Anomaly detection as a test of new physics phenomena in CERN ATLAS experiment data

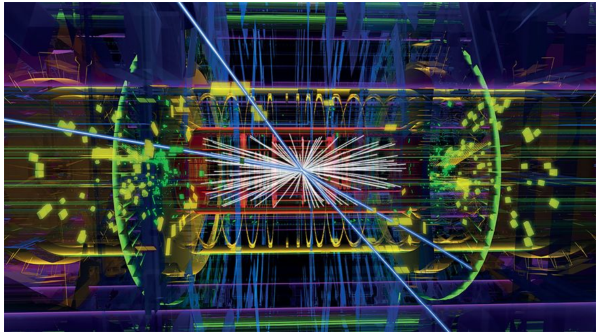
Students: Rui Ferreira and Ed Reeves
Advisers: Nuno Castro and Rute Pedro

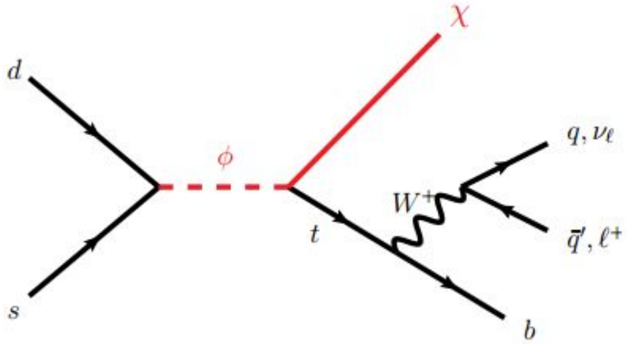
Aims

- Gain an understanding of the use of anomaly detection methods in high energy and beyond standard model physics searches
- Investigate previously suggested approaches in greater detail

Signal Events

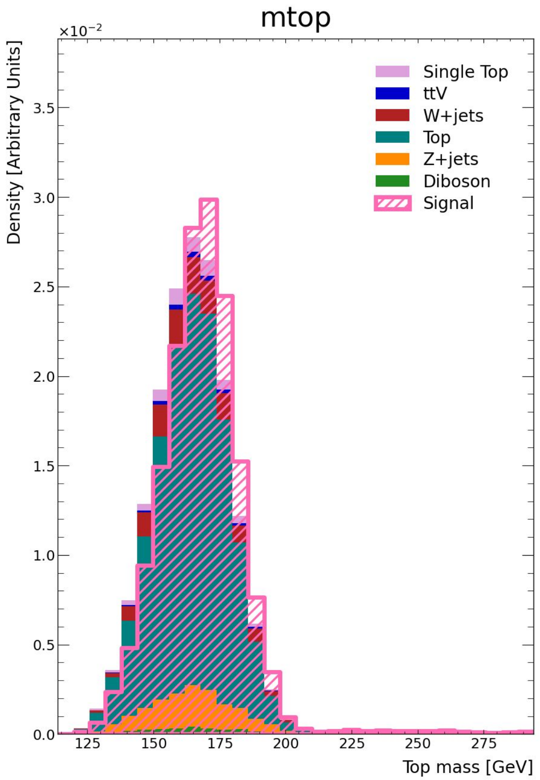
- Proton-proton collisions at LHC can generate many different processes
- Result of interest is an event defined by the presence of a single top quark (Monotop events) and large missing transverse momentum.
- Different theoretical processes can generate similar events
- Possibility that some of those are generating invisible particles that may correspond to dark matter
- As a benchmark signal we use resonant production of coloured scalar φ that decays into a DM particle and a top-quark





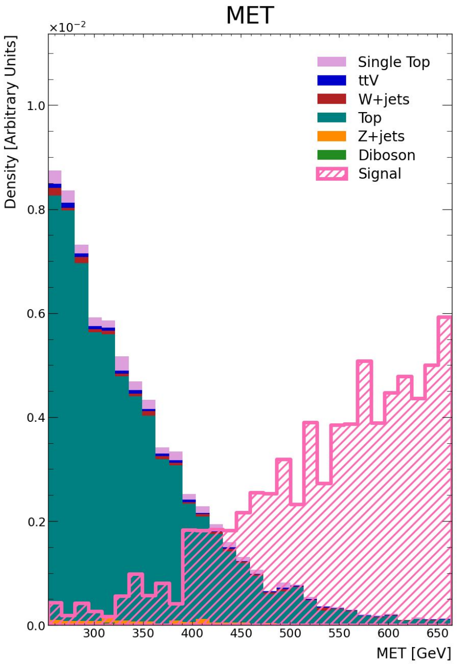
Background Events

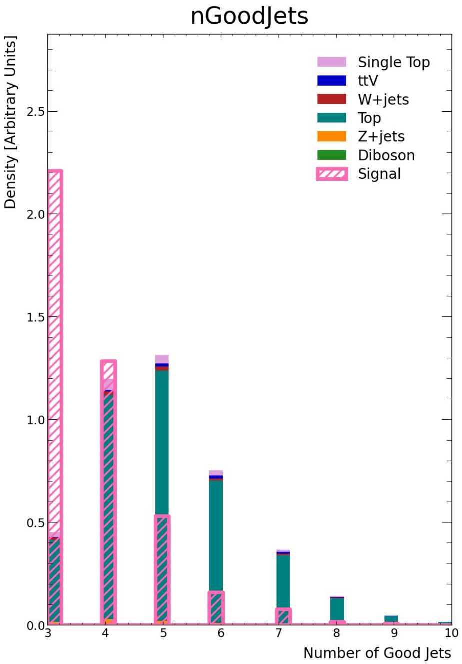
- Select SM background events compatible with signal topology
- Bkg: Single top, ttV, W+jets, Top, Z+jets, Diboson
- Sig: Monotop Res mPhi 2000
- All samples generated using ATLAS Monte Carlo simulations



Features - what do we measure in the detector?

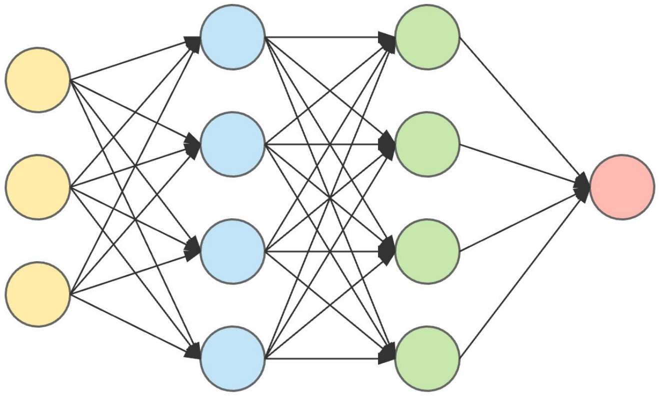
- Some features have better discriminating power than others
- Some are continuous, some are discrete
- Features have a different range of values, so are standardized before use in ML





Neural Networks and Supervised Learning

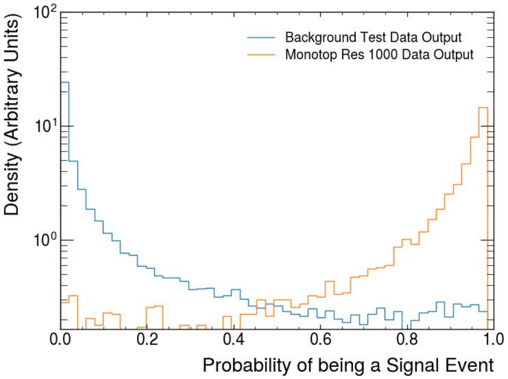
- Supervised learning: the machine is fed two or more sets of labeled data and is tasked with finding relations between them.
- Neural Network: sequences of layers which receive input and delivers the output to another layer.
- Classifier: a model trained to categorize events

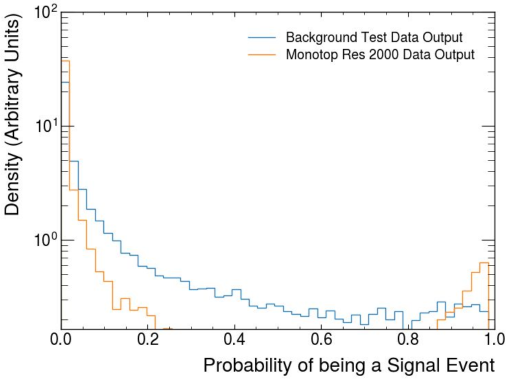


Shortcomings of the Supervised Learning

- A supervised model was trained using a specific type of signal - Monotop1000
- The model should return 0 for background data, and 1 for signal data.
- After training, the model was tested with two different signals - Monotop1000 (the one present during the training) and Monotop2000.

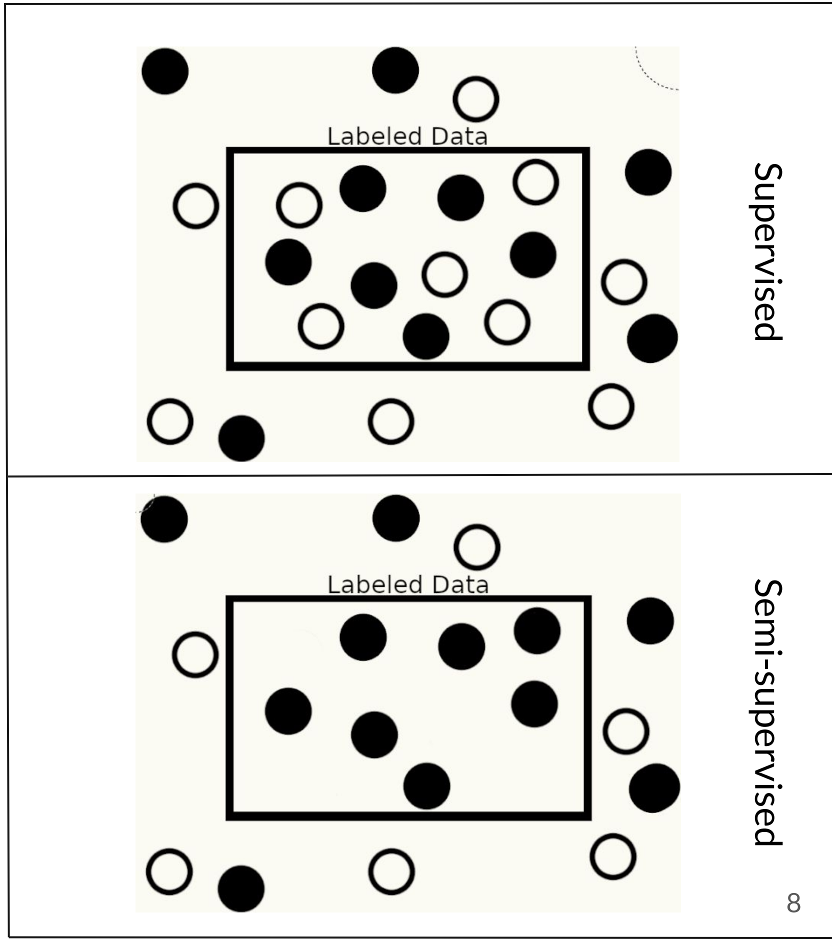
Outputs





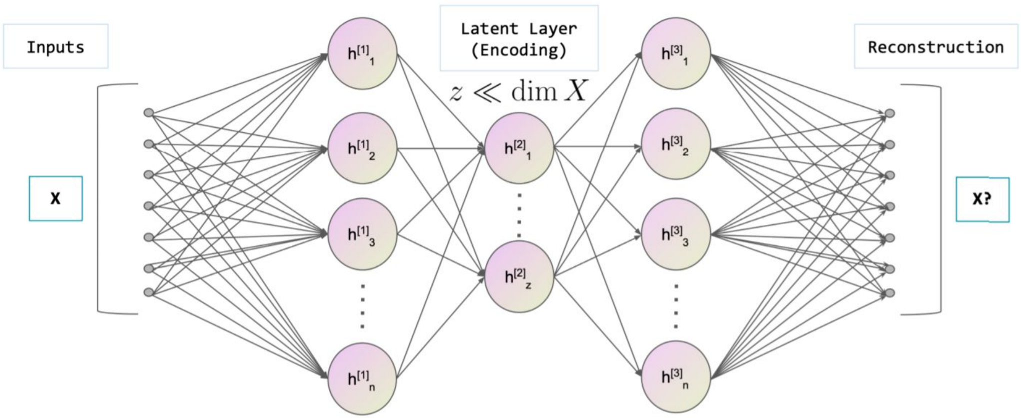
Semi-supervised learning

- Supervised Learning isn’t optimal to solve this classification problem
- We understand well the background events, but poorly the signal events
- Semi-supervised Learning: the machine is fed one labeled dataset
- The semi-supervised learning methods we used were the **Autoencoder** and the deep **SVDD**



Autoencoders

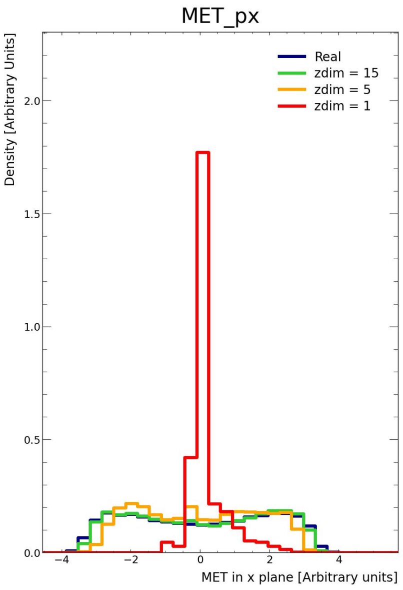
- Minimise reconstruction error during optimisation
- Anomaly score obtained by **log10(Reconstruction Error)**
- **Optuna** (a hyperparameter optimization python library) was used for the optimization of some hyperparameters.
- Investigate altering **zdim**
- Use full feature set of **50 features**

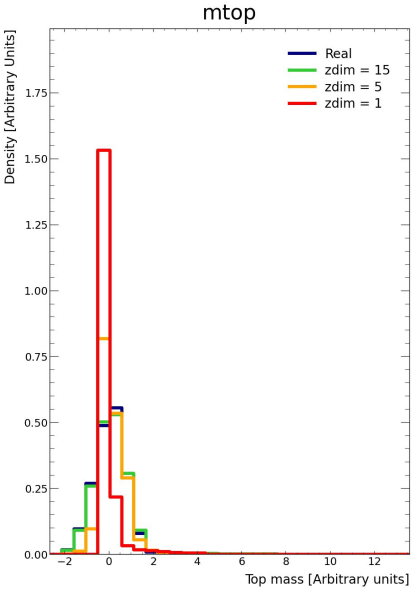


$$\min_{\mathcal{W}} \frac{1}{n} \sum_i ||AE(\mathbf{x}_i, \mathcal{W}) - \mathbf{x}_i||^2$$

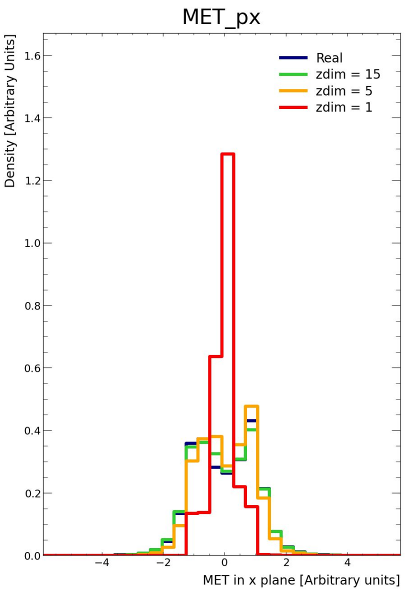
Reconstructed Features

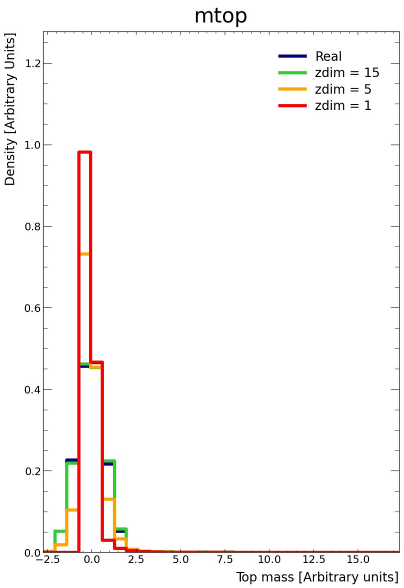
SIG - Monotop Res mPhi 2000



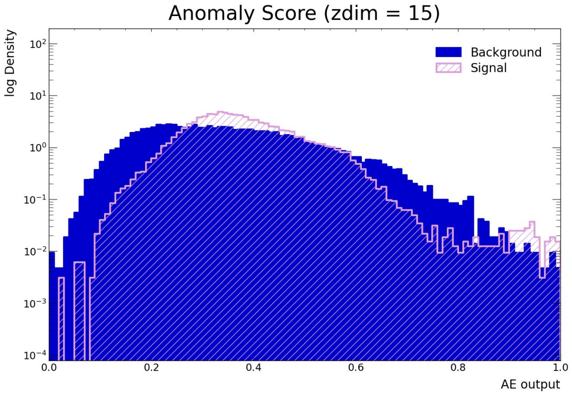


BKG

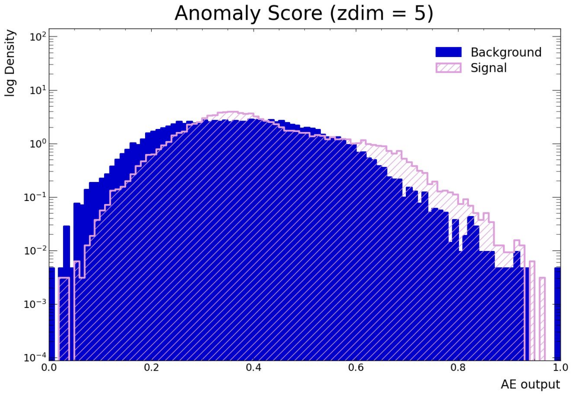




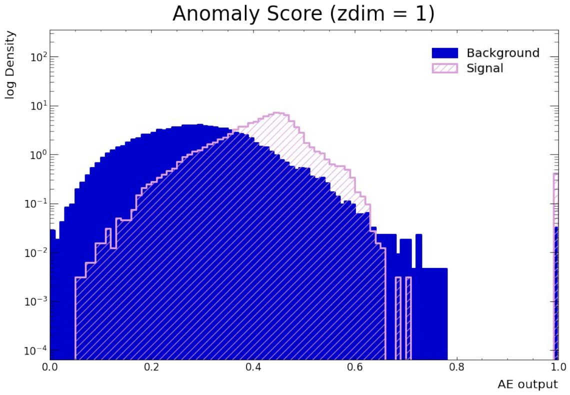
AE Total Reconstruction Error



Bkg mean squared error : **0.016**

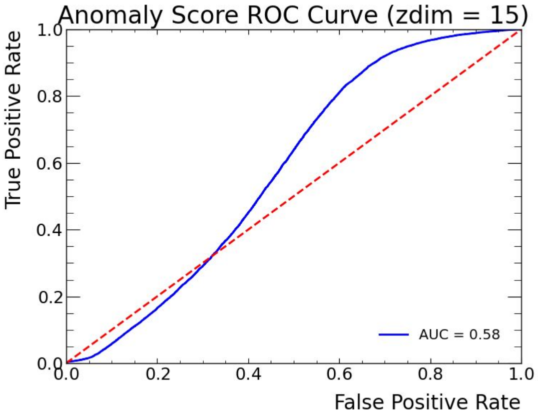


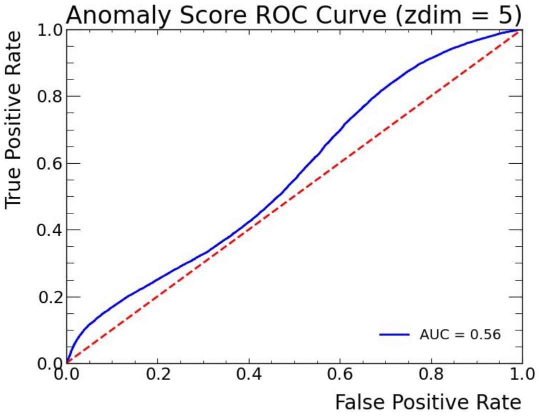
Bkg mean squared error : **0.131**

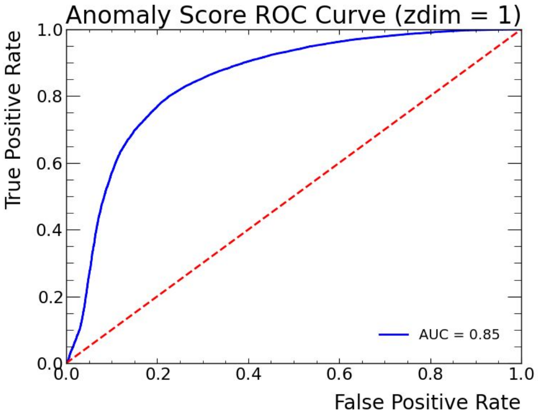


Bkg mean squared error : **0.382**

Classifier Performance







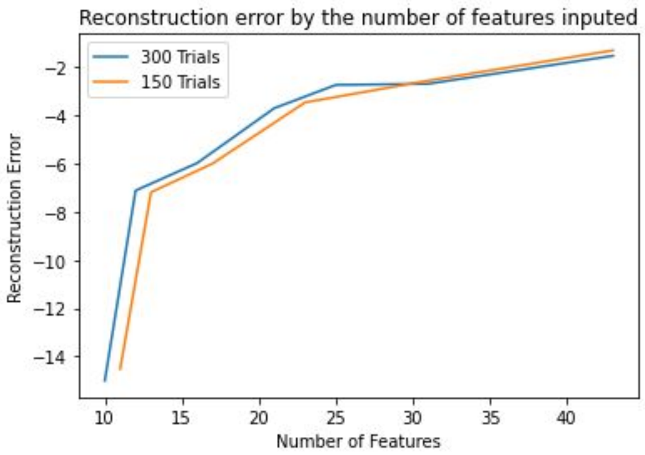
Feature Removal System

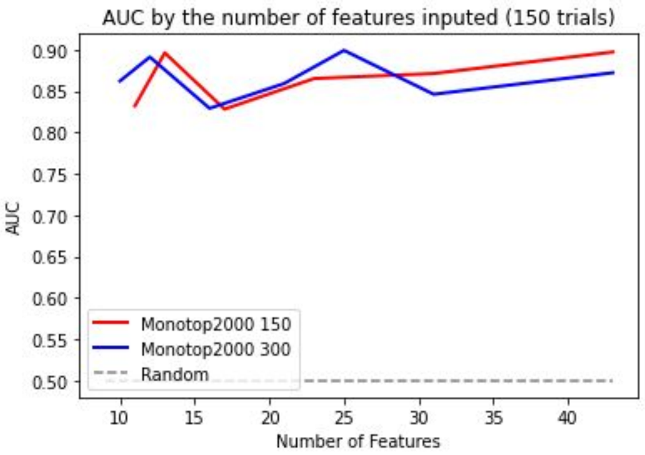
- The model was initially trained with 43 features (integer variables were removed from the 50).
- After the training, the worse reconstructed features were removed
- The model was trained again, without the removed features

- The process was cyclical until the remaining features were less than 9 (a fifth of the original number)
- The latent dimension layer was kept at a fourth of the number of features used in each model

Performance

- Below, a brief overview of the variation of the Reconstruction Error of the Background Data and the ROC curve of two similar trained models.
- As expected, the reconstruction error decreases (by many degrees of magnitude).
- The AUC only presents some marginal variation.

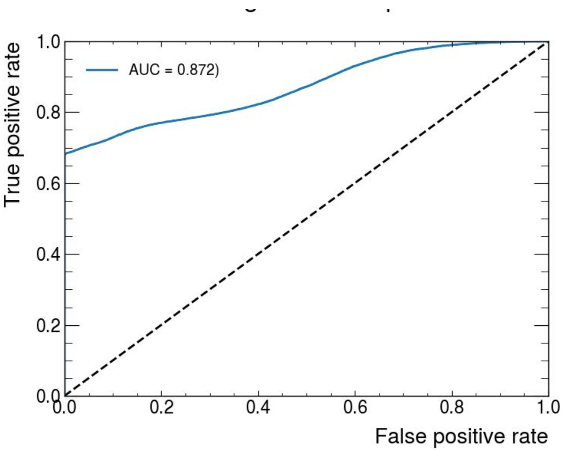




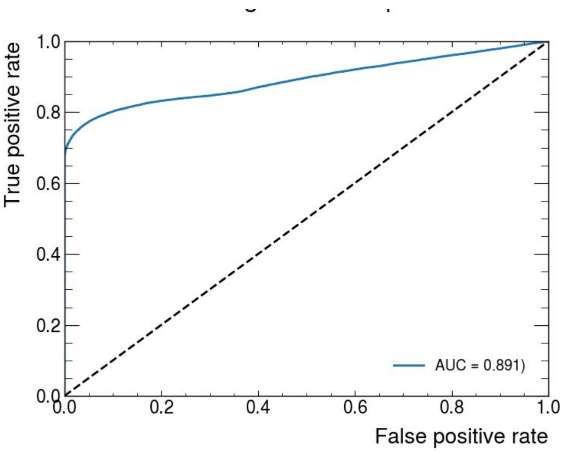
Deeper Look at the ROC Curves

- The shape of the ROC Curve maintains the form.
- Instead of a smooth curve, the ROC is almost linear

43 features

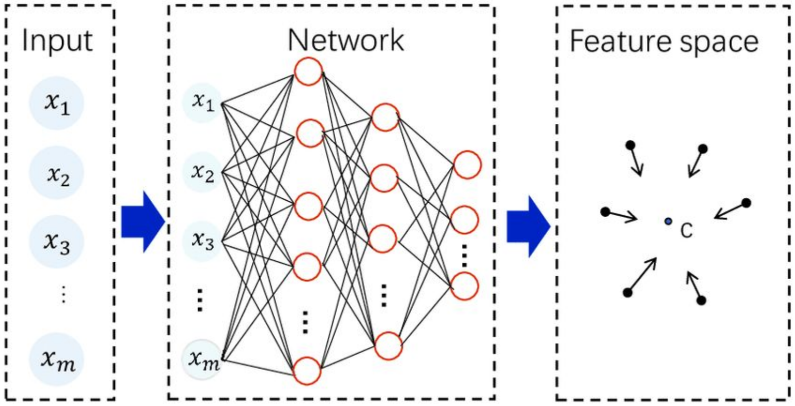


12 features



Deep SVDD - another anomaly detection method

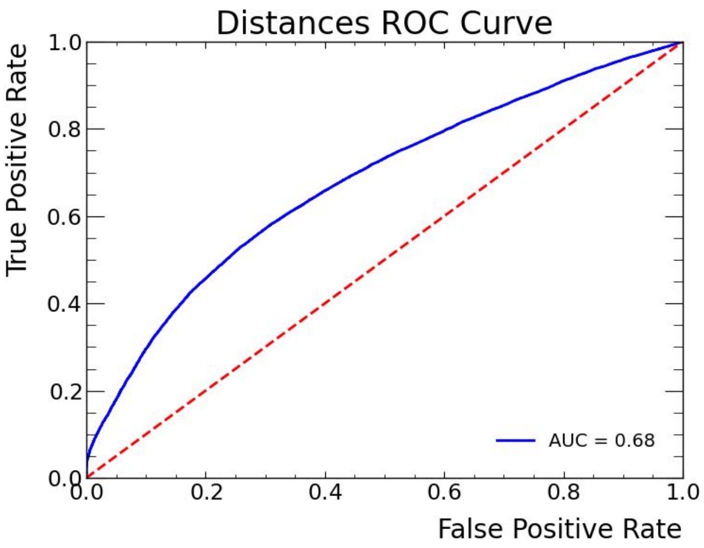
- Feature space defined by dimensionality of last DNN layer
- Minimise distance of data to centre of distribution in feature space
- Anomaly score = **log10(distance)**
- Investigate altering the feature space

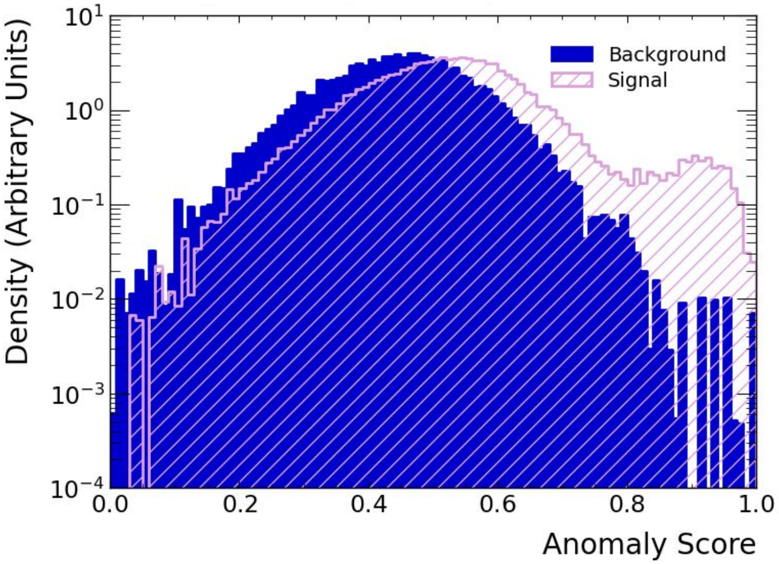


$$\min_{\mathcal{W}} \frac{1}{n} \sum_i ||\mathrm{DNN}(\mathbf{x}_i, \mathcal{W}) - \mathbf{c}||^2$$

Deep SVDD Performance

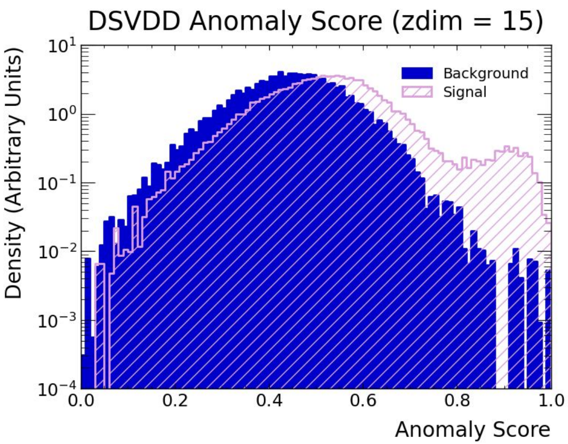
Monotop Res mPhi 2000 mPhi



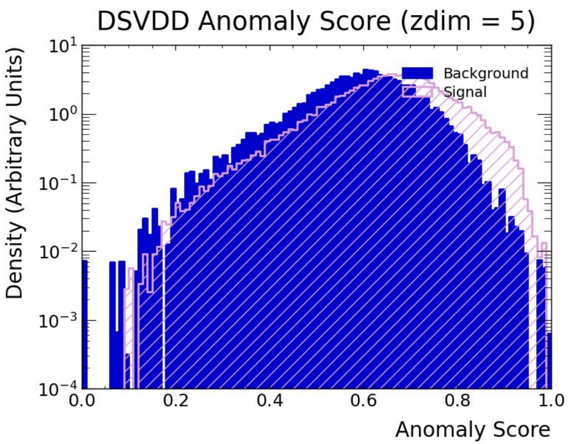


Altering feature space dimension

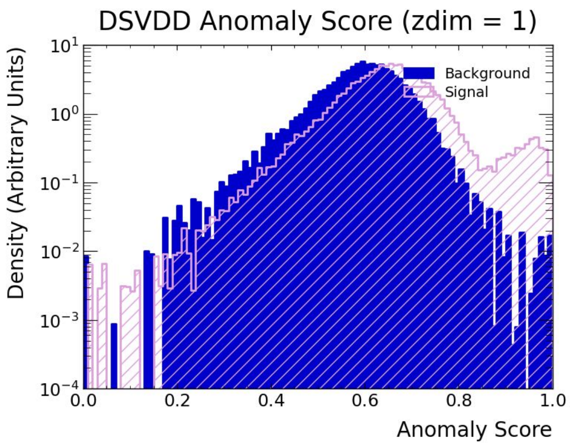
Monotop Res mPhi 2000



Bkg mean distance: 5.87e-06



Bkg mean distance: 6.21e-07

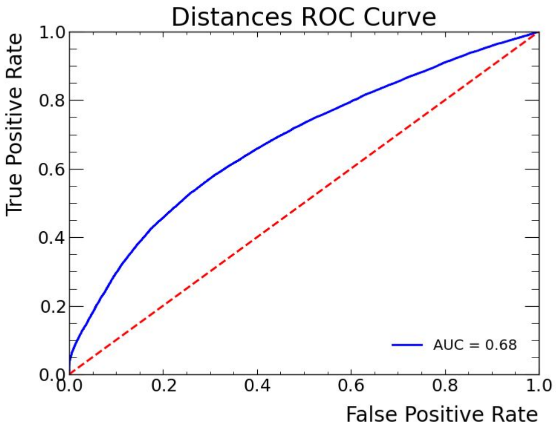


Bkg mean distance: 2.29e-07

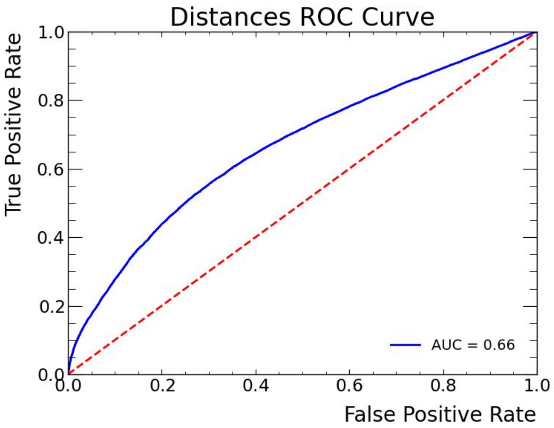
Altering feature space dimension

Monotop Res mPhi 2000

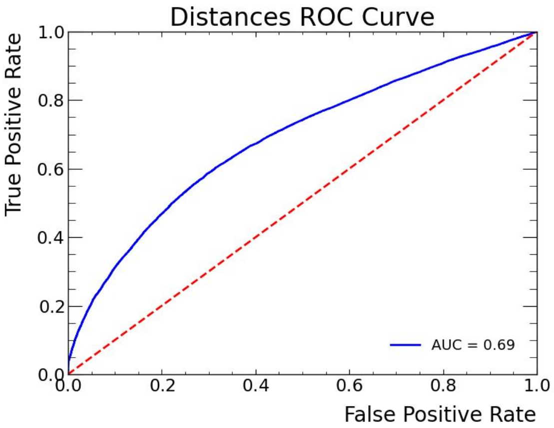
Zdim = 15



Zdim = 5



Zdim = 1



Conclusions

- Poor performance of supervised learning on unseen data motivates need for new anomaly detection methods
- Compressing data in Autoencoder to smaller latent layer dimension produces better performance as a classifier
- Removing features from the input of the model didn’t improve the model as a classifier. The ROC displays a strange shape that might worth some more study.
- Deep SVDD shows promising performance and is more stable in response to changes in architecture
- More testing with different signal samples is needed

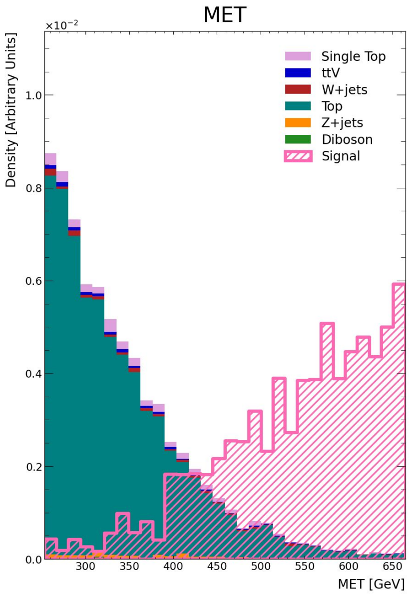
Thank you

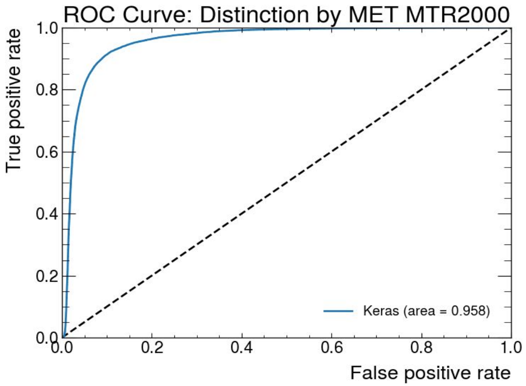
BACK UP

Missing Energy Transverse (MET)

- Quantity that describes the symetrical of the norm of the sum of the vectors of all objects detected after the collision in the transverse plane of the collision

- It is expected that the researched events (signal) have higher values for the MET





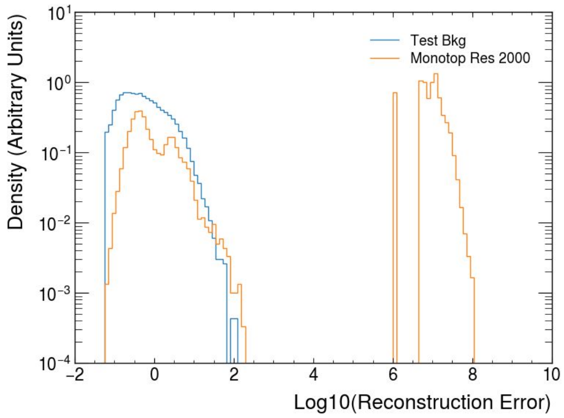
Optuna

- The training of the model optimizes the model parameters, like the weights of the relations between neurons
- The model depends on some parameters that define the model, and that aren't changed in the training - hyperparameters
- Some examples: number of layers, number of neurons in each layer, activation function, learning rate, etc.
- Optuna is a Python library that allows to optimize the hyperparameters through a training loop and a sampler that chooses the hyperparameters between lists of values (defined by the user) according to the metric chosen by the user.

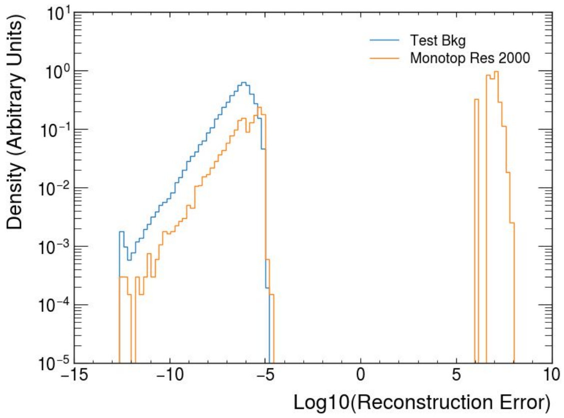
Deeper Look at the Error distributions

- The error distributions have different shapes
- Both exhibit the same two peaks
- The separation between peaks is bigger for the 12 features trying

43 features

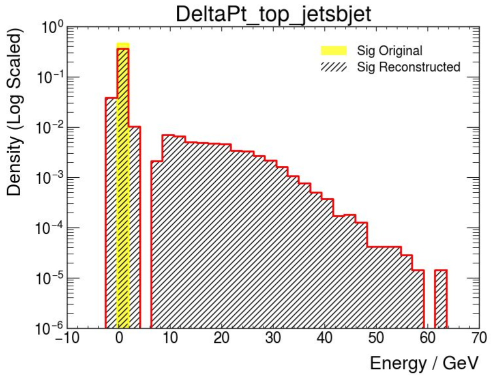


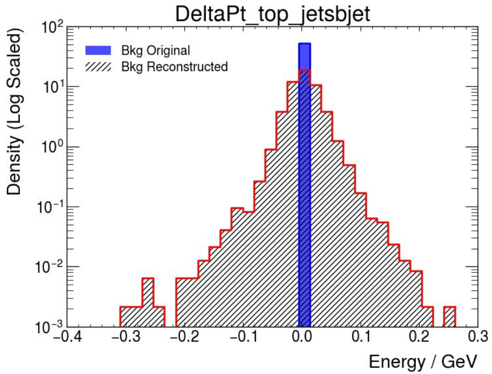
12 features



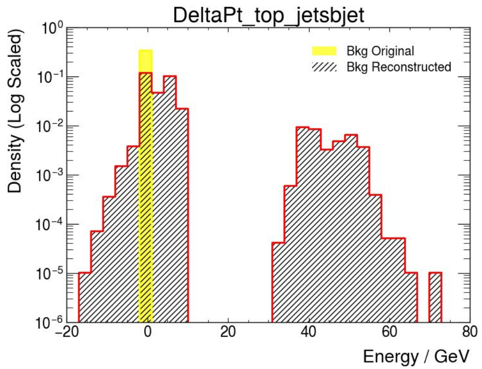
Reconstucted Features

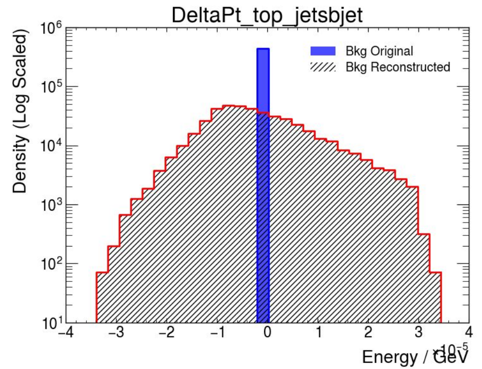
12 features





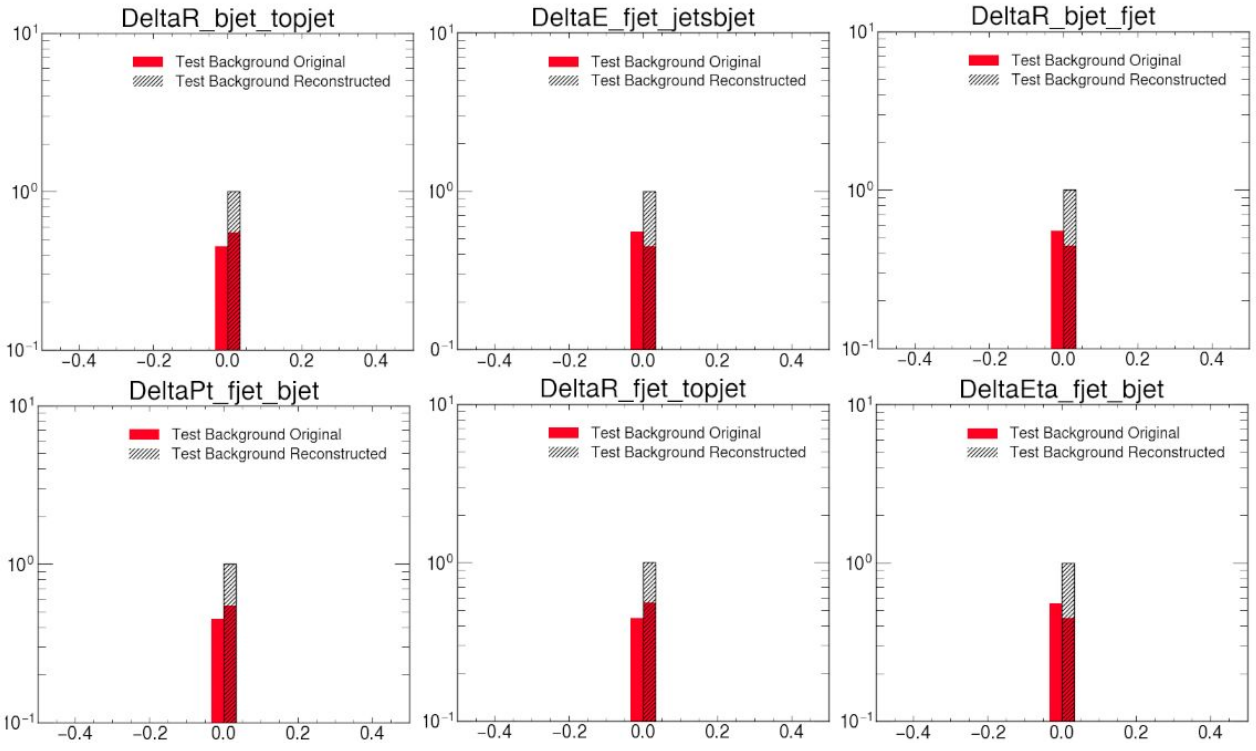
43 features



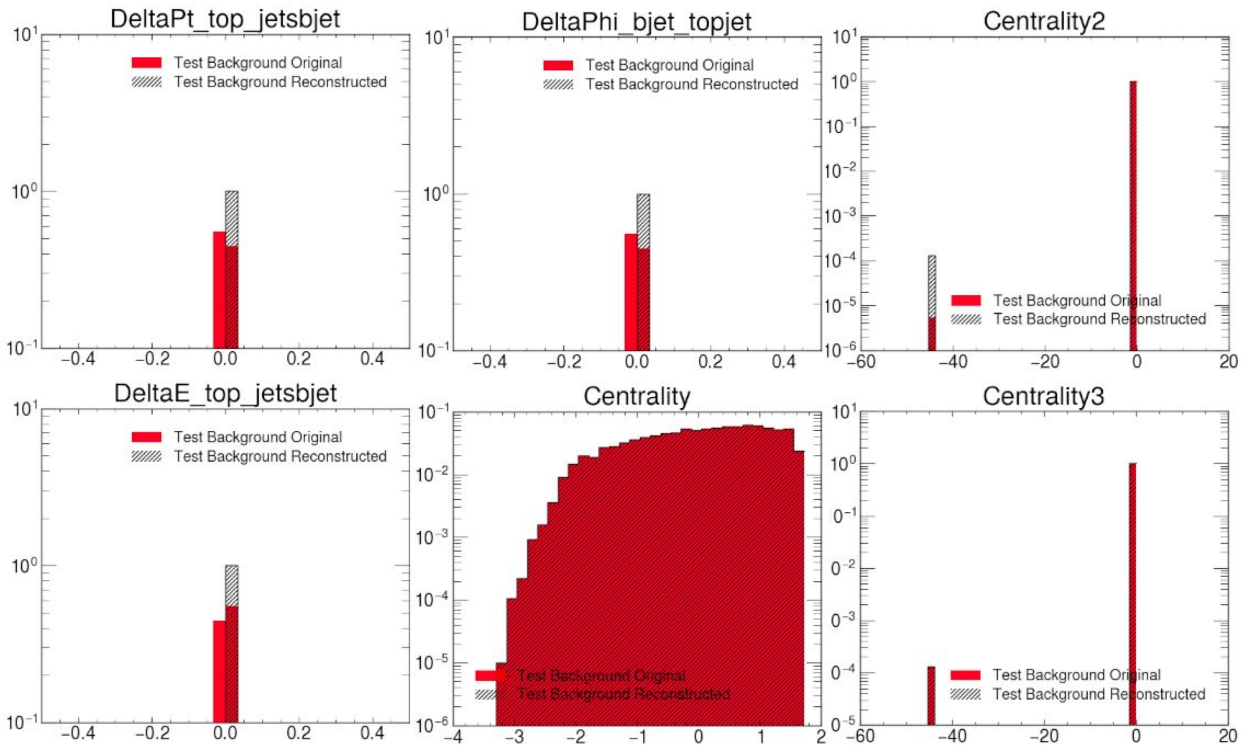


The Features Remaining

- Most features remaining have a very small range of values on the background data



The Features Remaining



- The small columns very far from the large columns are the missing values

- Centrality2 and Centrality3 returned the highest reconstruction error

Feature Removal Summary

- The performance of the AE is maintained in relation to the number of features used
- The reconstruction error decreases as the less well reconstructed features are removed from training
- Removing the features with the highest reconstruction error tends to remove features with high value ranges, disregarding how relevant they might be