LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

Search for New Phenomena using Anomaly Detection in the ATLAS/LHC Experiment:

Anomaly Detection on Monotop Simulated Events

Gledson Monteiro

Supervisor:
Rute Pedro

Anomaly Detection on Monotop Simulated Events
Physics Motivation: Why Dark Matter…matters?

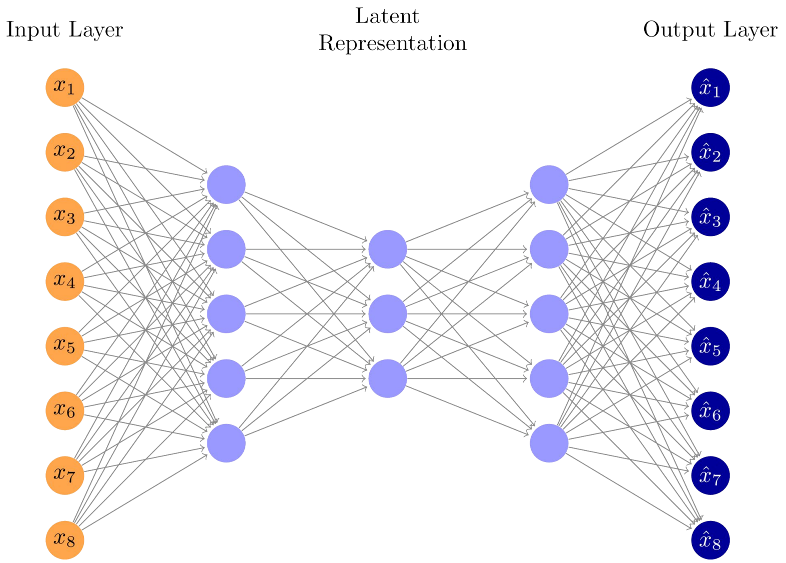
LABORATÓRIO DE INSTRUMENTAÇÃO E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

- Dark matter makes up to ~27% of the universe’s energy content but have only observed it indirectly through it’s gravitational effects: galaxy rotation curves, CMB
- Monotop signature: top quark + large MET
- This project: search for monotop events using simulated ATLAS Run 2 samples across several DM mediator benchmarks (sclar/non-resonant), with ML-based anomaly detection methods to identify signal-like deviations from background

Anomaly Detection on Monotop Simulated Events
What is Anomaly Detection?

LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

- Anomaly detection: train a neural network only on SM background
- Autoencoder: a type of NN used to compress an event to a small latent vector then reconstructs it back . Background reconstructs well, BSM events don't → **reconstruction error = anomaly score**
- Training: weighted MSE loss, Adam optimizer, background only with train/val/test = ⅓ each
- Latent dimension is the key hyperparameter



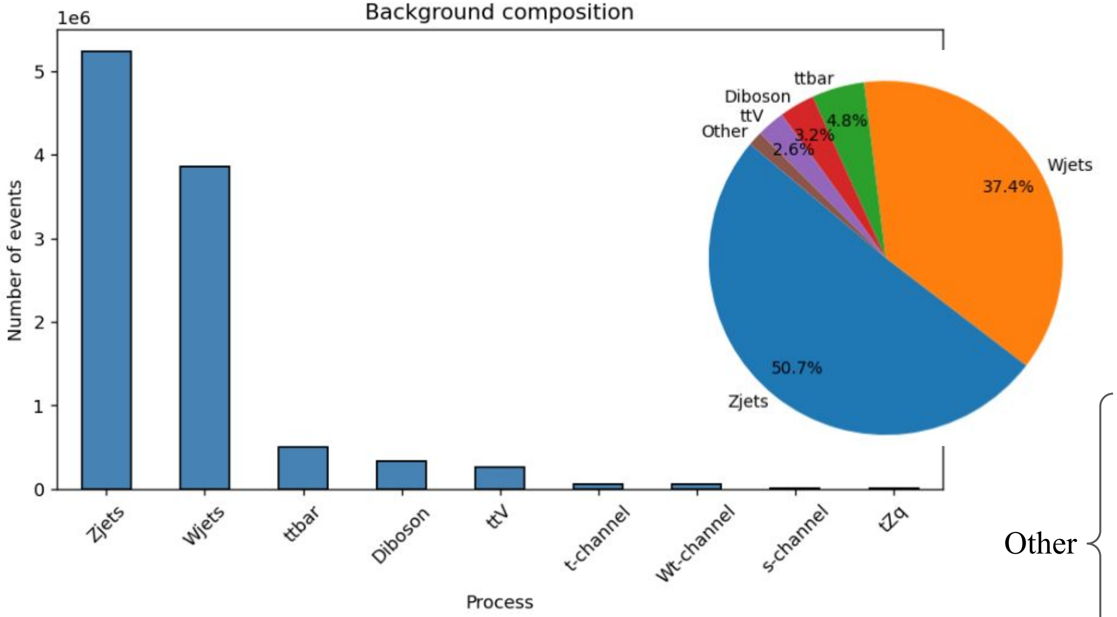
- Architecture: input space → encoder → latent space → decoder → output space

LABORATÓRIO DE INSTRUMENTAÇÃO E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

The goal of the project:

- ❏ Study the influence of pre-processing techniques and other hyper parameters on the models discriminative power and anomaly flagging capability
- ❏ Evaluate the efficiency and performance of Auto Encoders on Anomaly Detection applied to different BSM signal models related to Monotop signature

Anomaly Detection on Monotop Simulated Events
Dataset Composition: Background

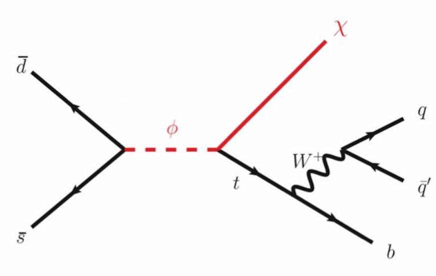


merge sample	count	percentage (%)
Zjets	5232094	50.72
Wjets	3853369	37.35
ttbar	496949	4.82
Diboson	331359	3.21
ttV	265265	2.57
t-channel	58500	0.57
Wt-channel	58311	0.57
s-channel	12872	0.12
tZq	7401	0.07

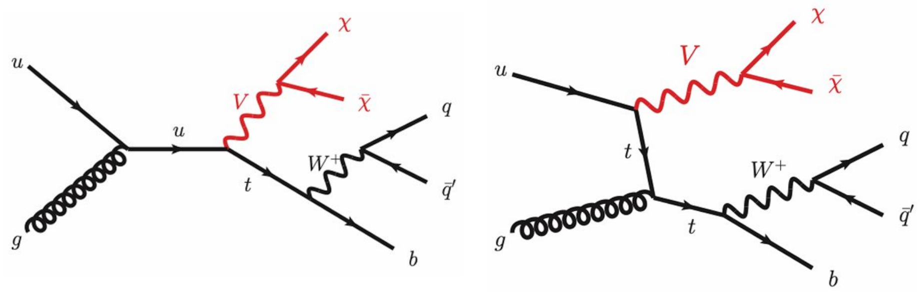
Other

Dataset Composition: Signal types

- Monotop signals: scalar/vector mediator DM + top quark
- Dark Matter via scalar mediator: Resonant (ResmMed)
- Dark Matter via vector mediator: Non-Resonant (NonResmMed)



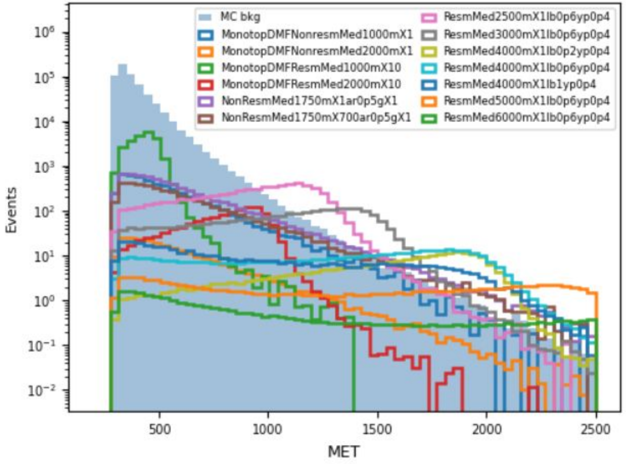
(ResmMed)

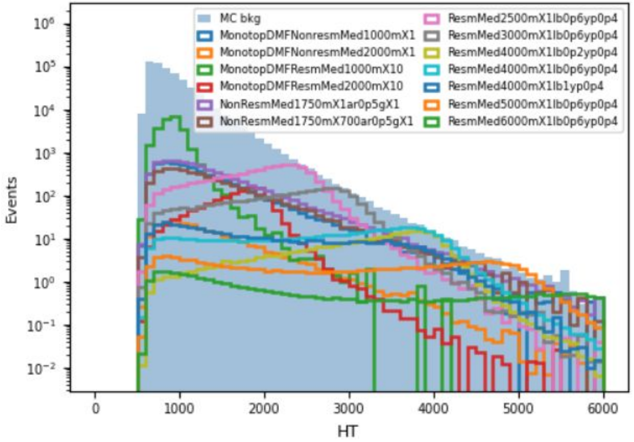


(NonResmMed)

Anomaly Detection on Monotop Simulated Events
Pre-selection cuts

- ❏ MET > 200 GeV
- ❏ nGoodJets >= 1
- ❏ ljet_pt >= 250 GeV
- ❏ DeltaPhi_MET_calojets > 0.2
- ❏ nGoodJets <= 4





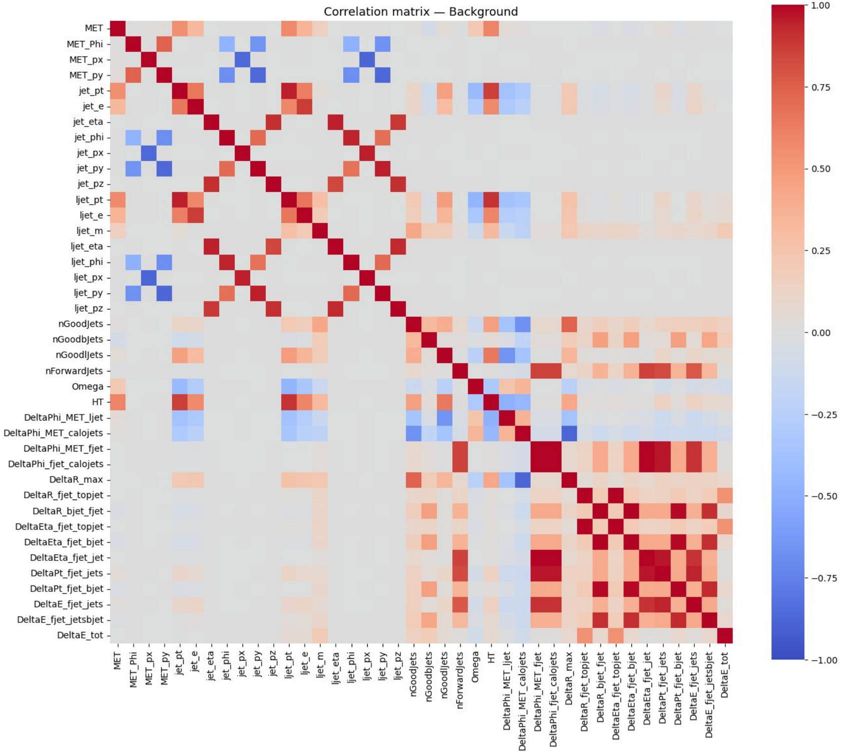
- ❏ MET and HT (sum of MET + all small R-jet_pt) histograms for background vs signal hypothesis

Anomaly Detection on Monotop Simulated Events
Chosen Features & Feature Grouping

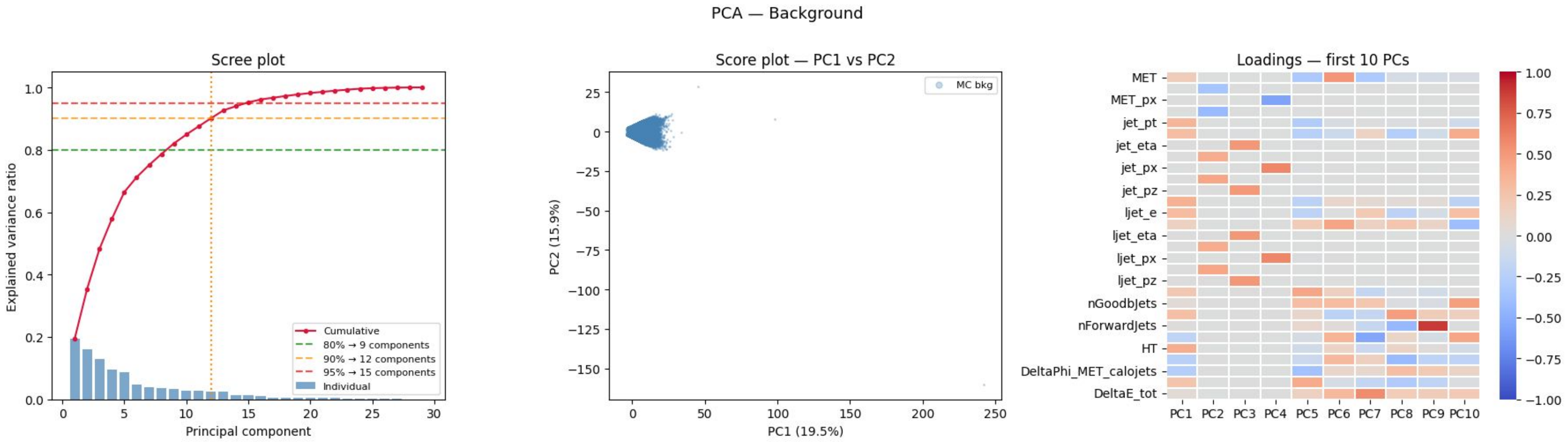
MET	MET	MET-Phi	MET_m	MET_eta	MET_px	MET_py					
jet	jet_pt	jet_e	jet_eta	jet_phi	jet_px	jet_py	jet_pz				
ljet	ljet_pt	ljet_e	ljet_m	ljet_eta	ljet_phi	ljet_px	ljet_py	ljet_pz			
nGood	nGoodJets	nGoodbJets	nGoodlJets	nForwardJets							
event	Omega	HT									
DeltaPhi	MET_ljet	MET_calojets	MET_fjet	fjet_calojets							
Delta_other	DeltaR_max	DeltaR_fjet_topjet	DeltaR_bjet_fjet	DeltaEta_fjet_topjet	DeltaEta_fjet_bjet	DeltaEta_fjet_jet	DeltaPt_fjet_jets	DeltaPt_fjet_bjet	DeltaE_fjet_jets	DeltaE_fjet_jetsbjet	DeltaE_tot

Anomaly Detection on Monotop Simulated Events
Background Correlation Matrix

- components from same 4-vector are correlated (expected)
- not expected: correlations between different features → can be dropped to reduce training cost



Anomaly Detection on Monotop Simulated Events
Pre-processing: PCA



- 80% variance → 9 components
- 90% variance → 12 components
- 95% variance → 15 components
- 99% variance → 23 components

Anomaly Detection on Monotop Simulated Events
Model Configurations Tested

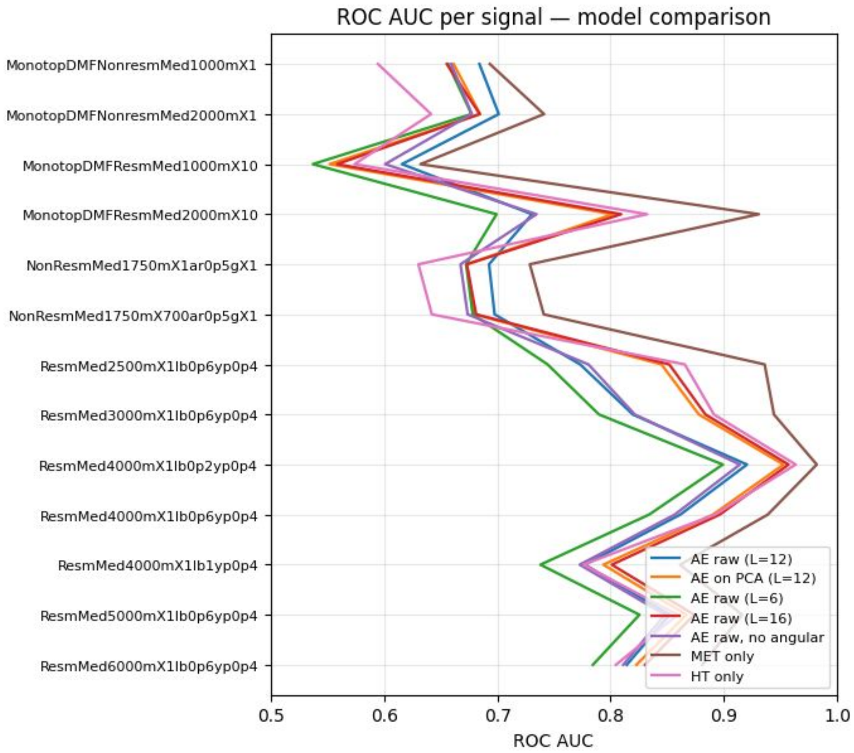
LABORATÓRIO DE INSTRUMENTAÇÃO E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

#	Name	Input	Latent dim	Isolates
A	raw_L	Raw features	12	Reference point
B	pca_L	~15 PCA components	12	PCA vs raw representation, latent held constant
C	raw_Llo	Raw features	6 (bellow 80% var)	Latent-dim sensitivity (more compression)
D	raw_Lhi	Raw features	16 (99% var)	Latent-dim sensitivity (less compression)
E	raw_noang	Raw, angular(eta/phi) dropped	12	Does dropping angular features help?

Anomaly Detection on Monotop Simulated Events Performance: Summary Plot

LABORATÓRIO DE INSTRUMENTAÇÃO E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

- Average AUC across all signals from biggest to lowest: Raw L=16 (0.781) > PCA L=12 (0.777) > Raw L=12 (0.765) > Raw, no angular (0.757) > Raw L=6 (0.736)
- No configuration dominates across all signals
- Heavy compression (L=6) underperforms every other config on every single signal
- MET has very high discriminating power for these set of signals, but it's not always the case → the models have similar discriminating power without any a priori information



Anomaly Detection on Monotop Simulated Events
Conclusions and Further Work

LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

Conclusions:

- ❏ No single configuration wins everywhere, model choice should depend on which part of signal-mass space the search prioritizes
- ❏ Heavy compression degrades the performance across every signal tested
- ❏ Latent dimension matters more than input representation; increasing the latent capacity improves performance almost as switching to PCA-compressed input, and does so with a simpler pipeline

Further Work:

- ❏ Implement real data
- ❏ Try normalizing flows models
- ❏ Use Run3 data