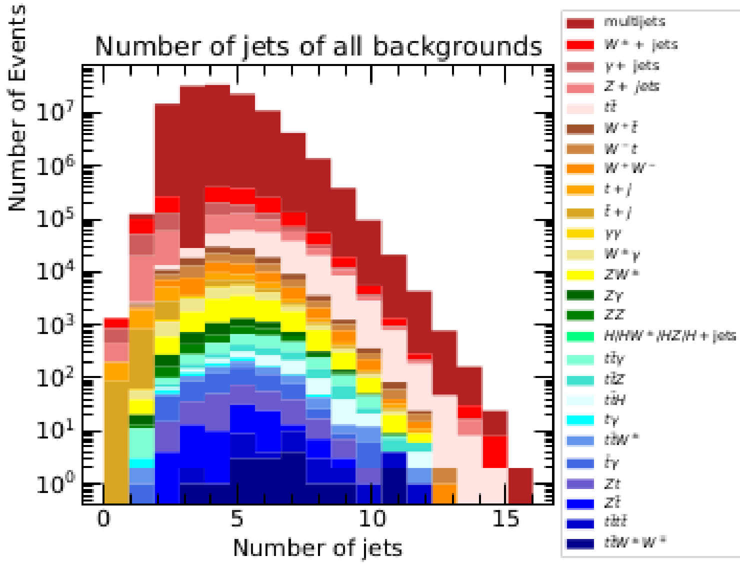
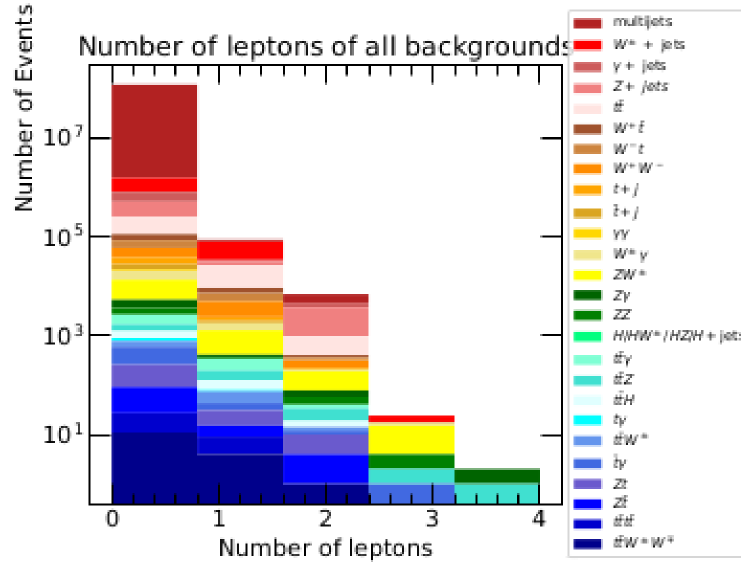




The Dark Machines Anomaly Score Challenge

Based on arXiv: 2105.14027

https://github.com/bostdiek/DarkMachines-UnsupervisedChallenge

Joe Davies on behalf of the Dark Machines Anomaly Detection Group

Queen Mary University of London

Paper Authors

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

---

T. Aarrestad[CERN] M. van Beekveld[Ox] M. Bona[QMUL] A. Boveia[OSU] S. Caron[HEF, Nikhef] J. Davies[QMUL] A. De Simone[SISSA, INFN] C. Doglioni[Lund] J.M. Duarte[UCSD] A. Farbin[UnivArlington] H. Gupta[GSoC] L. Hendriks[HEF, Nikhef] L. Heinrich[CERN] J. Howarth[Glasgow] P. Jawahar[WPI, CERN] A. Jueid[UnivKonkuk] J. Lastow[Lund] A. Leinweber[UnivAdelaide] J. Mamuzic[IFIC] E. Merényi[UnivRice] A. Morandini[RWTH] P. Moskvitina[HEF, Nikhef] C. Nellist[HEF, Nikhef] J. Ngadiuba[FNAL, Caltech] B. Ostdiek[Harvard, AIFI] M. Pierini[CERN] B. Ravina[Glasgow] R. Ruiz de Austri[IFIC] S. Sekmen[KNU] M. Touranakou[NKUA, CERN] M. Vaškevičiūte[Glasgow] R. Vilalta[UnivHouston] J.-R. Vlimant[Caltech] R. Verheyen[UCL] M. White[UnivAdelaide] E. Wulff[Lund] E. Wallin[Lund] K.A. Wozniak[UniVie, CERN] Z. Zhang[HEF, Nikhef]

https://arxiv.org/abs/2105.14027

Submitted to SciPost

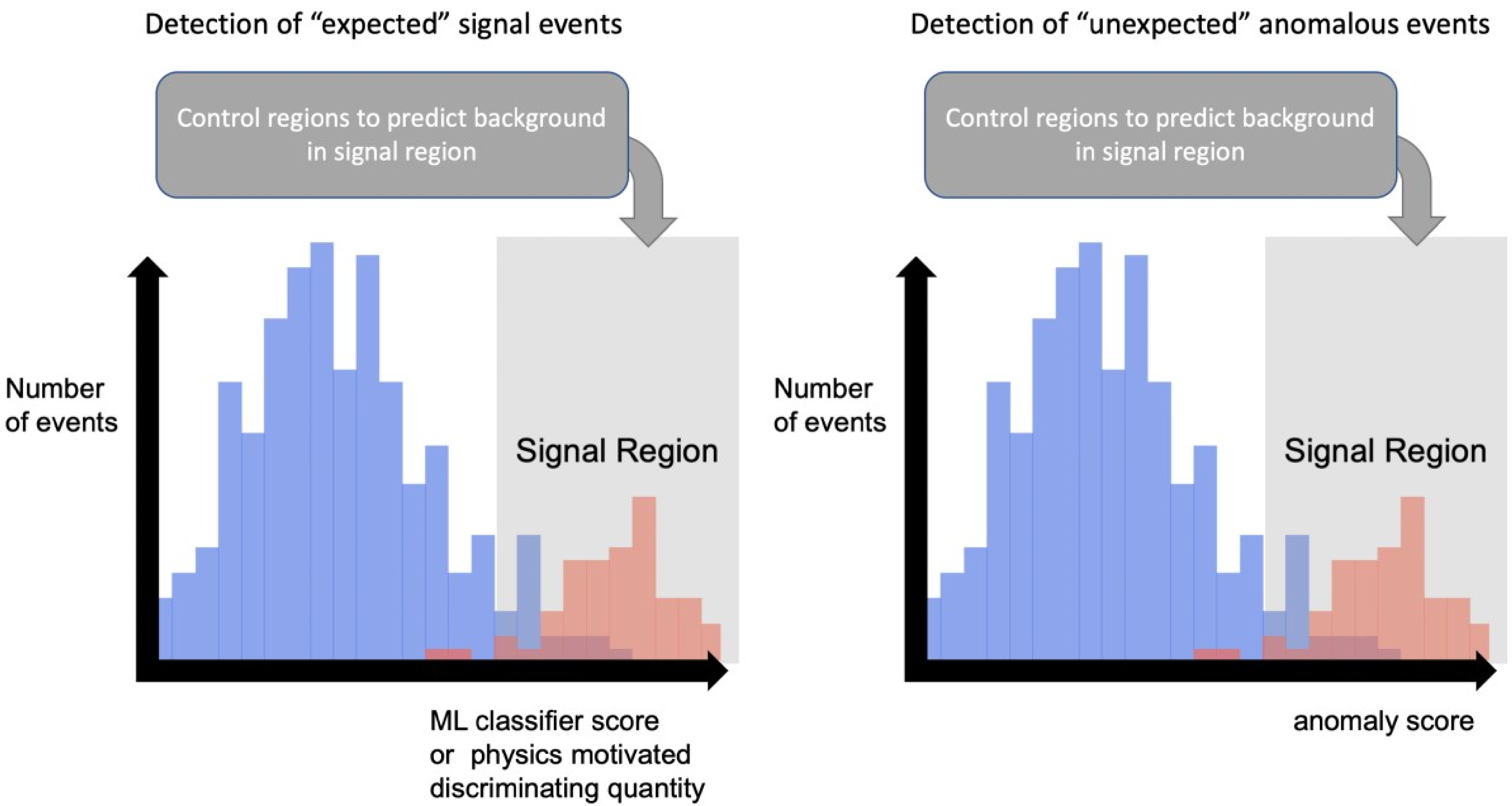
Challenge Justification and Goals

- Goal is to perform model-agnostic searches
- Already examples of similar searches:
  - DØ Collaboration at Tevatron using SLEUTH
  - H1 Collaboration at HERA using 1-D signal detection algorithm
  - CDF Collaboration at Tevatron (using similar to above)
- Searching for localized excesses in events can be done by Machine Learning
  - We look at anomaly detection techniques
- Unlike LHC Olympics which looked at overdensities as signals in black box data
  - https://arxiv.org/abs/2101.08320

Challenge Outline

- Dataset of > 1 Billion SM Events (Les Houches:https://arxiv.org/pdf/2002.12220.pdf)
  - https://zenodo.org/record/3685861
- Hackathon Dataset: (https://zenodo.org/record/3961917)
  - 4 different channels (channels here defined as distinct datasets based on selection cuts)
  - 11 different BSM signals (19 total mass points)
  - 34 unique signal/channel combinations
- Train each method 4 times (once per channel) using SM
- Select ML methods which perform best to apply to blinded Secret Dataset: https://zenodo.org/record/4443151

General Strategy

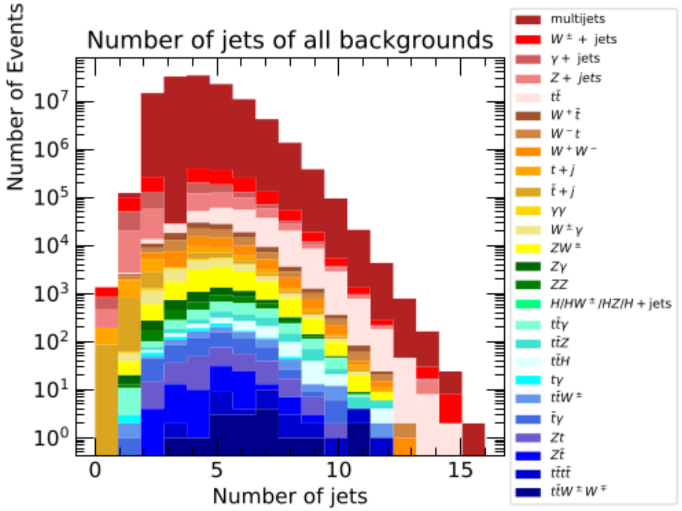


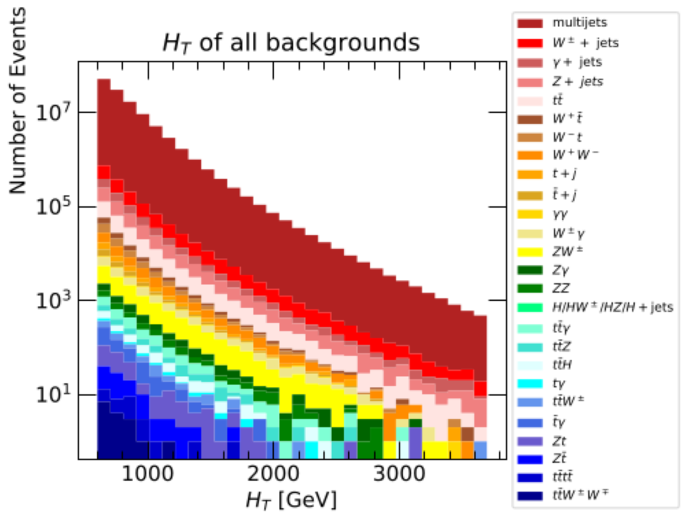
Challenge object is an event-by-event anomaly score and we use this to define a signal region

The Standard Model Datasets

SM processes			
Physics process	Process ID	σ (pb)	N_{tot} ($N_{10\,\mathrm{fb}^{-1}}$)
$pp \to jj(+2j)$	njets	$19718_{H_T>600\mathrm{GeV}}$	415331302 (197179140)
$pp \to l^{\pm}\nu_l(+2j)$	w_jets	$10537_{H_T>100\mathrm{GeV}}$	135692164 (105366237)
$pp \to \gamma j(+2j)$	gam_jets	$7927_{H_T>100\mathrm{GeV}}$	123709226 (79268824)
$pp \to l^+l^-(+2j)$	z_jets	$3753_{H_T>100\mathrm{GeV}}$	60076409 (37529592)
$pp \to t\bar{t}(+2j)$	ttbar	541	13590811 (5412187)
$pp \to t + \mathrm{jets}(+2j)$	single_top	130	7223883 (1297142)
$pp \to \bar{t} + \mathrm{jets}(+2j)$	single_topbar	112	7179922 (1116396)
$pp \to W^+W^-(+2j)$	ww	82.1	17740278 (821354)
$pp \to W^{\pm}t(+2j)$	wtop	57.8	5252172 (577541)
$pp \to W^{\pm}\bar{t}(+2j)$	wtopbar	57.8	4723206 (577541)
$pp \to \gamma\gamma(+2j)$	2gam	47.1	17464818 (470656)
$pp \to W^{\pm}\gamma(+2j)$	Wgam	45.1	18633683 (450672)
$pp \to ZW^{\pm}(+2j)$	zw	31.6	13847321 (315781)
$pp \to Z\gamma(+2j)$	Zgam	29.9	15909980 (299439)
$pp \to ZZ(+2j)$	zz	9.91	7118820 (99092)
$pp \to h(+2j)$	single_higgs	1.94	2596158 (19383)
$pp \to t\bar{t}\gamma(+2j)$	ttbarGam	1.55	95217 (15471)
$pp \to t\bar{t}Z$	ttbarZ	0.59	300000 (5874)
$pp \to t\bar{t}h(+1j)$	ttbarHiggs	0.46	200476 (4568)
$pp \to \gamma t(+2j)$	atop	0.39	2776166 (3947)
$pp \to t\bar{t}W^{\pm}$	ttbarW	0.35	279365 (3495)
$pp \to \gamma\bar{t}(+2j)$	atopbar	0.27	4770857 (2707)
$pp \to Zt(+2j)$	ztop	0.26	3213475 (2554)
$pp \to Z\bar{t}(+2j)$	ztopbar	0.15	2741276 (1524)
$pp \to t\bar{t}t\bar{t}$	4top	0.0097	399999 (96)
$pp \to t\bar{t}W^+W^-$	ttbarWW	0.0085	150000 (85)

$$H_T = \sum_i |p_{T,j_i}|$$





Madgraph+Pythia+Delphes | jets, b-jets, electrons, muons, photons

The Analysis Channels

Channel 1: 214K SM Events

- $H_T \geq 600$ GeV
- MET ≥ 200 GeV
- MET/$H_T \geq 0.2$
- At least 4 (b)-jets with $p_T > 50$ GeV
- At least 1 (b)-jets with $p_T > 200$ GeV

Channel 2b: 340K SM Events

- $H_T \geq 50$ GeV
- MET ≥ 50 GeV
- At least 2 μ/e with $p_T > 15$ GeV

Channel 2a: 20K SM Events

- MET ≥ 50 GeV
- At least 3 μ/e with $p_T > 15$ GeV
- At least 1 (b)-jets with $p_T > 200$ GeV
- <u>Few training events, many ML methods struggle</u>

Channel 3: 8.5M SM Events

- $H_T \geq 600$ GeV
- MET ≥ 100 GeV
- <u>Large dataset, timed out training on some methods</u>

The Methods

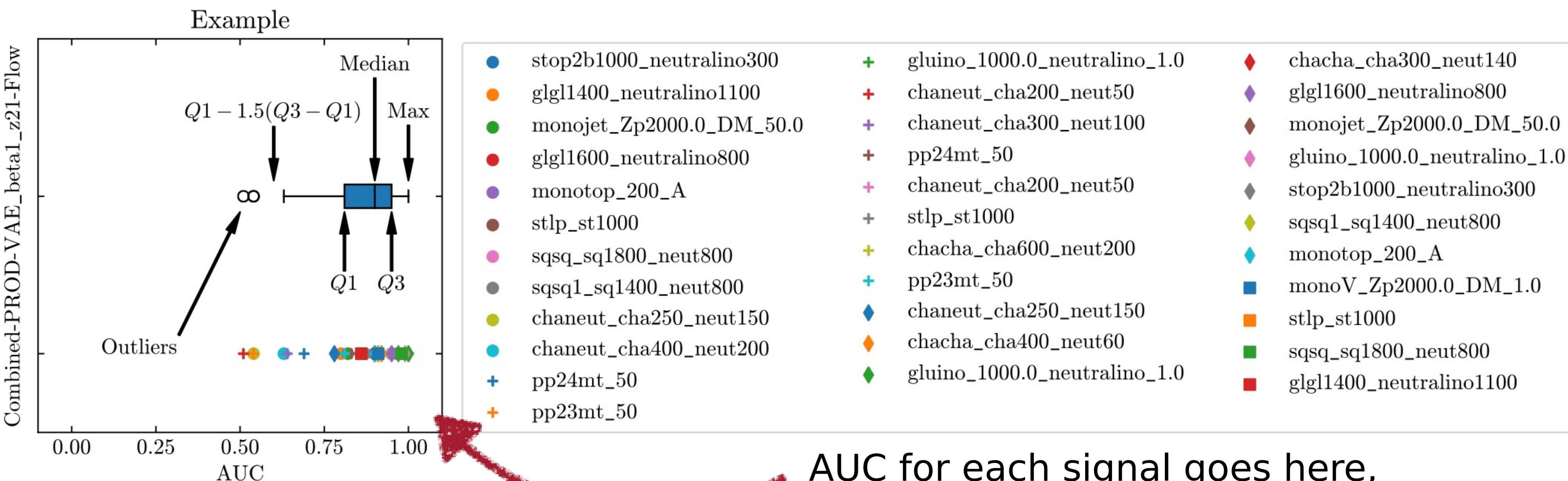
Abbreviation	Algorithm	Section	Hyperparameters	# Submitted
SimpleAE	Autoencoders	4.1	Tab. 6	1
VAEs	Variational Autoencoders	4.2	Tab. 7	140
DeepSetVAE	Deep Set Variational Autoencoders	4.3	Tab. 8	4
ConvVAE (NoF)	Convolutional Variational Autoencoders	4.4	Tab. 9	1
Planar	ConvVAE+Planar Flows	4.5.1	Tab. 10	1
SNF	ConvVAE+Sylvester Normalizing Flows	4.5.2	Tab. 11	3
IAF	ConvVAE+Inverse Autoregressive Flows	4.5.3	Tab. 12	1
ConvF	ConvVAE+Convolutional Normalizing Flows	4.5.4	Tab. 13	1
CNN	Convolutional (β)VAE	4.6		2
KDE	Kernel Density Estimation	4.7	Tab. 14	36
Flow	Spline autoregressive flow	4.8	Tab. 15	2
Deep SVDD	Deep SVDD	4.9	Tab. 16 & 17	80
Combined (Deep SVDD & Flow)	Spline autoregressive flow with Deep SVDD	4.10		8
DAGMM	Deep Autoencoding Gaussian Mixture Model	4.11	Tab. 19	384
ALAD	Adversarial Anomaly Detection	4.12	Tab. 21	96
Latent	Anomaly Detection in the Latent Space	4.13	Tab. 22	288

- Top 9 and the last use some form of encoding - decoding with a recon error anomaly score
- Planar, SNF, IAF, ConvF, Flow and Combined use some form of flow based likelihoods
- KDE, DAGMM and Latent use clustering or density estimation
- # submitted refers to number of methods of this type that were created

Metrics for Results

Figures of Merit:

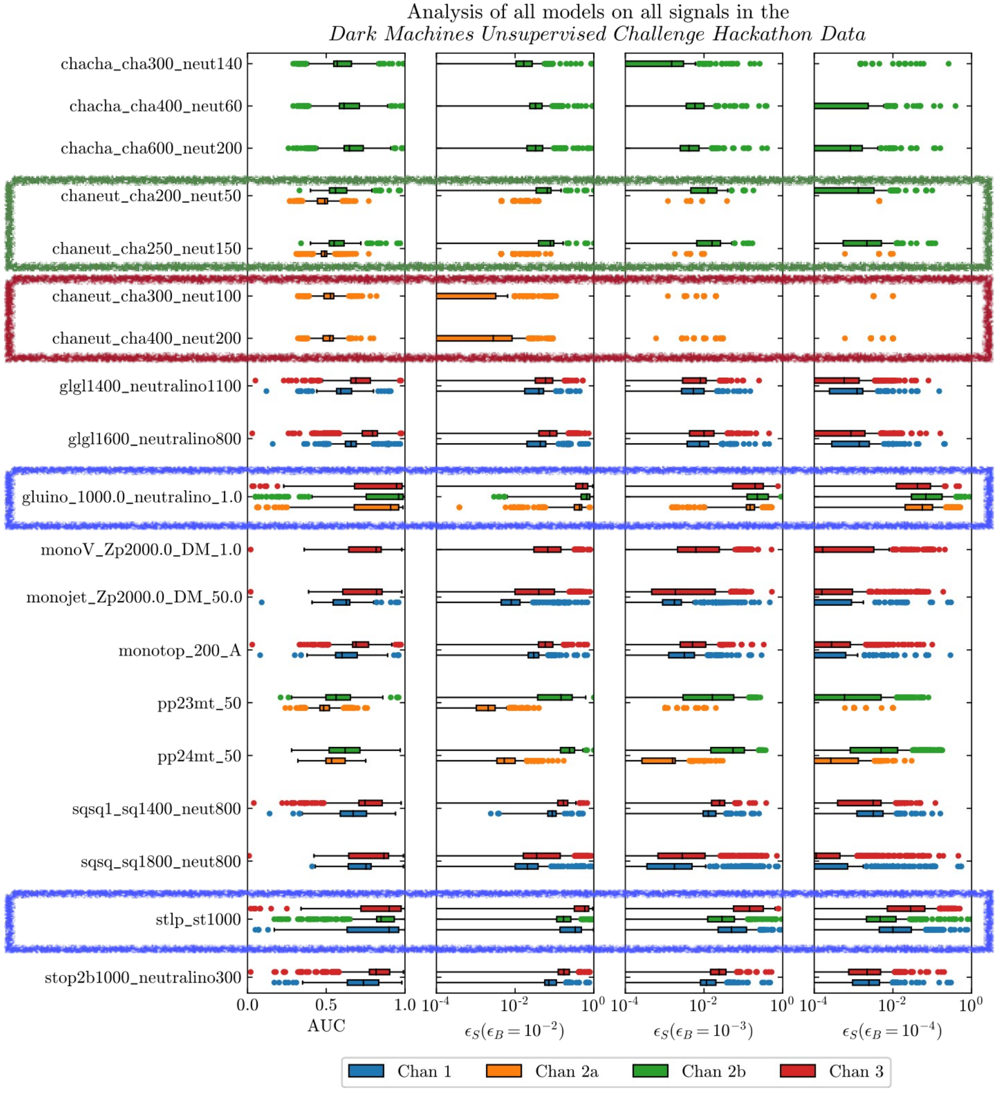
- Area under the ROC curve (AUC)
- The signal efficiency at a background efficiency of 10^{-2}
- The signal efficiency at a background efficiency of 10^{-3}
- The signal efficiency at a background efficiency of 10^{-4}



AUC for each signal goes here, summarized by box-and-whisker plot

Summary Results

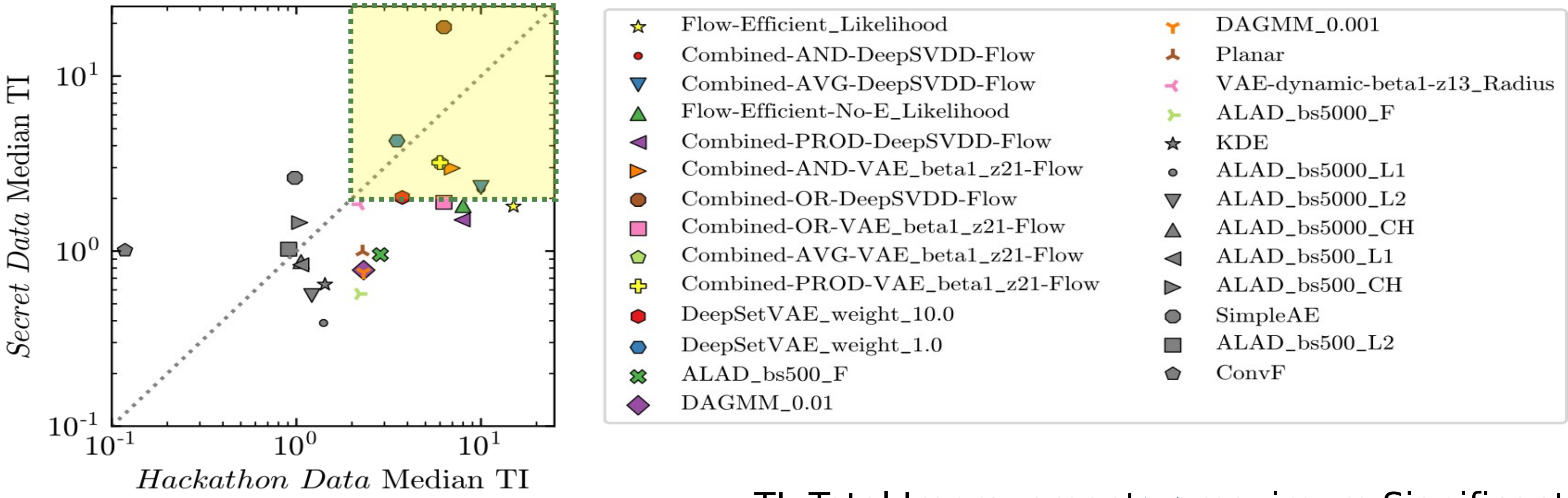
- Each method is a point on the box-and-whisker plot
- Each row is a BSM signal
- Some BSM easy for most methods
- Some BSM challenging for all methods
- Some BSM are easier than others to get a good anomaly score with



Each figure of merit has its own top methods, can we combine to form a single metric?

Best Performing Methods

Compare *Hackathon* and *Secret Data* results



Methods with Median TI > 2 on both datasets

Model	Hackathon Data	Secret Data
Combined-OR-DeepSVDD-Flow	6.30	19.02
DeepSetVAE_weight_1.0	3.50	4.27
Combined-AVG-VAE_beta1_z21-Flow	6.00	3.21
Combined-PROD-VAE_beta1_z21-Flow	6.00	3.20
Combined-AND-VAE_beta1_z21-Flow	7.00	2.98
Combined-AVG-DeepSVDD-Flow	10.00	2.31
Combined-AND-DeepSVDD-Flow	10.00	2.26
DeepSetVAE_weight_10.0	3.75	2.03

- TI: Total Improvement → maximum Significant Improvement over all background rejections over all channels

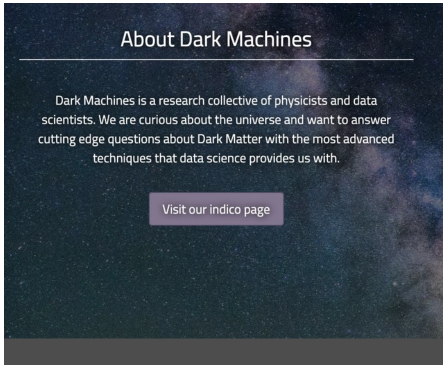
$$SI \equiv \frac{\epsilon_S}{\sqrt{\epsilon_B}}$$

- Apply the models to the same signals on the hackathon and secret datasets

- Each of the best performing models has some fixed target component (Deep SVDD, bVAE with b=1) and latent space seems to be important

Conclusion

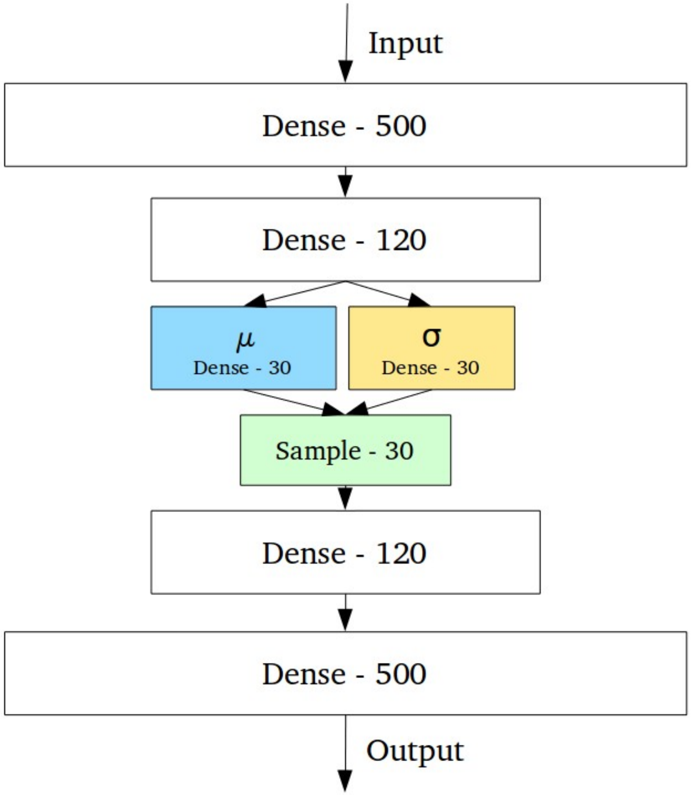
- Model-agnostic searches
- Primarily use Variational Auto-Encoders
- Variety of channels and signals
- Best methods use some form of fixed target
- Anomaly Detection is hard: seems even the Median metric doesn’t generalize well!
- https://twitter.com/dark_machines?s=20
- https://darkmachines.org/



Backups

Variational Autoencoder

- Same structure as an Autoencoder (encoder, bottleneck, decoder) except the latent space is continuous by design
- Sampling can be done on latent vectors to produce a continuous set of outputs
- (Generally) Minimum Squared Error (MSE) + Kullback-Liebler Divergence used as error



$$\sum_{i=1}^{N} \frac{1}{2}(t_i - y_i)^2$$

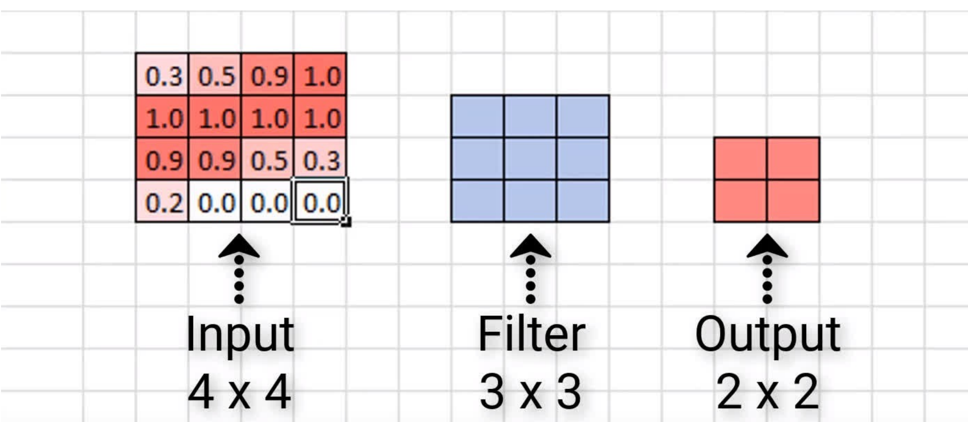
Typical MSE

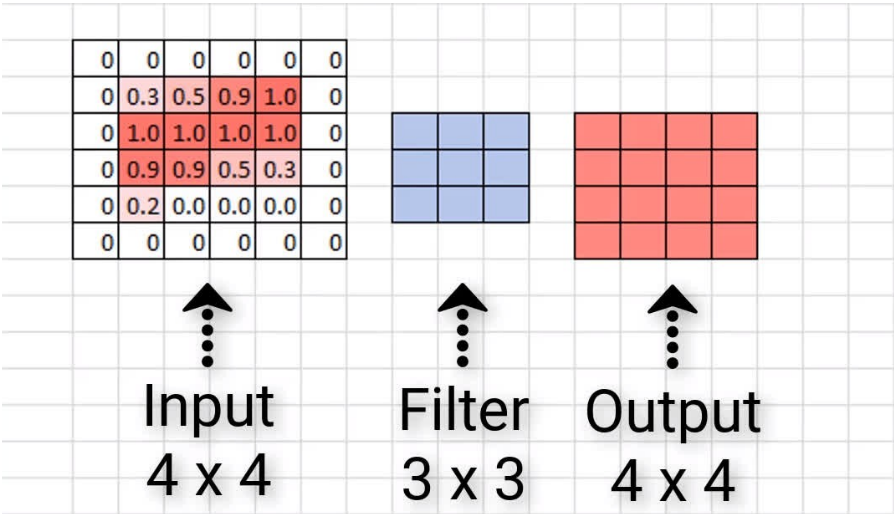
$$\sum_{i=1}^{n} \sigma_i^2 + \mu_i^2 - \log(\sigma_i) - 1$$

KL-Divergence

Challenges with the VAE

- Should the events be zero padded?
- Should we take a smaller number of objects?
- Which anomaly score to use:
  - Just one or the other of reconstruction or KL
  - Radius in the latent space
  - Beta parameters (and how to tweak them)





The BSM Physics

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

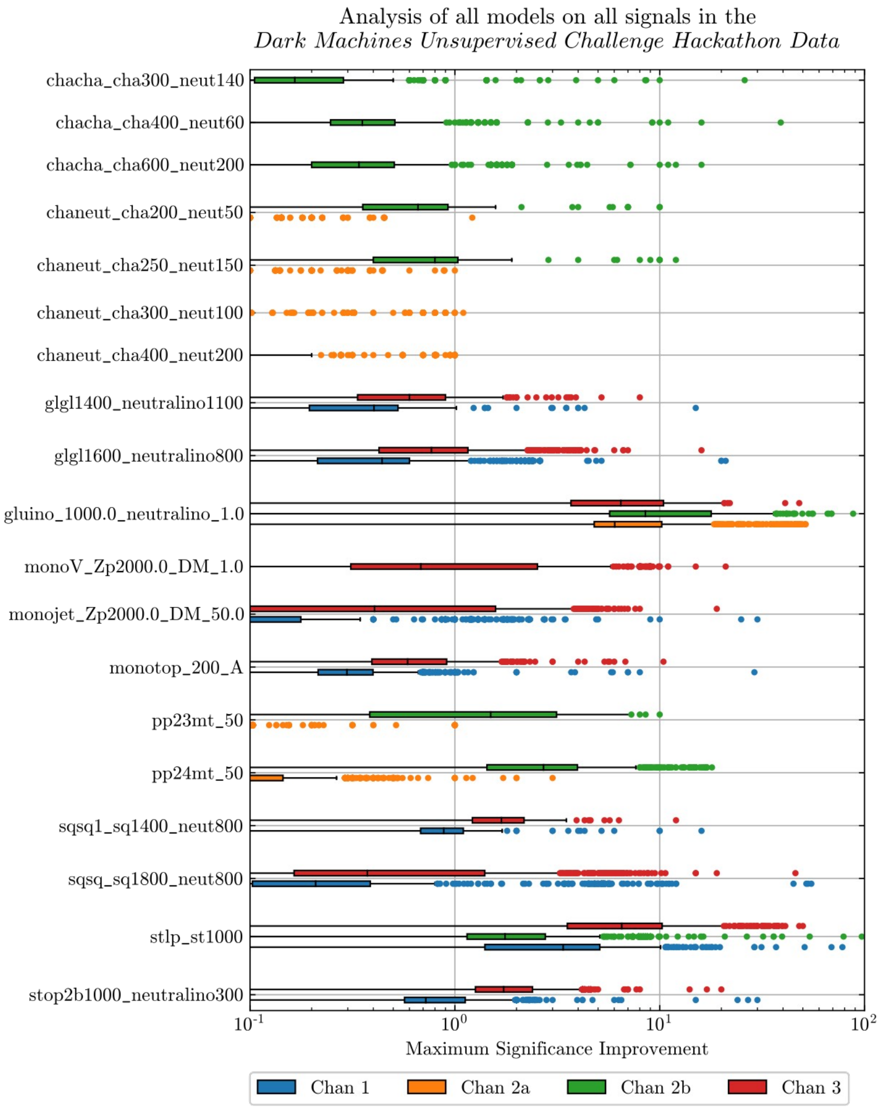
Some processes have different mass spectra or decay modes: 19 signals, 34 Signal-Channel combinations

Maximum Significant Improvement

$$\sigma_S = \frac{S}{\sqrt{B}}$$

$$\sigma_{AD} = \frac{S'}{\sqrt{B'}} = \frac{\epsilon_S \; S}{\sqrt{\epsilon_B \; B}} = \frac{\epsilon_S}{\sqrt{\epsilon_B}} \; \sigma_S$$

$$SI \equiv \frac{\epsilon_S}{\sqrt{\epsilon_B}}$$



Total Improvement Across Signals