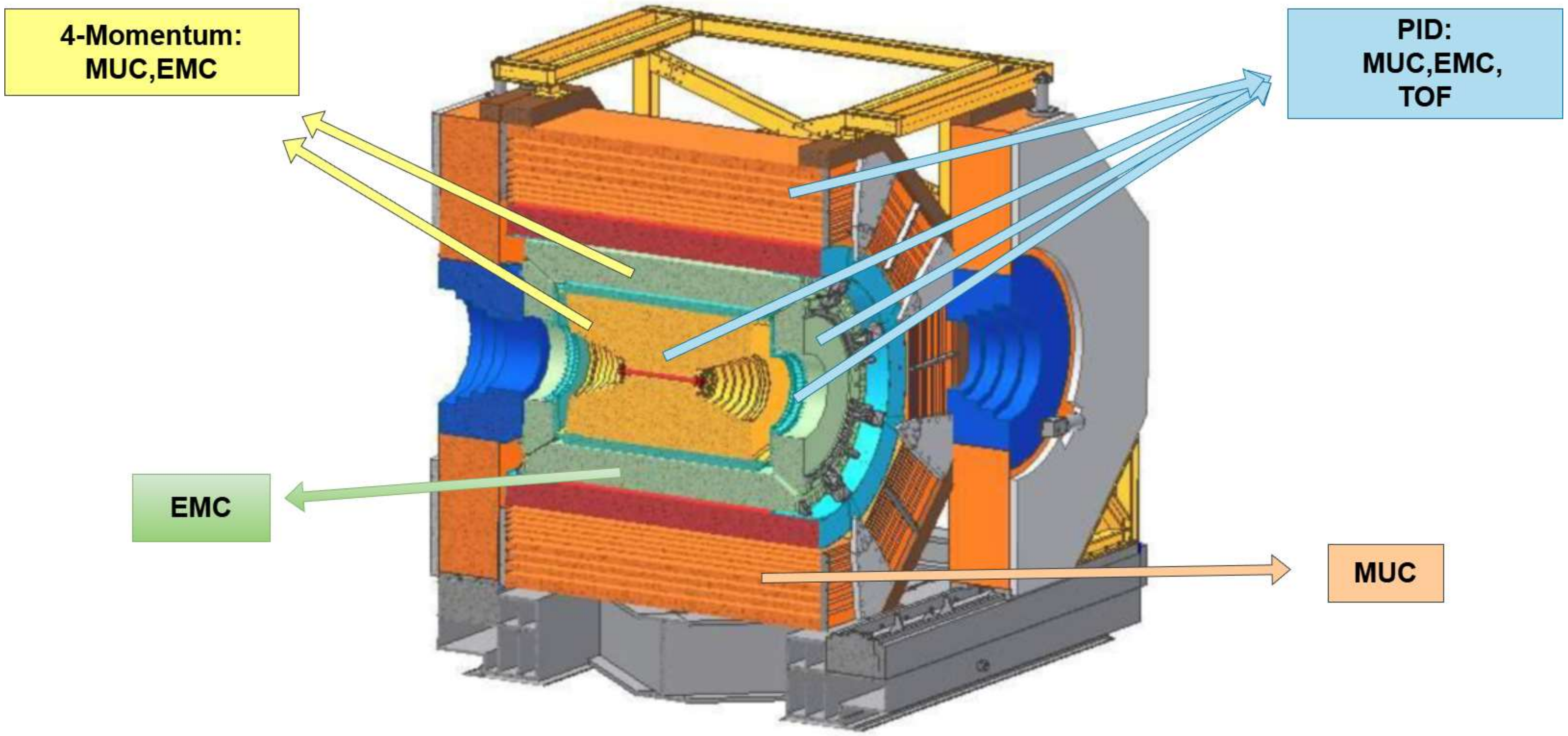
Innovation:

·Particle Transformer

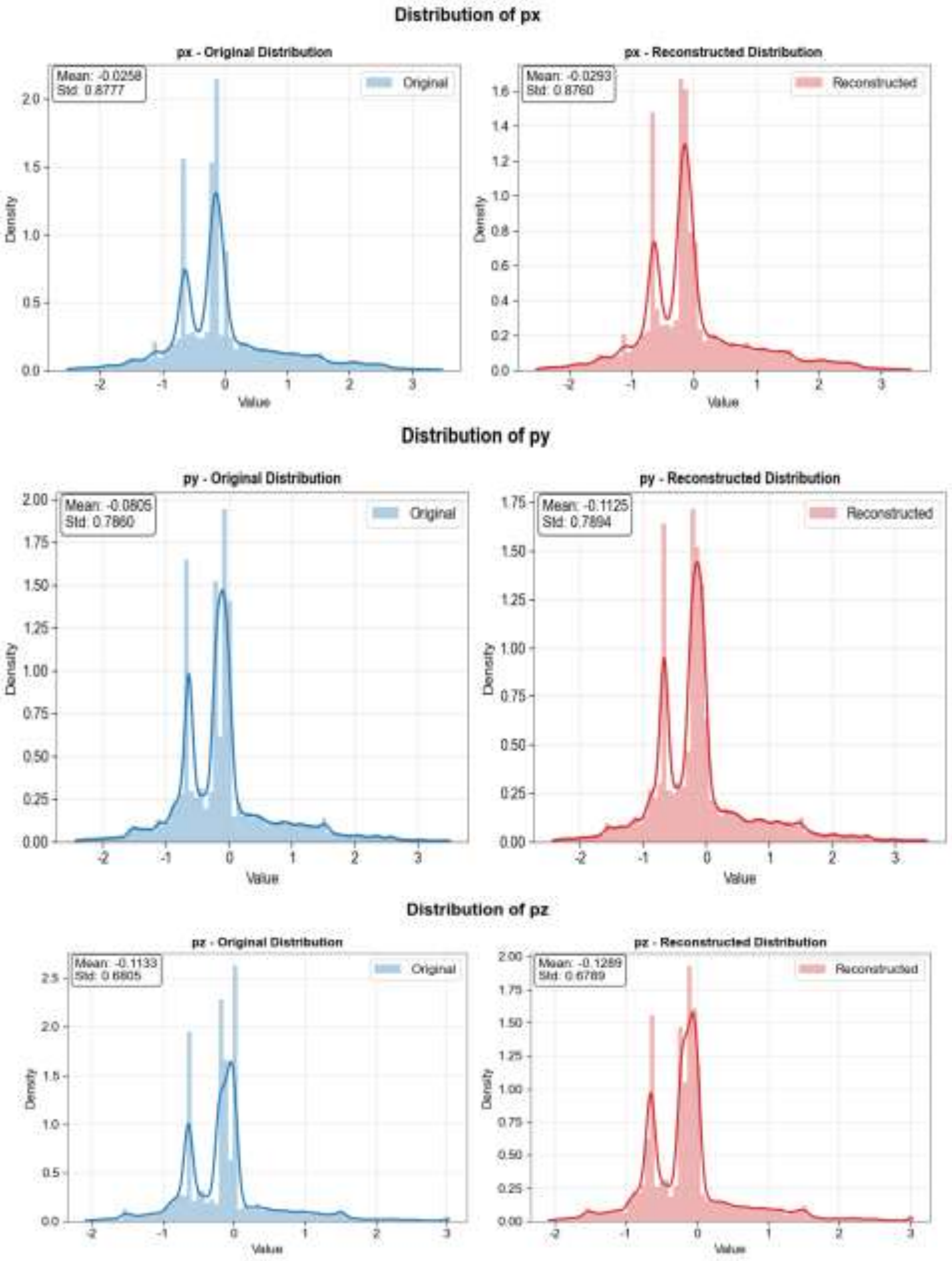
·Gated attention

·Conditional Layernorm

Data: Boss 7.0.3

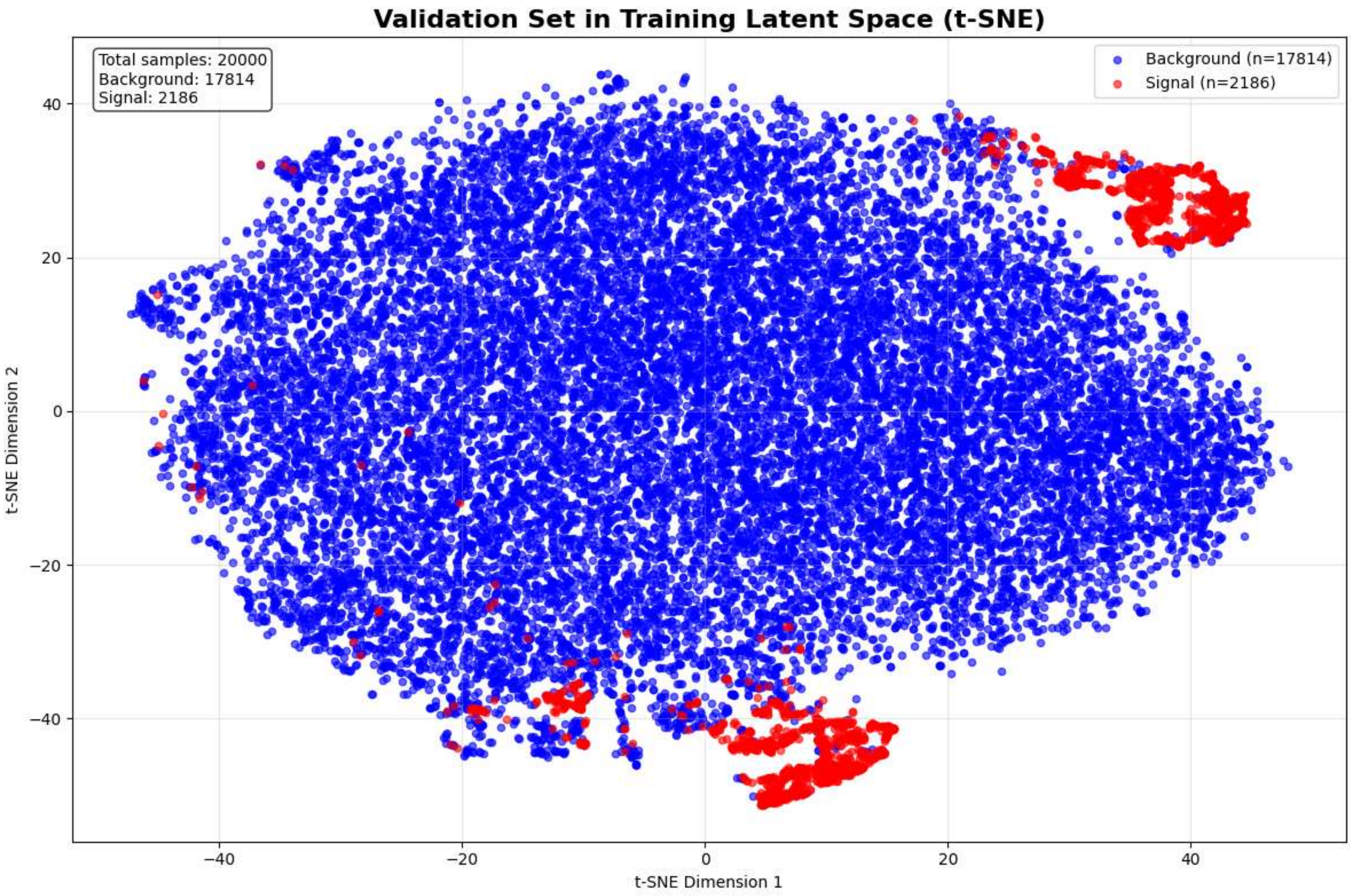


Loss: MSE



$$Loss = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_i^{'})^2$$

Anomaly Score

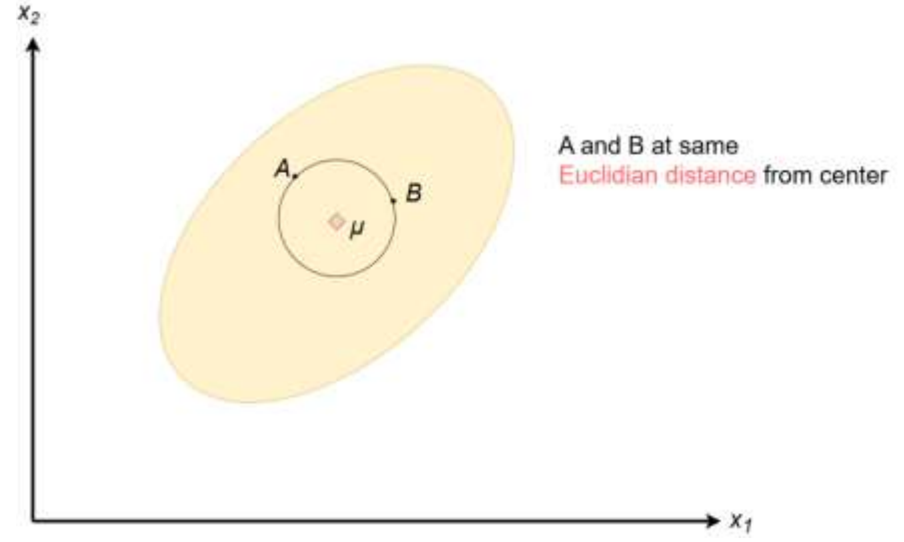


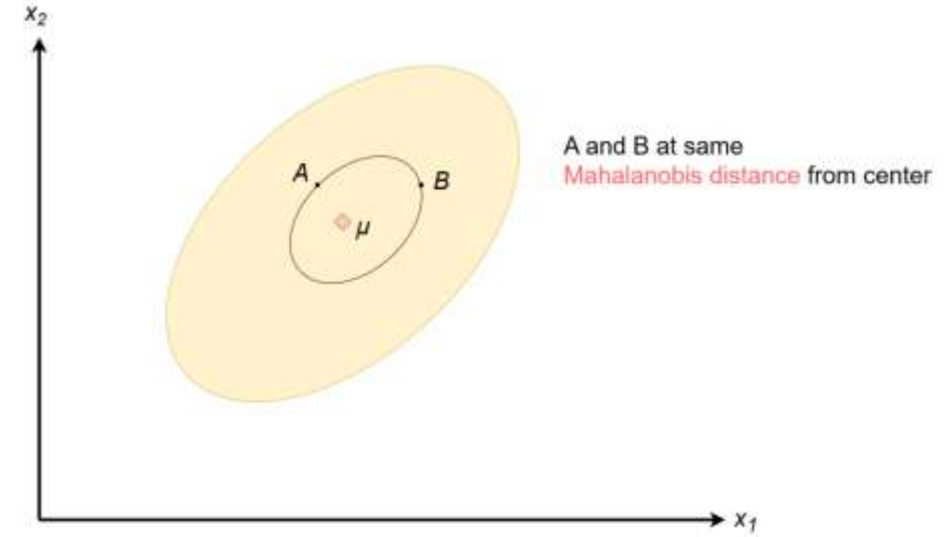
Mahalanobis Distance:
Mahalanobis distance is a statistical measure that accounts for the covariance structure of data, measuring the standardized distance between a point and a distribution.

$$D^2 = (x - \mu)^T \Sigma^{-1} (x - \mu)$$

μ: *Training set mean vector*
Σ: *Training set covariance matrix*

Anomaly Score





Mahalanobis Distance:

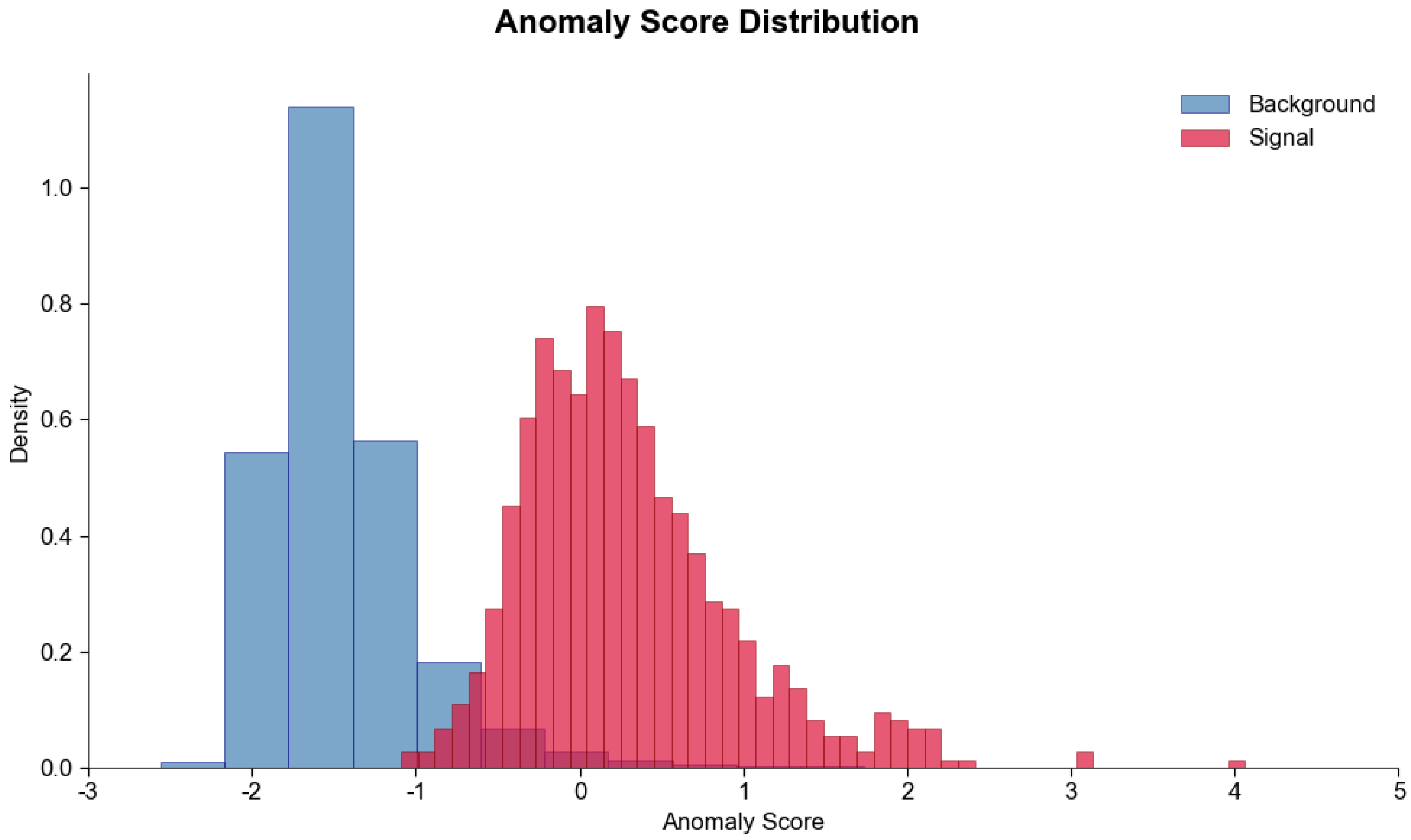
$$D^2 = (x-\mu)^T \Sigma^{-1} (x-\mu)$$

μ: *Training set mean vector*

Σ: *Training set covariance matrix*

	Euclidean	Mahalanobis
Core Concept	Straight-line distance	Statistical distance
Variable Correlation	Ignored	Adjusted via covariance matrix
Scale Sensitivity	Sensitive	Insensitive
Data Distribution Assumption	Sensitive	Anisotropic

Anomaly Score

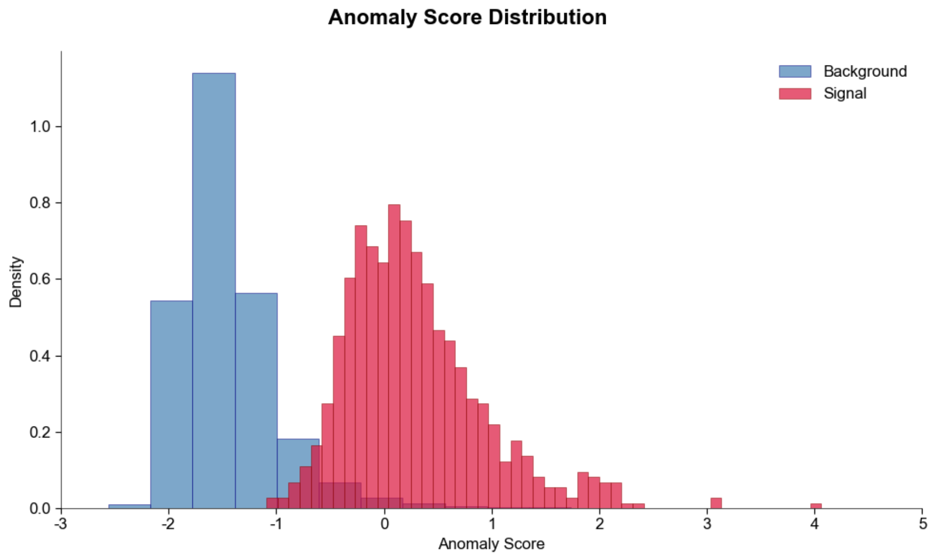


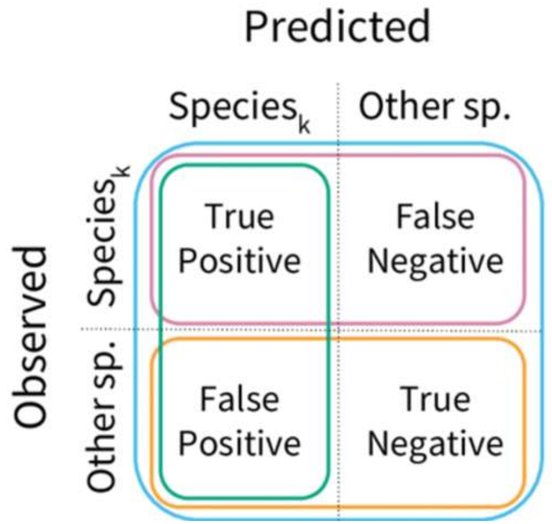
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_i^{'})^2$$

$$D^2 = (x-\mu)^T \Sigma^{-1} (x-\mu)$$

$$Anomaly\ Score = \ln MSE \times D$$

Performance Evaluation



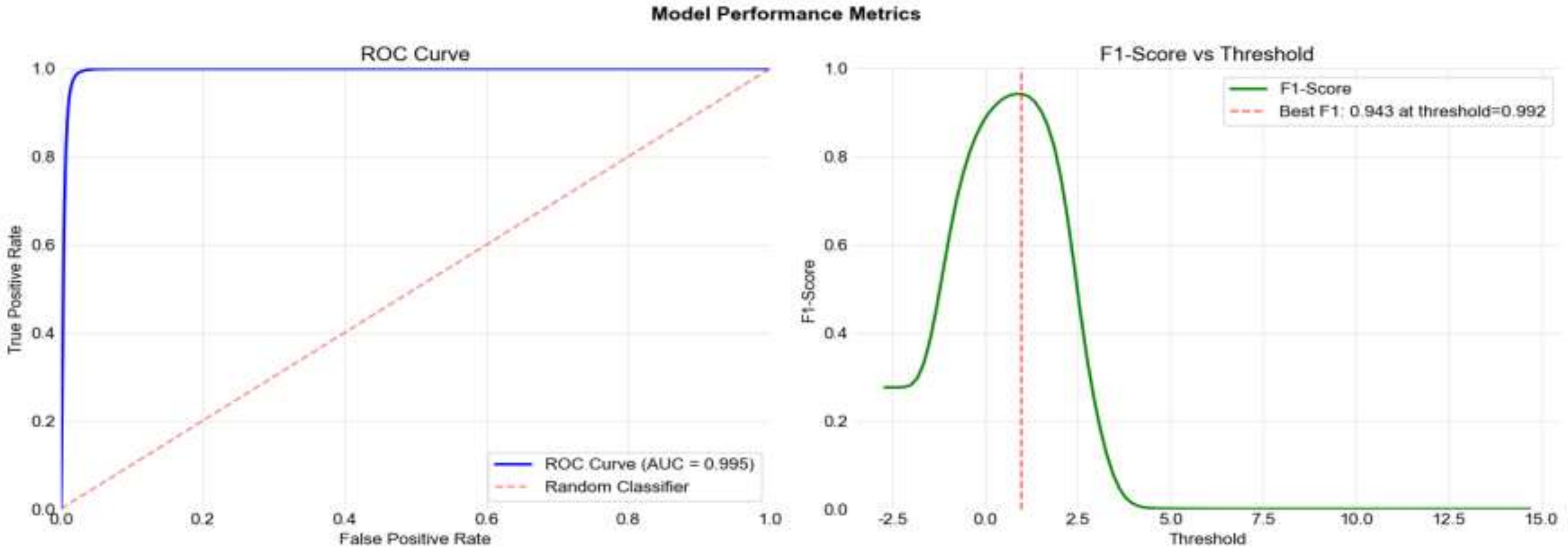


$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$



$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Training Plan

We currently do not use actual experimental data.
We only use MC to replace real data.

Stage 1
Baseline test

Stage 2
Generalization Ability

Stage 3
Anomaly detection

Stage 4
Cross-energy Capability

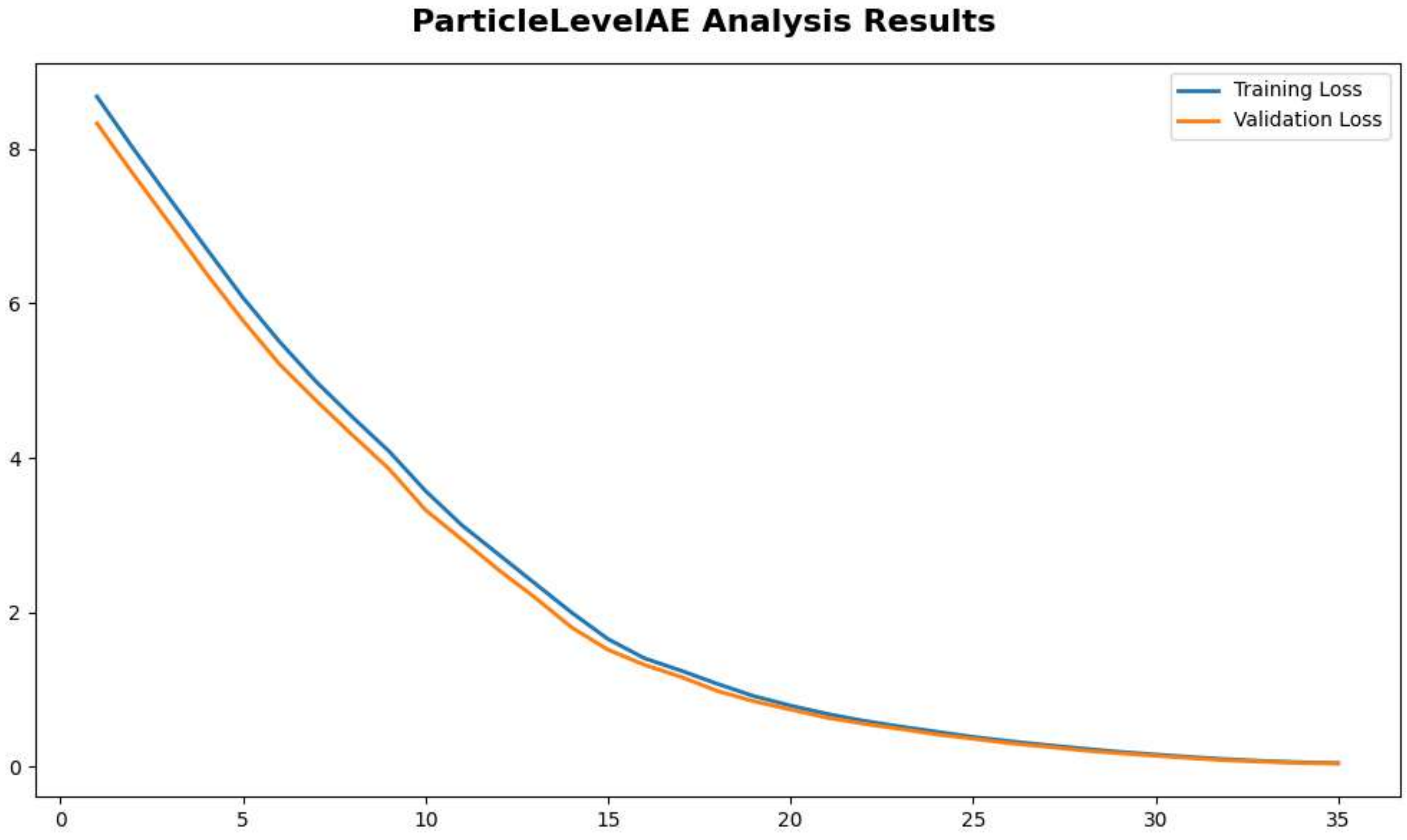
Stage 5
Method Validation

Results

Stage 1 : Baseline Test

Train Set: 500000 4260 inclusive MC

Test Set : 500000 4260 inclusive MC +
1000 X(3872) signal MC



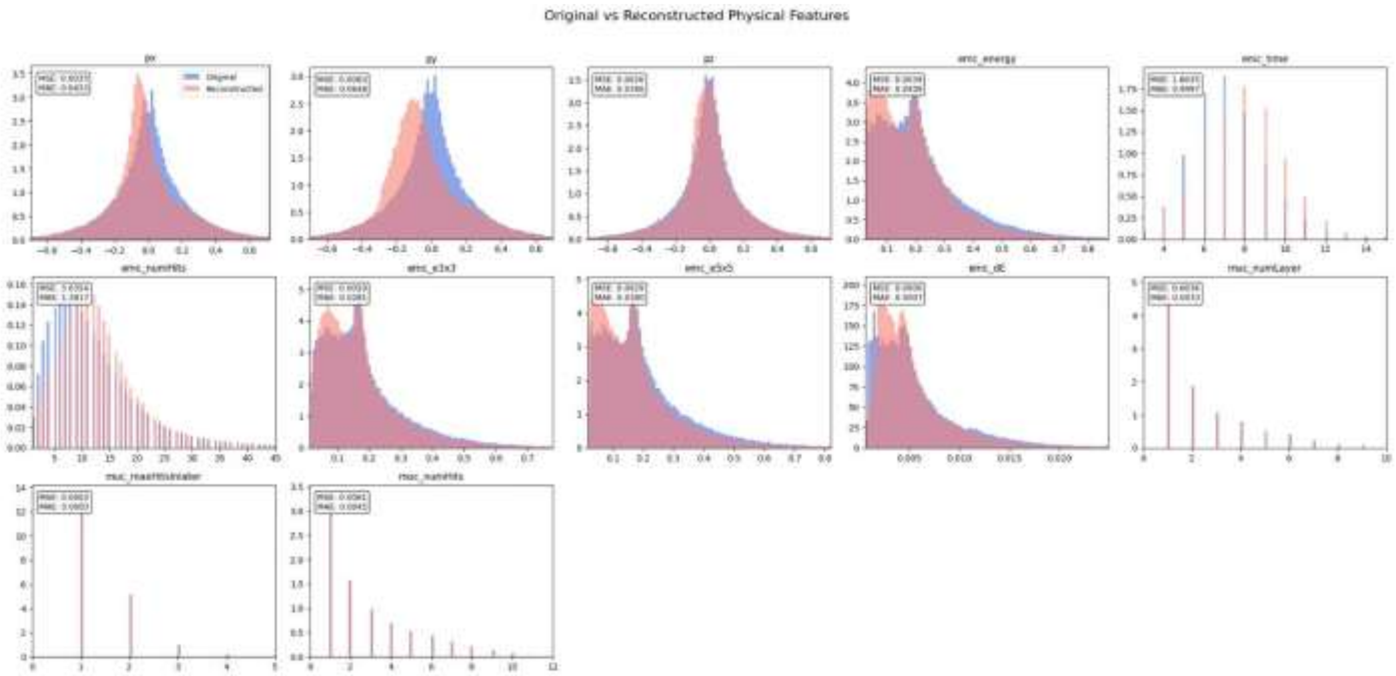
Tips:
The MC in the train set and the MC in the test set are different.
All the following data is like this.

Results

Stage 1 : Baseline Test

Train Set: 500000 4260 inclusive MC

Test Set : 500000 4260 inclusive MC +
1000 X(3872) signal MC

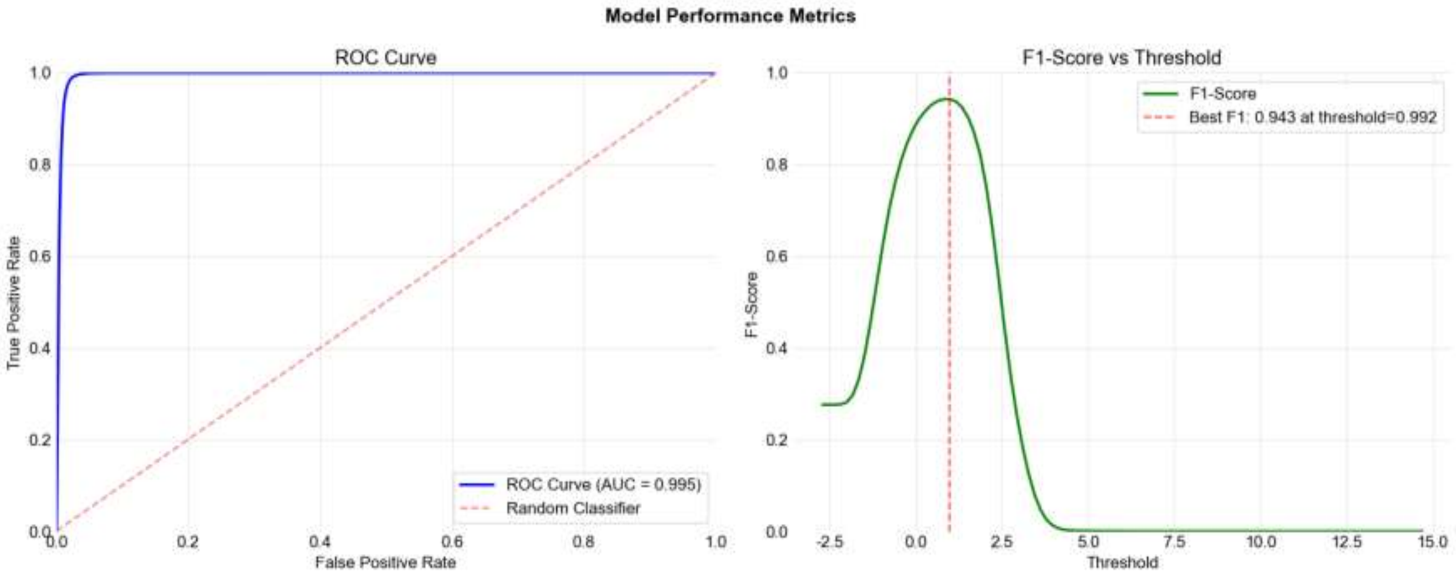


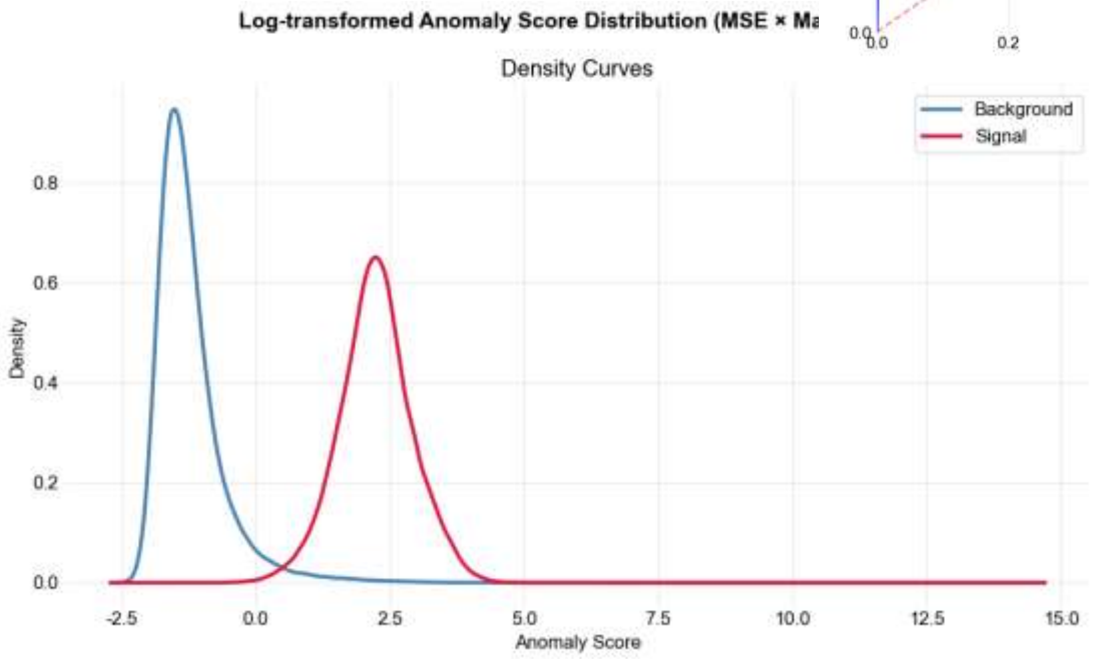
Results

Stage 1 : Baseline Test

Train Set: 500000 4260 inclusive MC

Test Set : 500000 4260 inclusive MC +
1000 X(3872) signal MC





Conclusion:

If threshold = 0.992,

TP = 94.3%, TN = 99.989%

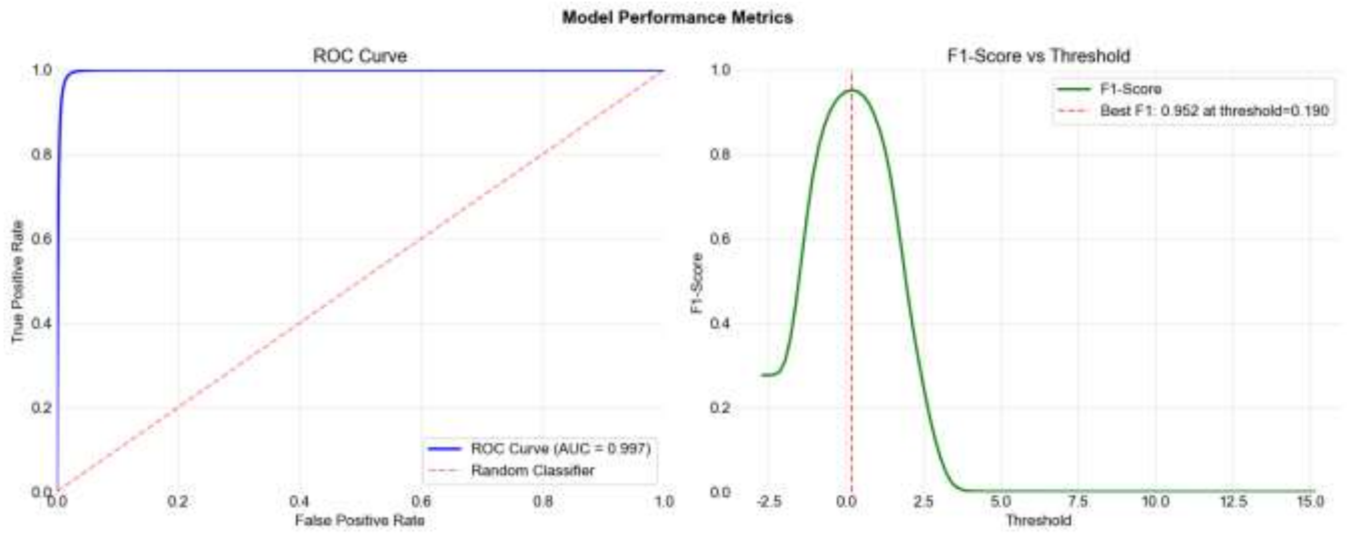
Train on a single background, detect a known signal.
Validate basic anomaly detection capability.

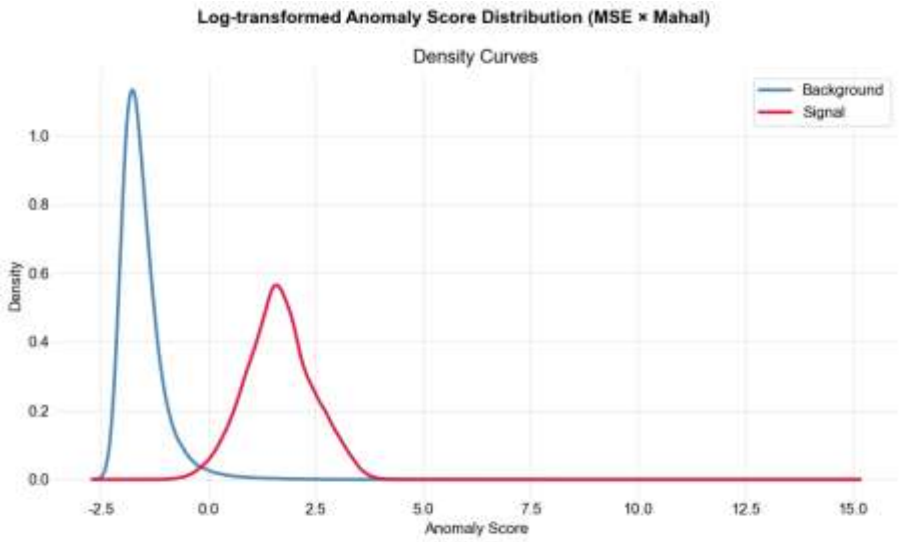
Results

Stage 2 : Generalization Ability
①Background Generalization

Train Set: 500000 4260 inclusive MC
500000 4230 inclusive MC

Test Set : 500000 4260 inclusive MC +
1000 X(3872) signal MC





Conclusion:
If threshold = 0.190,
TP = 96.0%, TN = 99.986%
Train on mixed backgrounds, detect a known signal.
Test model's robustness to complex backgrounds.

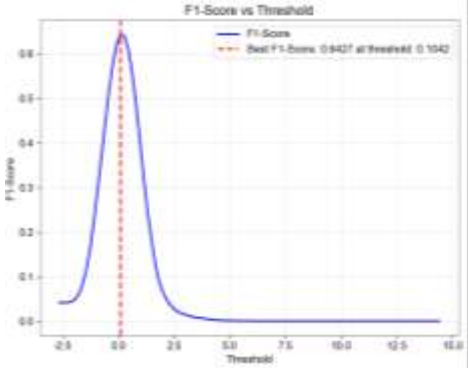
Results

Stage 2 : Generalization Ability
②Signal Generalization

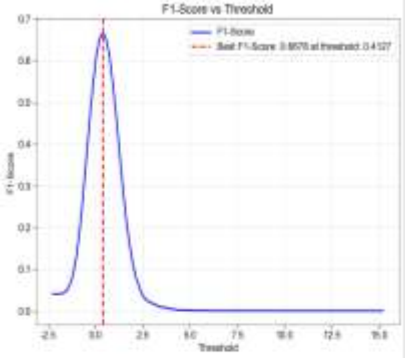
Train Set: 500000 4260 inclusive MC
500000 4230 inclusive MC

Test Set : 500000 4260 inclusive MC +
1000 X(3872) signal MC

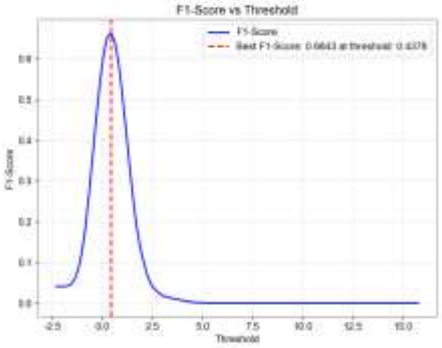
The Signal MC data for 9 groups of X particles exhibit varying masses and widths.



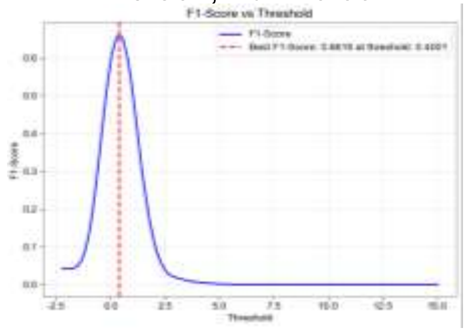
M=3.862,width=0.002



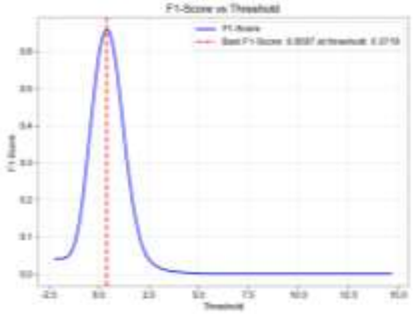
M=3.862,width=0.003



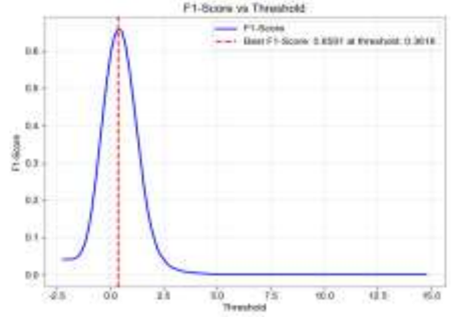
M=3.862,width=0.004



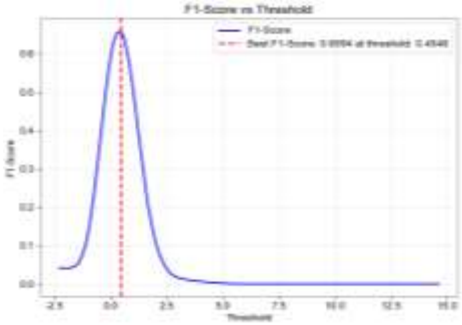
M=3.872,width=0.002



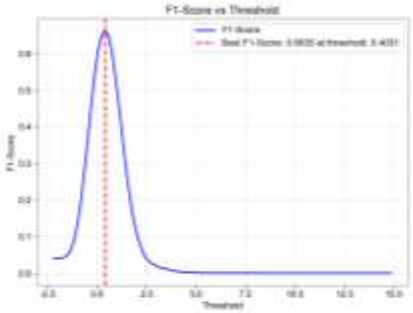
M=3.872,width=0.003



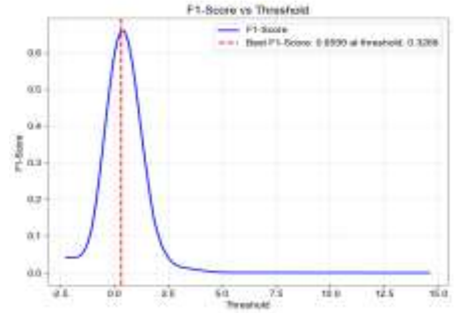
M=3.872,width=0.004



M=3.882,width=0.002



M=3.882,width=0.003



M=3.882,width=0.004

Results

Stage 2 : Generalization Ability

② Signal Generalization

Train Set: 500000 4260 inclusive MC
500000 4230 inclusive MC

Test Set : 500000 4260 inclusive MC +
1000 X(3872) signal MC

The Signal MC data for 9 groups of X particles exhibit varying masses and widths.

	Threshold	TP	TN
M=3.862,width=0.002	0.1042	91.0%	99.825%
M=3.862,width=0.003	0.4127	89.2%	99.883%
M=3.862,width=0.004	0.4378	86.3%	99.883%
M=3.872,width=0.002	0.4001	85.9%	99.885%
M=3.872,width=0.003	0.3719	85.7%	99.872%
M=3.872,width=0.004	0.3616	86.2%	99.873%
M=3.882,width=0.002	0.4548	85.5%	99.875%
M=3.882,width=0.003	0.4031	85.5%	99.871%
M=3.882,width=0.004	0.3268	84.3%	99.875%

Conclusion:

Train on the same mixed backgrounds, detect signals with varying parameters. Assess model's capability to explore new physics.

Results

Stage 3: Anomaly Detection

Train Set: ①+②+④

Test Set : ③

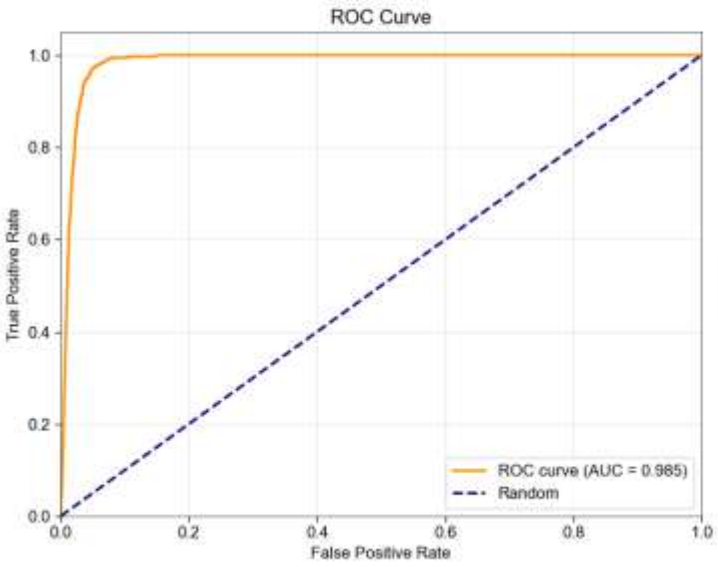
Data:

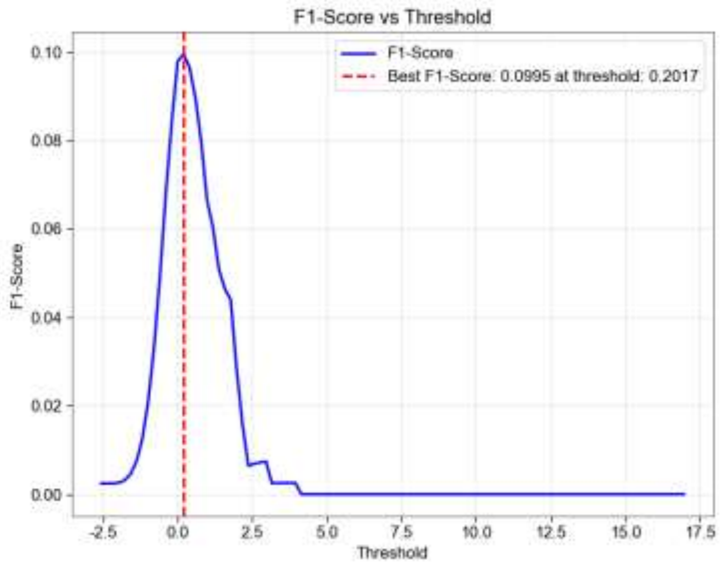
①500000 4230 inclusive MC

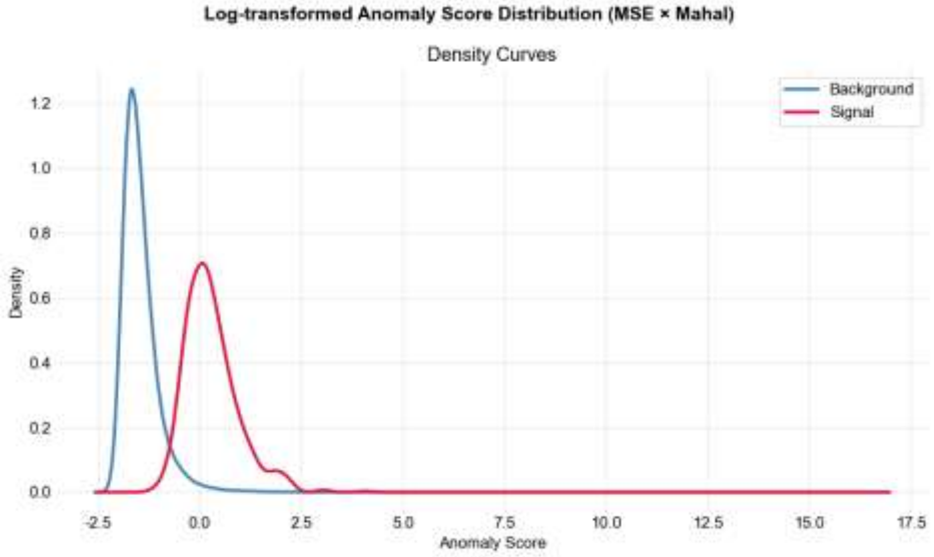
②500000 4230 inclusive MC *(Cut out 10 channels)*

③500000 4260 inclusive MC *(Signal is the 10 channels)*

④500000 4260 inclusive MC *(Cut out 10 channels)*







Conclusion:

If threshold = 0.2017,

TP = 93.2%, TN = 99.8995%

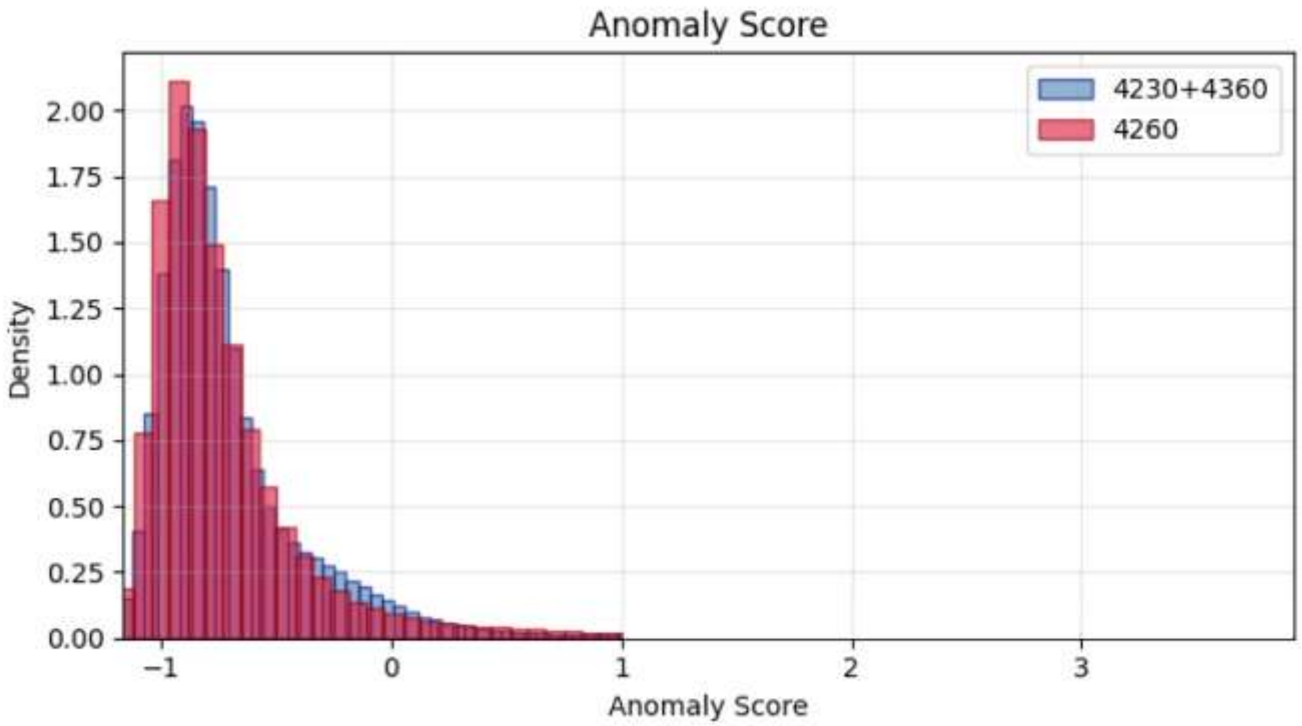
Train on datasets with artificial "missing" data, find specific local anomalies. Simulate the search for rare decay channels or data contamination.

Results

Stage 4: Cross-energy Capability

Train Set: 500000 4230 inclusive MC +
500000 4360 inclusive MC

Test Set : 500000 4260 inclusive MC



We have verified that the model exhibits good learning capabilities for data at different energy points.

Results

Stage 5: Method Validation

①Anomaly score test

Train Set: 500000 4260 inclusive MC
500000 4230 inclusive MC

Test Set : 500000 4260 inclusive MC +
1000 X(3872) signal MC

TPR	Background Rejection				
	Mse	Mse + Mahalanobis	Mse + Euclid	Mse + Manhattan	Mse + Chebyshev
70%	76.34	100.00	68.03	68.03	64.52
80%	61.35	78.13	51.55	51.28	50.51
90%	39.68	47.62	37.07	36.64	36.10

Conclusion:

Mse + Mahalanobis distance is the best anomaly score we have found so far

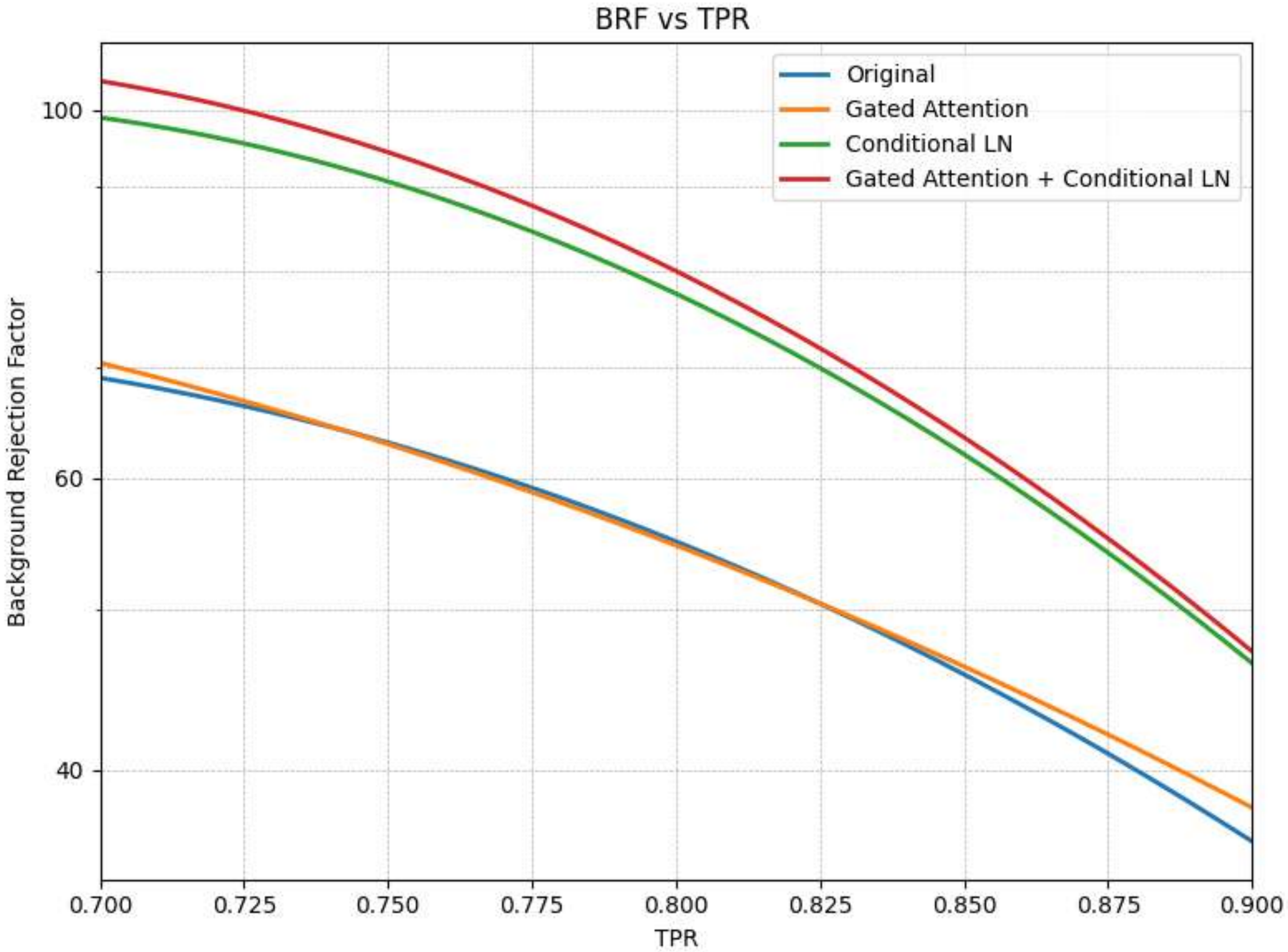
Results

Stage 5: Method Validation

②Transformer upgrade

Train Set: 500000 4260 inclusive MC
500000 4230 inclusive MC

Test Set : 500000 4260 inclusive MC +
1000 X(3872) signal MC



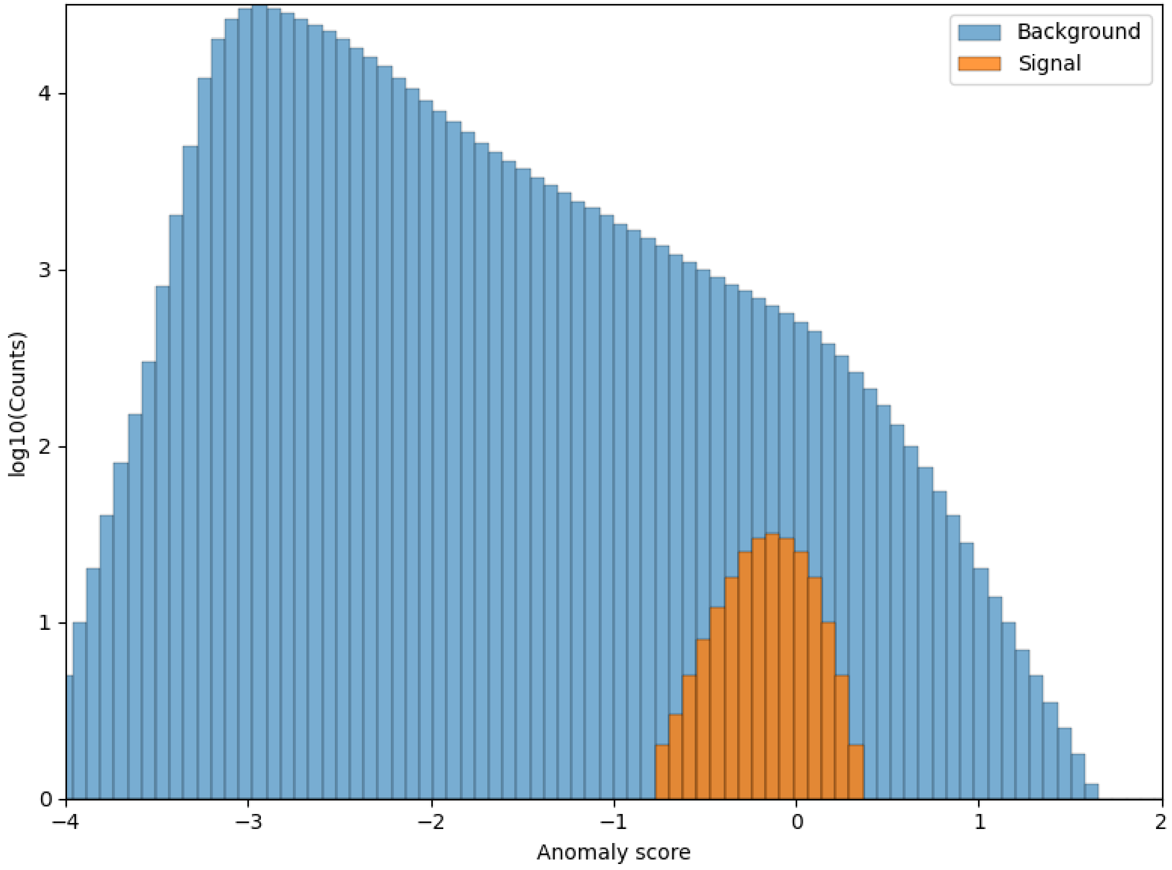
Conclusion:
The two modifications, Gated attention and Condintional layernorm, have indeed improved the performance of Transformer

Results

Stage 5: Method Validation
③Statistical Significance

Train Set: 500000 4260 inclusive MC
500000 4230 inclusive MC

Test Set : 500000 4260 inclusive MC +
3000 X(3872) signal MC



Conclusion:

For a representative configuration with 100 bins and 3000 signal events, the result is $Z=6.0\sigma$.
Previous studies on X(3872) have reported a significance of approximately 6.6σ.
Our model already possesses the ability to identify abnormal signals, but there is still room for improvement.

Summary

Completed work:
We developed a model-independent anomaly detection framework for BESIII data analysis.

Innovation :
- *Cross-energy dataset comparison*
- *Physics-aware Transformer*
- *Combined anomaly score*

Physics impact:
Enables direct search for unknown particles across different energy datasets.

Future work:
- Apply to real experimental data
- Further improve model performance

Thank you!