Machine Learning for BSM Searches

A New Perspective

Charanjit Kaur Khosa

Università di Genova

Theory Seminar, 4th November 2020

Outline

A. Why to use Machine Learning for BSM searches?

B. Basics of Machine Learning methods

C. Anomaly detection

D. SMEFT analysis

E. Summary

Physics Beyond the SM

Standard Model: the mysteries

- Dark matter
- Neutrino masses
- Baryon Asymmetry of the Universe

Standard Model: the unexplained

- Flavour puzzle
- Hierarchy problem
- Fourth fundamental force (gravity) is not included

Current Status of BSM

We have not seen any signal for New Physics (NP) so far!

Nature of BSM

- New Physics signature is beyond the reach of the current colliders
- Known BSM models do not include the "correct" model

How do we search for it?

- Traditional analysis strategies are not suitable
- Model dependent searches (pre-bias)

Need to go beyond these limitations

1. Large volumes of data
2. High dimensionality of the data sets
3. Large number of model parameters

"Hidden" correlations in the data can be explored using "powerful" ML techniques

nature COMMUNICATIONS

ARTICLE

Received 19 Feb 2014 | Accepted 4 Jun 2014 | Published 2 Jul 2014 DOI: 10.1038/ncomms5308

Searching for exotic particles in high-energy physics with deep learning

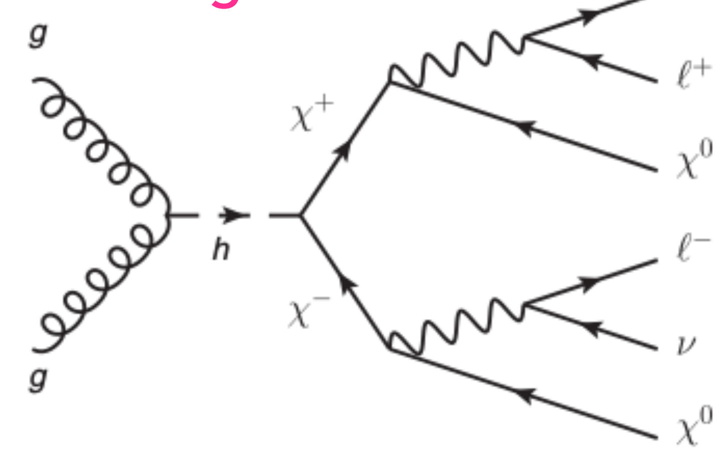
P. Baldi[1], P. Sadowski[1] & D. Whiteson[2]

Collisions at high-energy particle colliders are a traditionally fruitful source of exotic particle discoveries. Finding these rare particles requires solving difficult signal-versus-background classification problems, hence machine-learning approaches are often used. Standard approaches have relied on 'shallow' machine-learning models that have a limited capacity to learn complex nonlinear functions of the inputs, and rely on a painstaking search through manually constructed nonlinear features. Progress on this problem has slowed, as a variety of techniques have shown equivalent performance. Recent advances in the field of deep learning make it possible to learn more complex functions and better discriminate between signal and background classes. Here, using benchmark data sets, we show that deep-learning methods need no manually constructed inputs and yet improve the classification metric by as much as 8% over the best current approaches. This demonstrates that deep-learning approaches can improve the power of collider searches for exotic particles.

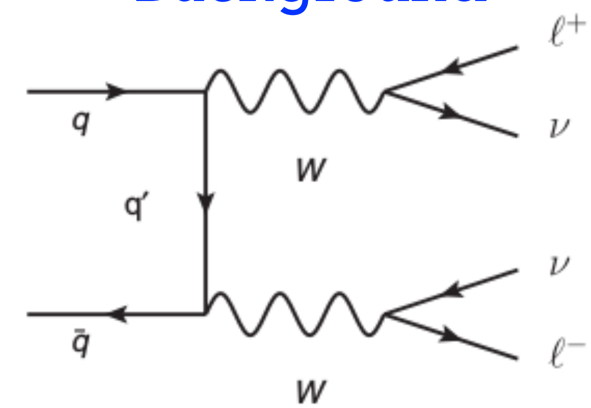
Low-level features: $p_T^{l_1}, p_T^{l_2}, \sum p_T^j, MET, N_j$

High-level features: Axial MET, M_{T_2}, razor quantities

Signal



Background



SUSY benchmark:

chargino production (lepton+MET final state)

Technique	Low-level	High-level	Complete
AUC			
BDT	0.850 (0.003)	0.835 (0.003)	0.863 (0.003)
NN	0.867 (0.002)	0.863 (0.001)	0.875 (<0.001)
$NN_{dropout}$	0.856 (<0.001)	0.859 (<0.001)	0.873 (<0.001)
DN	0.872 (0.001)	0.865 (0.001)	0.876 (<0.001)
$DN_{dropout}$	0.876 (<0.001)	0.869 (<0.001)	0.879 (<0.001)
Discovery significance			
NN	6.5σ	6.2σ	6.9σ
DN	7.5σ	7.3σ	7.6σ

BDT, boosted decision tree; DN, deep neural network; NN, shallow neural network; SUSY, supersymmetry particle.

Neural Networks in a Nutshell

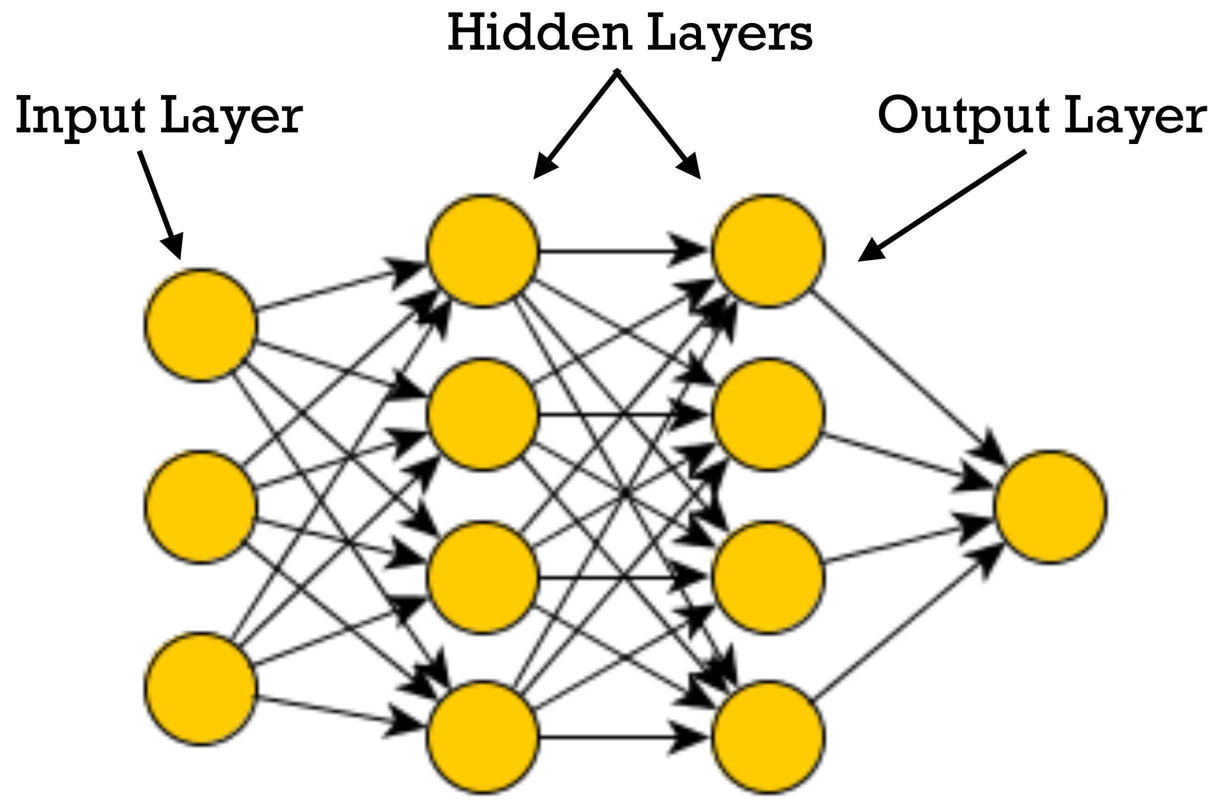
Neural Network (NN)
Basic Structure

Training sample validation sample test (real) sample

A. Input layer nodes: set of observables(kinematical features)/images

B. Number of hidden layers (shallow or deep NN)

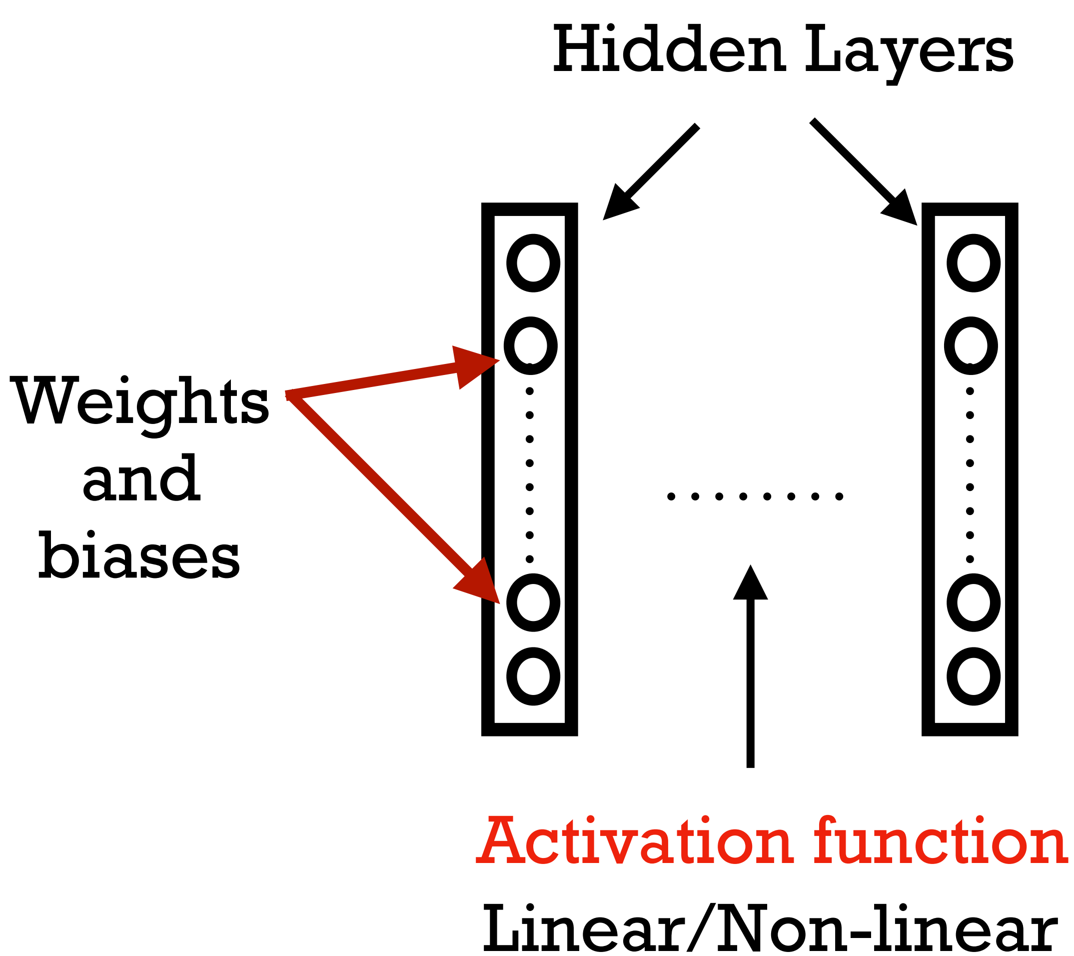
C. Output layer: predictions

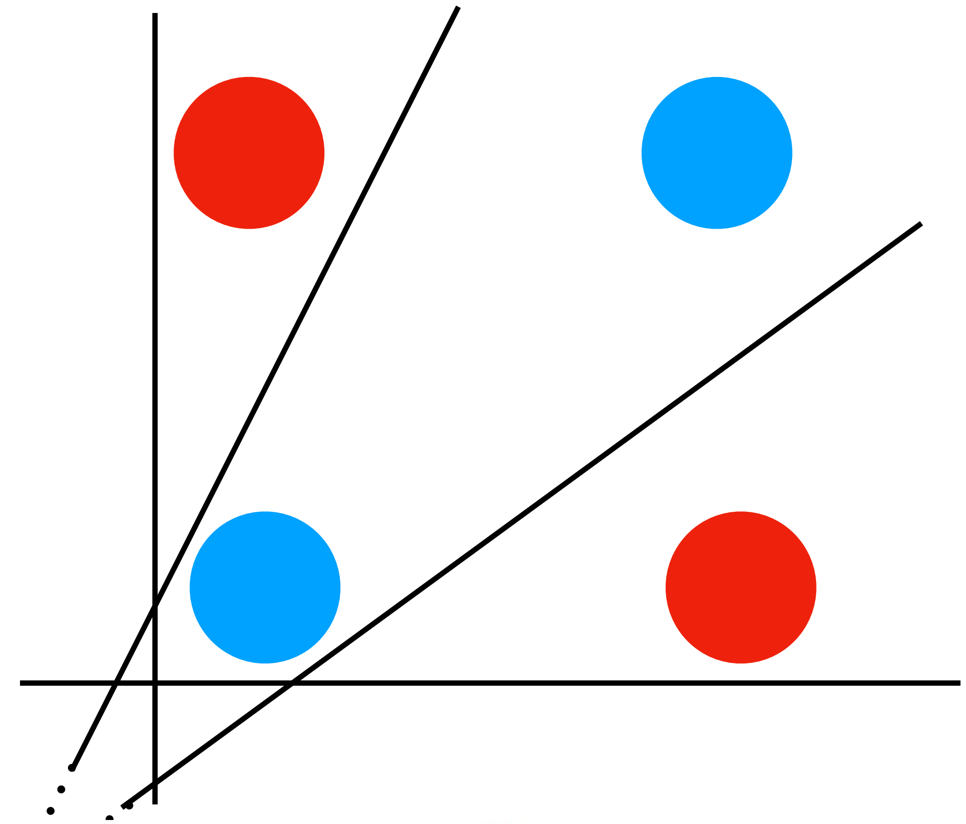


Schematic of a Neural Network

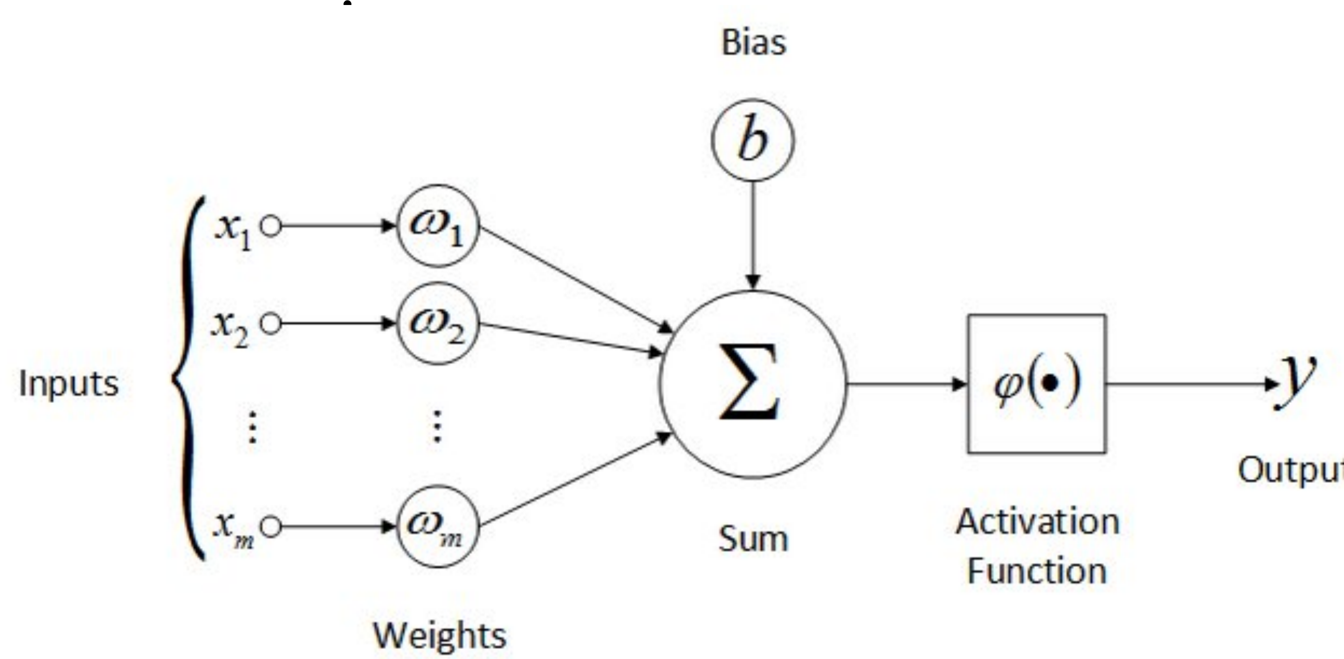
Train the network using training sample and make predictions for the test (real) dataset

How To Train Your NN?





More layers enables network
to learn
more complex features



NN Parameters and Concepts

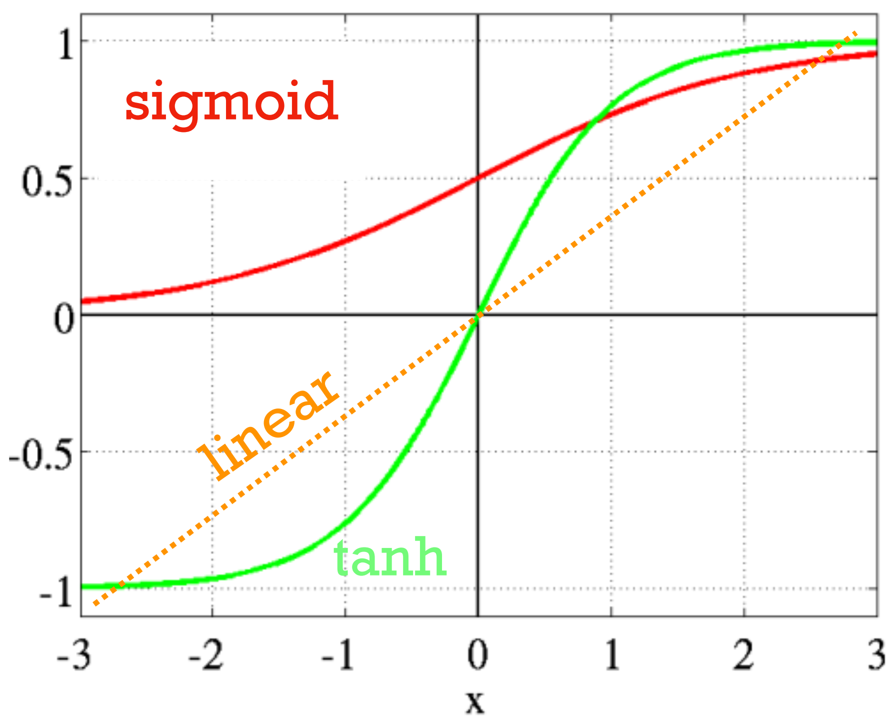
Loss function: $|\hat{y} - y|^2$ $\quad \hat{y} = f(x, \theta) \approx y$

NN parameters

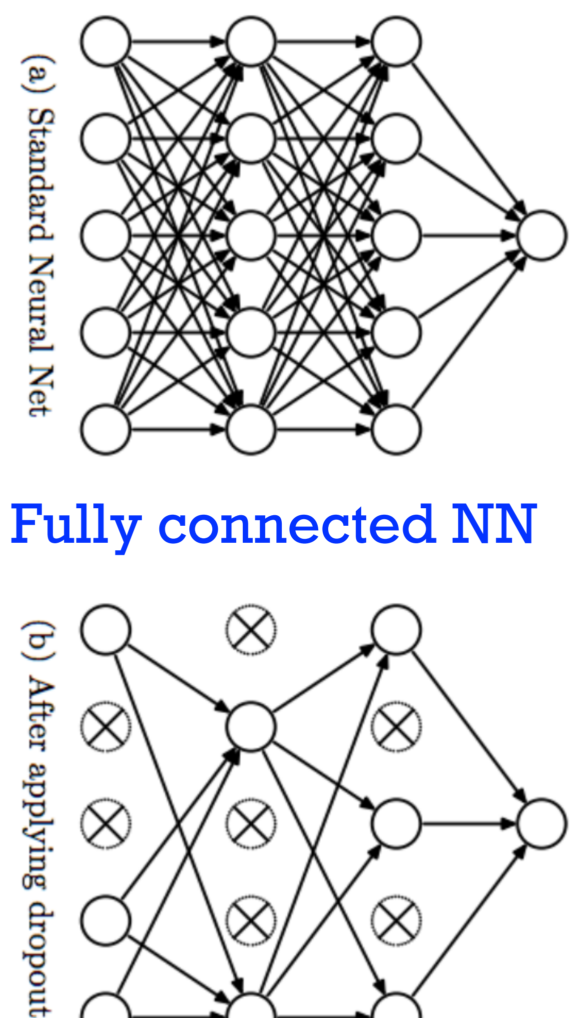
Epochs:
Counter for number of times complete data set is explored by the algorithm

Activation functions

common examples



Dropouts (regularisation)

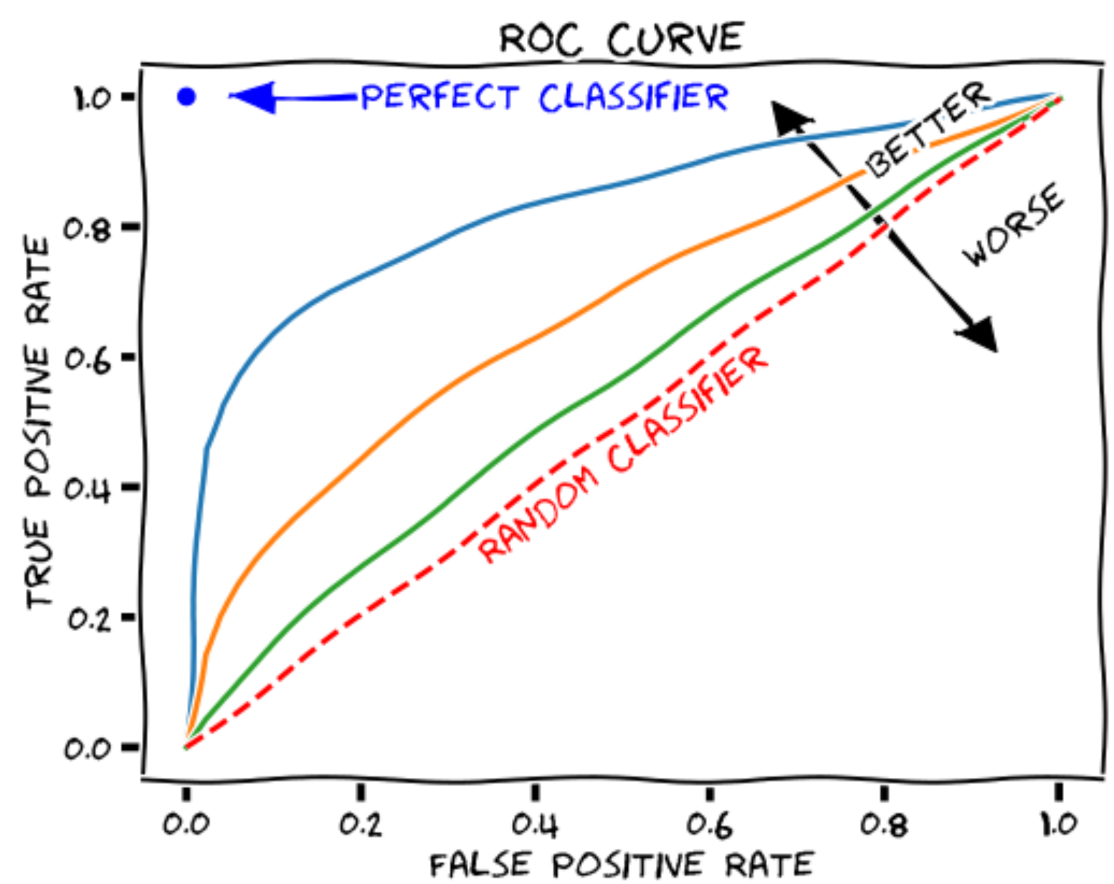


Regularised NN

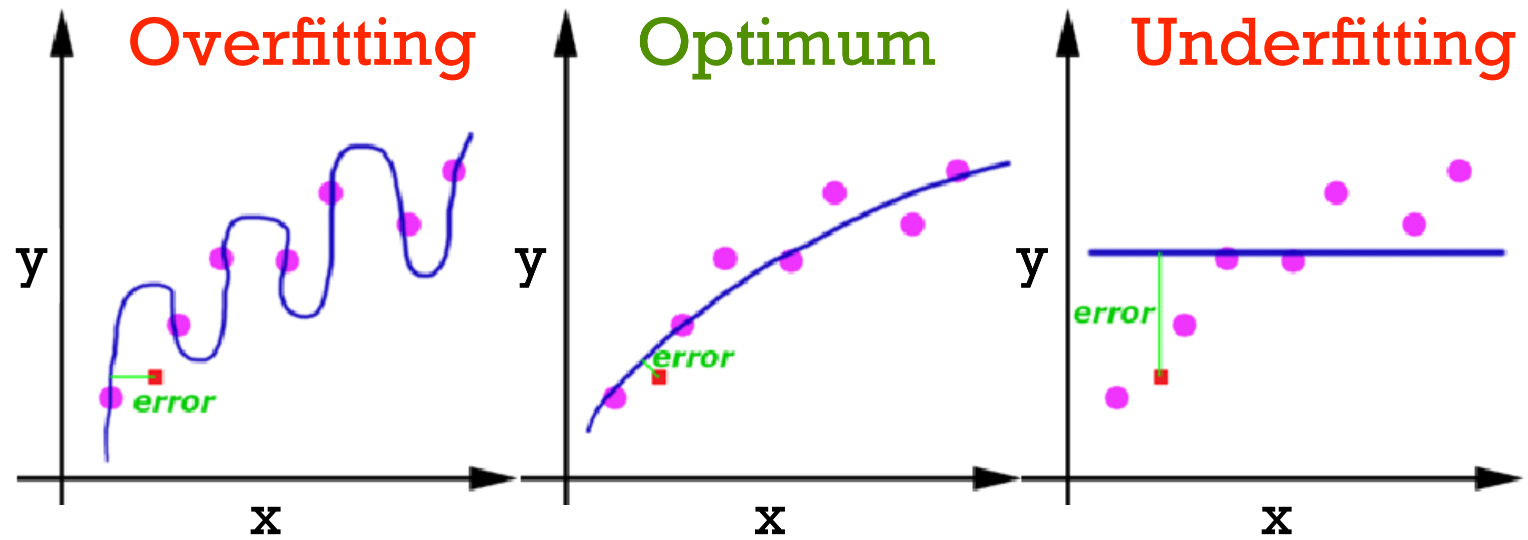
ROC Curve

True positive (signal selection)

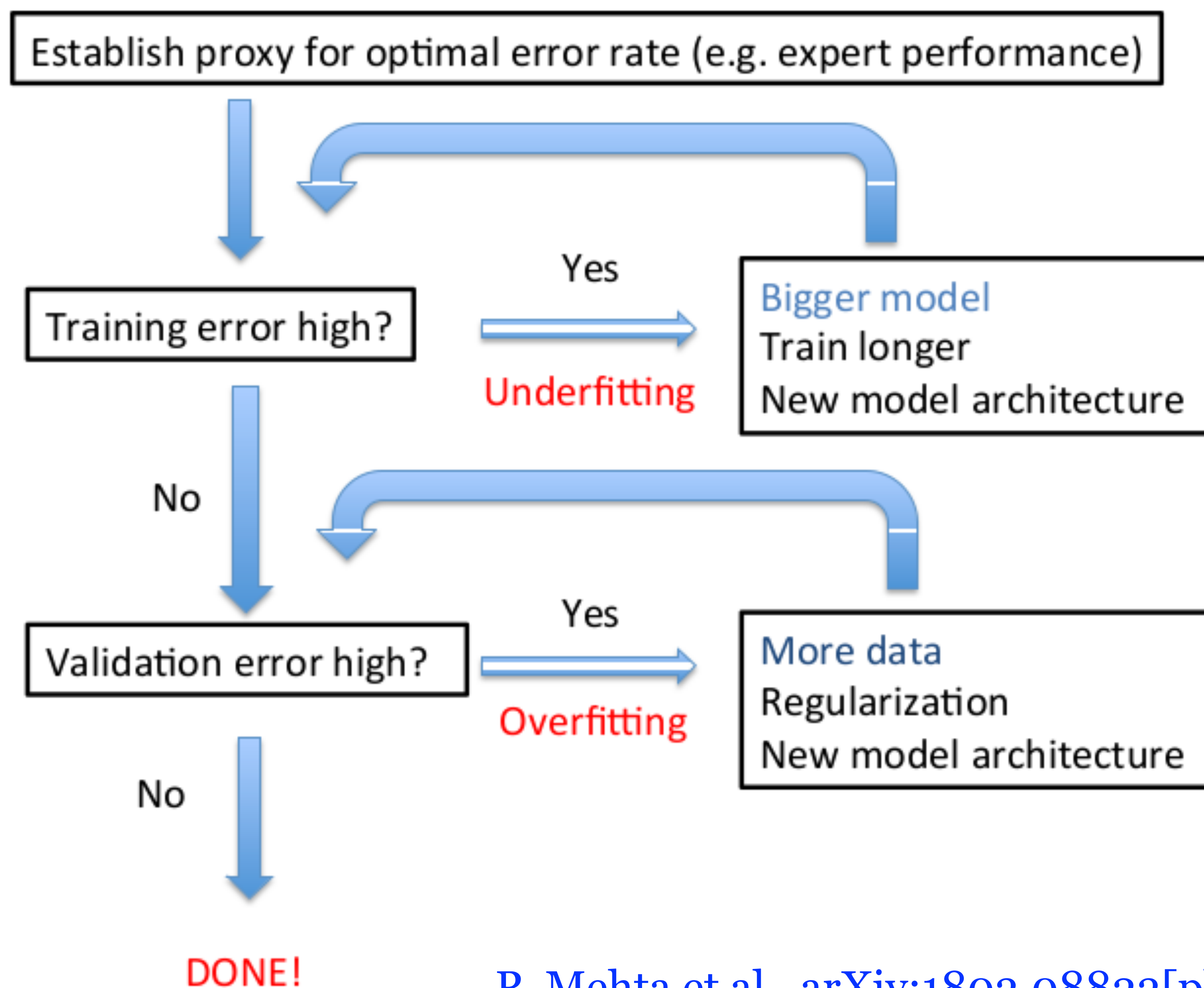
False positive (background selection)



Bias-Variance Trade-Off

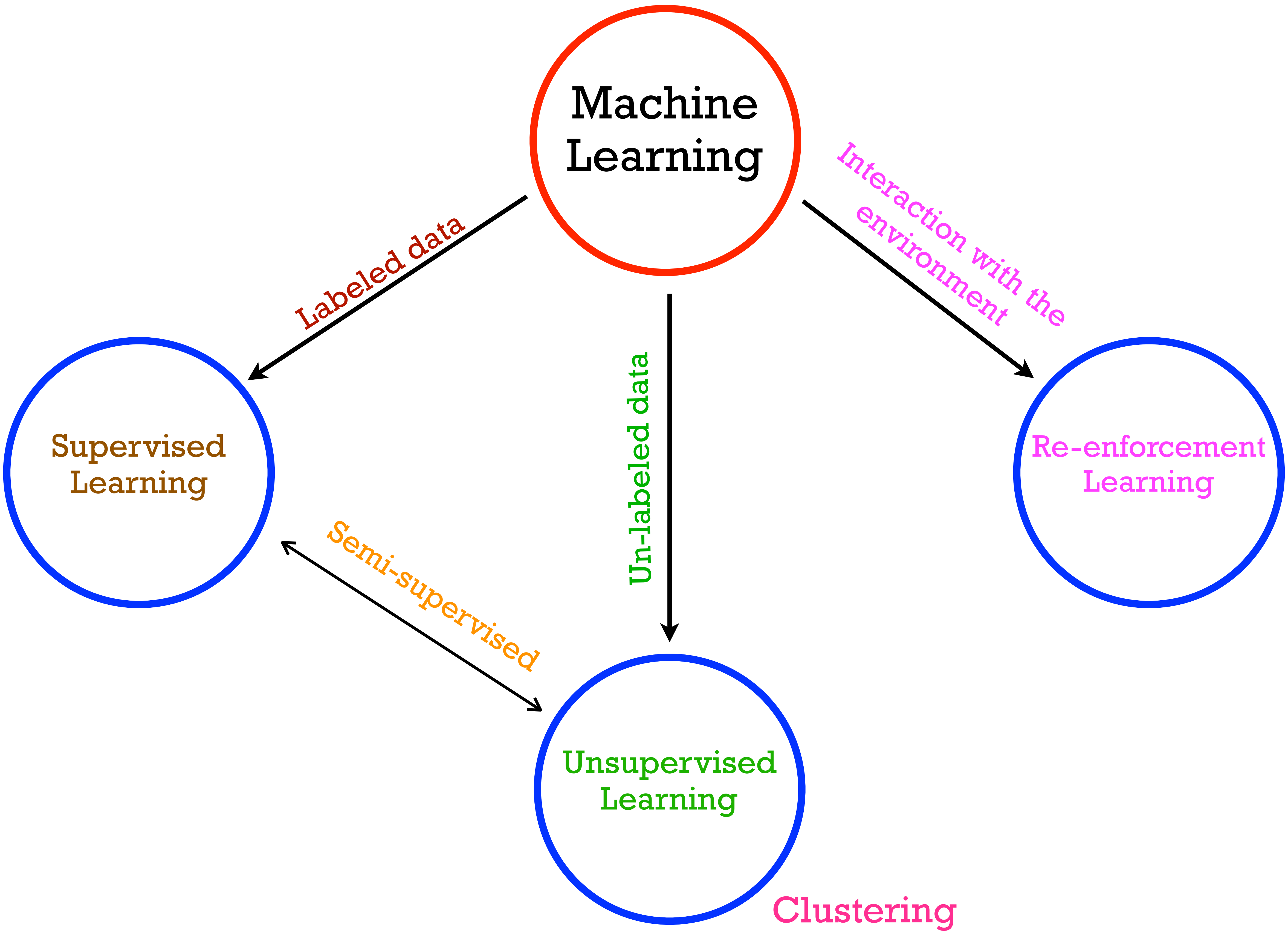


“Huge” datasets pose a computational challenge

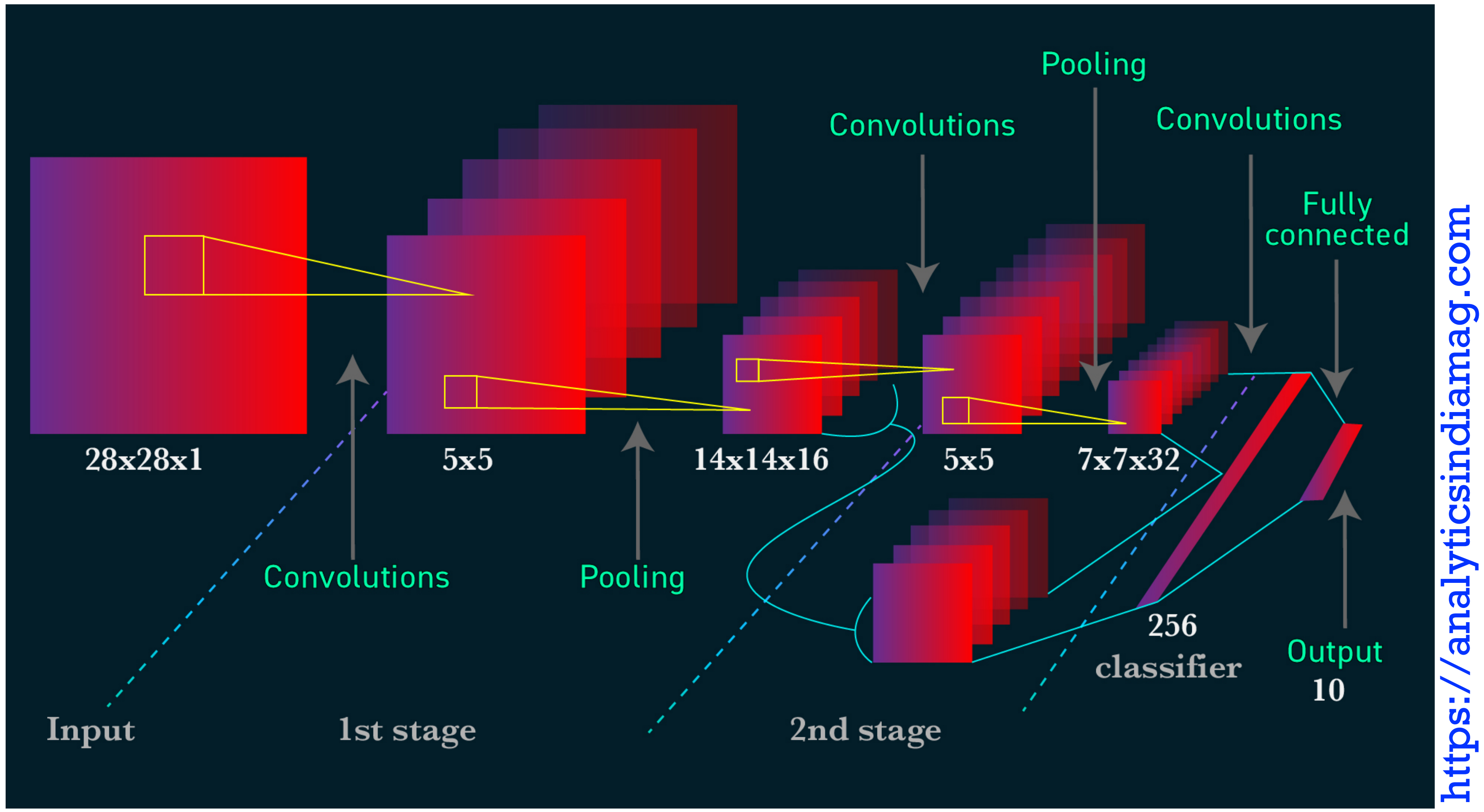


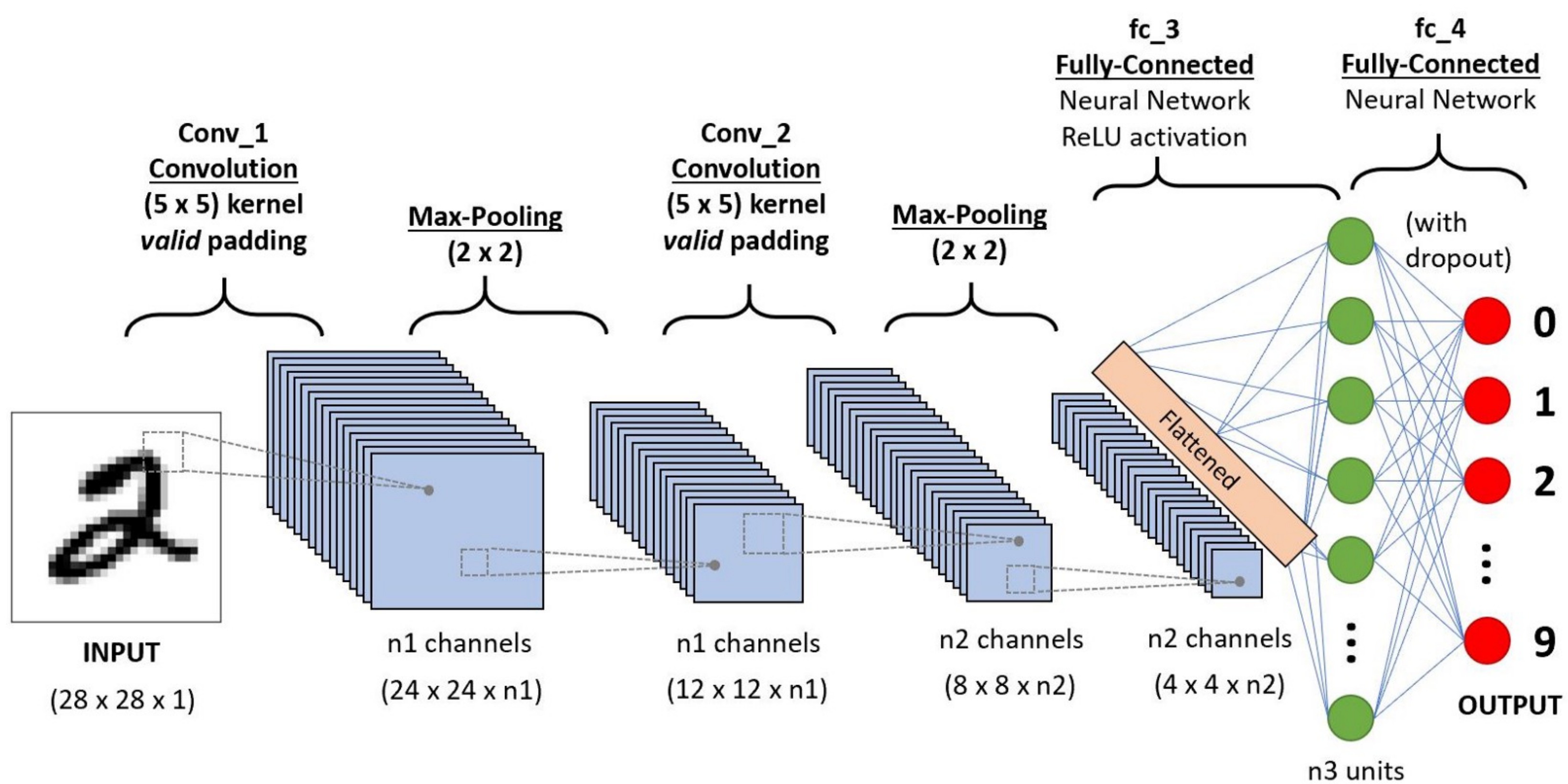
P. Mehta et al., arXiv:1803.08823[physics.comp-ph]

Broad Categories of Machine Learning

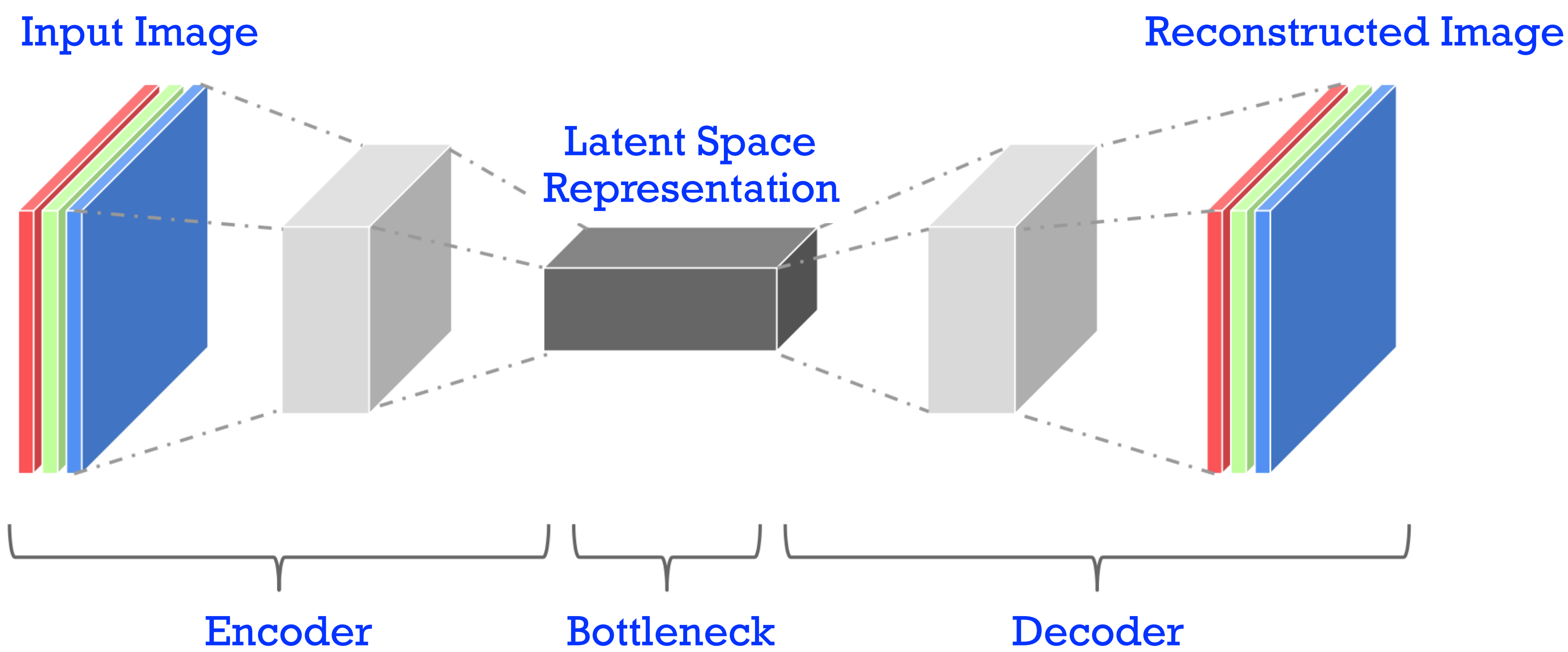


Convolutional Neural Networks





Autoencoders



Could be used as anomaly detector:

1. Train with the background sample.
2. Compare how the reconstructed output is different from the input (reconstructed error).

Reconstructed error will be more for the anomalous event.

T.Heimel, G.Kasieczka, T.Plehn and J.M.Thompson, SciPost Phys.6 (2019), 030
M. Farina, Y. Nakai and D. Shih, arXiv: 1808.08992

nature COMMUNICATIONS

ARTICLE

Received 19 Feb 2014 | Accepted 4 Jun 2014 | Published 2 Jul 2014 DOI: 10.1038/ncomms5308

Searching for exotic particles in high-energy physics with deep learning

P. Baldi[1], P. Sadowski[1] & D. Whiteson[2]

Deep Learning methods improve the reach of the collider searches for “new physics searches”

Collisions at high-energy particle colliders are a traditionally fruitful source of exotic particle discoveries. Finding these rare particles requires solving difficult signal-versus-background classification problems, hence machine-learning approaches are often used. Standard approaches have relied on 'shallow' machine-learning models that have a limited capacity to learn complex nonlinear functions of the inputs, and rely on a painstaking search through manually constructed nonlinear features. Progress on this problem has slowed, as a variety of techniques have shown equivalent performance. Recent advances in the field of deep learning make it possible to learn more complex functions and better discriminate between signal and background classes. Here, using benchmark data sets, we show that deep-learning methods need no manually constructed inputs and yet improve the classification metric by as much as 8% over the best current approaches. This demonstrates that deep-learning approaches can improve the power of collider searches for exotic particles.

Signal

Background

SUSY benchmark:
(chargino pair production, dilepton+MET final state)

Technique	Low-level	High-level	Complete
AUC			
BDT	0.850 (0.003)	0.835 (0.003)	0.863 (0.003)
NN	0.867 (0.002)	0.863 (0.001)	0.875 (<0.001)
$NN_{dropout}$	0.856 (<0.001)	0.859 (<0.001)	0.873 (<0.001)
DN	0.872 (0.001)	0.865 (0.001)	0.876 (<0.001)
$DN_{dropout}$	0.876 (<0.001)	0.869 (<0.001)	0.879 (<0.001)
Discovery significance			
NN	6.5σ	6.2σ	6.9σ
DN	7.5σ	7.3σ	7.6σ

BDT, boosted decision tree; DN, deep neural network; NN, shallow neural network; SUSY, supersymmetry particle.

Low-level features: $p_T^{l_1}, p_T^{l_2}, \sum p_T^j, MET, N_j$

High-level features: Axial MET, M_{T_2}, razor quantities

More Recent Review

REVIEW

https://doi.org/10.1038/s41586-018-0361-2

Machine learning at the energy and intensity frontiers of particle physics

Alexander Radovic[1]*, Mike Williams[2]*, David Rousseau[3], Michael Kagan[4], Daniele Bonacorsi[5,6], Alexander Himmel[7], Adam Aurisano[8], Kazuhiro Terao[4] & Taritree Wongjirad[9]

“Machine Learning techniques increase the discovery potential of the experiments.”

Table 1 | E[illegible] study of the Higgs boson

Analysis	Years of data collection	Sensitivity without machine learning	Sensitivity with machine learning	Ratio of P values	Additional data required
CMS[24] $H \to \gamma\gamma$	2011–2012	2.2σ, $P = 0.014$	2.7σ, $P = 0.0035$	4.0	51%
ATLAS[43] $H \to \tau^+\tau^-$	2011–2012	2.5σ, $P = 0.0062$	3.4σ, $P = 0.00034$	18	85%
ATLAS[99] $VH \to bb$	2011–2012	1.9σ, $P = 0.029$	2.5σ, $P = 0.0062$	4.7	73%
ATLAS[41] $VH \to bb$	2015–2016	2.8σ, $P = 0.0026$	3.0σ, $P = 0.00135$	1.9	15%
CMS[100] $VH \to bb$	2011–2012	1.4σ, $P = 0.081$	2.1σ, $P = 0.018$	4.5	125%

A. Radovic et al., Natur

REVIEW

https://doi.org/10.1038/s41586-018-0361-2

Machine learning at the energy and intensity frontiers of particle physics

Alexander Radovic[1]*, Mike Williams[2]*, David Rousseau[3], Michael Kagan[4], Daniele Bonacorsi[5,6], Alexander Himmel[7], Adam Aurisano[8], Kazuhiro Terao[4] & Taritree Wongjirad[9]

Our knowledge of the fundamental particles of nature and their interactions is summarized by the standard model of particle physics. Advancing our understanding in this field has required experiments that operate at ever higher energies and intensities, which produce extremely large and information-rich data samples. The use of machine-learning techniques is revolutionizing how we interpret these data samples, greatly increasing the discovery potential of present and future experiments. Here we summarize the challenges and opportunities that come with the use of machine learning at the frontiers of particle physics.

Table 1 | Effect of machine learning on the discovery and study of the Higgs boson

Analysis	Years of data collection	Sensitivity without machine learning	Sensitivity with machine learning	Ratio of P values	Additional data required
CMS[24] $H \to \gamma\gamma$	2011–2012	2.2σ, $P = 0.014$	2.7σ, $P = 0.0035$	4.0	51%
ATLAS[43] $H \to \tau^+\tau^-$	2011–2012	2.5σ, $P = 0.0062$	3.4σ, $P = 0.00034$	18	85%
ATLAS[99] $VH \to bb$	2011–2012	1.9σ, $P = 0.029$	2.5σ, $P = 0.0062$	4.7	73%
ATLAS[41] $VH \to bb$	2015–2016	2.8σ, $P = 0.0026$	3.0σ, $P = 0.00135$	1.9	15%
CMS[100] $VH \to bb$	2011–2012	1.4σ, $P = 0.081$	2.1σ, $P = 0.018$	4.5	125%

A. Radovic et al., Nature 560(2018) no. 7716,41

ML Techniques at Various Experimental Analysis Steps

A. Decision-making for data storage
B. Jet reconstruction and heavy flavor tagging
C. Track reconstruction
D. Signal and background classification
E. Event generation

New Potential for BSM searches

- SMEFT (new physics deformations)
- Top tagging
- Anomaly detection
- Dark matter searches
- DNN likelihood
- Decoding black box
- Many more..

[Incomplete list of references]

- J. Brehmer et al., Phys. Rev. Lett. 121 (2018) 111801, Phys. Rev. D 98 (2018) 052004, F.F. Freitas, **CKK**, V. Sanz, Phys. Rev. D 100 (2019) no.3, 035040 [arXiv:1902.05803 [hep-ph]].
- G.Kasieczka, T.Plehn, M.Russell and T.Schell, JHEP 1705 (2017) 006.
- arXiv: 1807.10261, Shih et al. 1808.08992[hep-ph]. Heimel et al., SciPost Phys.6 (2019), 030. **CKK** and Veronica Sanz, arXiv:2007.14462 [cs.LG] and others
- **CKK**, L. Mars, J. Richards and V. Sanz, J. Phys. G 47 (2020) no.9, 095201. **CKK**, V. Sanz and M. Soughton [arXiv:1910.06058 [hep-ph]].
- A.Coccaro, M.Pierini, L.Silvestrini and R. Torre, Eur. Phys. J. C 80 (2020) no.7, 664 [arXiv:1911.03305 [hep-ph]].
- G. Kasieczka, S. Marzani, G. Soyez and G.Stagnitto, JHEP 09 (2020), 195 [arXiv:2007.04319 [hep-ph]]. T. Faucett, J. Thaler and D. Whiteson, [arXiv:2010.11998 [hep-ph]].

(Deep)Neural Networks, CNNs, (V)Autoencoders, Clustering, GNNs,..

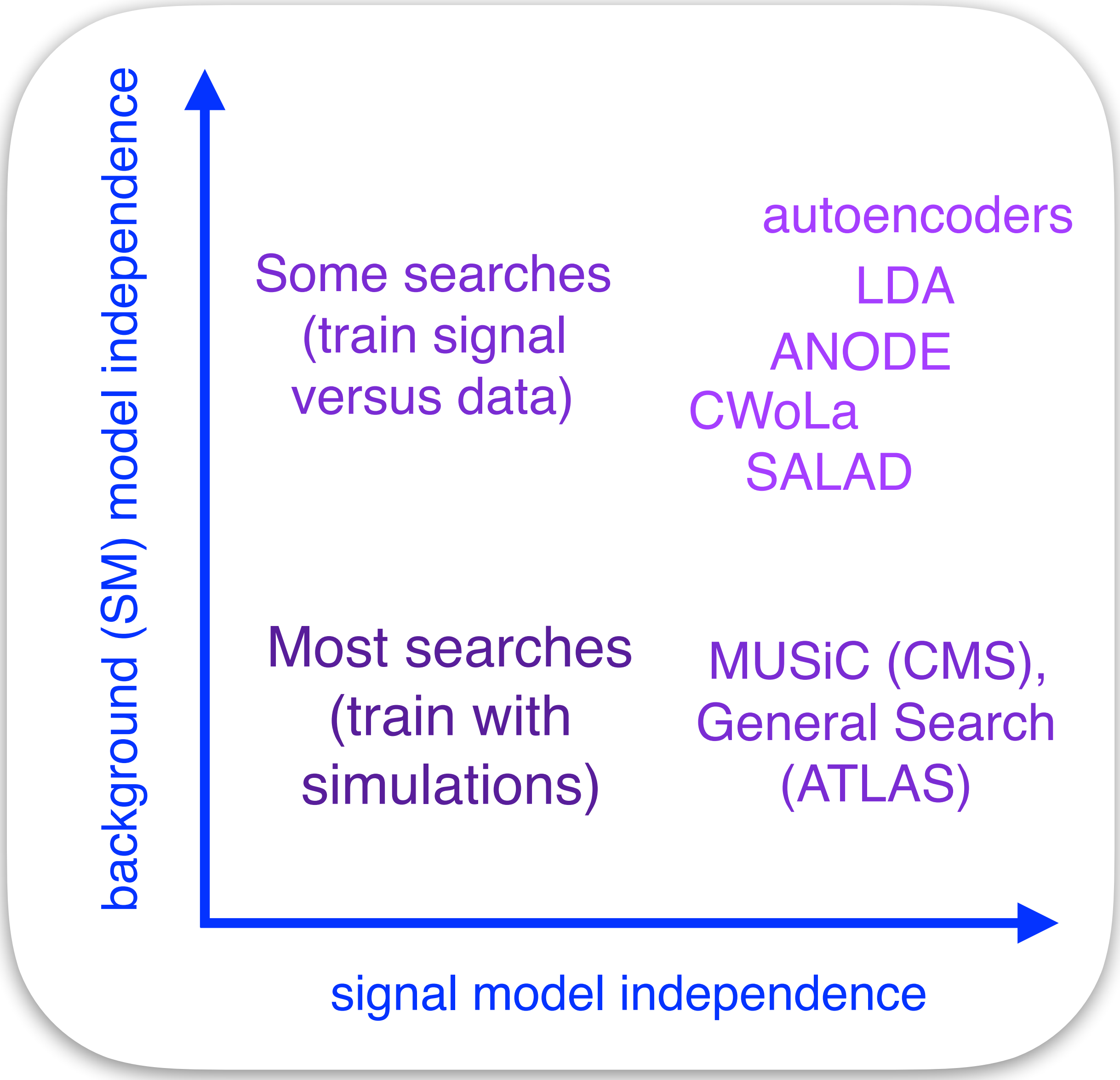
↓

Classification, Jet tagging, Anomalous Jet, Anomalous Events, Limit setting, Resonance,..

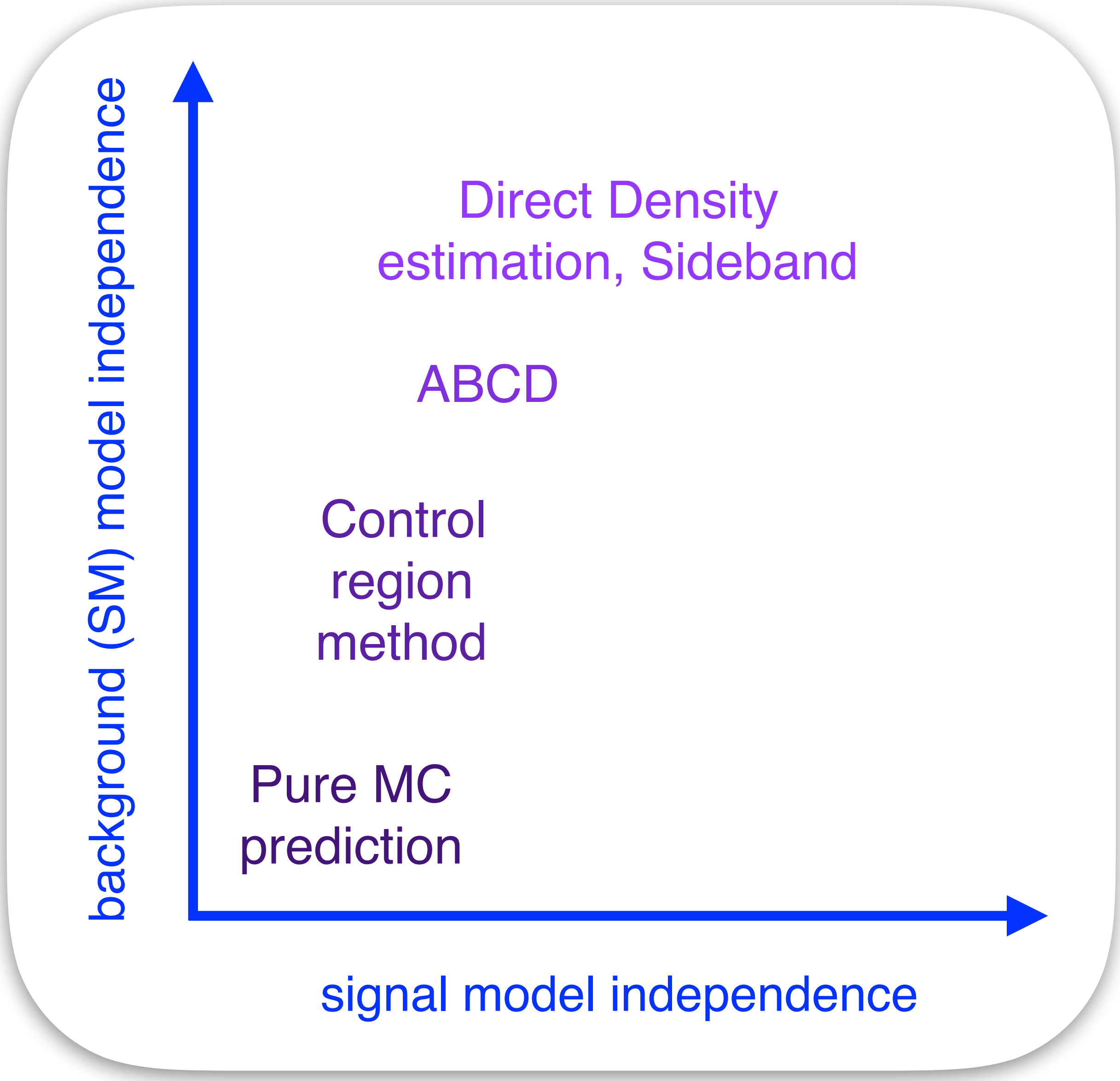
↓

Assisting BSM detection

Model Independent Searches



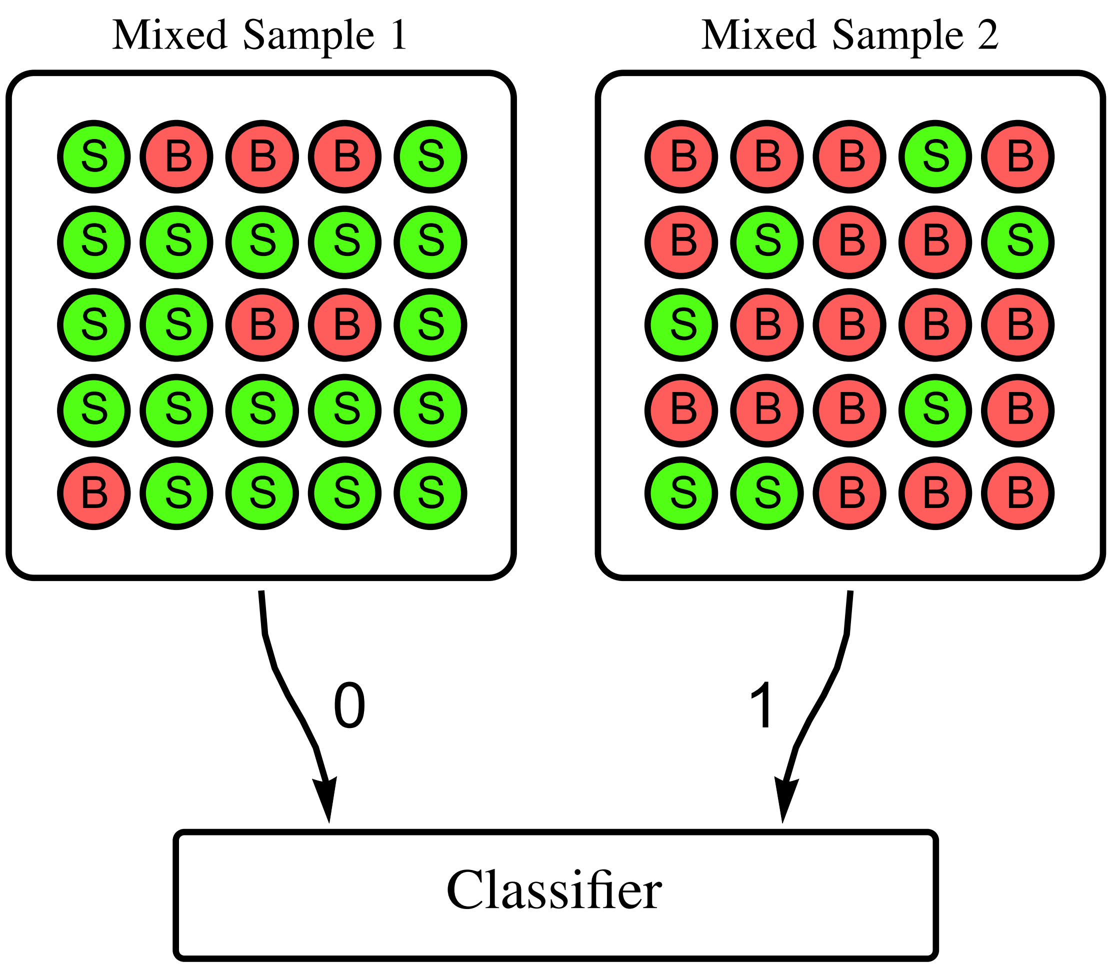
(a) Signal sensitivity

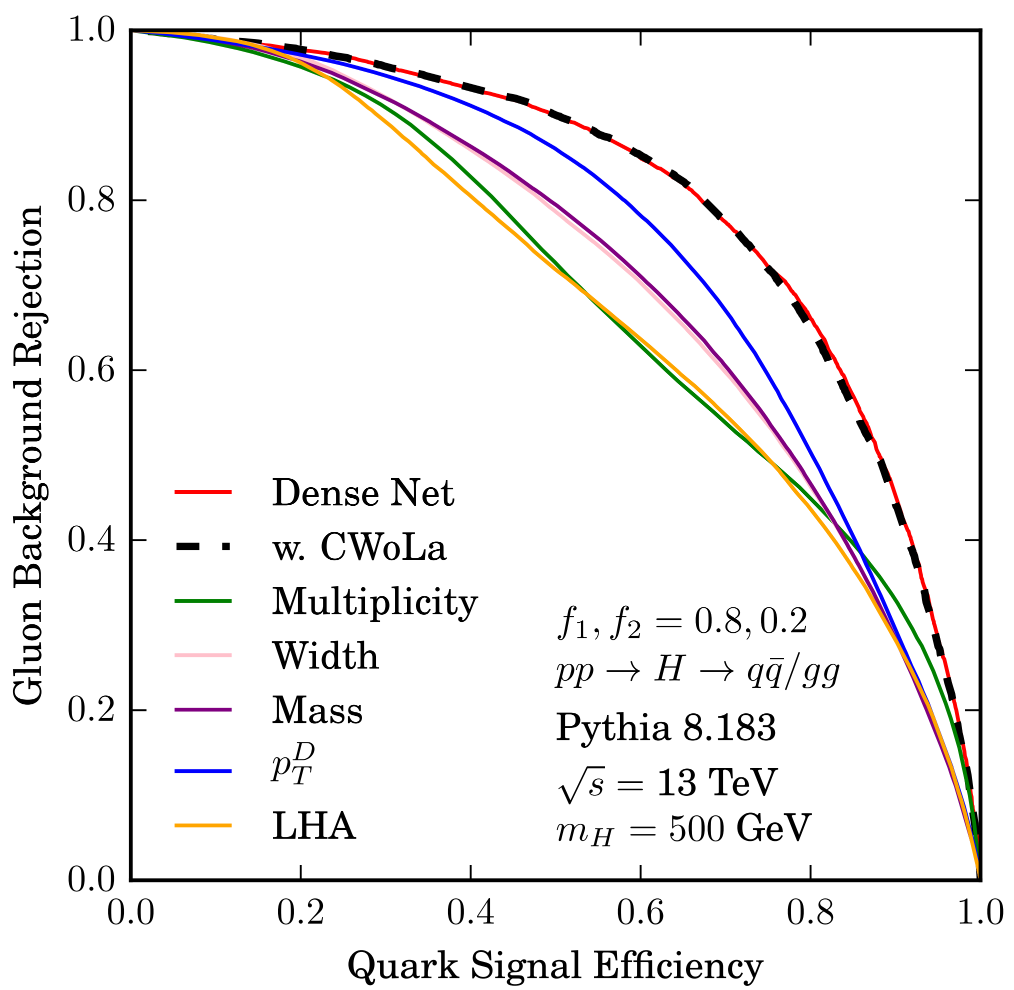


(b) Background specificity

Image taken from 2001.04990[arxiv: hep-ph]

Classification Without Labels (CWoLa)





E.M.Metodiev, B.Nachman and J.Thaler, JHEP 10 (2017), 174, [arXiv:1708.02949 [hep-ph]]

Other Recent Developments

- Learning New Physics from a Machine: R.T.D'Agnolo and A.Wulzer, Phys. Rev. D **99** (2019) no.1, 015014
- Guiding New Physics Searches with Unsupervised Learning: A. De Simone et al., Eur. Phys. J. C 79 (2019) no.4, 289, [arXiv:1807.06038].
- Using Variational Autoencoders: Cerri et al., JHEP 05 (2019), 036, [arXiv:1811.10276 [hep-ex]]. Cheng et al.,[arXiv:2007.01850].
- Uncovering latent jet substructure: B.M.Dillon et al., Phys. Rev. D 100 (2019) no.5, 056002)[arXiv:1904.04200].
- Tag N' Train: O.Amram and C.M.Suarez, [arXiv:2002.12376].
- Anti QCD tagger: J. A.Aguilar-Saavedra, J. H. Collins and R. K. Mishra, JHEP 11 (2017), 163, [arXiv:1709.01087].
- Anomaly Detection with Density Estimation: B. Nachman and D.Shih, Phys. Rev. D 101 (2020), 075042, [arXiv:2001.04990]
- Simulation Assisted Likelihood-free Anomaly Detection: A.Andreassen et al., Phys. Rev. D 101 (2020) no.9, 095004 [arXiv:2001.05001].

For other proposals, see recent review by Nachman:
Anomaly Detection for Physics Analysis and Less than Supervised Learning[arXiv:2010.14554 [hep-ph]].

Creating Anomaly Aware Methods

Anomaly Awareness (AA)
a new algorithm for anomaly detection
(C. K. Khosa & V. Sanz. 2020)

Our algorithm learns about normal events (SM) while being made aware of an array of anomalies (BSM) in a way that it becomes sensitive to unseen BSM anomalies.

Which type of BSM we consider?

Boosted Regime

What type of Input data?

Jet Images

Which model do we use?

CNN

- Lets see how AA works in a well-known topology for new physics searches: Fat Jets
- Demonstrate its use against an array of BSM scenarios:

1. EFT Higgs,
2. Resonances $\longrightarrow$ leading jet with 2, 3 or 4 subsets

The "Algorithm"

Algorithm 1 Anomaly Awareness (AA).
Important parameters are λ_{AA}, p_{An}^{min}, p_{An}^{max}.

Prior Run

Initialize test:train splitting of *Normal* (N) dataset
Initialize hyper parameters
Initialize Model (CNN architecture)
for Training over the epochs **do**
  Cross entropy loss
  Update model parameters.
end for
Get accuracy for D_{test} and D_{train}
This run sets the hyper-parameters for the AA run

Anomaly Detection Run

Load the *Anomaly* (An) dataset
Initialize amount of data w.r.t. the *Normal* dataset
Initialize λ_{AA}
for Training over the epochs **do**
  l_1 = Cross entropy loss (*Normal* dataset)
  l_2 = Cross entropy loss (*Anomaly* dataset with Uniform Distribution)
  Loss = $l_1 + \lambda_{AA} l_2$
end for
Get softmax probabilities for all the datasets,
p_i, $i = N, An$
Select datapoints in a range $[p_{An}^{min}, p_{An}^{max}]$,
range optimized to select anomaly over normal events

Top and QCD Jets

the input dataset

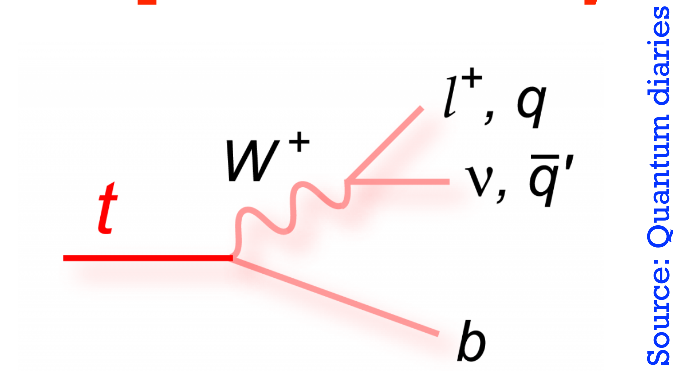
SM $t\bar{t}$ and QCD diJet production, $\sqrt{s} = 13\ TeV$

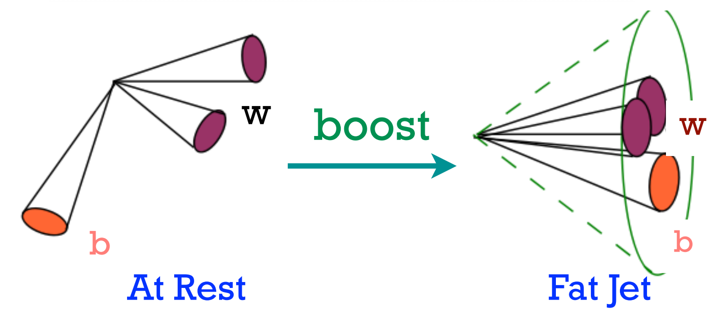
Madgraph + Pythia

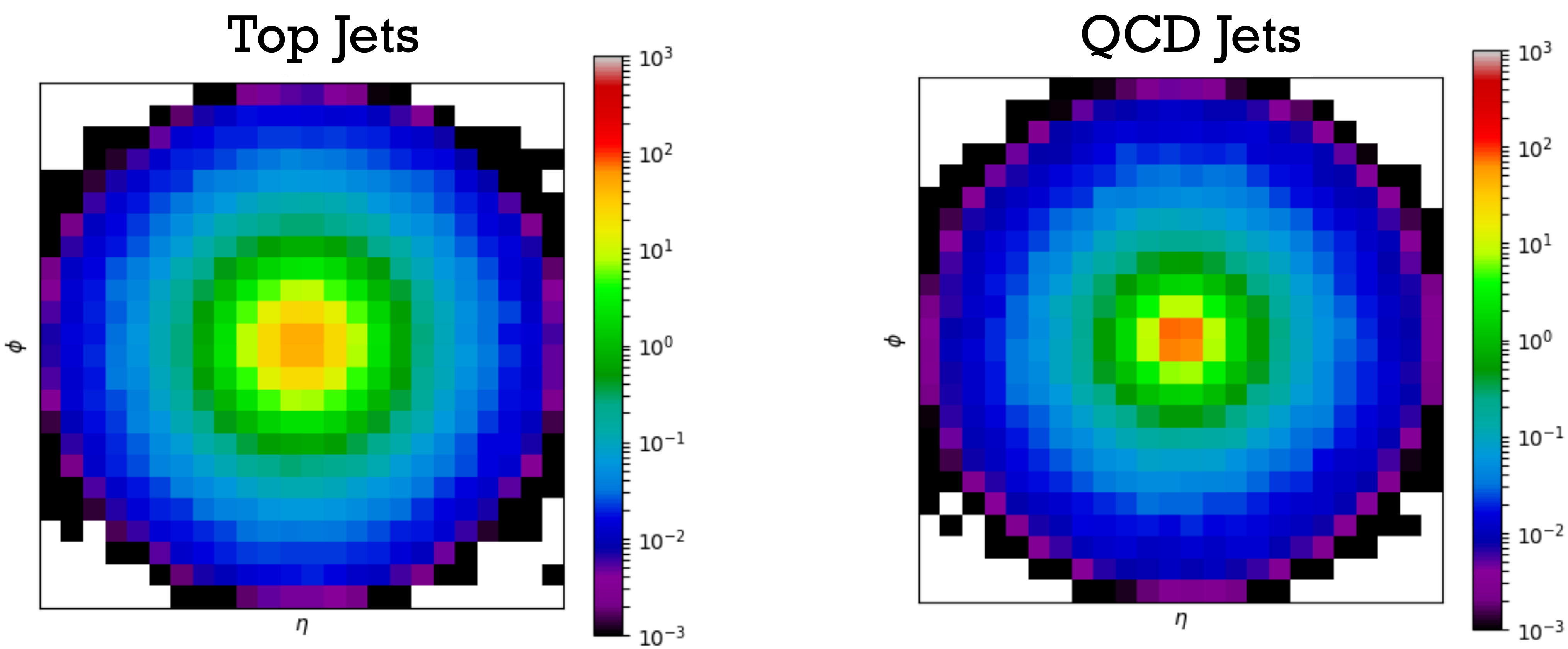
Leading jet with $p_t > 750\, GeV$, R=1 Anti-kt jet

$\Delta\eta = 0.087, \Delta\phi = 0.087$

Top Quark Decay

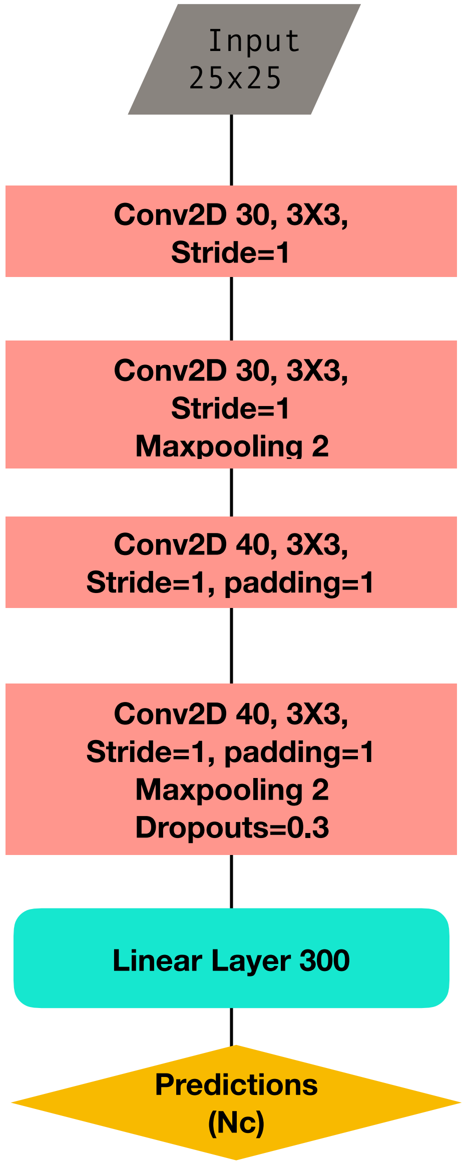






Averaged over 50,000 events

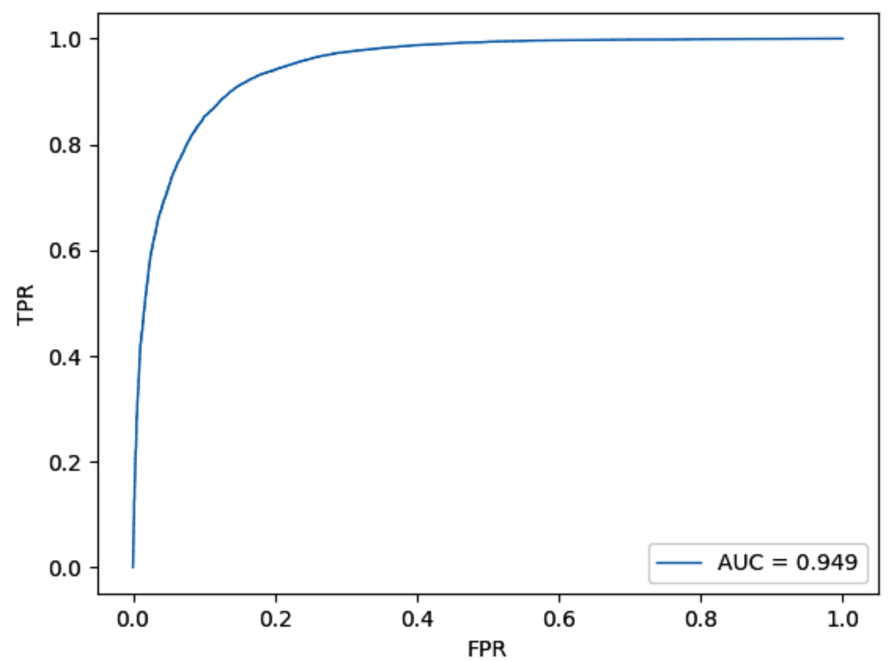
CNN for Top vs QCD Classification
(2C Baseline Classification)

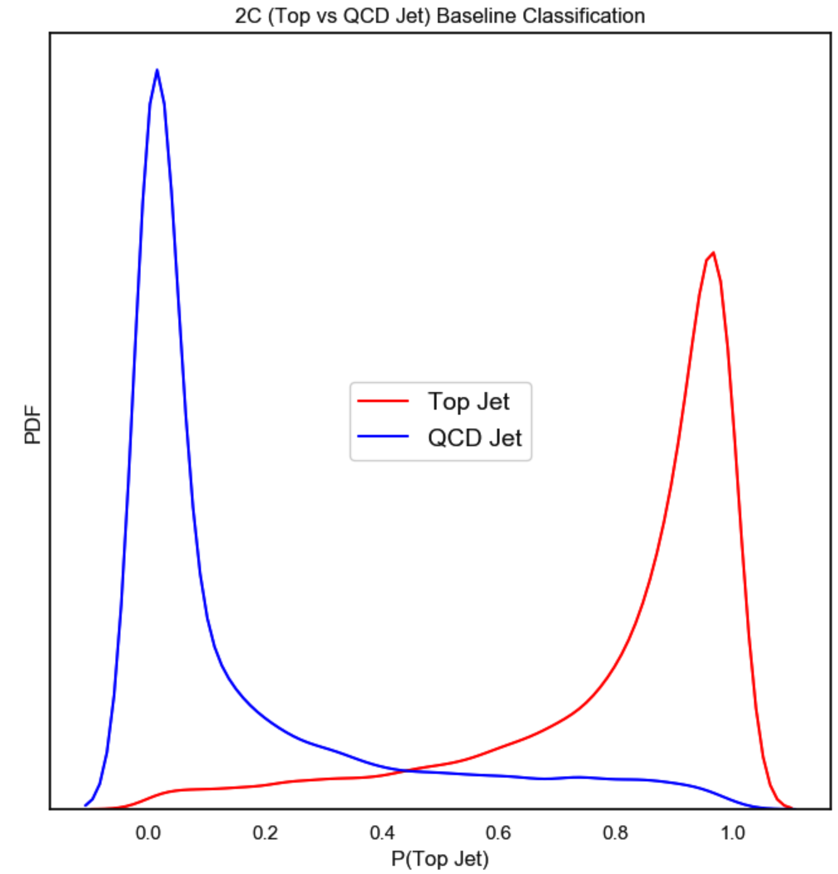


100K Events (balanced data)
Training: Test data= 70:30%

Batch Size=100
Epochs=100

Cross-Entropy Loss function





Related work
G.Kasieczka, T.Plehn, M.Russell and T.Schell, JHEP 05 (2017), 006
S.Macaluso and D.Shih, JHEP 10(2018), 121

BSM Benchmarks

SM

$$W jet : pp \rightarrow W^+W^-, W \rightarrow jj$$

HEFT C_{HW} $([D^\mu H^\dagger T_{2k} D^\nu H] W^k_{\mu\nu})$

$$\mathrm{EFT} : pp \rightarrow HZ, H \rightarrow b\bar{b}, Z \rightarrow l^+l^-$$

Resonance (Graviton) decay, mY=3 TeV

$$R_2 : pp \rightarrow Y \rightarrow ZZ, Z \rightarrow jj$$

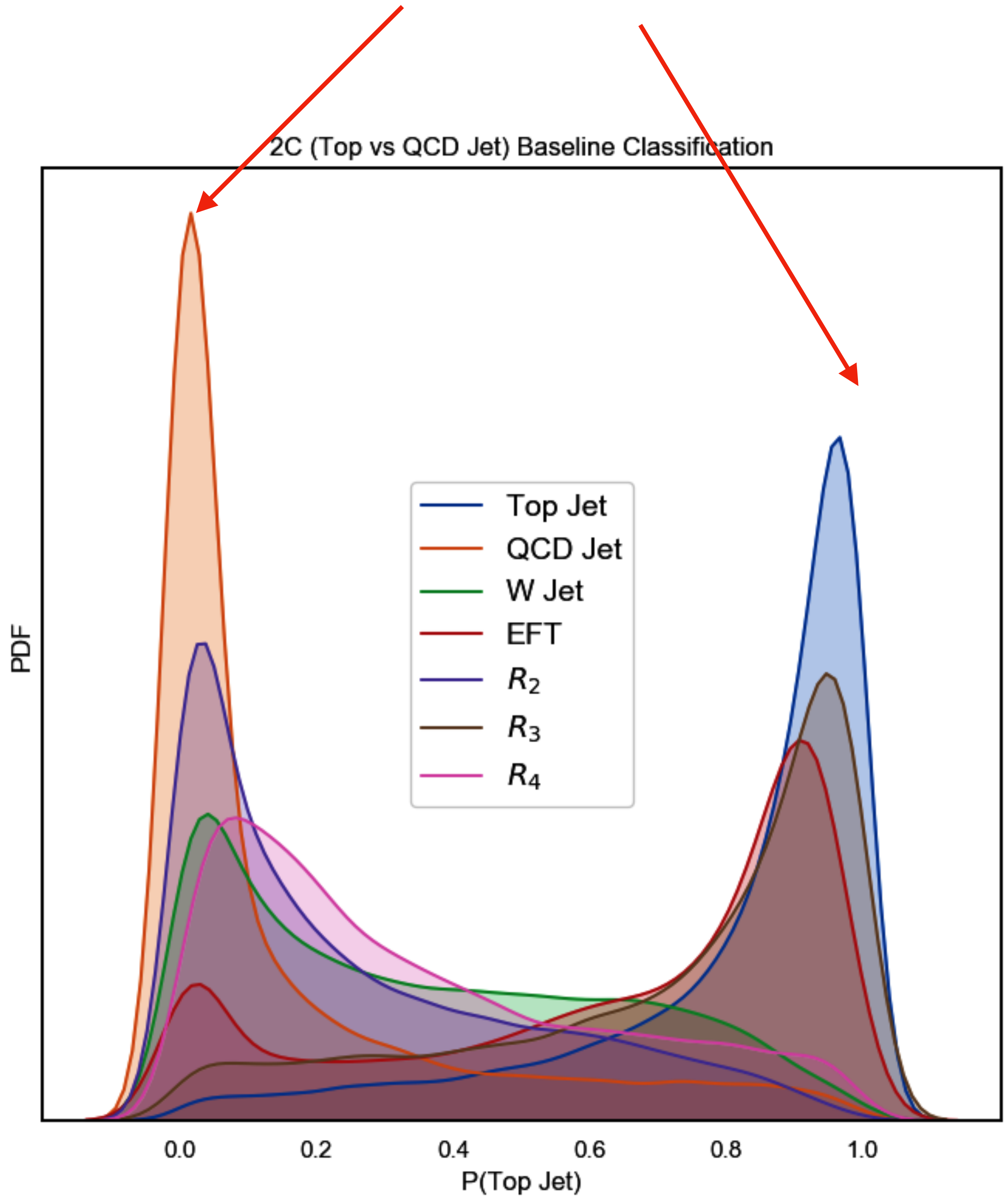
$$R_3 : pp \rightarrow Y \rightarrow t\bar{t}, t \rightarrow bW, W \rightarrow jj$$

$$R_4 : pp \rightarrow Y \rightarrow HH, H \rightarrow W^+W^-, W \rightarrow jj$$

Leading jet, same cuts as baseline jets

Top vs QCD Jet (Prior Run)

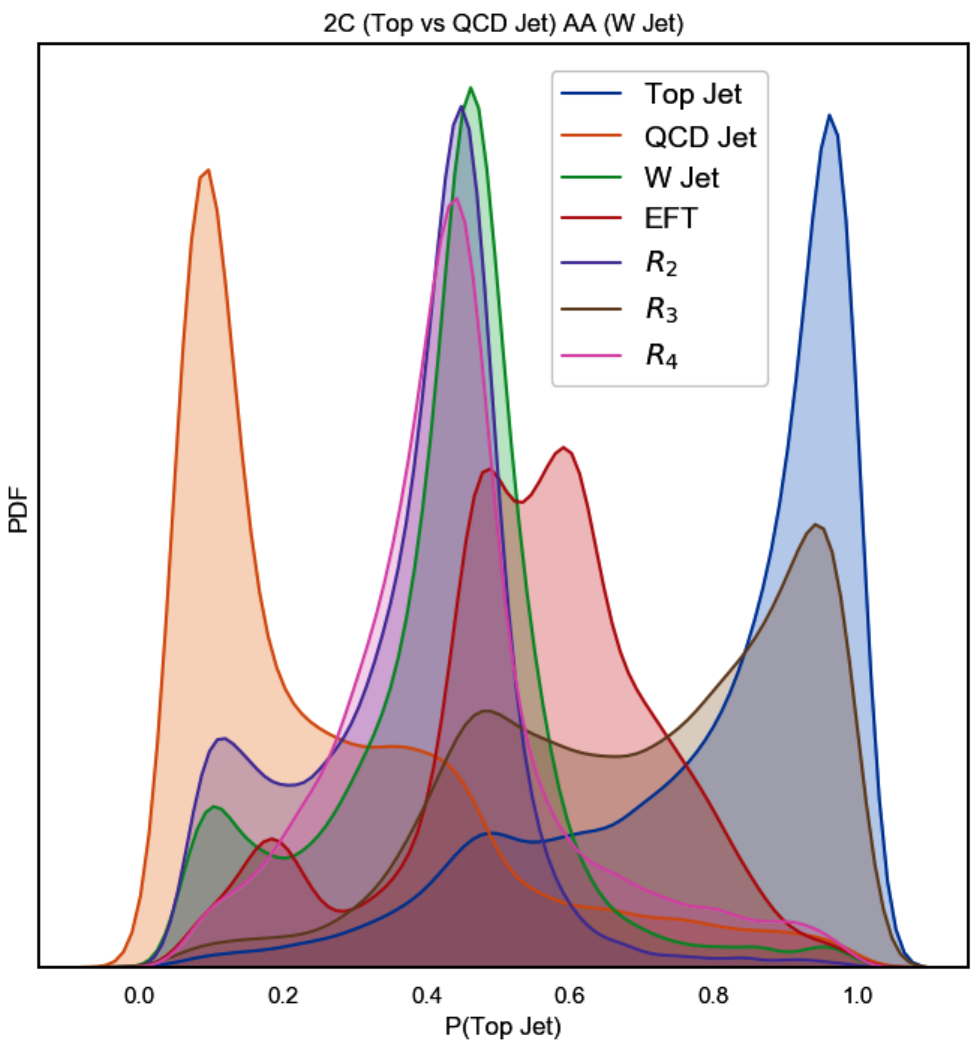
Training on



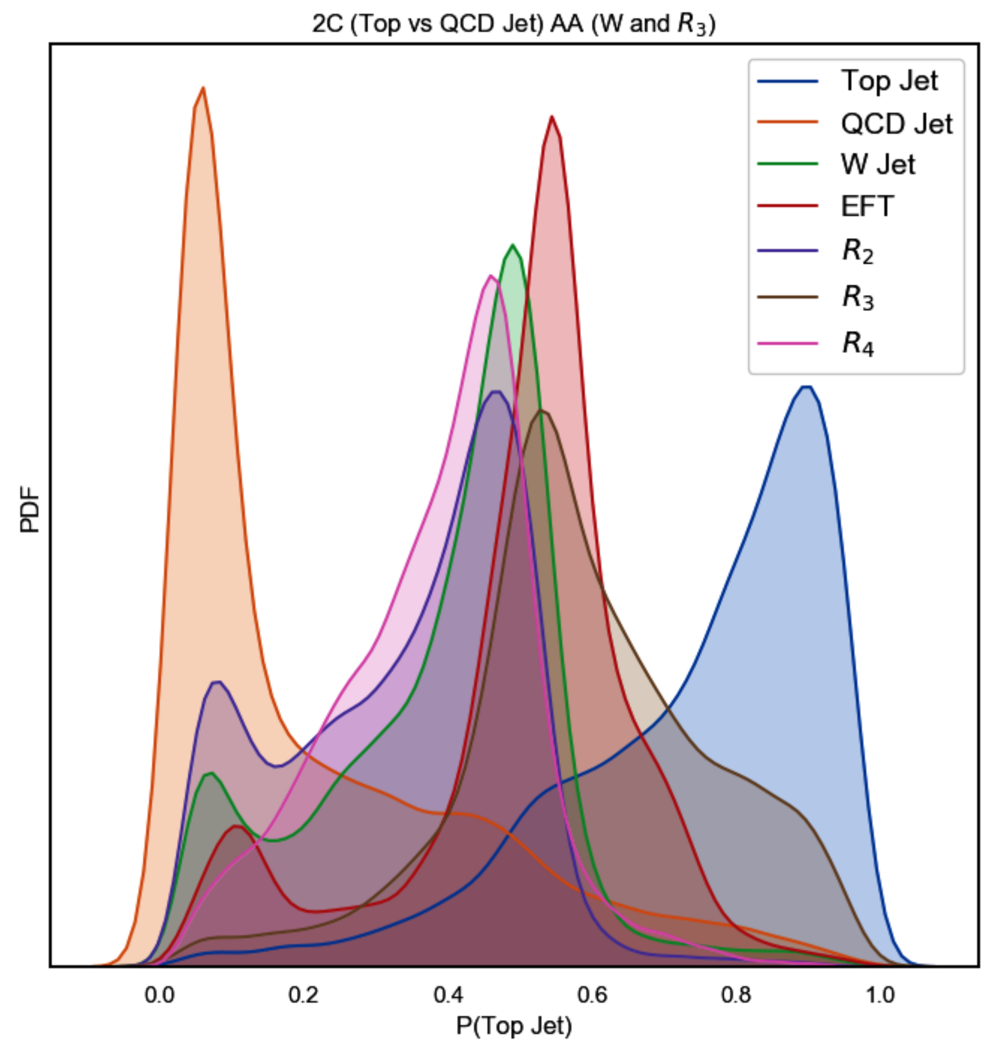
Anomaly Awareness for 2C Data

We see the effect of adding awareness to the classification task

One type of anomaly data set



Two types of anomaly data set

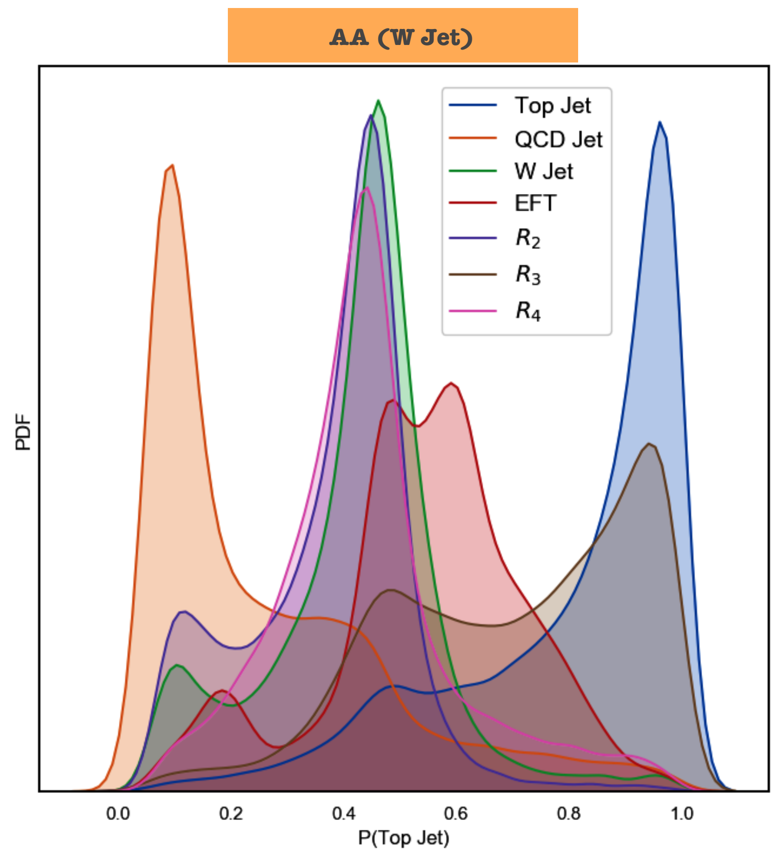


As we add more types of BSM examples, ALL BSMs gather in the centre

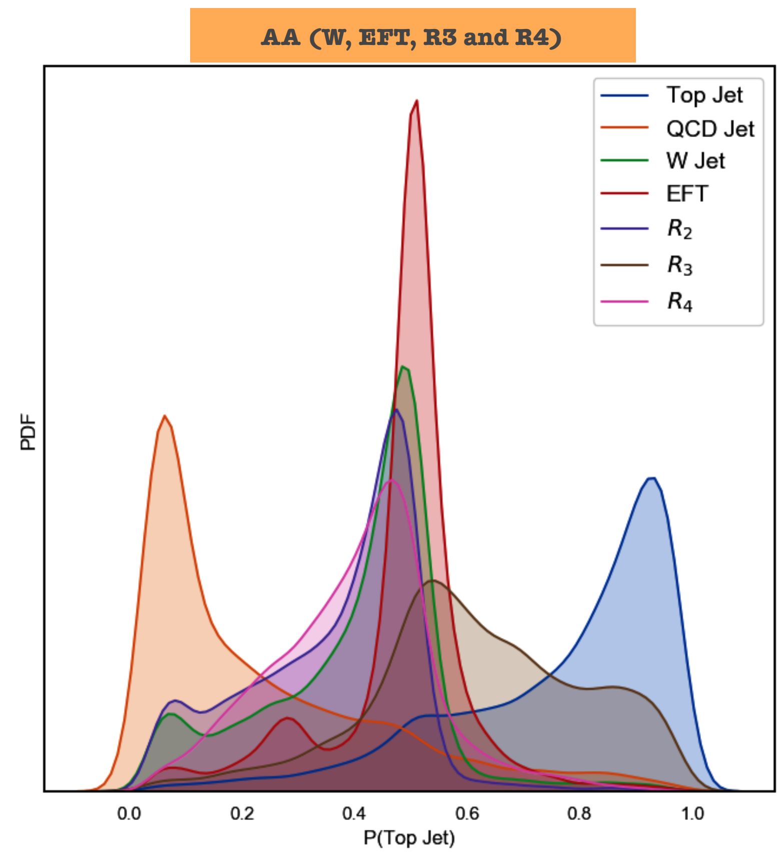
Uniform Distribution over the baseline classes for all the BSM events

Robust Anomaly Detector

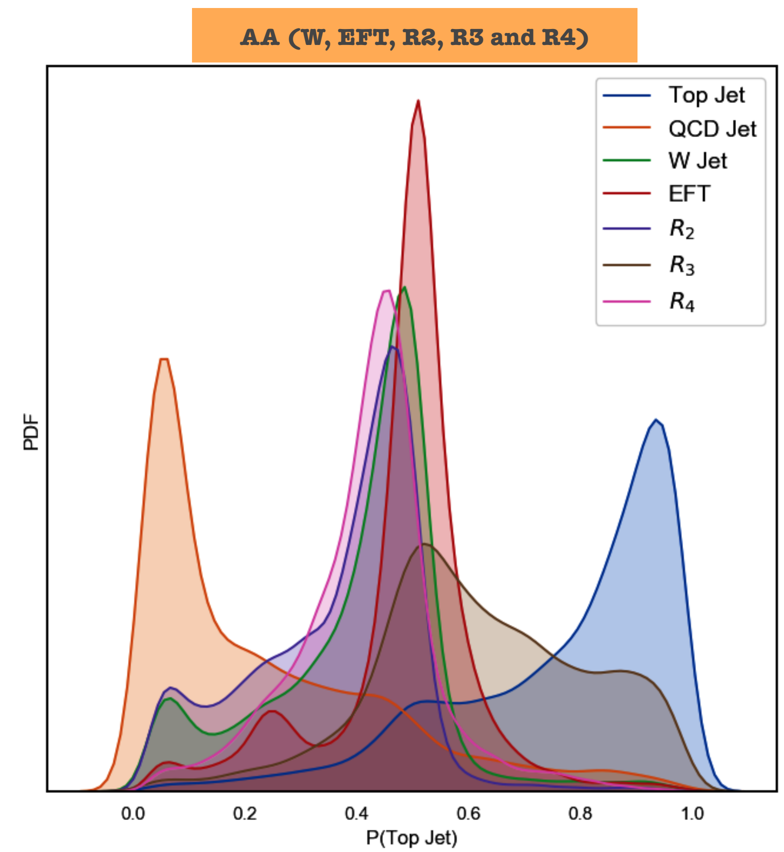
AA term with
1 type of anomaly



AA term with
4 types of anomalies

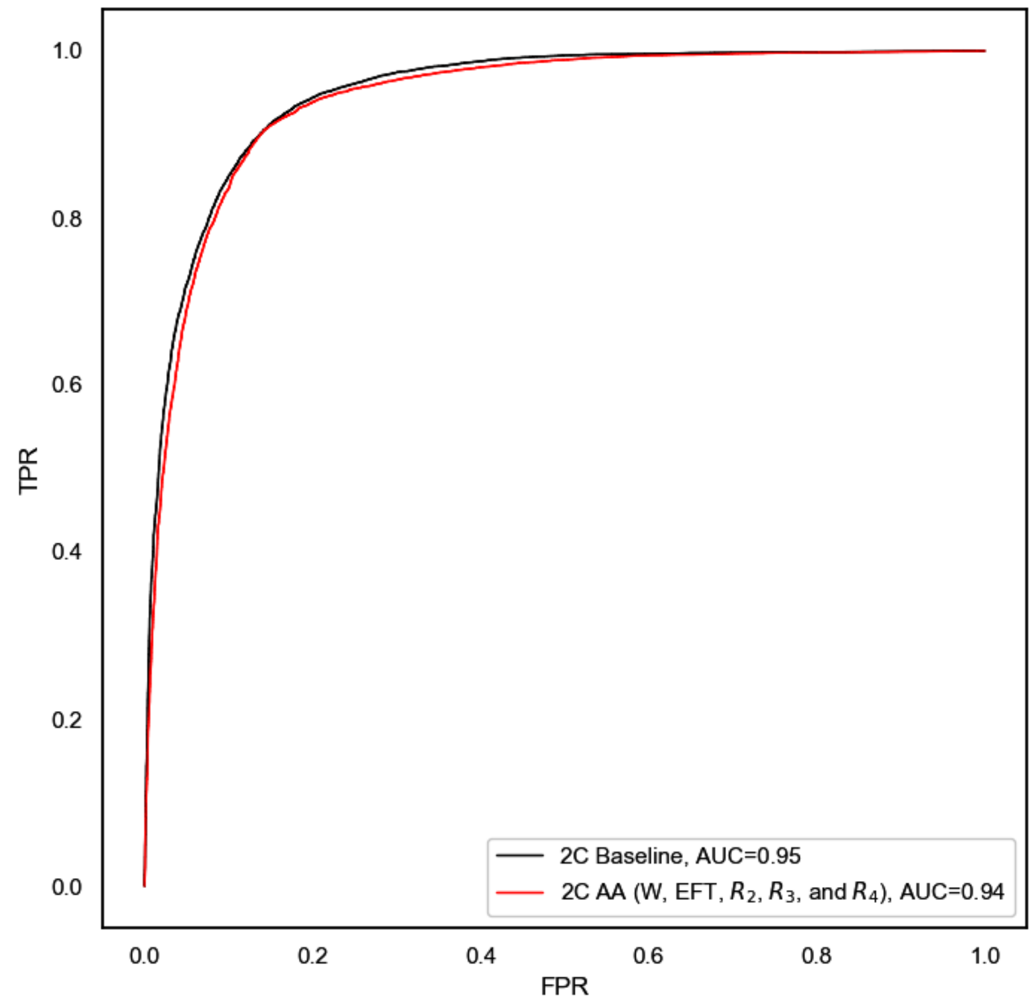


AA term with
5 types of anomalies



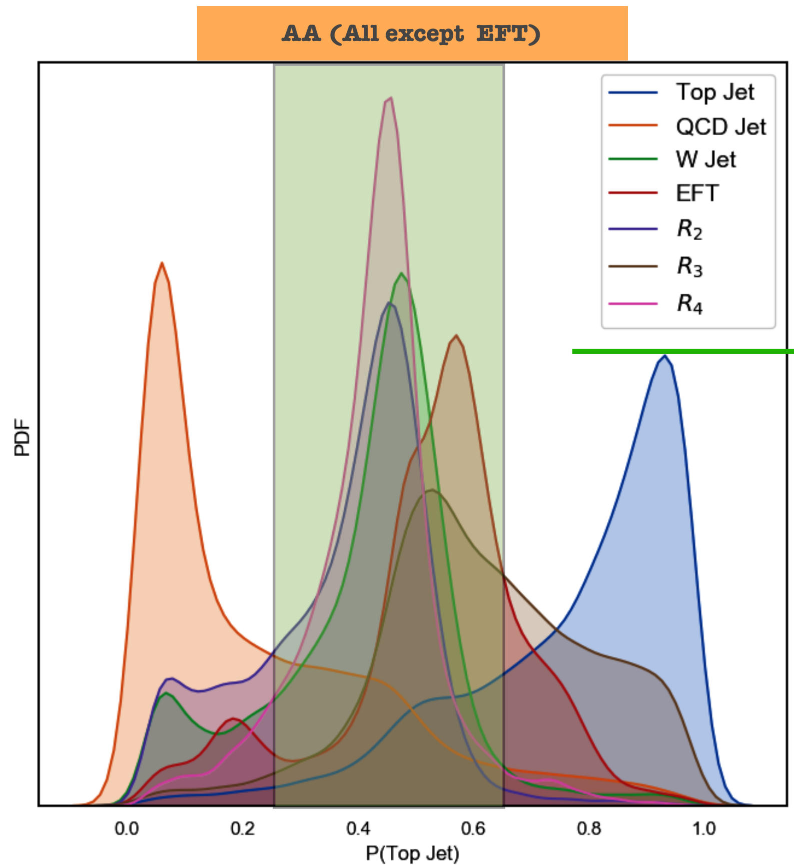
Awareness of a variety of anomalies → **Robust Anomaly Detector**

Baseline vs AA Comparison

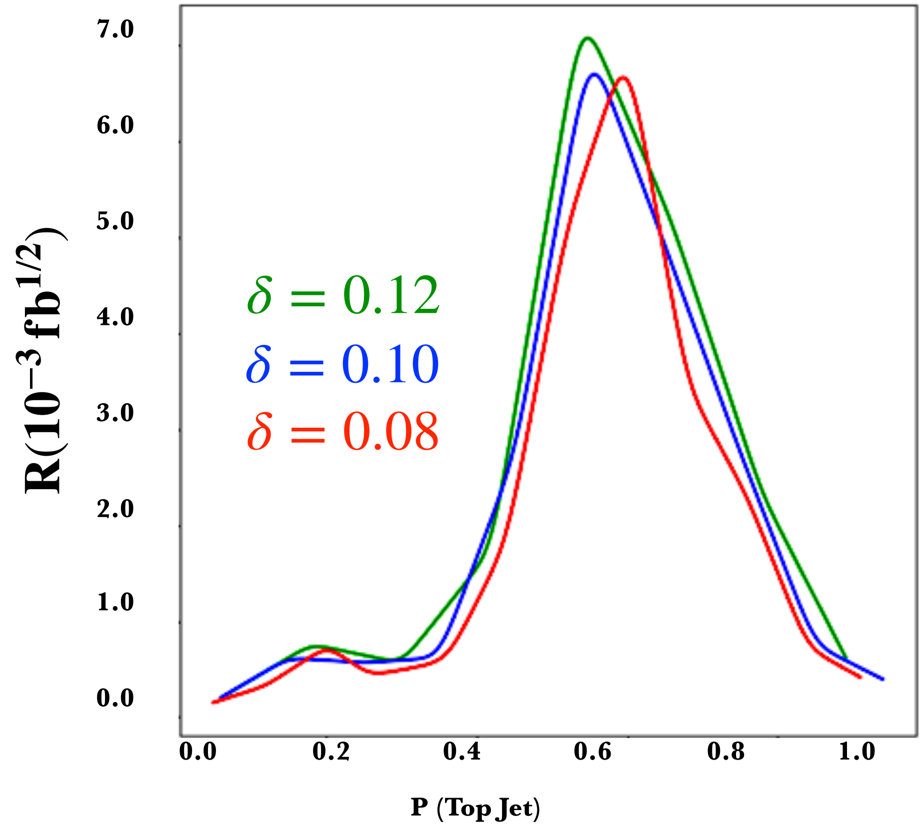


The addition of the AA term does not degrade the baseline classification but adds the ability to use its output for anomaly detection

Signal Cross-section Reach



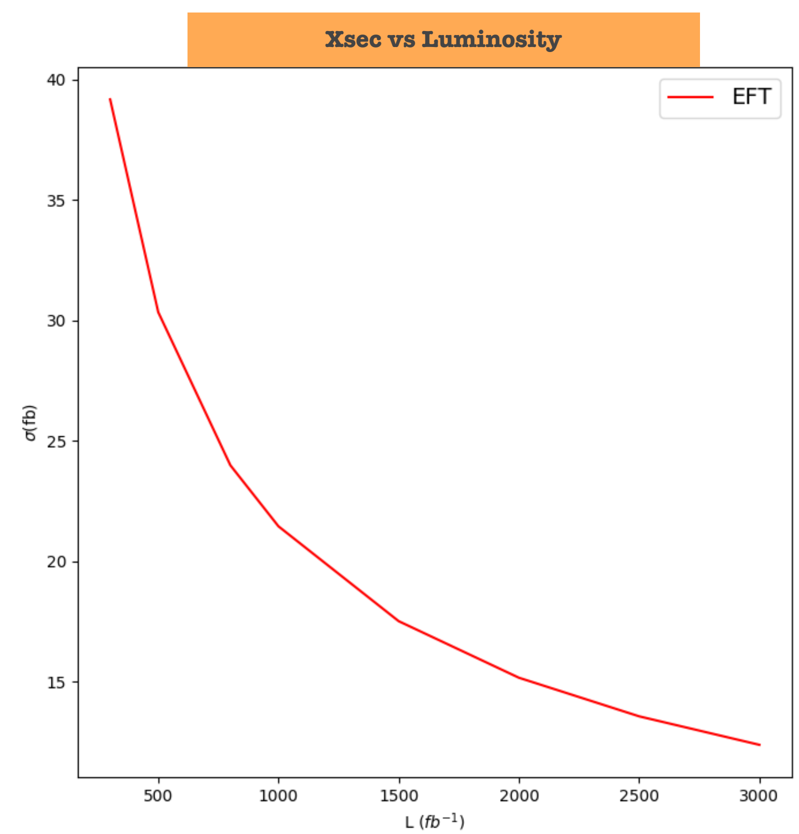
$$R = \frac{\epsilon_S}{\sqrt{\sigma_B \epsilon_B}}$$



We scan on windows of the classifier output
Cutting a small window around 0.5
anomaly detection is enhanced

$$\frac{S}{\sqrt{B}} = 5$$

Translation into a generic Anomaly xsec



We use $S/\sqrt{B}$ as an example of quantity to maximise (S=BSM, B=SM)

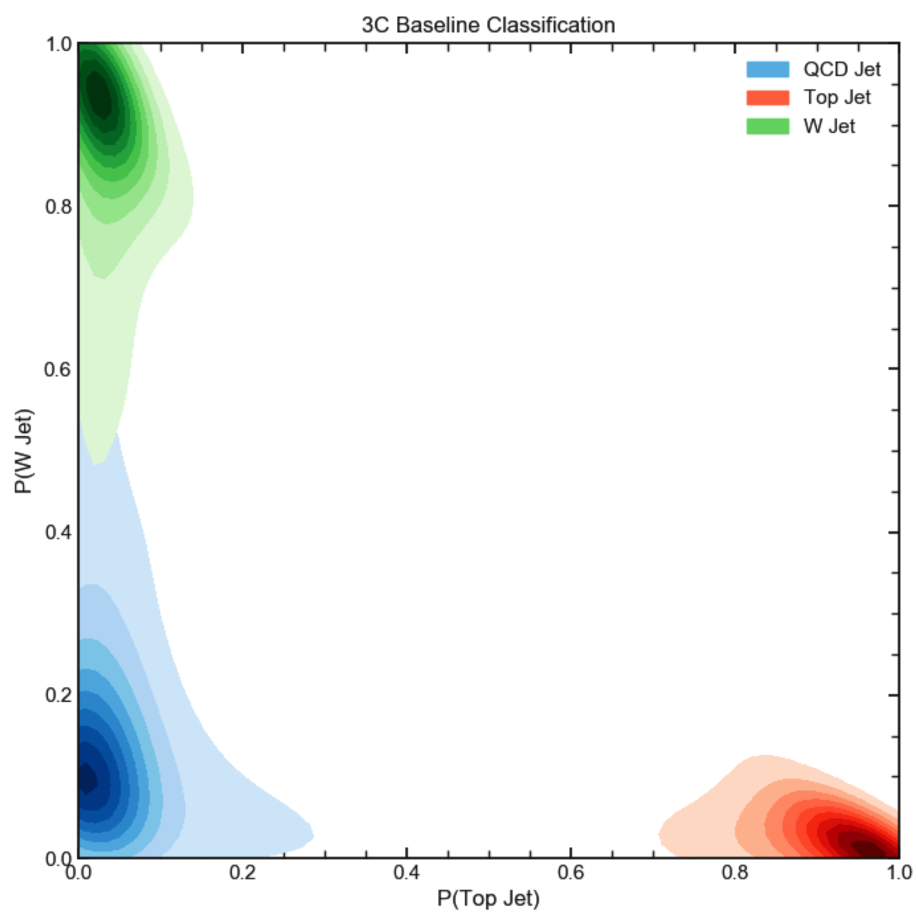
Three-class Example

This procedure can be generalized beyond binary classification

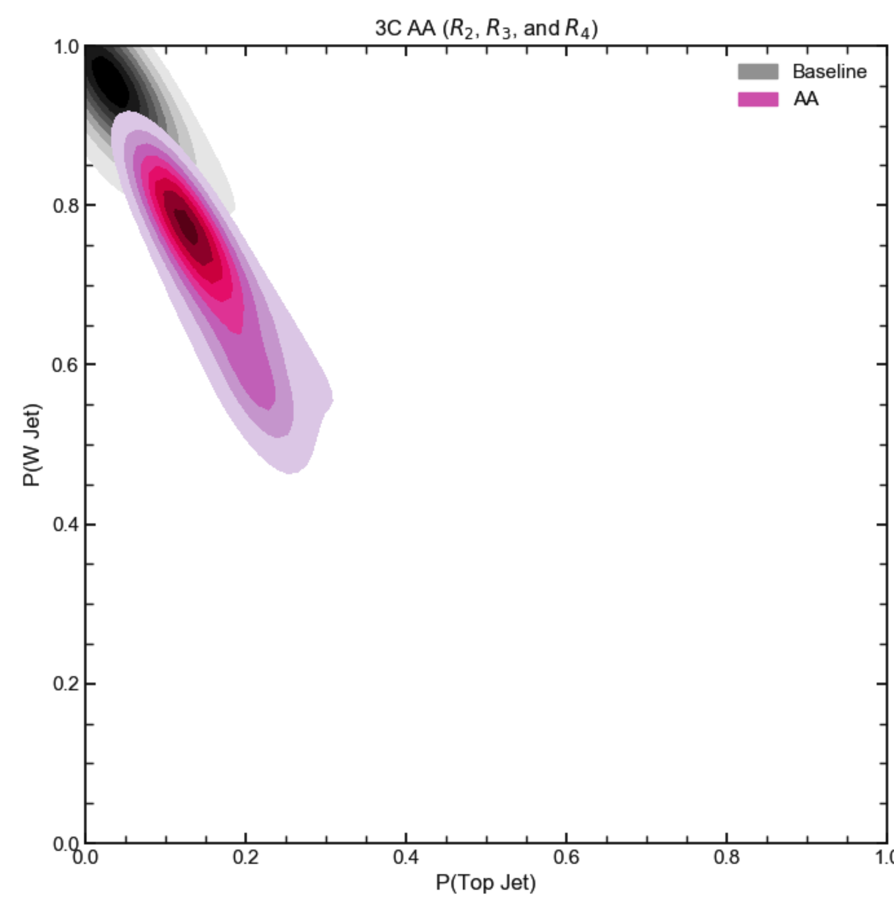
Top Jet, QCD jet, W-jet

150 K images (balanced data set), training:test=70:30%

Prior Run



With AA



Unseen Data Set: EFT

Key Message

- All methods have their "advantages" and "limitations".
- No "Generic" method: Each method operates in a limited region of parameter space.
- "None" of the current methods are suitable for whole phase space and can not target the anomalous events of all types.
- Simulated benchmarks are also different, so it is "non-trivial" task to compare one with other.

LHC Olympics 2020 was organised to address this issue
https://lhco2020.github.io/homepage/

SMEFTs

Model independent framework to parametrize the new physics

Higher dimensional operators respecting SM symmetries and involving SM fields

$$\mathscr{L}^{d=6} = \mathscr{L}_{SM} + \sum_i \frac{C_i}{\Lambda^2}\mathcal{O}_i$$

SMEFT fit to Higgs, diboson and EW data

$$\begin{aligned}
\mathcal{L}_{\rm SMEFT}^{\rm SILH} \supset\; & \frac{\bar{c}_W}{m_W^2}\frac{ig}{2}\left(H^\dagger\sigma^a \overleftrightarrow{D}^\mu H\right) D^\nu W^a_{\mu\nu} + \frac{\bar{c}_B}{m_W^2}\frac{ig'}{2}\left(H^\dagger \overleftrightarrow{D}^\mu H\right)\partial^\nu B_{\mu\nu} + \frac{\bar{c}_T}{v^2}\frac{1}{2}\left(H^\dagger \overleftrightarrow{D}_\mu H\right)^2 \\
& + \frac{\bar{c}_{ll}}{v^2}(\bar{L}\gamma_\mu L)(\bar{L}\gamma^\mu L) + \frac{\bar{c}_{He}}{v^2}(iH^\dagger \overleftrightarrow{D}_\mu H)(\bar{e}_R\gamma^\mu e_R) + \frac{\bar{c}_{Hu}}{v^2}(iH^\dagger \overleftrightarrow{D}_\mu H)(\bar{u}_R\gamma^\mu u_R) \\
& + \frac{\bar{c}_{Hd}}{v^2}(iH^\dagger \overleftrightarrow{D}_\mu H)(\bar{d}_R\gamma^\mu d_R) + \frac{\bar{c}'_{Hq}}{v^2}(iH^\dagger\sigma^a \overleftrightarrow{D}_\mu H)(\bar{Q}_L\sigma^a\gamma^\mu Q_L) \\
& + \frac{\bar{c}_{Hq}}{v^2}(iH^\dagger \overleftrightarrow{D}_\mu H)(\bar{Q}_L\gamma^\mu Q_L) + \frac{\bar{c}_{HW}}{m_W^2} ig(D^\mu H)^\dagger\sigma^a(D^\nu H)W^a_{\mu\nu} + \frac{\bar{c}_{HB}}{m_W^2} ig'(D^\mu H)^\dagger(D^\nu H)B_{\mu\nu} \\
& + \frac{\bar{c}_{3W}}{m_W^2} g^3\epsilon_{abc}W^{a\,\nu}_\mu W^b_{\nu\rho}W^{c\,\rho\mu} + \frac{\bar{c}_g}{m_W^2} g_s^2|H|^2 G^A_{\mu\nu}G^{A\mu\nu} + \frac{\bar{c}_\gamma}{m_W^2} g'^2|H|^2 B_{\mu\nu}B^{\mu\nu} \\
& + \frac{\bar{c}_H}{v^2}\frac{1}{2}(\partial^\mu|H|^2)^2 + \sum_{f=e,u,d}\frac{\bar{c}_f}{v^2} y_f|H|^2\bar{F}_L H^{(c)} f_R \\
& + \frac{\bar{c}_{3G}}{m_W^2} g_s^3 f_{ABC}G^{A\nu}_\mu G^{B\rho}_\nu G^{C\mu}_\rho + \frac{\bar{c}_{uG}}{m_W^2} g_s y_u \bar{Q}_L H^{(c)}\sigma^{\mu\nu}\lambda_A u_R G^A_{\mu\nu}
\end{aligned}$$

J.Ellis, C.W.Murphy, V.Sanz and T.You JHEP06(2018)146

SMEFT : Global Analysis

- Precision electroweak data, LHC Run 1 & 2 data (Higgs production, pair of gauge bosons)

Observable	Measurement	Ref.	SM Prediction	Ref.
Γ_Z [GeV]	2.4952 ± 0.0023	[41]	2.4943 ± 0.0005	[40]
σ^0_{had} [nb]	41.540 ± 0.037	[41]	41.488 ± 0.006	[40]
R^0_ℓ	20.767 ± 0.025	[41]	20.752 ± 0.005	[40]
$A^{0,\ell}_{\mathrm{FB}}$	0.0171 ± 0.0010	[41]	0.01622 ± 0.00009	[120]
$\mathcal{A}_\ell\,(P_\tau)$	0.1465 ± 0.0033	[41]	0.1470 ± 0.0004	[120]
$\mathcal{A}_\ell$ (SLD)	0.1513 ± 0.0021	[41]	0.1470 ± 0.0004	[120]
R^0_b	0.021629 ± 0.00066	[41]	0.2158 ± 0.00015	[40]
R^0_c	0.1721 ± 0.0030	[41]	0.17223 ± 0.00005	[40]
$A^{0,b}_{\mathrm{FB}}$	0.0992 ± 0.0016	[41]	0.1031 ± 0.0003	[120]
$A^{0,c}_{\mathrm{FB}}$	0.0707 ± 0.0035	[41]	0.0736 ± 0.0002	[120]
$\mathcal{A}_b$	0.923 ± 0.020	[41]	0.9347	[120]
$\mathcal{A}_c$	0.670 ± 0.027	[41]	0.6678 ± 0.0002	[120]
M_W [GeV]	80.387 ± 0.016	[42]	80.361 ± 0.006	[120]
M_W [GeV]	80.370 ± 0.019	[100]	80.361 ± 0.006	[120]

LEP data
+
WW LEP2 data
+
M_W Tevatron

J.Ellis, C.W.Murphy, V.Sanz and T.You JHEP06(2018)146

LHC data

Production	Decay	Signal Strength	Production	Decay	Signal Strength
ggF	$\gamma\gamma$	$1.10^{+0.23}_{-0.22}$	Wh	$\tau\tau$	-1.4 ± 1.4
ggF	ZZ	$1.13^{+0.34}_{-0.31}$	Wh	bb	1.0 ± 0.5
ggF	WW	0.84 ± 0.17	Zh	$\gamma\gamma$	$0.5^{+3.0}_{-2.5}$
ggF	$\tau\tau$	1.0 ± 0.6	Zh	WW	$5.9^{+2.6}_{-2.2}$
VBF	$\gamma\gamma$	1.3 ± 0.5	Zh	$\tau\tau$	$2.2^{+2.2}_{-1.8}$
VBF	ZZ	$0.1^{+1.1}_{-0.6}$	Zh	bb	0.4 ± 0.4
VBF	WW	1.2 ± 0.4	tth	$\gamma\gamma$	$2.2^{+1.6}_{-1.3}$
VBF	$\tau\tau$	1.3 ± 0.4	tth	WW	$5.0^{+1.8}_{-1.7}$
Wh	$\gamma\gamma$	$0.5^{+1.3}_{-1.2}$	tth	$\tau\tau$	$-1.9^{+3.7}_{-3.3}$
Wh	WW	$1.6^{+1.2}_{-1.0}$	tth	bb	1.1 ± 1.0
pp	$Z\gamma$	$2.7^{+4.6}_{-4.5}$	pp	$\mu\mu$	0.1 ± 2.5

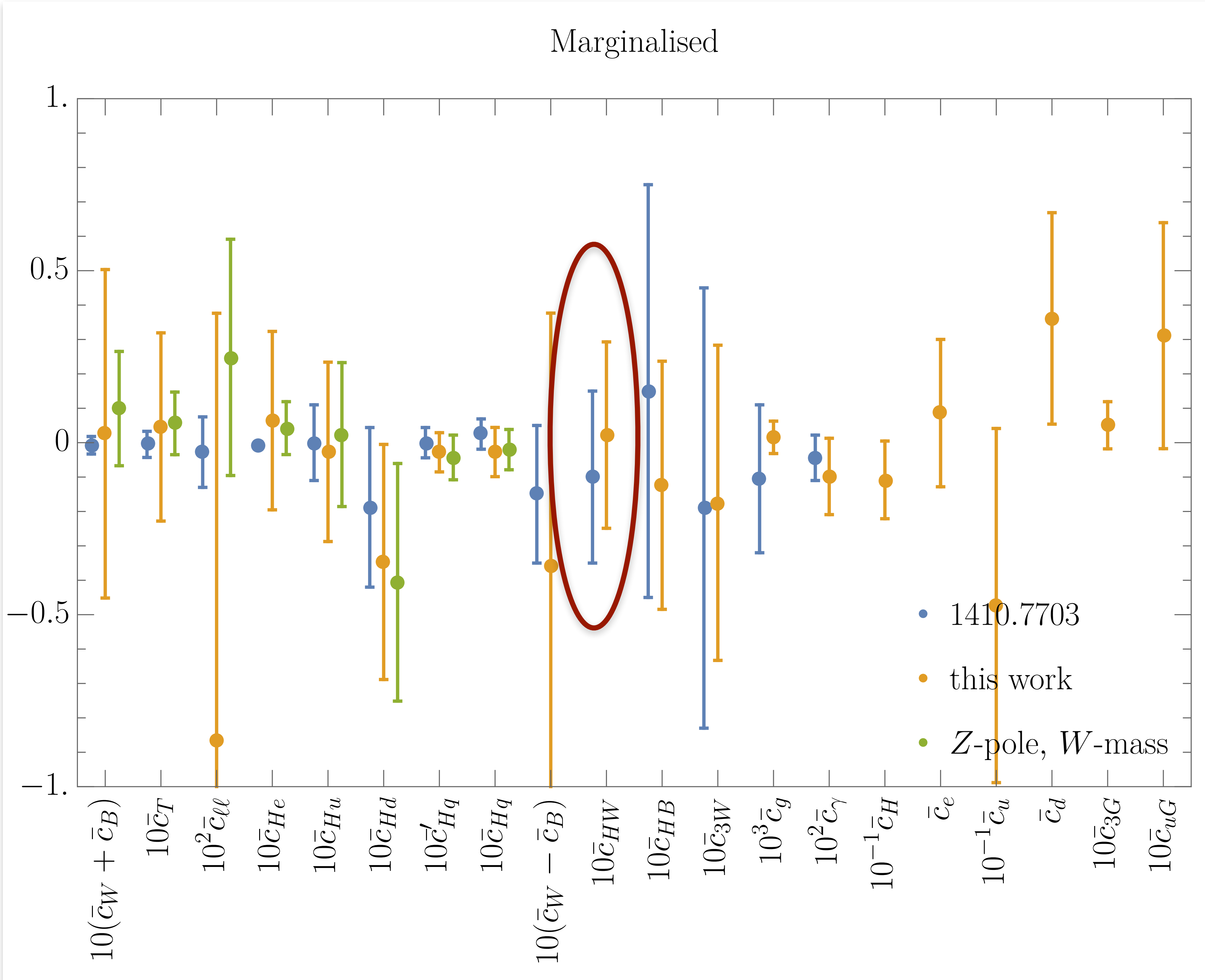
Run 1 data

(Early) Run 2 data

+

STXS

	Production	Decay	Sig. Stren.		Production	Decay	Sig. Stren.
[102]	1-jet, $p_T > 450$	$b\bar{b}$	$2.3^{+1.8}_{-1.6}$	[110]	pp	$\mu\mu$	-0.1 ± 1.5
[103]	Zh	$b\bar{b}$	0.9 ± 0.5	[111]	Zh	$b\bar{b}$	$1.12^{+0.50}_{-0.45}$
[103]	Wh	$b\bar{b}$	1.7 ± 0.7	[111]	Wh	$b\bar{b}$	$1.35^{+0.68}_{-0.59}$
[104]	$t\bar{t}h, \geq 1\ell$	$b\bar{b}$	0.72 ± 0.45	[112]	$t\bar{t}h$	$b\bar{b}$	$0.84^{+0.64}_{-0.61}$
[105]	$t\bar{t}h$	$1\ell + 2\tau_h$	$-1.52^{+1.76}_{-1.72}$	[113]	$t\bar{t}h$	$2\ell os + 1\tau_h$	$1.7^{+2.1}_{-1.9}$
[105]	$t\bar{t}h$	$2\ell ss + 1\tau_h$	$0.94^{+0.80}_{-0.67}$	[113]	$t\bar{t}h$	$1\ell + 2\tau_h$	$-0.6^{+1.6}_{-1.5}$
[105]	$t\bar{t}h$	$3\ell + 1\tau_h$	$1.34^{+1.42}_{-1.07}$	[113]	$t\bar{t}h$	$3\ell + 1\tau_h$	$1.6^{+1.8}_{-1.3}$
[105]	$t\bar{t}h$	$2\ell ss$	$1.61^{+0.58}_{-0.51}$	[113]	$t\bar{t}h$	$2\ell ss + 1\tau_h$	$3.5^{+1.7}_{-1.3}$
[105]	$t\bar{t}h$	3ℓ	$0.82^{+0.77}_{-0.71}$	[113]	$t\bar{t}h$	3ℓ	$1.8^{+0.9}_{-0.7}$
[105]	$t\bar{t}h$	4ℓ	$0.9^{+2.3}_{-1.6}$	[113]	$t\bar{t}h$	$2\ell ss$	$1.5^{+0.7}_{-0.6}$
[106]	0-jet DF	WW	$1.30^{+0.24}_{-0.23}$	[114]	ggF	WW	$1.21^{+0.22}_{-0.21}$
[106]	1-jet DF	WW	$1.29^{+0.32}_{-0.27}$	[114]	VBF	WW	$0.62^{+0.37}_{-0.36}$
[106]	2-jet DF	WW	$0.82^{+0.54}_{-0.50}$	[115]	$\mathrm{B}(h \to \gamma\gamma)/\ \mathrm{B}(h \to 4\ell)$		$0.69^{+0.15}_{-0.13}$
[106]	VBF 2-jet	WW	$0.72^{+0.44}_{-0.41}$	[115]	0-jet	4ℓ	$1.07^{+0.27}_{-0.25}$
[106]	Vh 2-jet	WW	$3.92^{+1.32}_{-1.17}$	[115]	1-jet, $p_T < 60$	4ℓ	$0.67^{+0.72}_{-0.68}$
[106]	Wh 3-lep	WW	$2.23^{+1.76}_{-1.53}$	[115]	1-jet, $p_T \in (60, 120)$	4ℓ	$1.00^{+0.63}_{-0.55}$
[107]	ggF	$\gamma\gamma$	$1.10^{+0.20}_{-0.18}$	[115]	1-jet, $p_T \in (120, 200)$	4ℓ	$2.1^{+1.5}_{-1.3}$
[107]	VBF	$\gamma\gamma$	$0.8^{+0.6}_{-0.5}$	[115]	2-jet	4ℓ	$2.2^{+1.1}_{-1.0}$
[107]	$t\bar{t}h$	$\gamma\gamma$	$2.2^{+0.9}_{-0.8}$	[115]	"BSM-like"	4ℓ	$2.3^{+1.2}_{-1.0}$
[107]	Vh	$\gamma\gamma$	$2.4^{+1.1}_{-1.0}$	[115]	VBF, $p_T < 200$	4ℓ	$2.14^{+0.94}_{-0.77}$
[108]	ggF	4ℓ	$1.20^{+0.22}_{-0.21}$	[115]	Vh lep	4ℓ	$0.3^{+1.3}_{-1.2}$
[109]	0-jet	$\tau\tau$	0.84 ± 0.89	[115]	$t\bar{t}h$	4ℓ	$0.51^{+0.86}_{-0.70}$
[109]	boosted	$\tau\tau$	$1.17^{+0.47}_{-0.40}$	[116]	Wh	WW	$3.2^{+4.4}_{-4.2}$
[109]	VBF	$\tau\tau$	$1.11^{+0.34}_{-0.35}$				
[106]	Zh 4-lep	WW	$0.77^{+1.49}_{-1.20}$				



J.Ellis, C.W.Murphy, V.Sanz and T.You JHEP06(2018)146

A.Biekötter, T.Corbett and T.Plehn, arXiv:1812.07587 [hep-ph]

E.da Silva Almeida et al.,Phys.Rev.D 99(2019)

Why VH channel?

VBF channel

J. Brehmer, K. Cranmer, G. Louppe and J Pavez,
Phys. Rev. Lett. 121 (2018) 111801, Phys. Rev. D 98 (2018) 052004

VH channel (higher statistics)

Felipe F. Freitas, CKK, Veronica Sanz, arXiv: 1902.05803 [hep-ph]

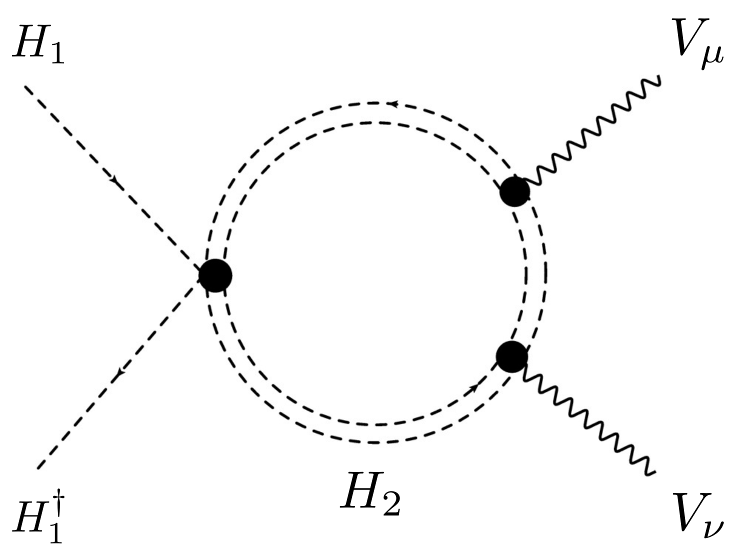
SMEFT via VH channel

Generic EFT effect

$$\eta_{\mu\nu}\, g m_V \Rightarrow \eta_{\mu\nu}\, g m_V - \frac{2\, g\, c_{HW}}{m_W}\, p^V_\mu\, p^V_\nu + \ldots$$

$$\mathscr{L}_{SM} \Rightarrow \mathscr{L}_{SM} + \frac{2igc_{HW}}{m_W^2}\,[D^\mu H^\dagger T_{2k} D^\nu H]\, W^k_{\mu\nu} + \ldots$$

Specific operator which produces it



Felipe F. Freitas, CKK, Veronica Sanz, arXiv: 1902.05803 [hep-ph]

Analysis set-up (VH Channel)

$p_T^{b_1}, p_T^{b_2}, p_T^{VH}, M_T^{VH}, p_T^{W/Z}, p_T^{H}, \eta^H, \phi^H$

0-lepton	$pp \to HZ, (H \to b\bar{b}, Z \to \nu\bar{\nu})$	$MET, \Delta\phi_{b_1 MET}$
1-lepton	$pp \to HW, (H \to b\bar{b}, W \to l\nu_l)$	$M_T^W, p_T^l, MET, \Delta R_{wl}, \Delta\phi_{b_1 l}, \Delta\phi_{lMET}$
2-lepton	$pp \to HZ, (H \to b\bar{b}, Z \to l^+l^-)$	$p_T^{l_1}, p_T^{l_2}, \Delta R_{ll}, \Delta\phi_{b_1 l_1}, \Delta\phi_{b_2 l_1}$

Feynrules Model :
Higgs Effective Lagrangian
arXiv:1310.5150

$\sqrt{s}$ =14 TeV, 100K events for both SMEFT and SM using MC@NLO Madgraph

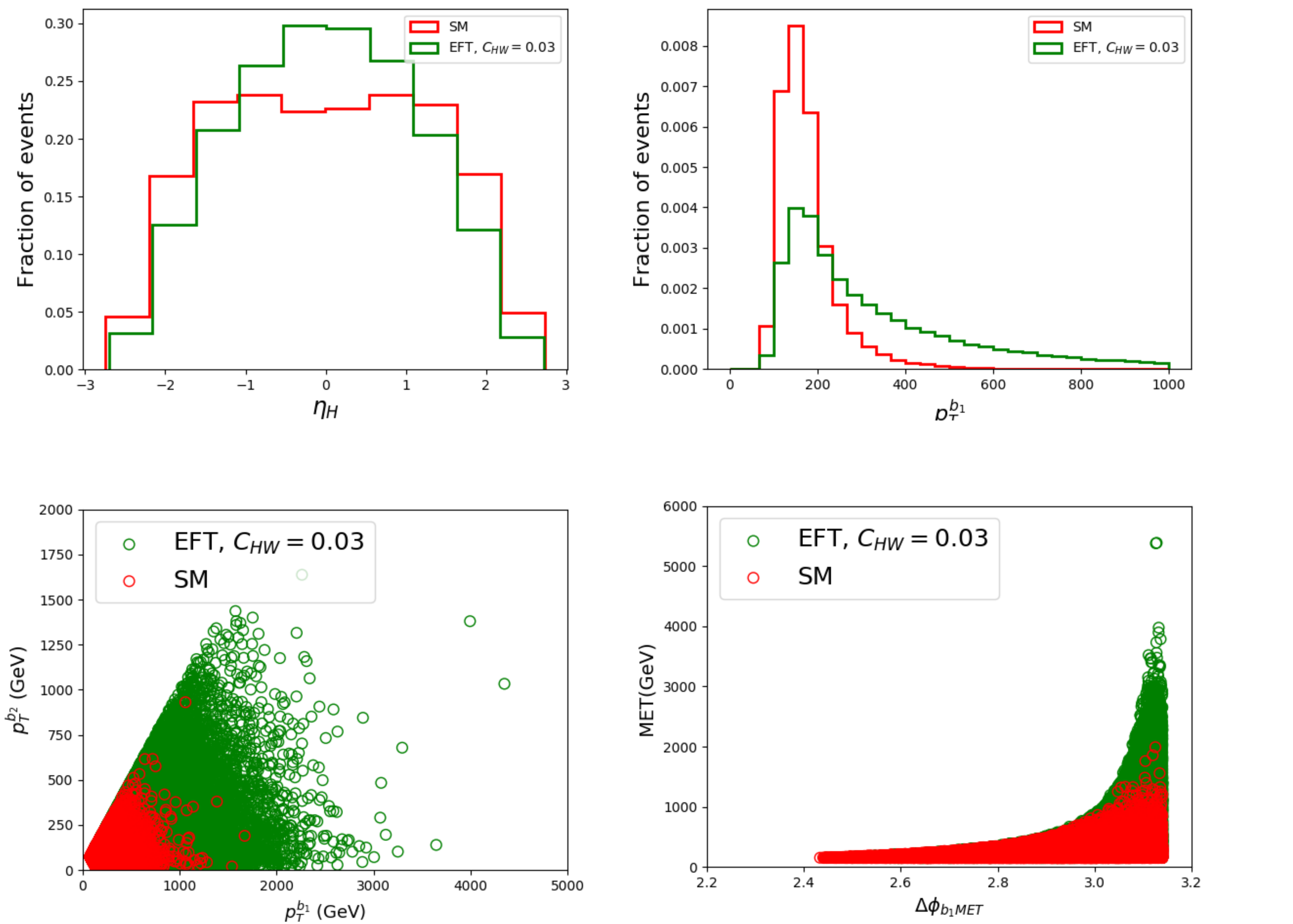
LO, parton level analysis

Channel	Inclusive
0L	$\not{E}_T > 150$ GeV
1L	$p_T^l > 25$ GeV, $\|\eta_l\| < 2.7$ $\not{E}_T > 30$ GeV, $p_T^V > 150$ GeV
2L	$p_T^l > 7$ GeV, $\|\eta_l\| < 2.7, p_T^V > 75$ GeV Leading lepton $p_T > 27$ GeV
0L,1L,2L	$p_T^b > 20$ GeV, $\|\eta_b\| < 2.5$, Leading b-jet $p_T > 45$ GeV

M. Aaboud et al., Phys. Lett. B 786(2018) 59

Felipe F. Freitas, CKK, Veronica Sanz, arXiv: 1902.05803 [hep-ph]

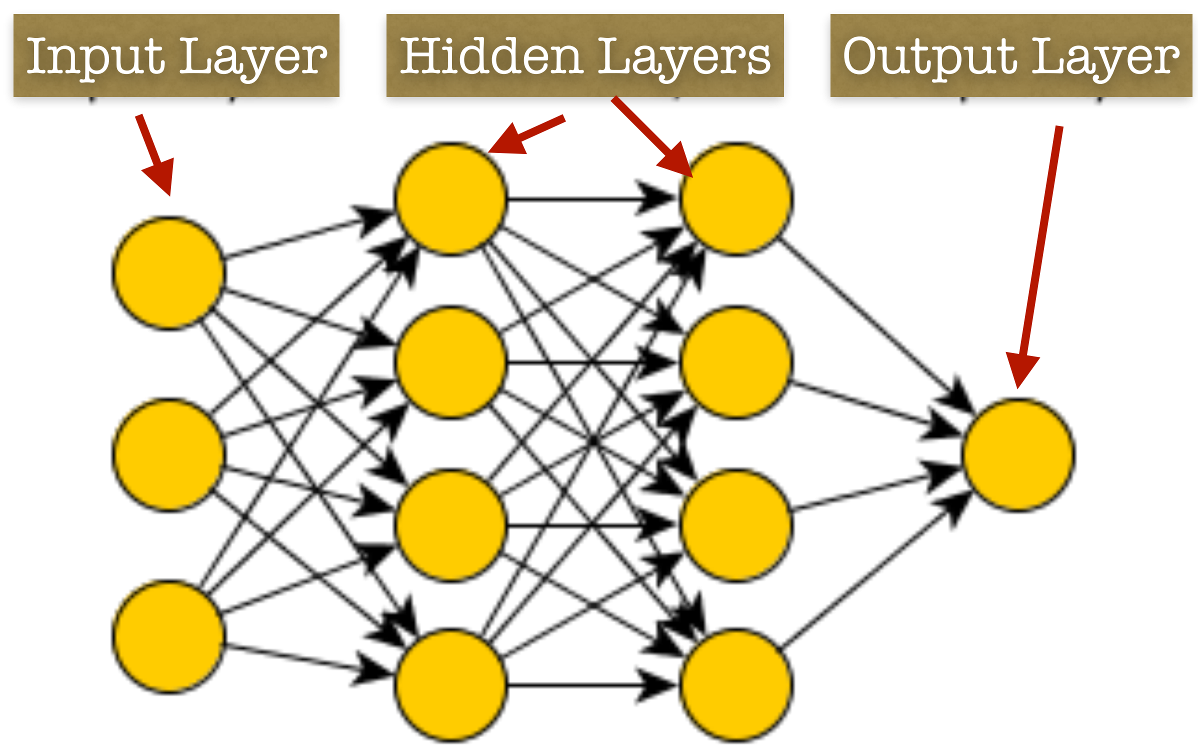
1D and 2D features



Neural Network

$p_T^{b_1}, p_T^{b_2}, p_T^{VH}, M_T^{VH}, p_T^{W/Z}, p_T^H, \eta^H, \phi^H$

$M_T^W, p_T^l, MET, \Delta R_{wl}, \Delta\phi_{b_1 l}, \Delta\phi_{lMET}$

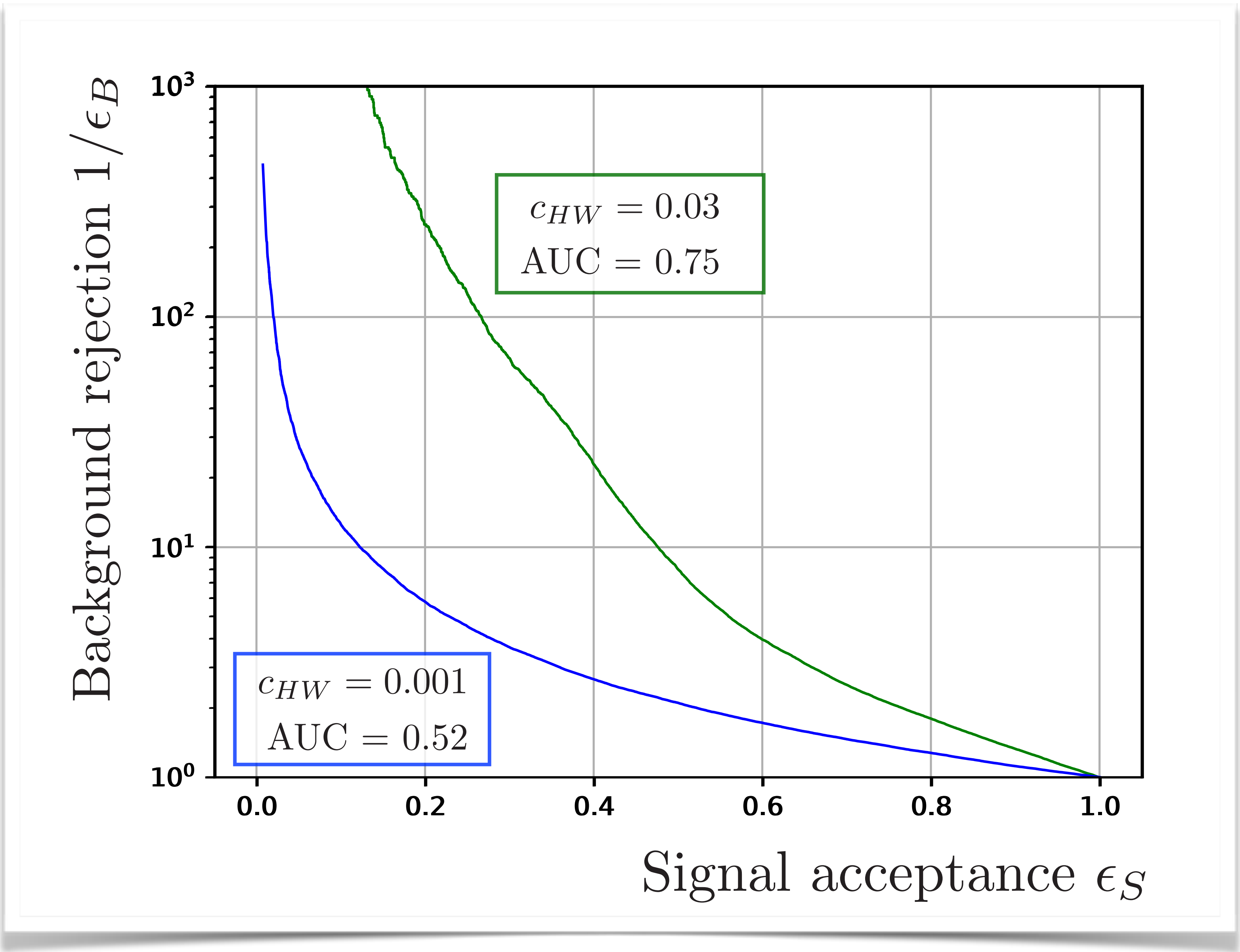


Y label

Neural Network Architecture

- Training set: Test set = 70 %: 30% (data scaling)
- Hidden layers: 1 (optimised)
- Activation function: ReLu
- Dropouts: 0.2
- Loss function: Asimov loss function (pre-training with MSE)

ROC curve (SMEFT vs SM Higgs background)



AUC: area under ROC

Appropriate performance measure for HEP analysis

Asimov Significance

$$Z_A = \left[2\left((s+b)\ln\left[\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right] - \frac{b^2}{\sigma_b^2}\ln\left[1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right]\right)\right]^{1/2}$$

$$s = W_s \sum_i^{N_{batch}} y_i^{pred} \times y_i^{true} \qquad b = W_b \sum_i^{N_{batch}} y_i^{pred} \times (1 - y_i^{true})$$

Specific(Asimov) Loss function

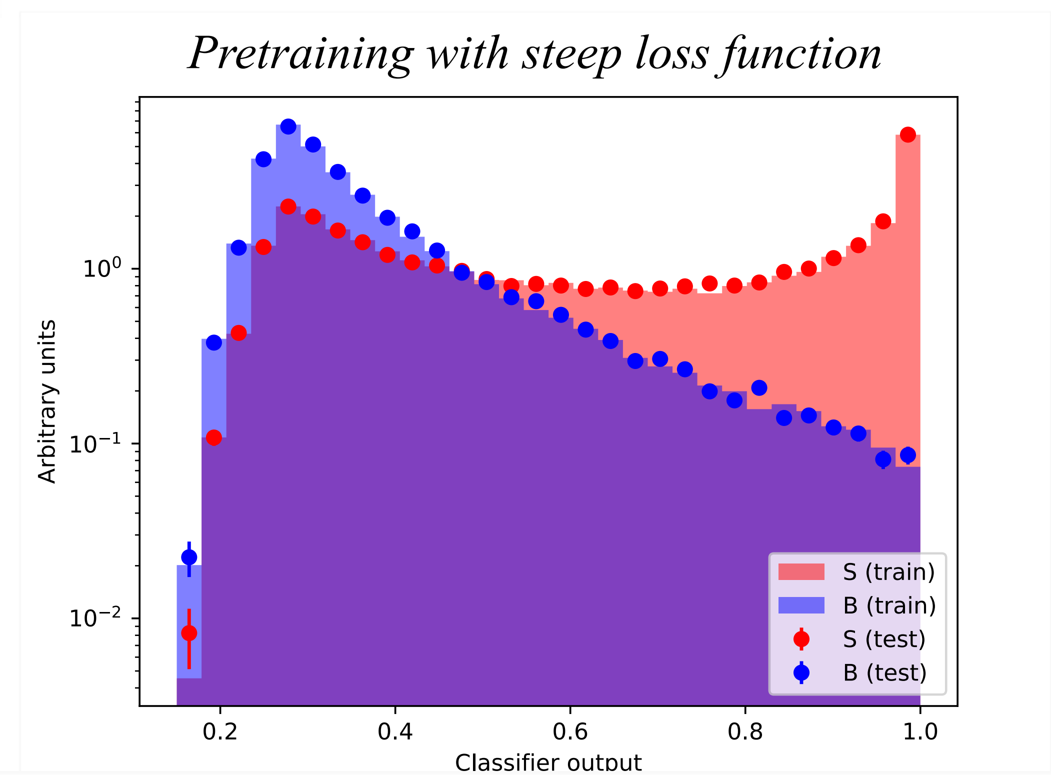
$$\ell_{Asimov} = 1/Z_A$$

Adam Elwood and Dirk Krücker, arXiv: 1806.00322[hep-ex]

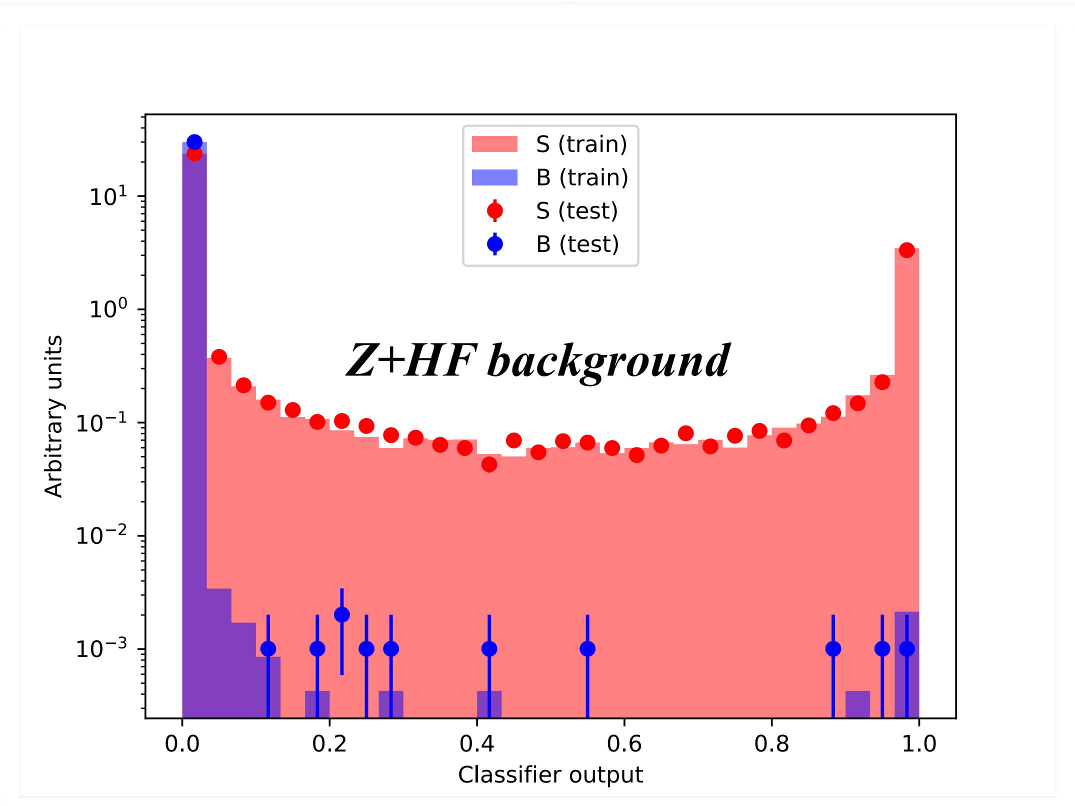
Glen Cowan, Kyle Cranmer, Eilam Gross, Ofer Vitells, arXiv:1007.1727[physics.data-an]

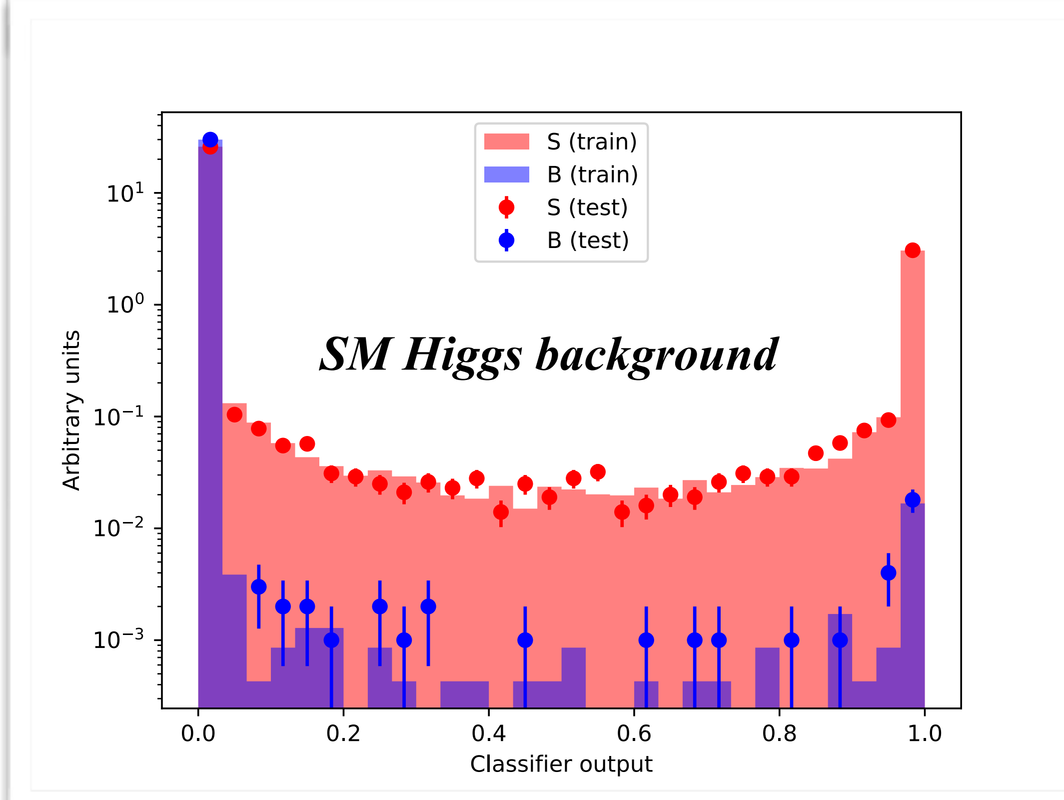
Classifier Output : 0L channel $(C_{HW} = 0.03)$

Pre-training with usual loss function and 5 epochs

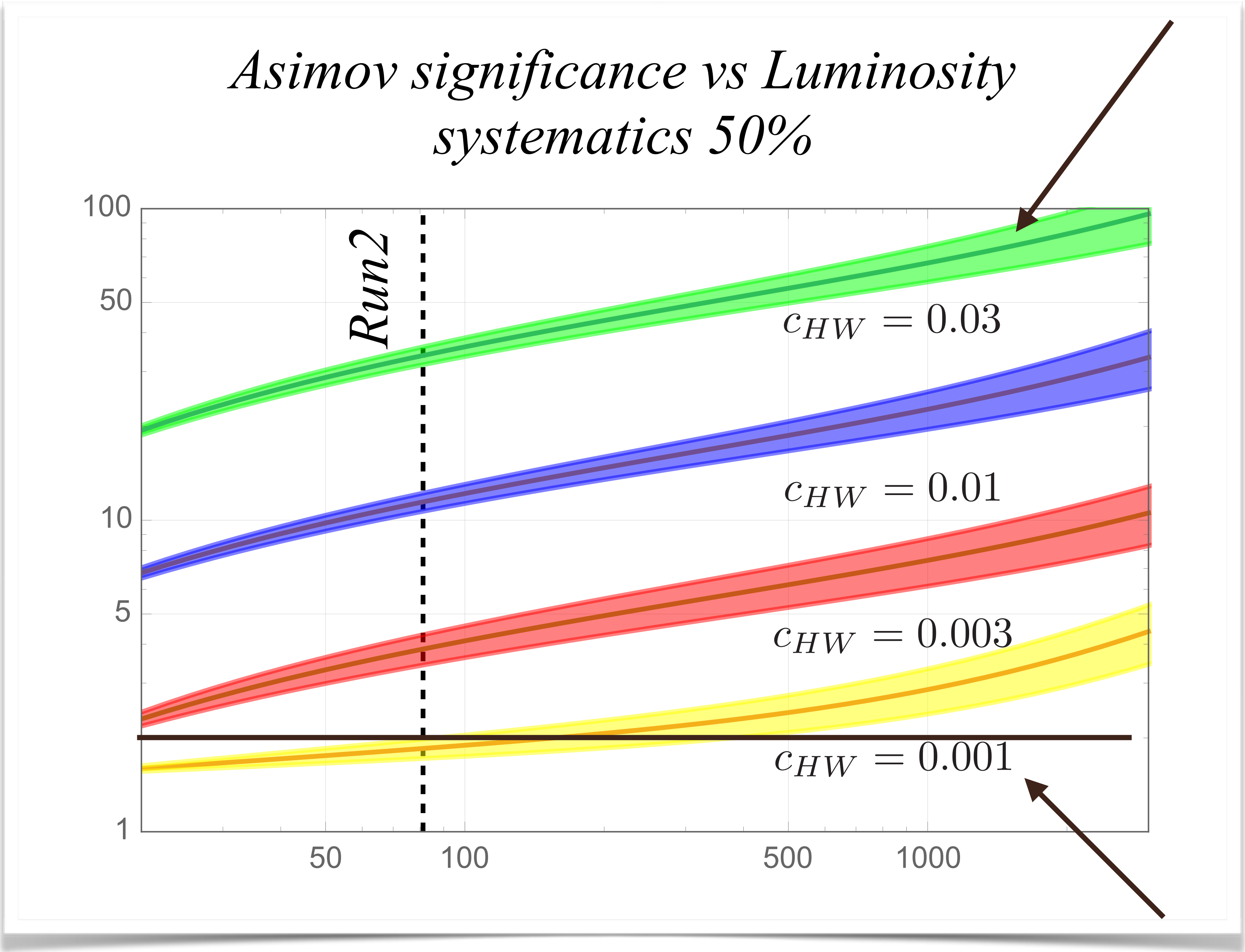


Final distributions with Asimov loss function



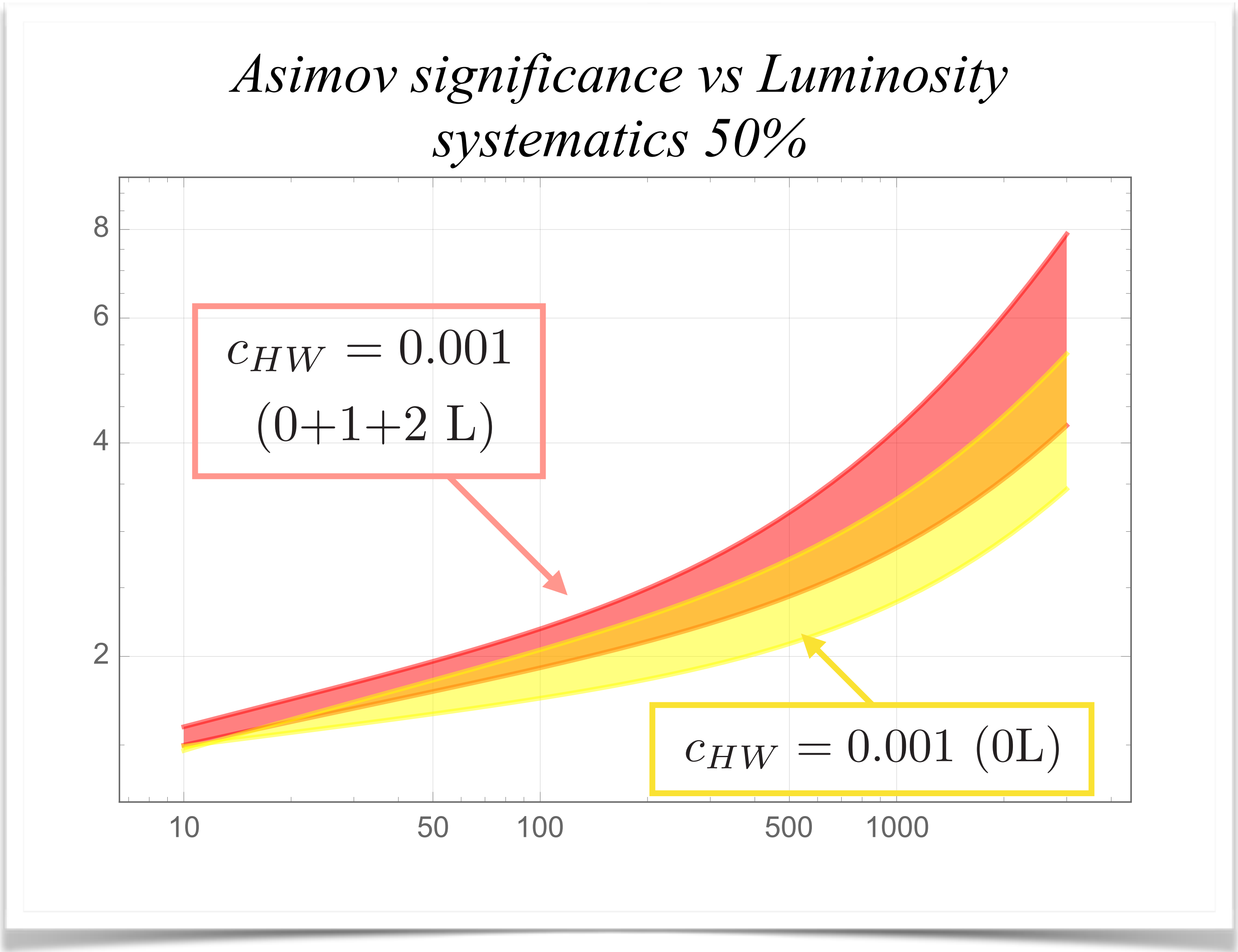


Current limit at 95% CL



Irreducible background only

Combining 0L+1L+2L for a limiting case



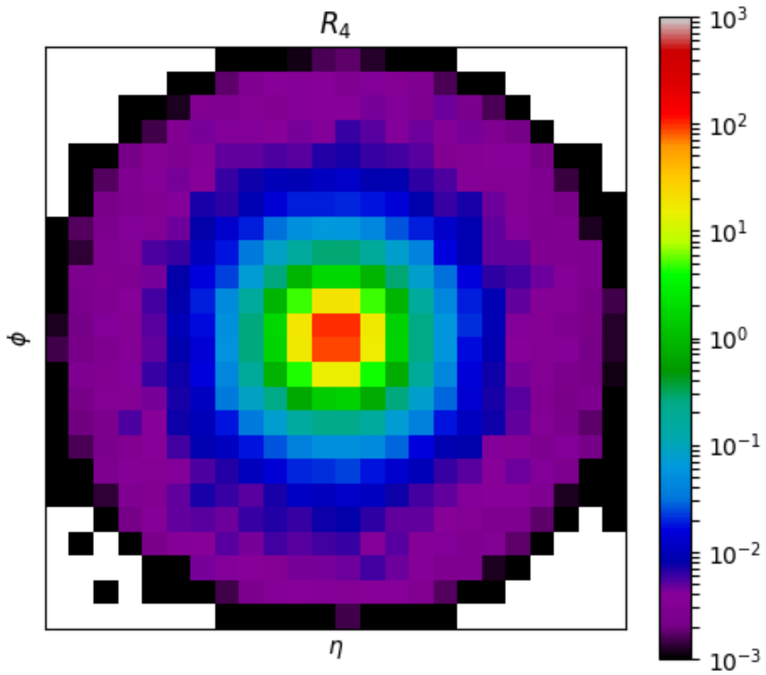
Irreducible background only

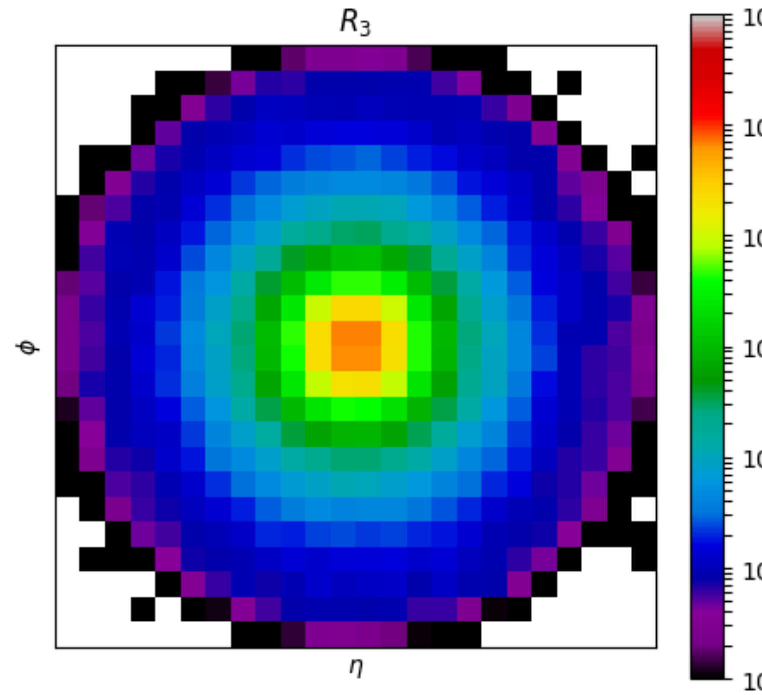
For a realistic analysis combining different channels may help

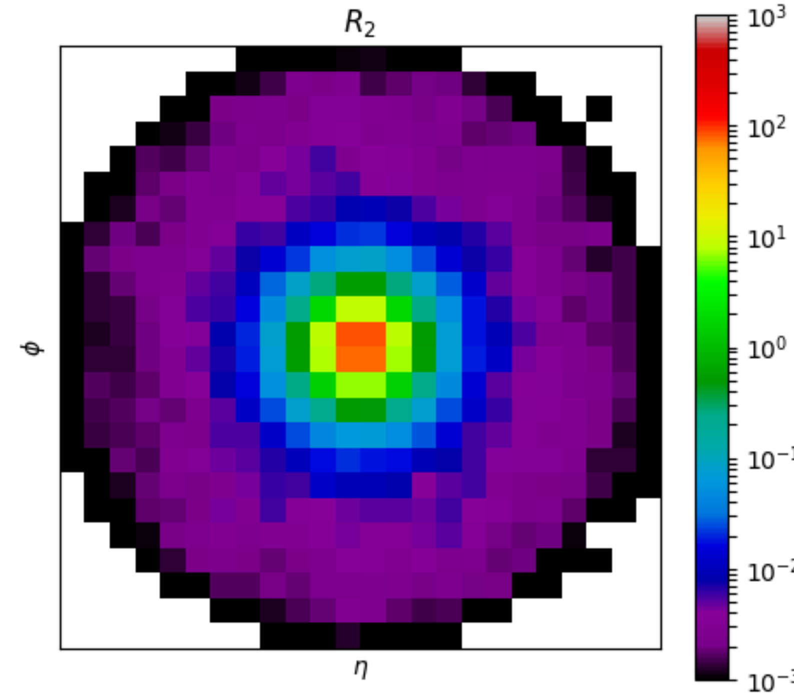
Summary and Outlook

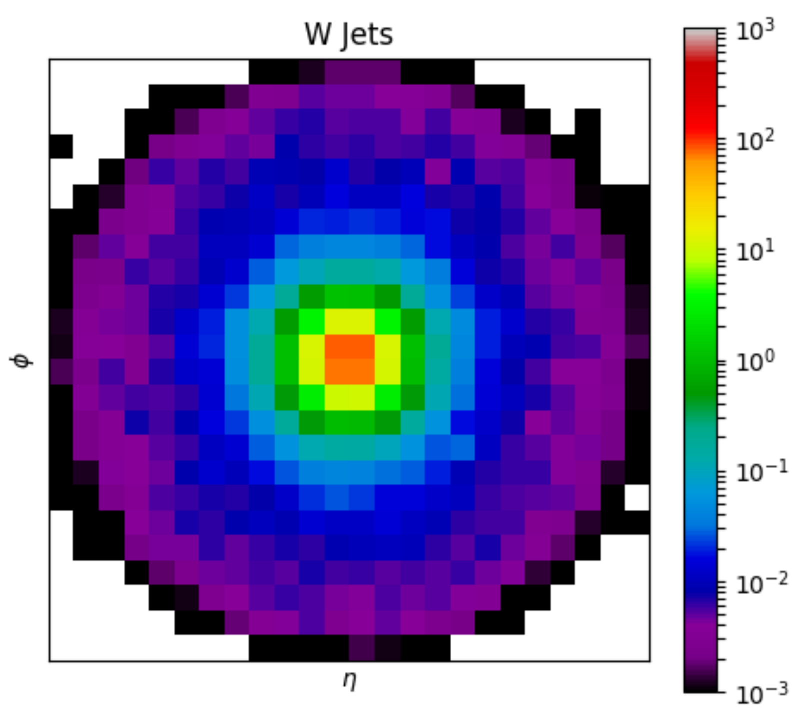
- **ML techniques are emerging as a competitive tool to look for new phenomena in the complex data**
- **HEP community is adapting these techniques for various tasks: trigger, heavy flavour tag, quark gluon discrimination, jet tagging etc.**
- **There is lot of activity to build methods for anomaly detection.**
- **We present a new algorithm for anomaly detection. It is based on the procedure of classifying 'normal' (SM) events, While the algorithm is made aware of the presence of anomalies (BSM) through a modification of the learning function.**
- **We used supervised learning techniques to exploit kinematic information in VH channel for SMEFT framework. This approach may provide a significant stronger bounds on EFT coefficients (scalability for more operators, realistic simulation).**
- **Finally, we need to test/use these approaches for LHC data.**

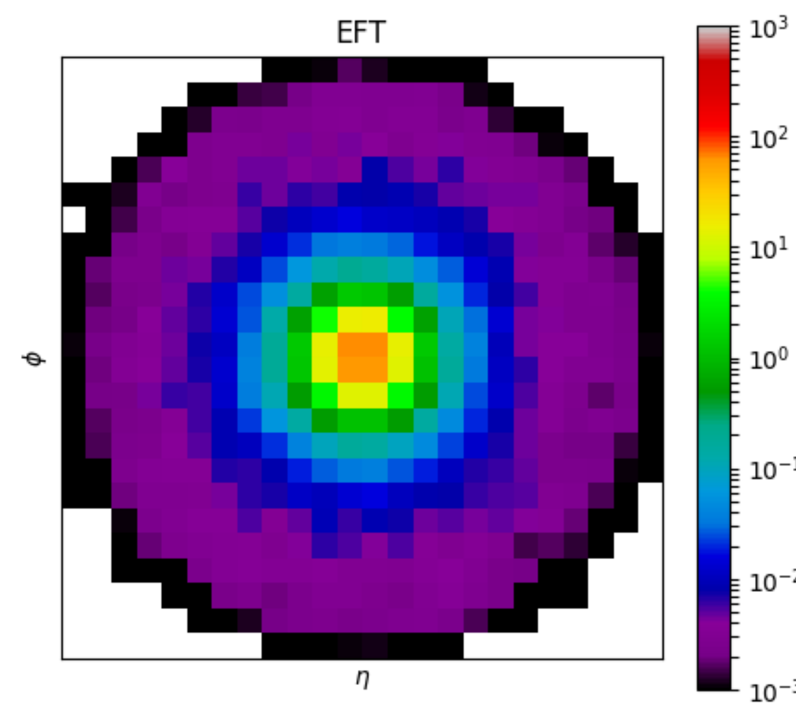
Thanks


R_4
ϕ
η
10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}


R_3
ϕ
η

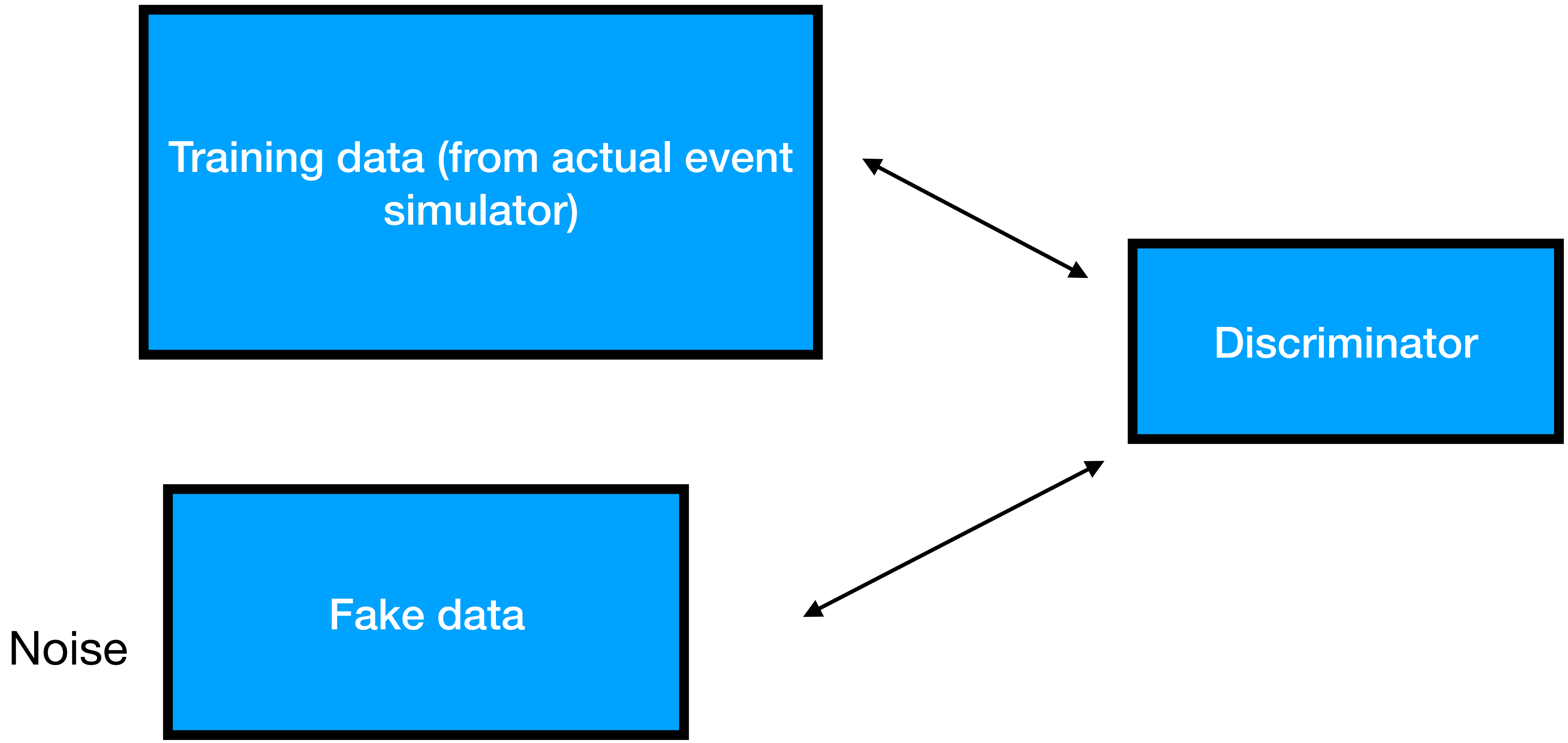

R_2
ϕ
η
10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}


W Jets
ϕ
η
10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}


EFT
ϕ
η
10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}

Generative Adversarial Networks (GANs)

Can interpolate the phase space for the event generation



Aim is to see if GANs could be used for fast simulations (dijetGAN, arxiv:1903.02433)