HOW CAN WE TURN CLASSIFIERS INTO ANOMALY DETECTORS?

Sascha Caron (Nikhef, Radboud U), José Enrique García Navarro (IFIC, CSIC-UV), María Moreno Llácer (IFIC, CSIC-UV), Polina Moskvitinaa (Nikhef, Radboud U), Mats Rovers (Radboud U., Nikhef), **Adrián Rubio Jiménez (IFIC, CSIC-UV)**, Roberto Ruiz de Austri (IFIC, CSIC-UV), Zhongyi Zhang (Bonn U.).

MOTIVATION AND STRATEGY

Motivation

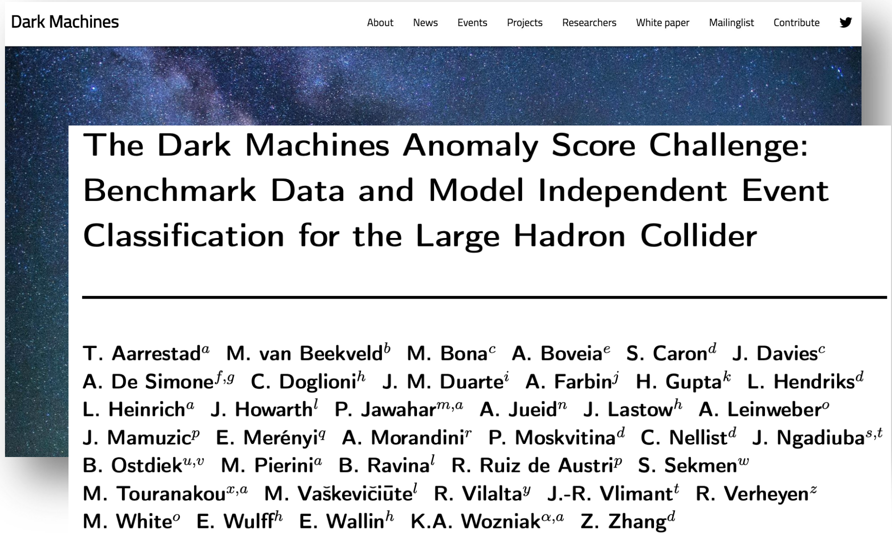
- The most powerful architectures for supervised classification learn the physical information more efficiently.
- But… **how can we turn them into anomaly detectors and how good are they?**

Strategy

- Adaptation of 2-3 different classifier architectures with 3 methods to detect anomalies (8 models).
- No network optimisation (or minimal) was performed to avoid biases.

DarkMachines dataset

- Open data: Zenodo link to dataset from anomaly score challenge.
- Event generation: *proton-proton* collisions at 13 TeV .
- Detector simulation: simplified card for ATLAS detector at CERN.
- Reconstructed particles (objects): jets, b-tagged jets, charged leptons, photons.
- Low level variables: object type, the four-momentum of the objects and the missing transverse momentum of the event.


Dark Machines About News Events Projects Researchers White paper Mailinglist Contribute

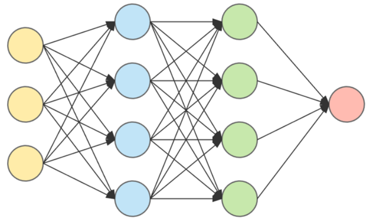
The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider

T. Aarrestad[a] M. van Beekveld[b] M. Bona[c] A. Boveia[e] S. Caron[d] J. Davies[c]
A. De Simone[f,g] C. Doglioni[h] J. M. Duarte[i] A. Farbin[j] H. Gupta[k] L. Hendriks[d]
L. Heinrich[a] J. Howarth[l] P. Jawahar[m,a] A. Jueid[n] J. Lastow[h] A. Leinweber[o]
J. Mamuzic[p] E. Merényi[q] A. Morandini[r] P. Moskvitina[d] C. Nellist[d] J. Ngadiuba[s,t]
B. Ostdiek[u,v] M. Pierini[a] B. Ravina[l] R. Ruiz de Austri[p] S. Sekmen[w]
M. Touranakou[x,a] M. Vaškevičiūte[l] R. Vilalta[y] J.-R. Vlimant[t] R. Verheyen[z]
M. White[o] E. Wulff[h] E. Wallin[h] K.A. Wozniak[α,a] Z. Zhang[d]

ARCHITECTURES AND TECHNIQUES

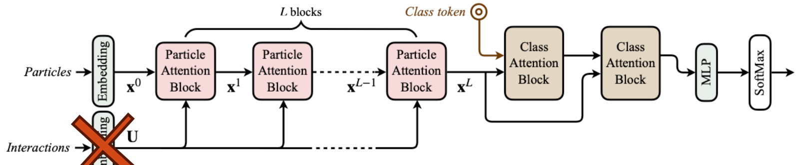
Architectures

Multi-Layer Perceptron (MLP)



Particle Transformer (ParT)

https://arxiv.org/abs/2211.05143



(a) Particle Transformer

No pairwise interactions

ParT+ SM couplings

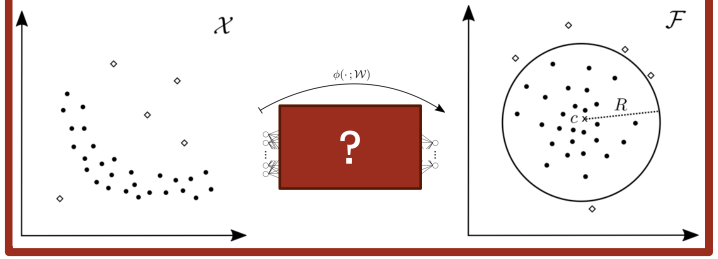
- Pairwise interactions
  - $\ln(m^2_{ij})$
  - $\ln(\Delta R_{ij})$
  - Physical information from Standard Model: couplings.

Developed by this group

Techniques

Deep Support Vector Data Description (dSVDD)

- Add an output layer with certain dimensions.
- Training: minimise distance to a centre in the hypersphere (anomaly score).
- Outliers are considered anomalies.
- Make ensemble for different dimensions.

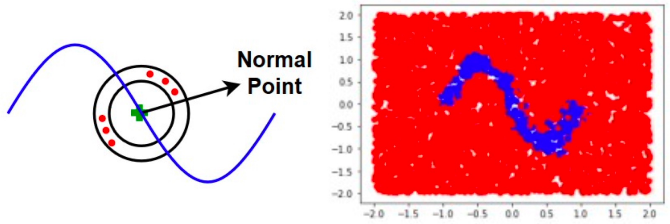


Deep Robust One-Class Classification (DROCC)

- Background is assumed to lie in a low-dimensional manifold.
- Anomalous background events are generated and their location in the manifold is searched with an adversarial training.

$$\sum_{i=1}^{n}[\ell(f_\theta(x_i),1)+\mu \max_{\tilde{x}_i \in N_i(r)} \ell(f_\theta(\tilde{x}_i),-1)]$$

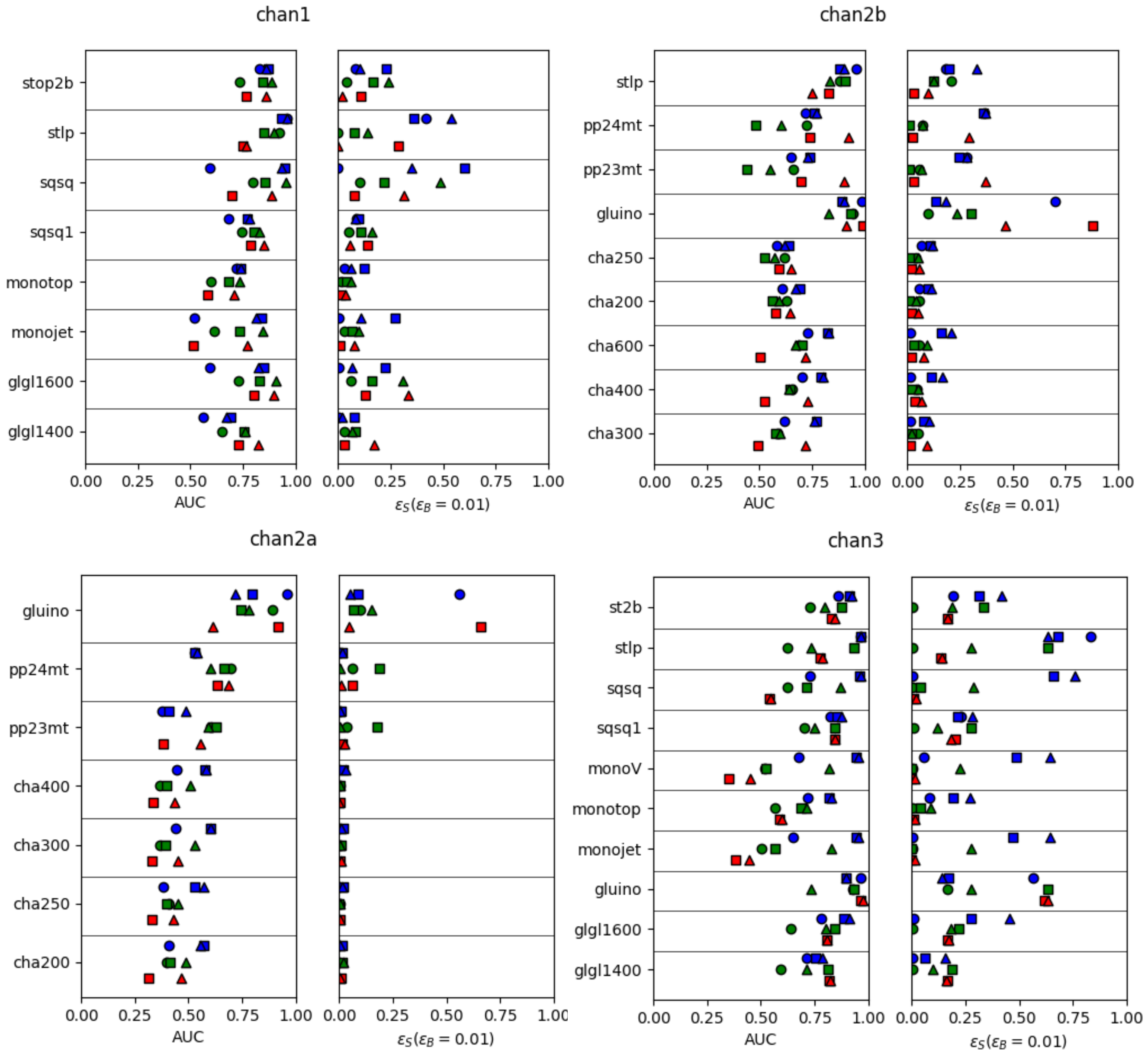
- Weakly supervised implementation

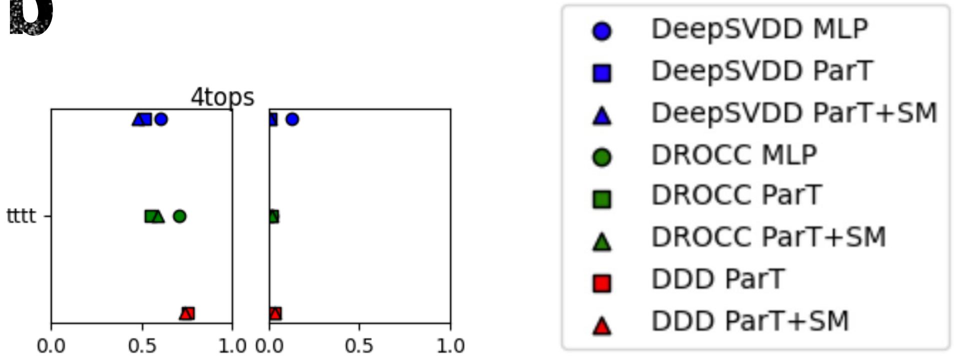


Discriminant distorsion detection (DDD)

- New technique developed for this study.
- Anomalies look like distorted background.
- Distorted training dataset is created:
  - Smearing kinematic variables with a gaussian.
  - Adding or removing objects.
- Train: discriminate *distorted bkg* vs *bkg.*
- Models with AUCs ~ 0.7-0.8 are picked up for testing on signals. Ensemble was made.

RESULTS AND CONCLUSIONS





- Shown that we can take a supervised classifier and transform it into a (good) anomaly detector.
- **The best classifiers are -on average- better anomaly detectors**: ParT+SM in this case.
- Similar performances among the 3 techniques. Compatible with anomaly score challenge.
- A recommendation could be to use dSVDD and DDD in combination (fully unsupervised).
- The new method DDD discriminates between data with and without distortions. This opens interesting future research directions.
- A more detailed recipe will be found in the paper (very soon in arXiv).

5

BACK-UP

CHANNELS AND SIGNALS

- Channel 1 (214k SM and 38k BSM):
  - $H_T \geq 600$ GeV .
  - $E_{Tmiss} \geq 200$ GeV.
  - $E_{Tmiss}/HT \geq 0.2$.
  - At least 4 (b)-jets with $p_T > 50$ GeV.
  - 1 (b)-jet with $p_T > 200$ GeV.
- Channel 2a (20k SM and 11k BSM):
  - $E_{Tmiss} > 50$ GeV.
  - $N_{lep} >= 3$ (where $p_{Tlep} > 15$ GeV).
- Channel 2b (340k SM and 90k BSM):
  - $E_{Tmiss} > 50$ GeV.
  - $N_{lep} >= 2$ (where $p_{Tlep} > 15$ GeV).
  - HT > 50 GeV.
- Channel 3 (8.5M SM and 1M BSM):
  - $E_{Tmiss} > 100$ GeV.
  - $H_T > 600$ GeV.

BSM process	Channel 1	Channel 2a	Channel 2b	Channel 3
Z' + monojet	×	×		×
$Z' + W/Z$				×
Z' + single top	×			×
Z' in lepton-violating $U(1)_{L_\mu - L_\tau}$		×	×	
$\not{R}$-SUSY stop-stop	×		×	×
$\not{R}$-SUSY squark-squark	×			×
SUSY gluino-gluino	×	×	×	×
SUSY stop-stop	×			×
SUSY squark-squark	×			×
SUSY chargino-neutralino		×	×	
SUSY chargino-chargino			×	