Machine Learning to Statistical Criteria

Charanjit Kaur Khosa

University of Manchester

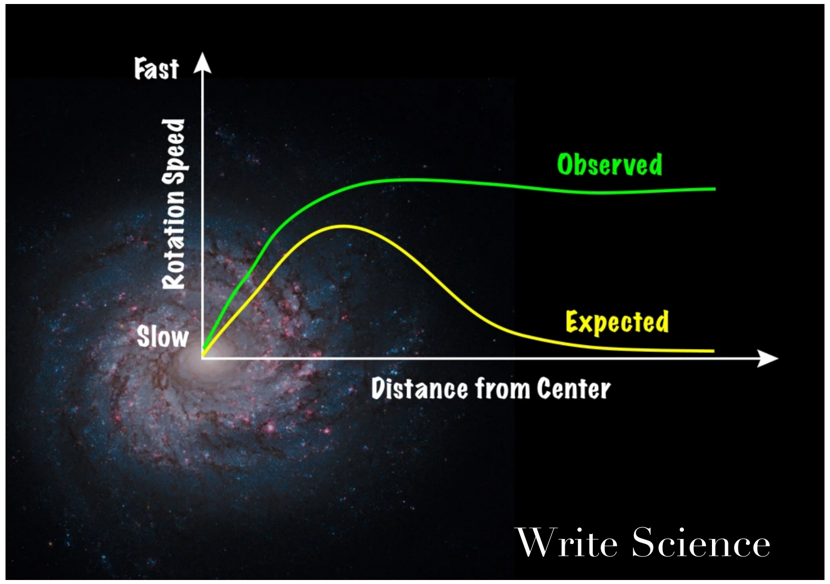
HEP Seminar, Oklahoma State University
4th August, 2022

Outline

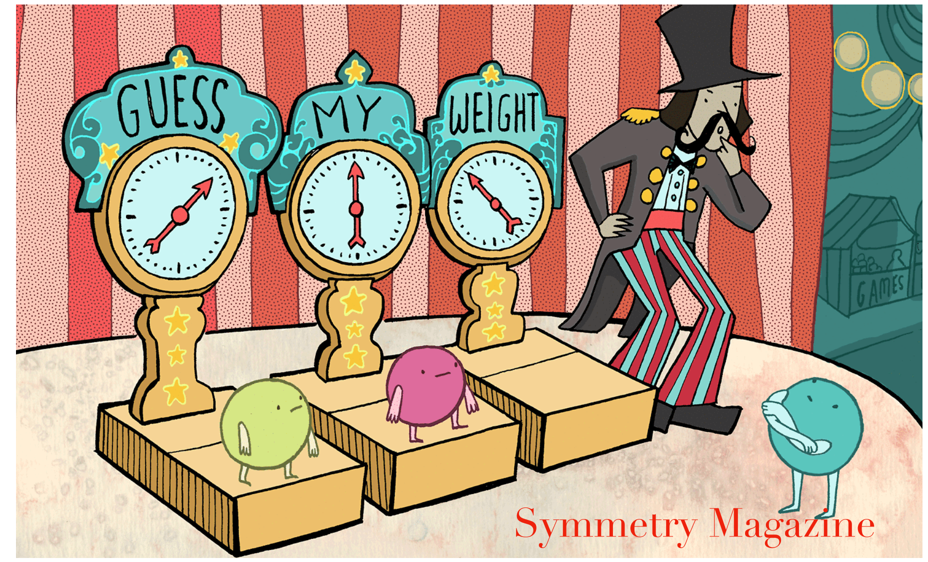
- Why to use Machine Learning in Particle Physics?
- How to implement these methods with statistical approaches?
- What are the current developments in this direction?
- What is a way forward?

Why Physics Beyond the Standard Model?

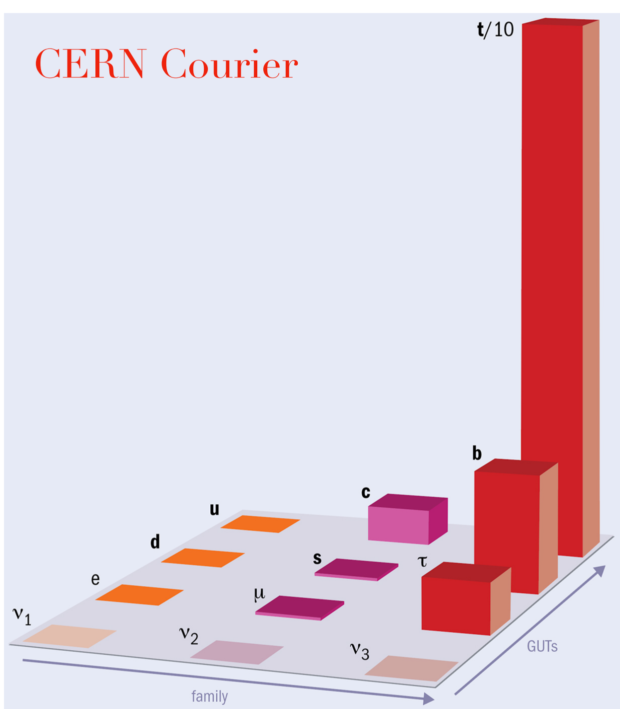
Dark Matter



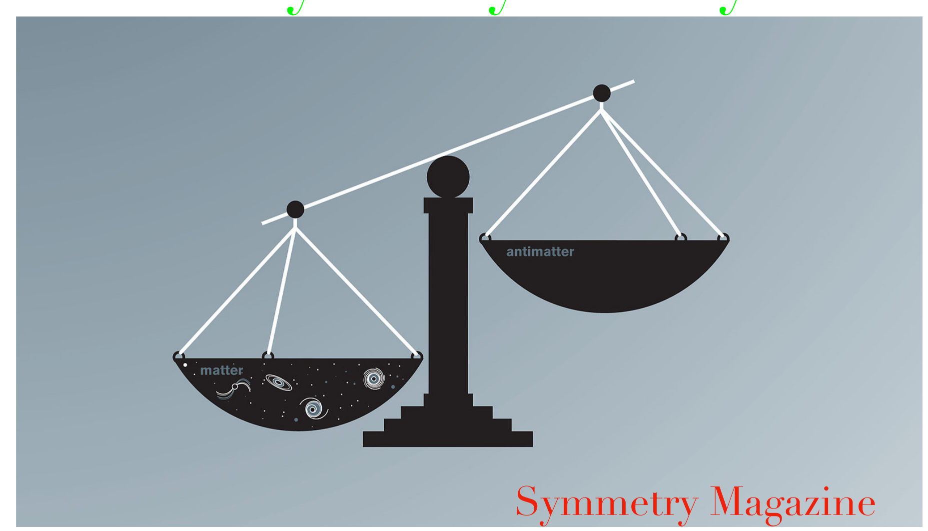
Neutrino Masses



Flavor Puzzle



Baryon Asymmetry



Hierarchy Problem? Gravity? ...

Current Status of BSM

Absence of direct signal for New Physics might be due to:

Nature of BSM

- New Physics signature is beyond the reach of the current experiments
- 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)

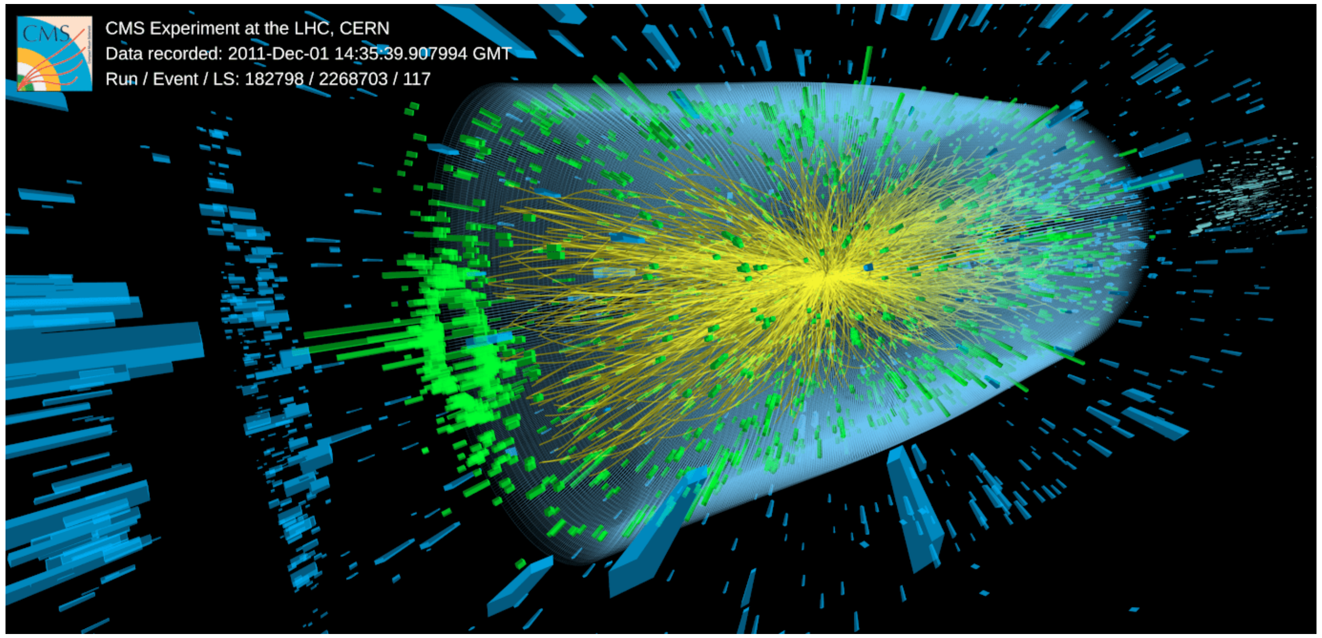
Need to go beyond these limitations

A New Perspective

Large Hadron Collider (LHC)

Big data brings challenges

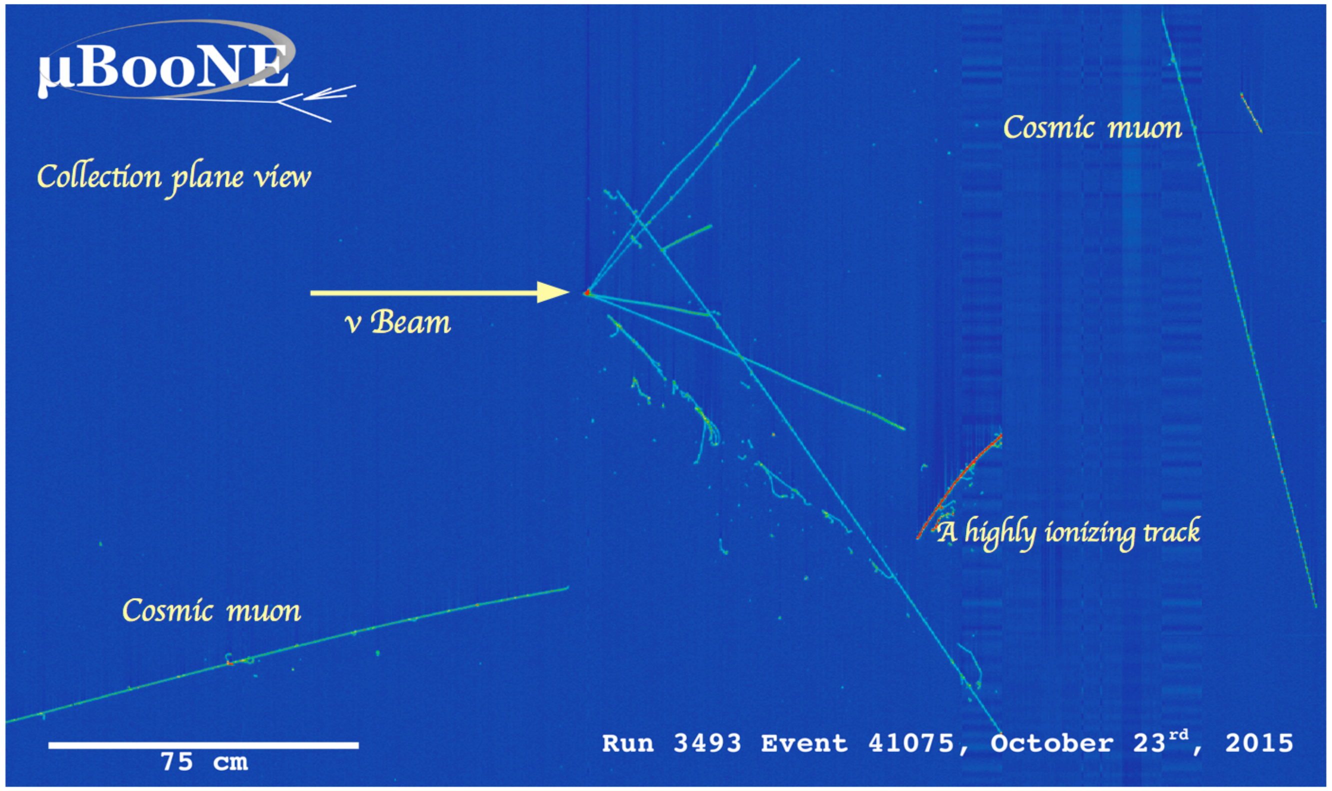
1. Large volumes of data
2. High dimensionality of the data sets
3. Large number of model parameters



Neutrino Experiments

Challenges that require new techniques

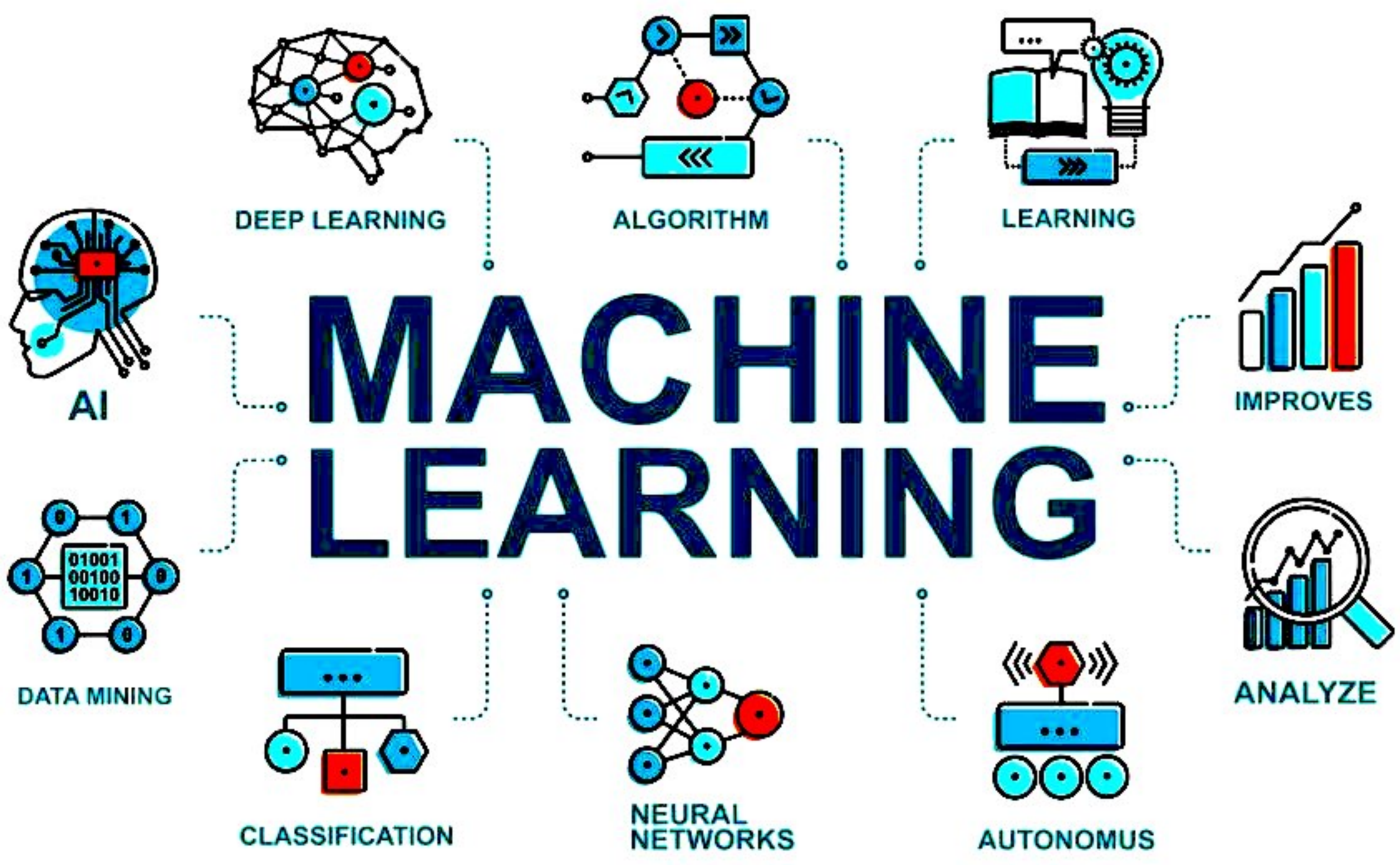
1. Data reconstruction
2. Physics inference
3. Physics modelling
4. Uncertainty quantification



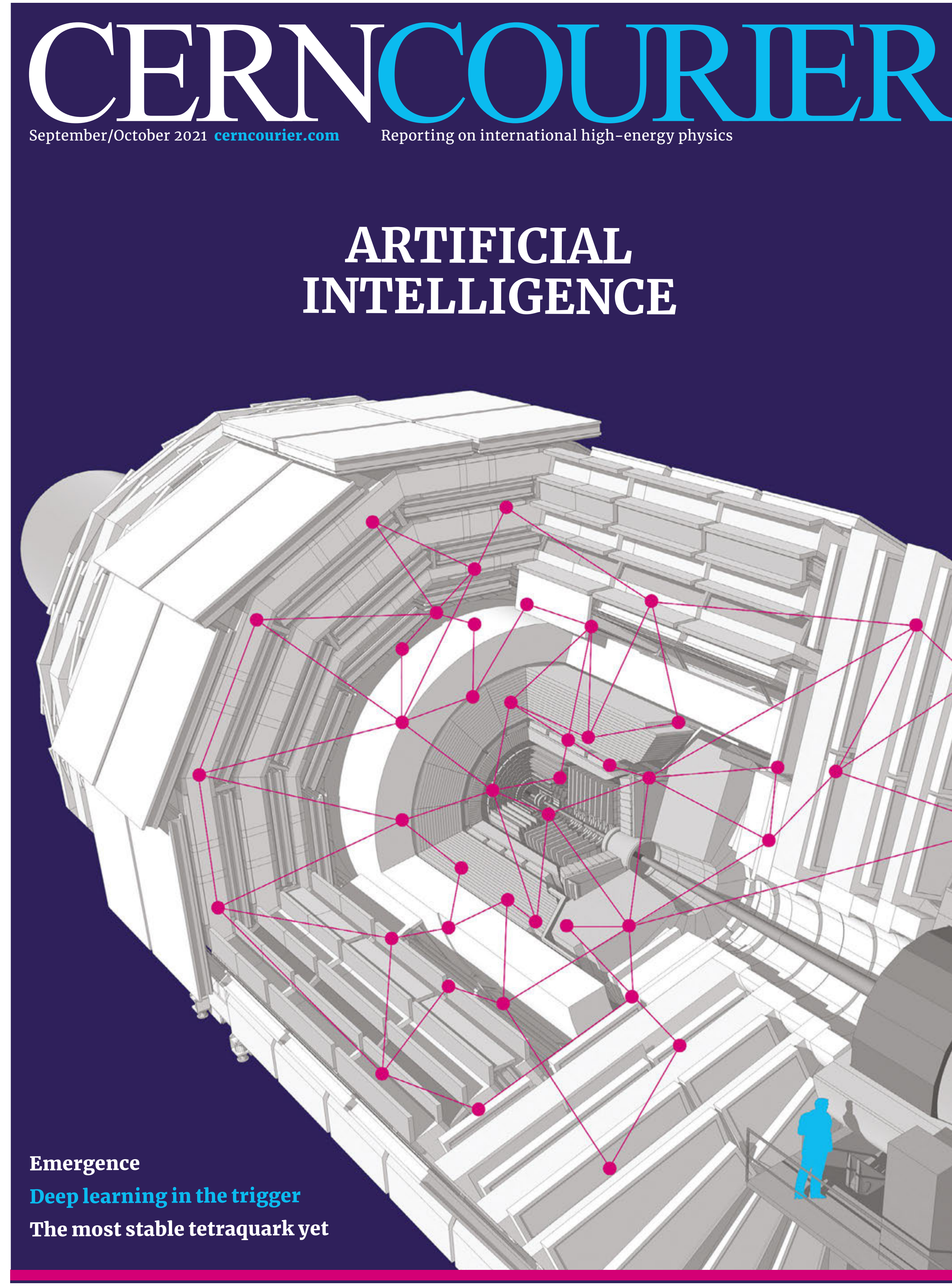
What Is Machine Learning?

A. Machine learning is a subdomain of artificial intelligence.

B. It uses statistical models and algorithms to identify patterns in and/or fit data without using explicit instructions.



Machine Learning for Particle Physics

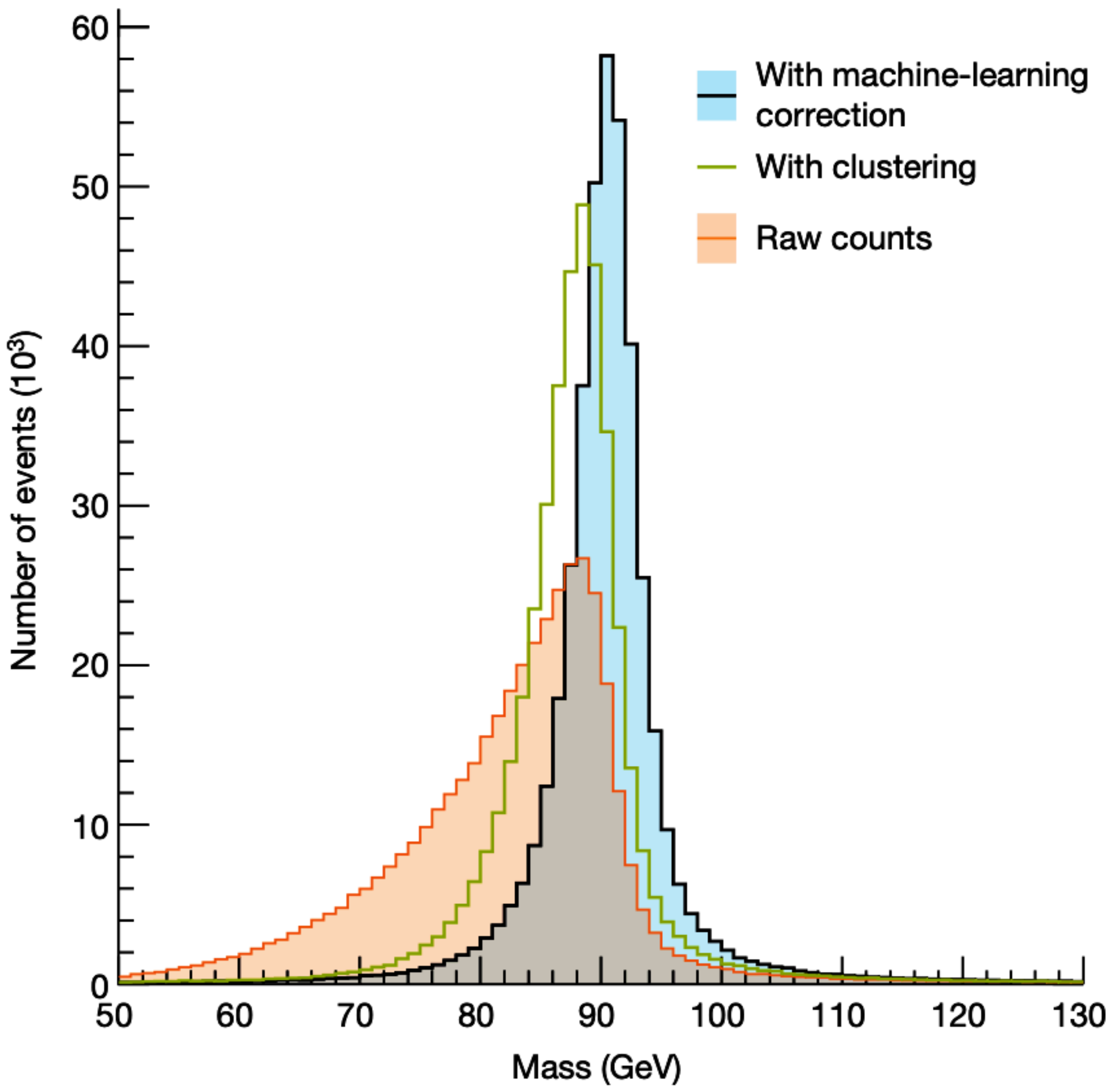


Machine Learning at the Energy and Intensity Frontiers

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

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

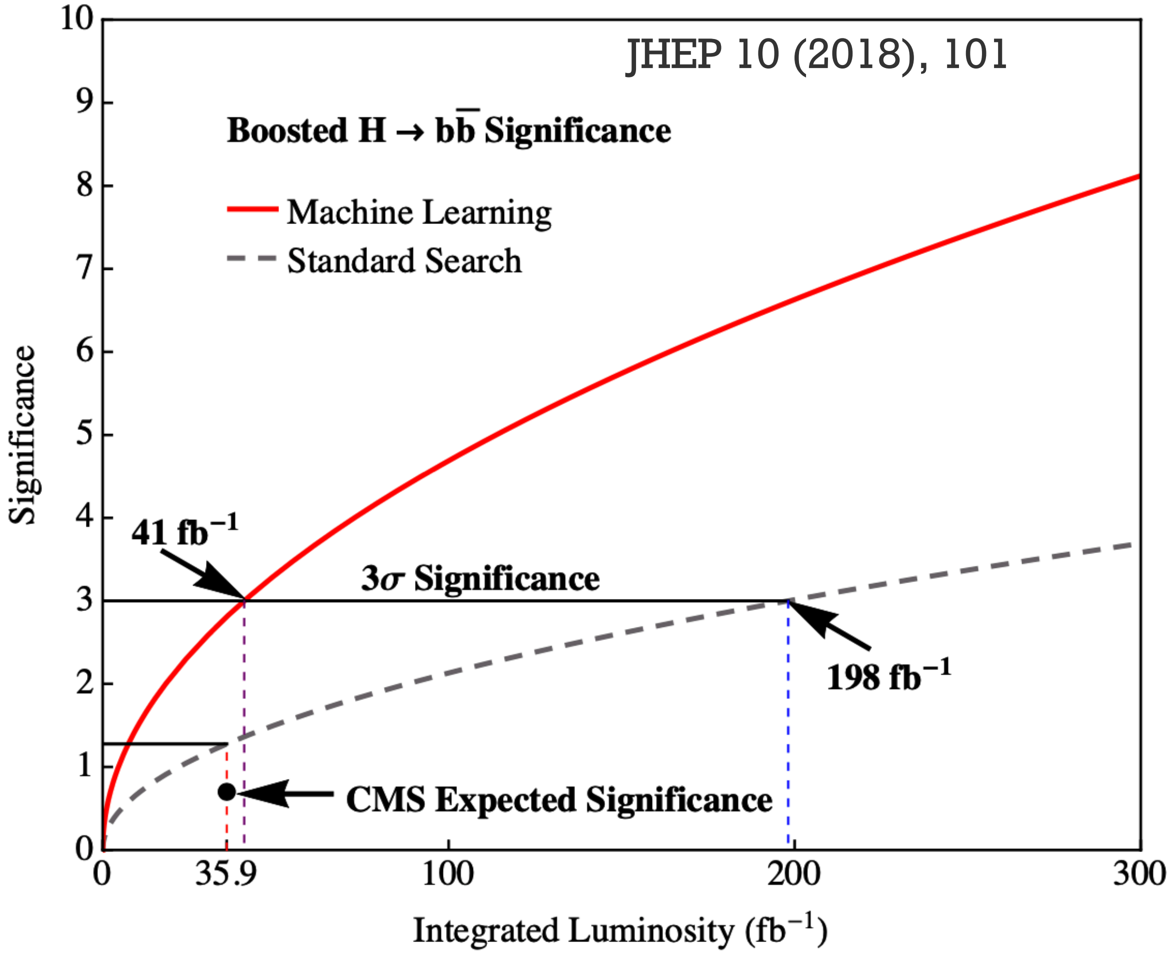
Analysis	Years of data collection	Sensitivity without machine learning	Sensitivity with machine learning	Ratio of P values	Additional data required
CMS[24] $H \to \gamma\gamma$	2011–2012	2.2σ, $P = 0.014$	2.7σ, $P = 0.0035$	4.0	51%
ATLAS[43] $H \to \tau^+\tau^-$	2011–2012	2.5σ, $P = 0.0062$	3.4σ, $P = 0.00034$	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%



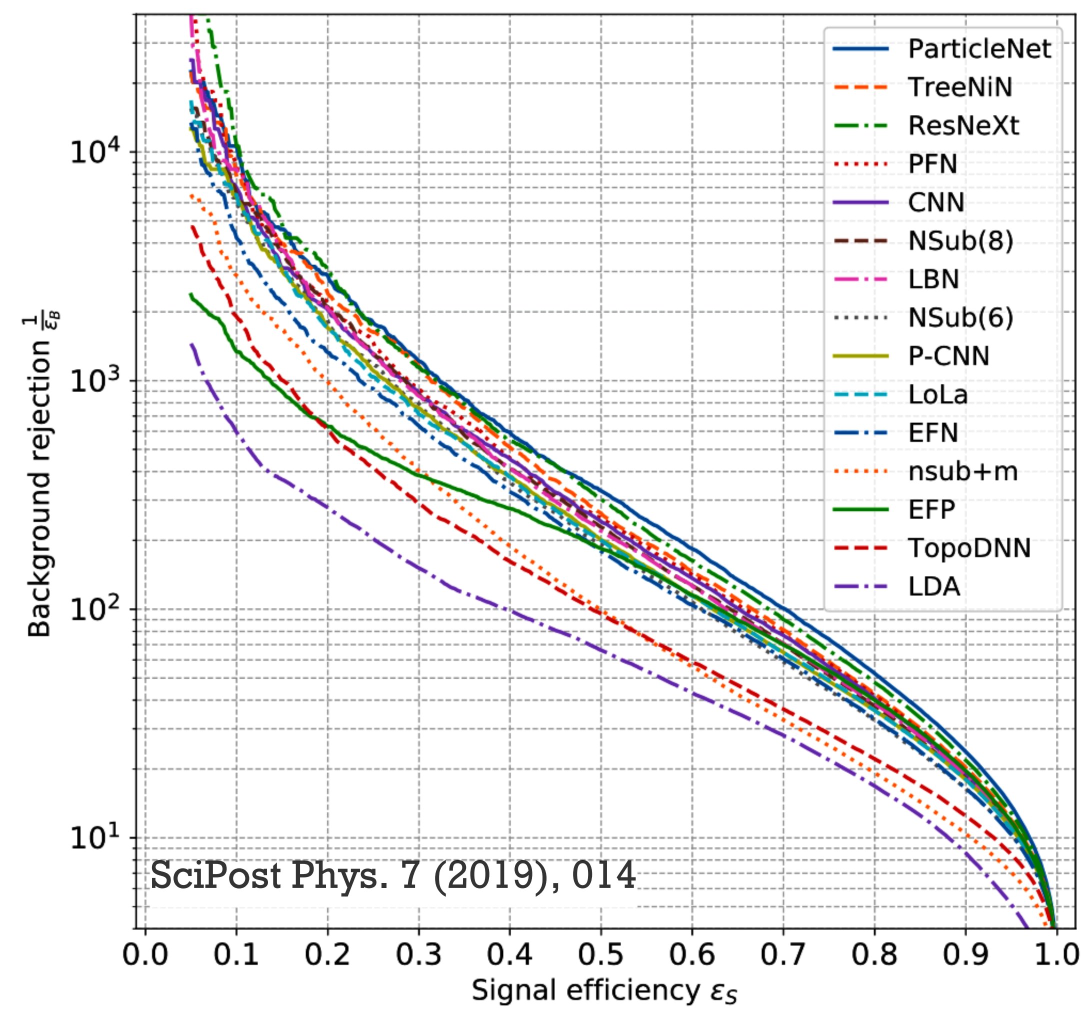
Early Successes

Complete list of references: [HEPML-LivingReview]

H → $b\bar{b}$ decay



Top Tagging



Machine Learning techniques increase the discovery potential of the experiments

Commonly Used ML Methods

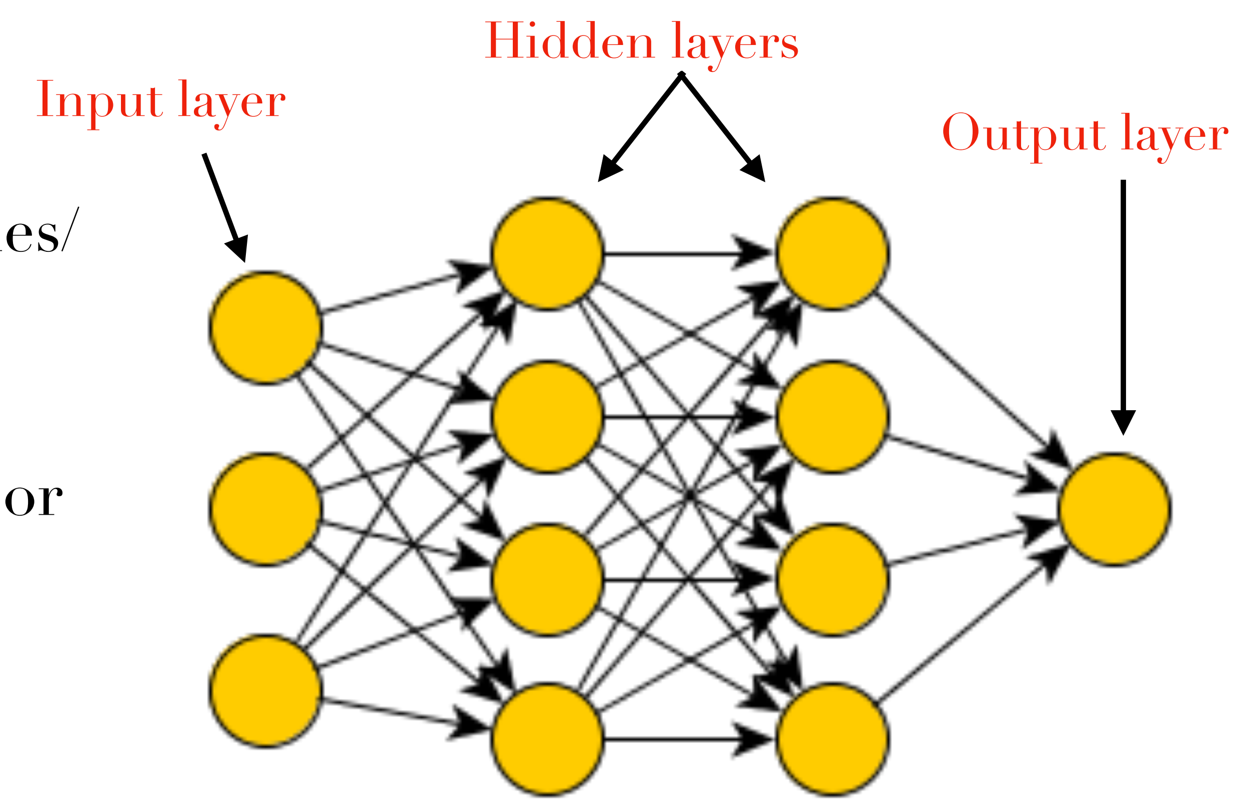
Neural Network (NN)

Basic Structure

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

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

C. Output layer: predictions

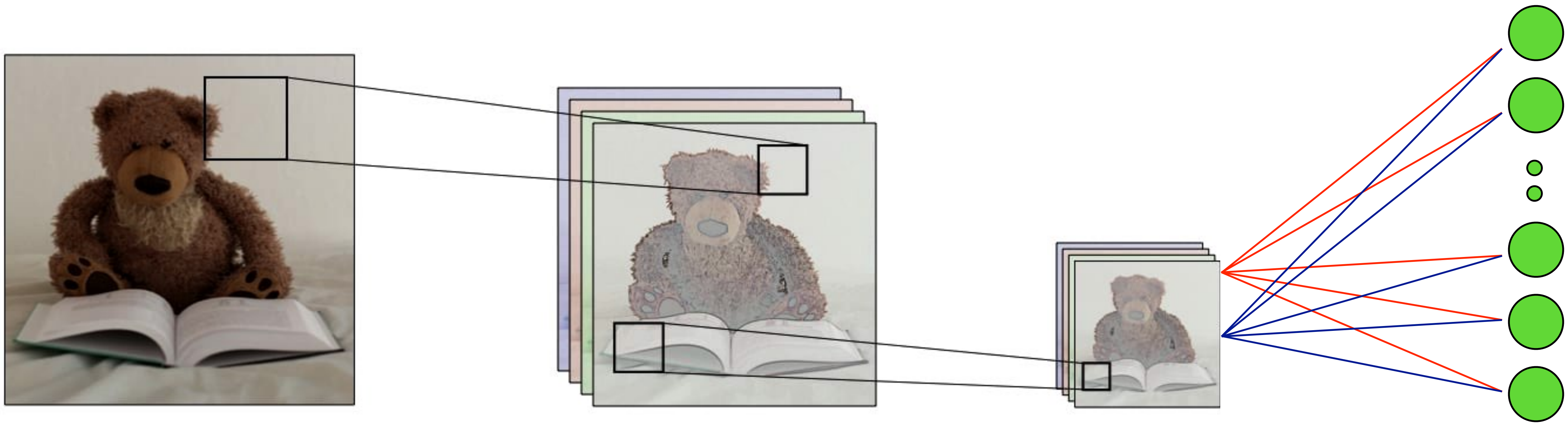


Schematic of a Neural Network

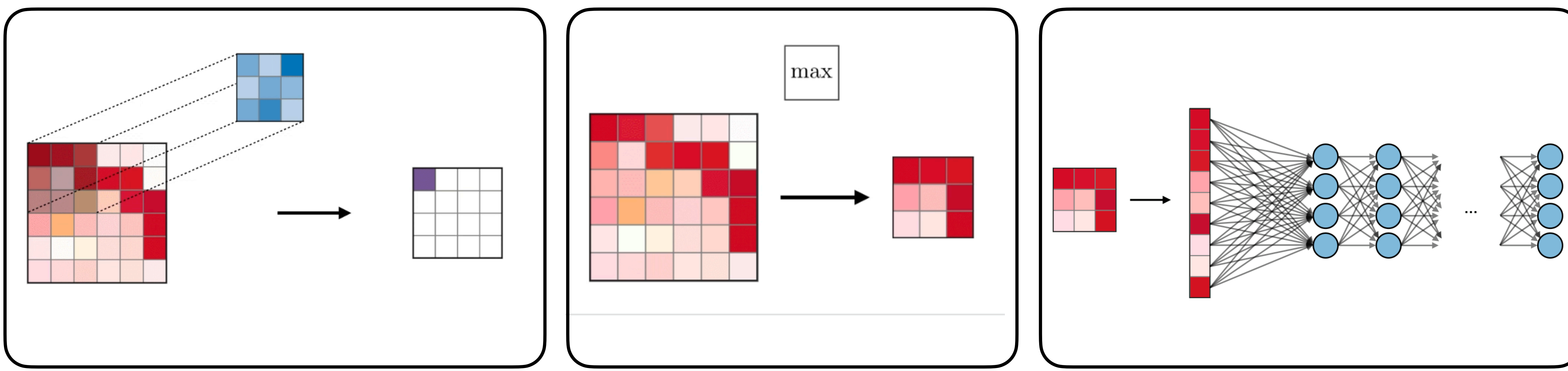
Training sample	Validation sample	Test sample

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

Convolutional Neural Networks (CNN)



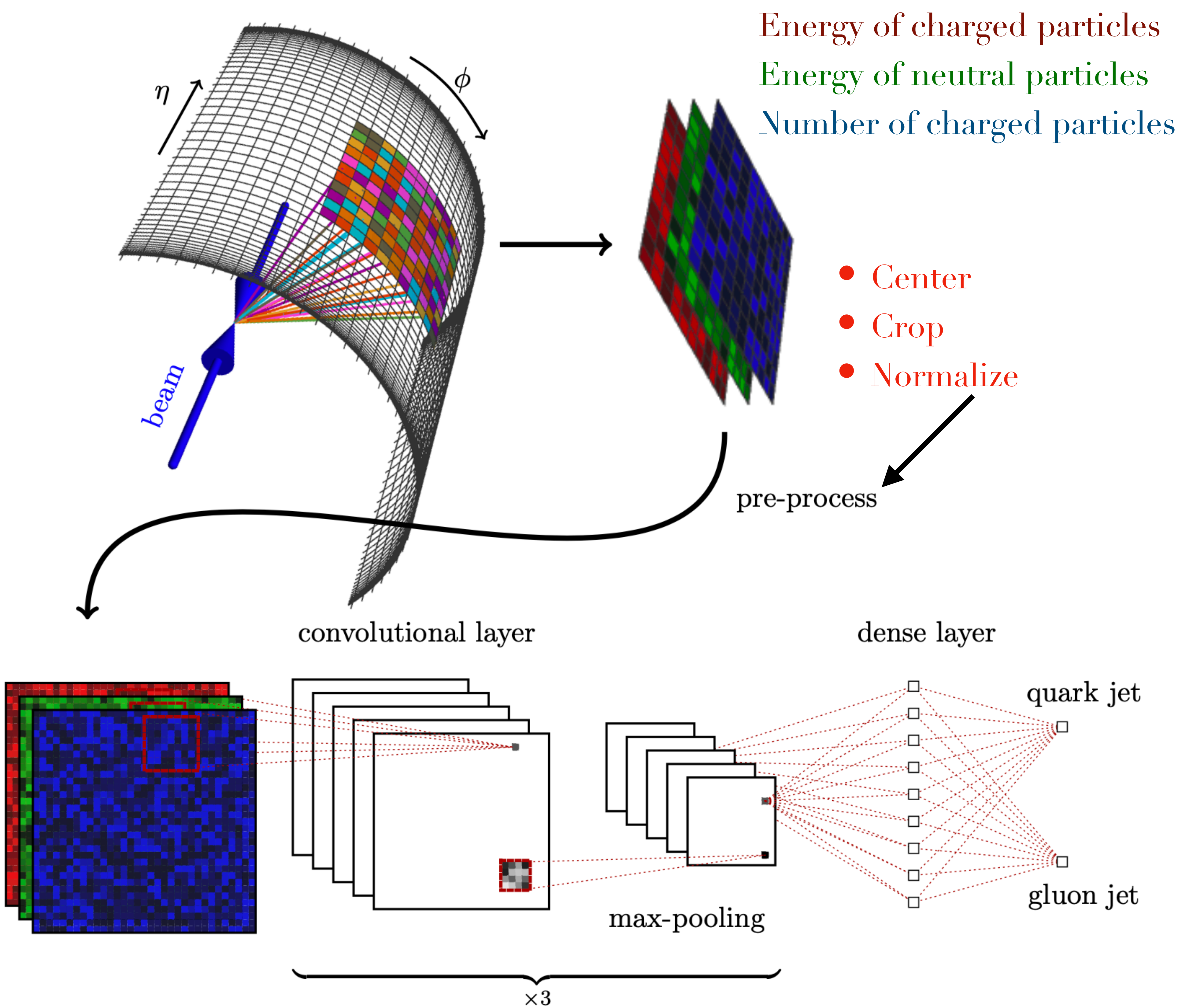
Input image Convolutions Pooling Fully connected



Convolution layer Pooling Fully connected

Source: stanford.edu

Jet Images as an Input to CNN



Benchmark 1: Top and QCD Jets

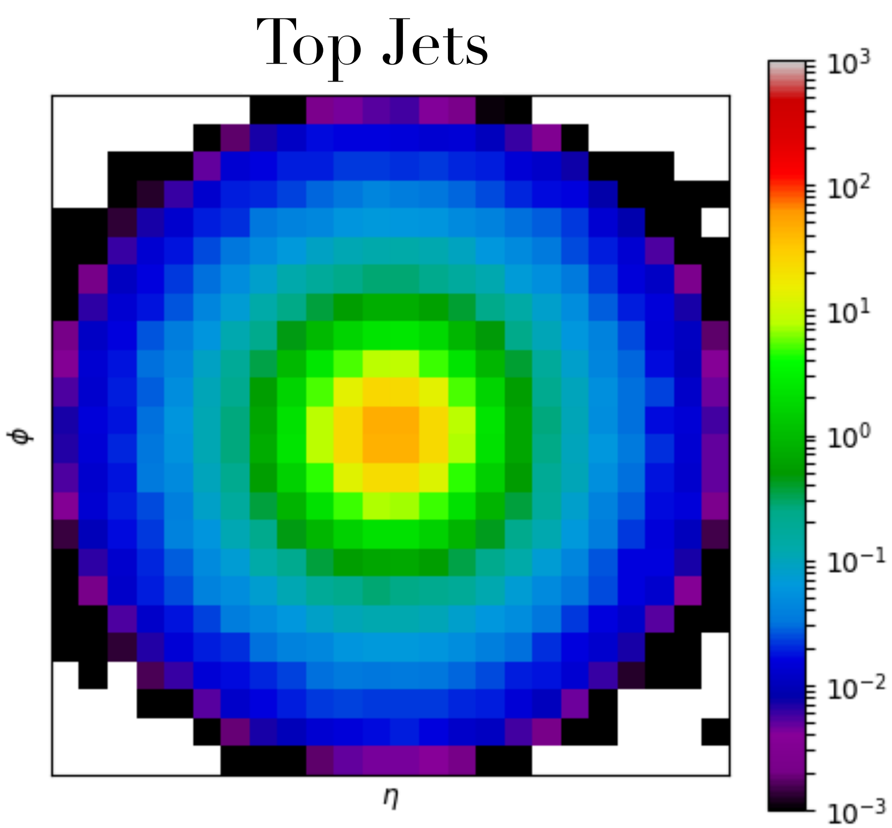
the input dataset

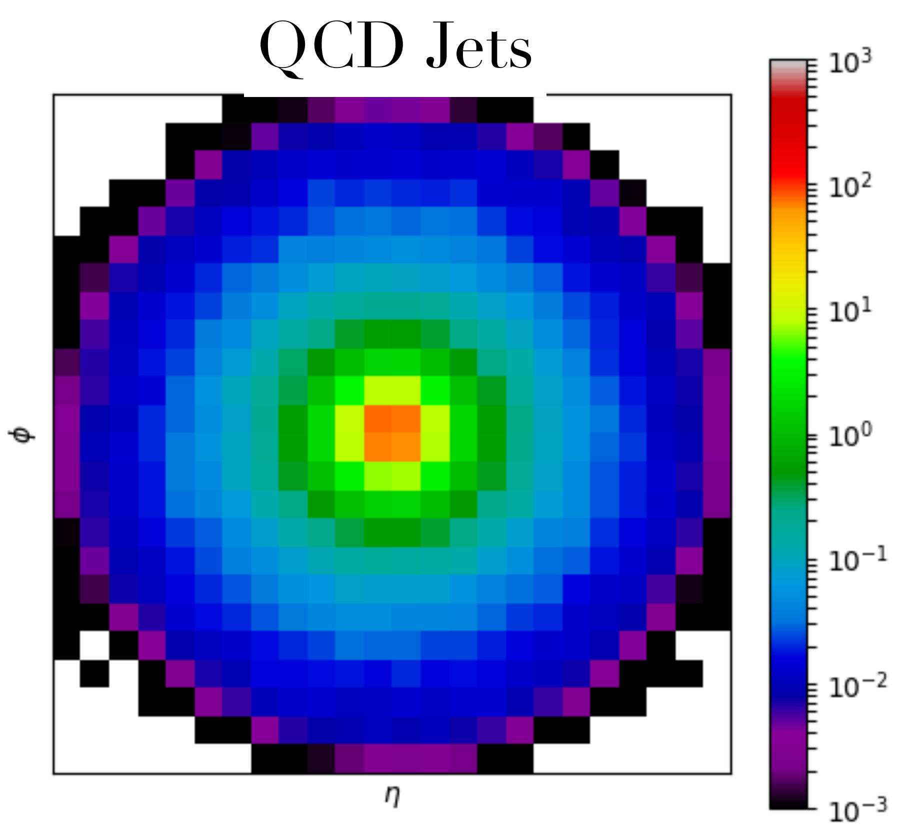
SM $t\bar{t}$ and QCD diJet production, $\sqrt{s} = 13\ TeV$

Madgraph + Pythia

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

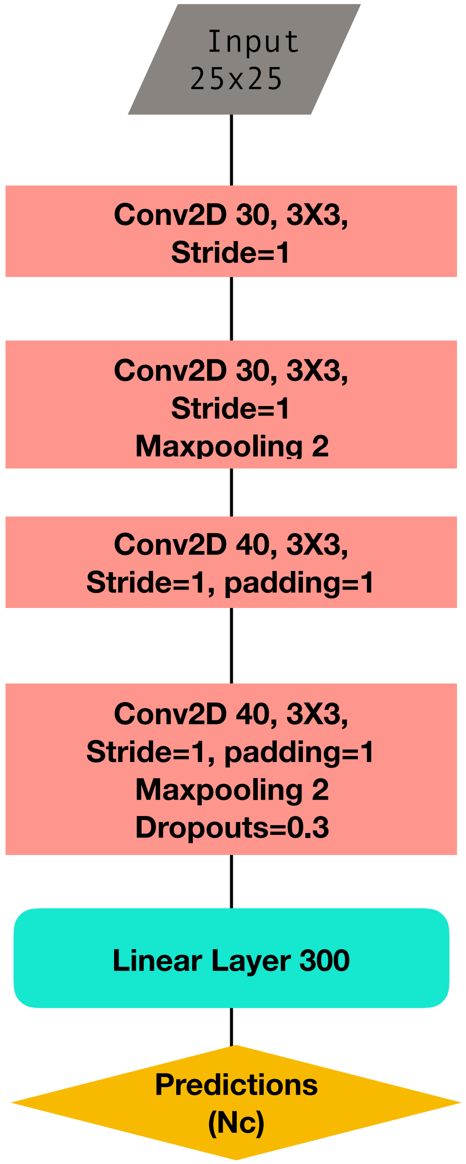
$\Delta\eta = 0.087, \Delta\phi = 0.087$





Averaged over 50K events

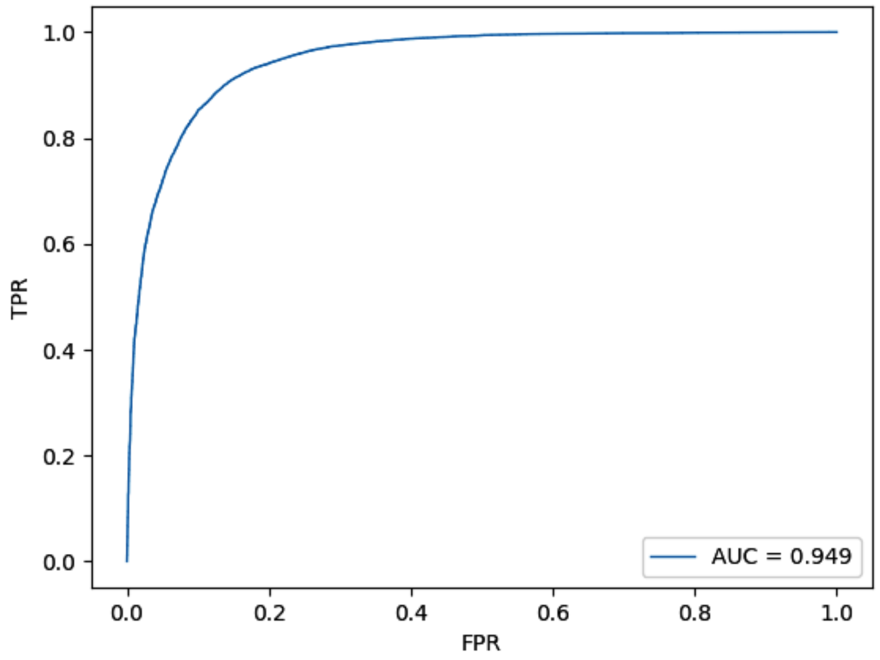
CNN for Top Vs QCD Classification (2C Baseline Classification)

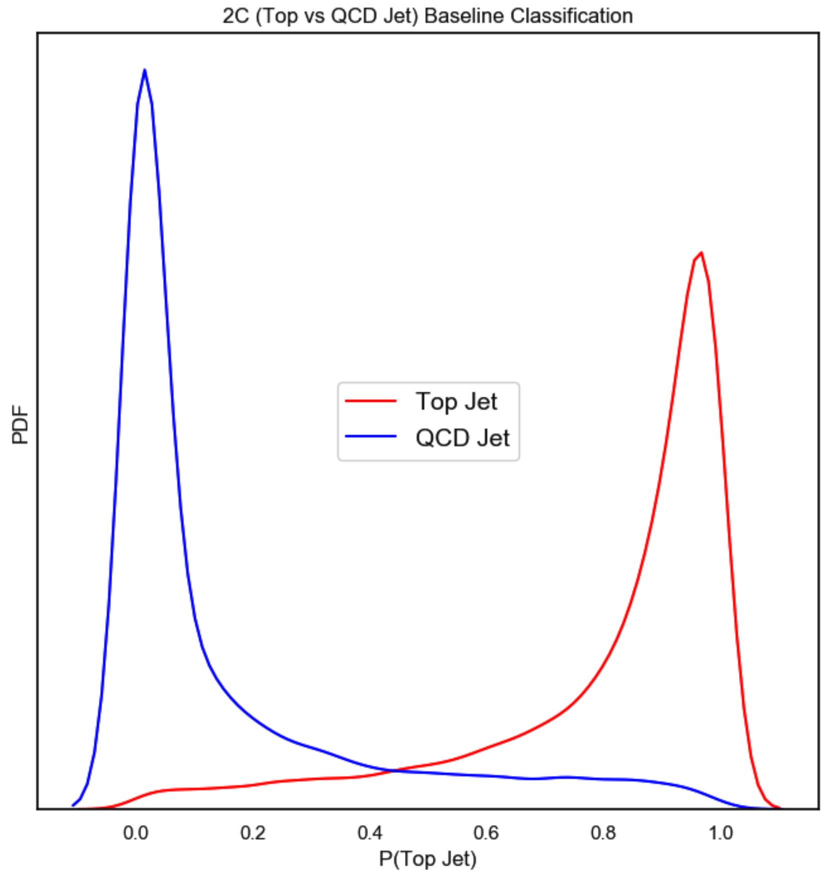


100K Events (balanced data)
Training:Test data= 70:30%

Batch Size=100
Epochs=100

Cross-Entropy Loss function

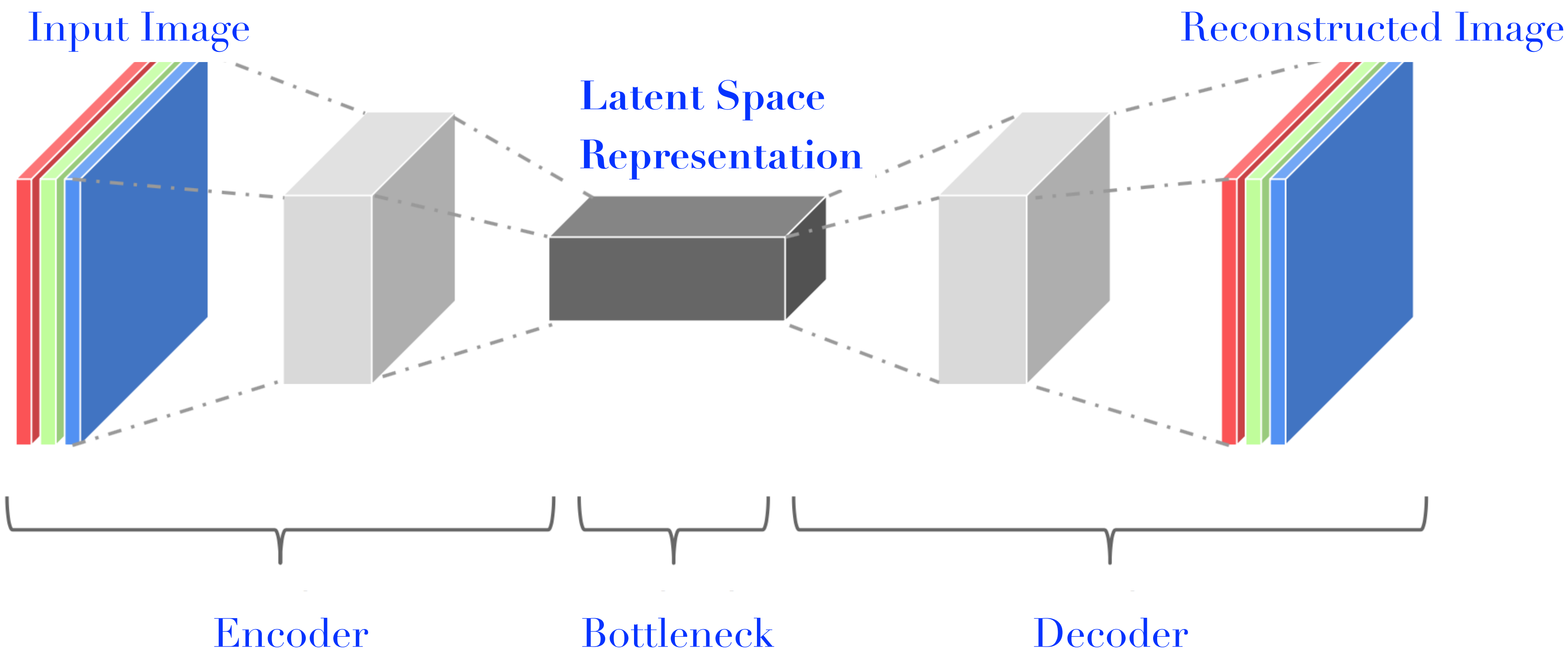




Related work

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

Auto-Encoders



Can 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, any others (see HEPML-LivingReview)

Statistics: Key Terms

- Probability Model $(p(x, y \,|\, \mu, \theta))$: Term used for the statistical model.
- Likelihood $L(\mu, \theta)$: Value of the probability for a fixed data set as a function of the parameters.
- Hypotheses (null (H_0) or alternative (H_1)): The null hypothesis is what one tries to test and the alternative hypothesis is what data is giving us.
- Test Statistics (t_μ): measures how well data favours either hypothesis.
- Significance: Quantify the discrepancy of observed data with the null hypothesis.

Standard Practice in Particle Physics

- Discovery Significance: discrepancy of the observed data with the null hypothesis. An actual experiment is conducted for which the null hypothesis predicts the probability of the observed data.
- Expected Significance: One considered signal hypothesis instead of actual data which predicts the probability of the observed outcome.
- Limit Setting: Role of null hypothesis and signal hypothesis is reversed. One rejects a parameter value if the discrepancy of the observed data with the signal hypothesis is larger than a threshold value.

Statistical Tests

Test statistic, profile likelihood and significance

$$t_\mu = -2\ln\lambda(\mu) \qquad \lambda(\mu) = \frac{L(\mu)}{L(\hat{\mu})} \qquad Z = \sqrt{t_\mu}$$

$$L(\mu) = \prod_{j=1}^{N} \frac{(\mu_j s_j + b_j)^{n_j}}{n_j!} e^{-(\mu_j s_j + b_j)}$$

s_j and b_j are the expected number of signal and background events in the j^{th} bin.

n_j is the total number of events in the j^{th} bin according to the alternate hypothesis H_1.

<u>Expected discovery significance</u>

$$E[Z_0] = \sqrt{-2\left[\sum_{j=1}^{N}\left(s_j + (b_j + s_j)\ln\left(\frac{b_j}{b_j + s_j}\right)\right)\right]}$$

<u>Expected exclusion significance</u>

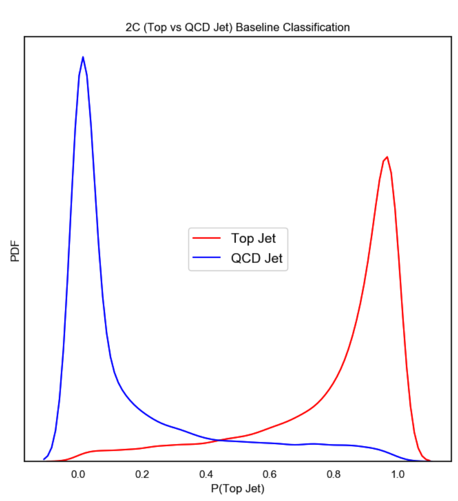
$$E[Z_e] = \sqrt{2\left[\sum_{j=1}^{N}\left(s_j + b_j\ln\left(\frac{b_j}{b_j + s_j}\right)\right)\right]}$$

Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells, Eur. Phys. J. C 71, 1554 (2011); Eur. Phys. J. C 73, 2501(E) (2013).

Our Approach

- A simple dictionary connecting the dots of this known knowledge about machine learning methods and statistical approaches.
- More like thinking about possible ways to connect the dots.
- An attempt to build a recipe for any type of output to statistical significance.

ML output



Directly use ROC?

Toy experiments and log-likelihood?

Should we use full distribution?

What are the equivalent options for un-supervised learning?

How to implement uncertainties?

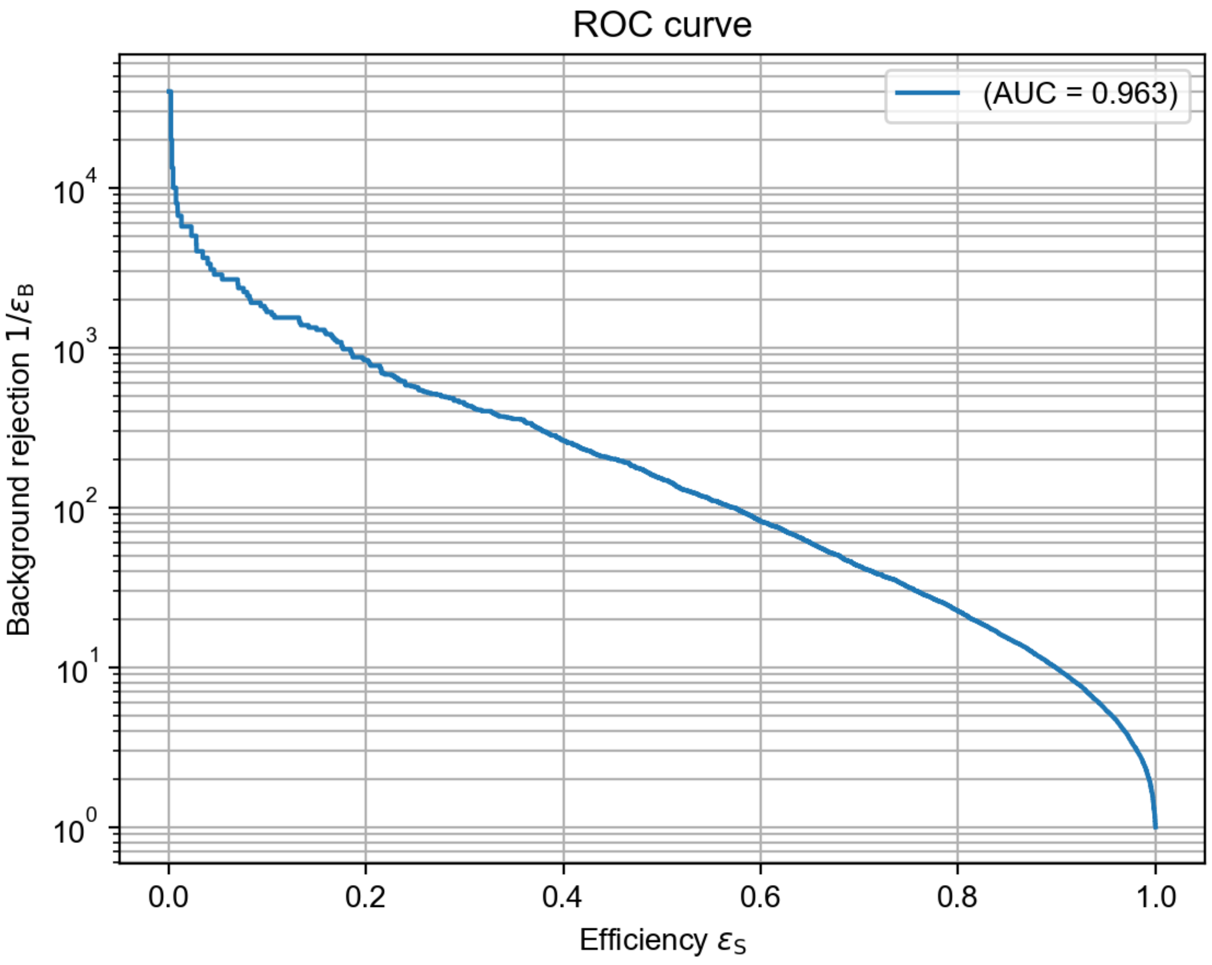
Significance

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

Asimov Significance

C. K. Khosa, V. Sanz and M. Soughton, A simple guide from Machine Learning outputs to statistical criteria, [arXiv:2203.03669 [hep-ph]].

Good First Approximation



- Choose a point on the ROC curve
- Calculate $S/\sqrt{B}$ for that point considering the cross-section

Log-likelihood Ratio (LLR) Test Statistics

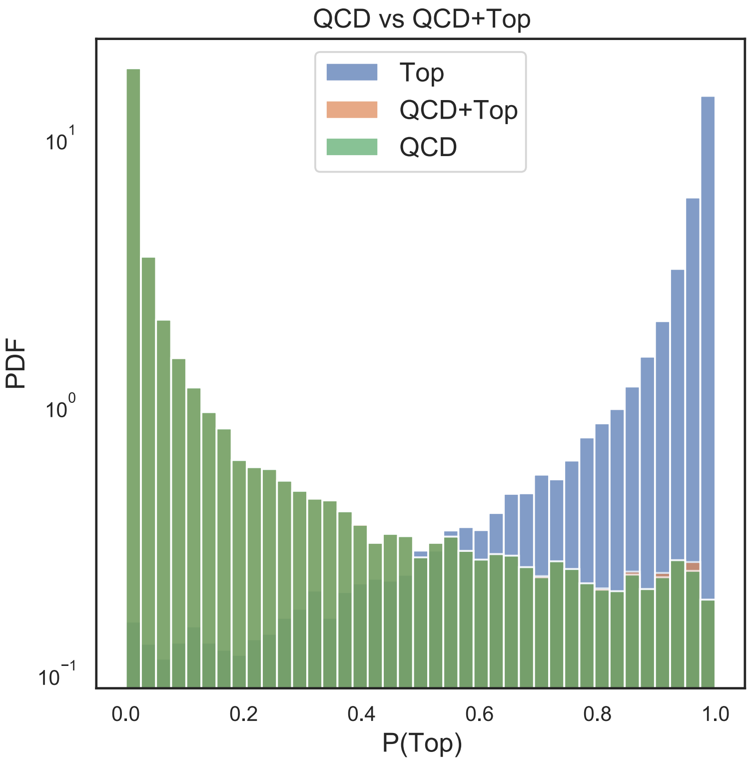
Likelihood for hypothesis H in a single pseudo-experiment

$$L_H = \prod_{1}^{events} pdf_H(x)$$

QCD vs top+QCD

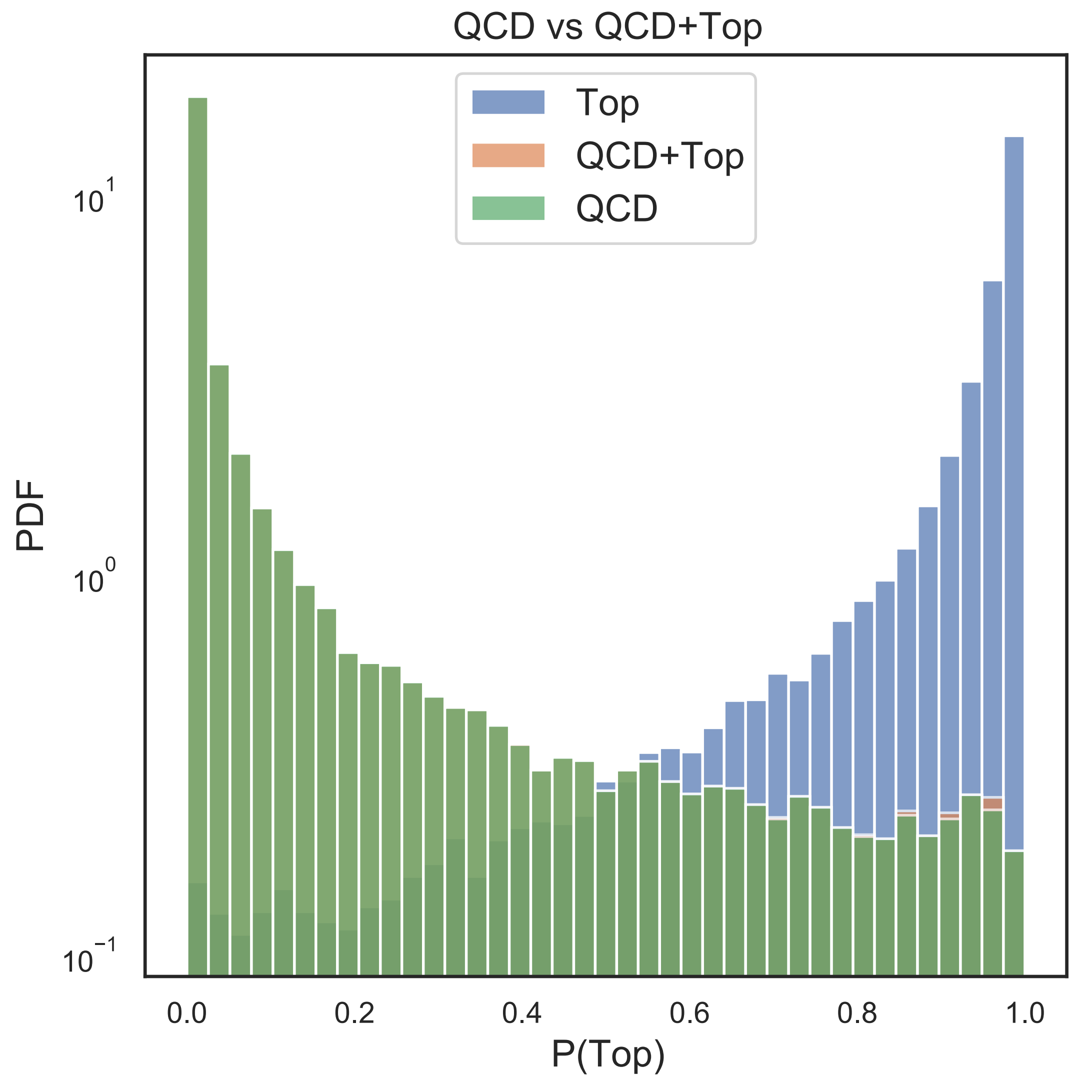
$$\Lambda_{QCD} = -2\ln\frac{L(QCD, QCD)}{L(top+QCD, QCD)}$$

$$\Lambda_{top+QCD} = -2\ln\frac{L(QCD, top+QCD)}{L(top+QCD, top+QCD)}$$

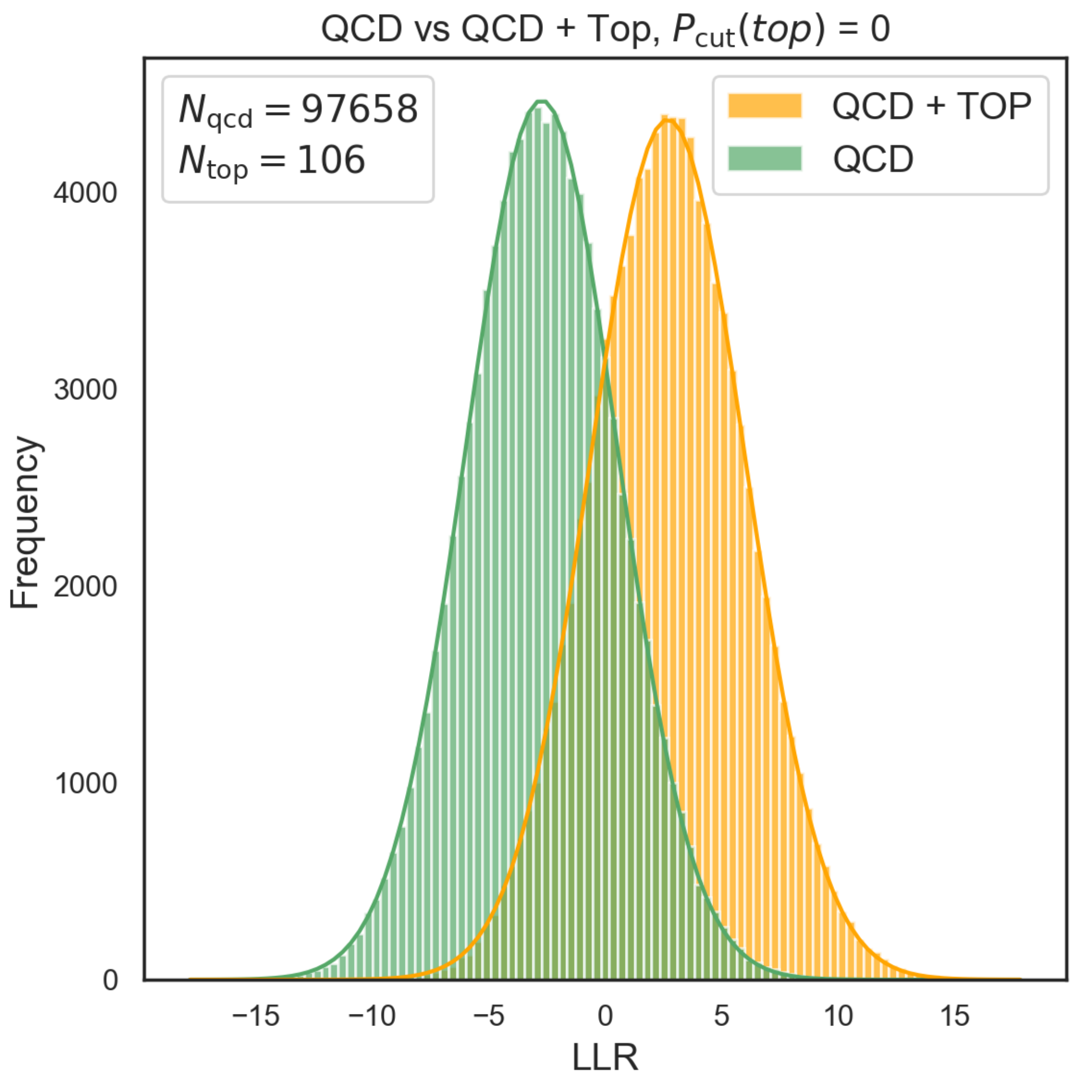


S. Baker and R. D. Cousins, Nucl. Instrum. Meth. 221, 437 (1984). R. Cousins, J. Mumford, J. Tucker and V. Valuev, Journal of High Energy Physics 2005(11), 046 (2005)

PDF and LLR



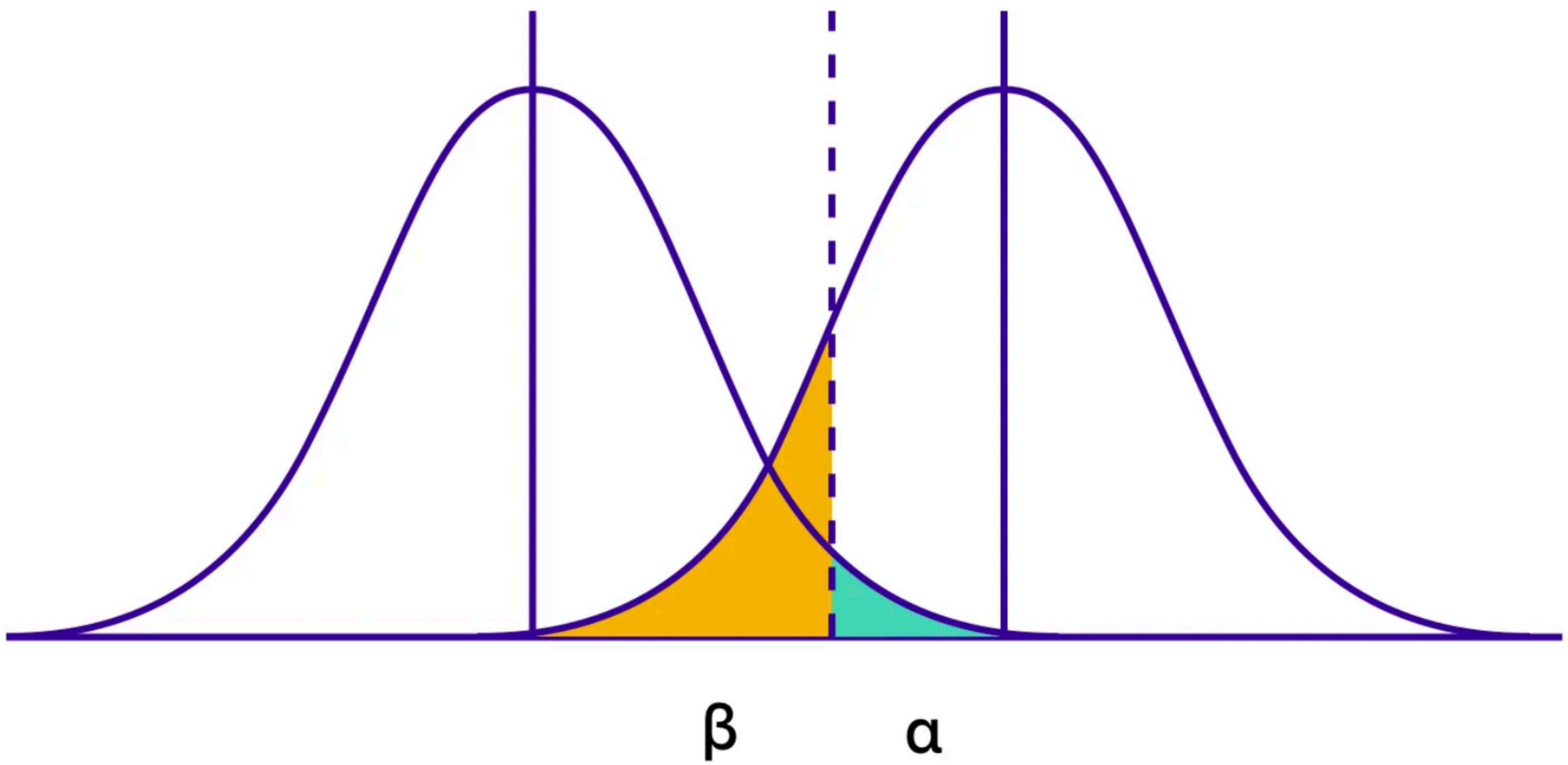
QCD vs QCD+Top
Top
QCD+Top
QCD
PDF
10^{1}
10^{0}
10^{-1}
0.0
0.2
0.4
0.6
0.8
1.0
P(Top)



QCD vs QCD + Top, $P_{\mathrm{cut}}(top)$ = 0
$N_{\mathrm{qcd}} = 97658$
$N_{\mathrm{top}} = 106$
QCD + TOP
QCD
Frequency
4000
3000
2000
1000
0
−15
−10
−5
0
5
10
15
LLR

Translating to Significance

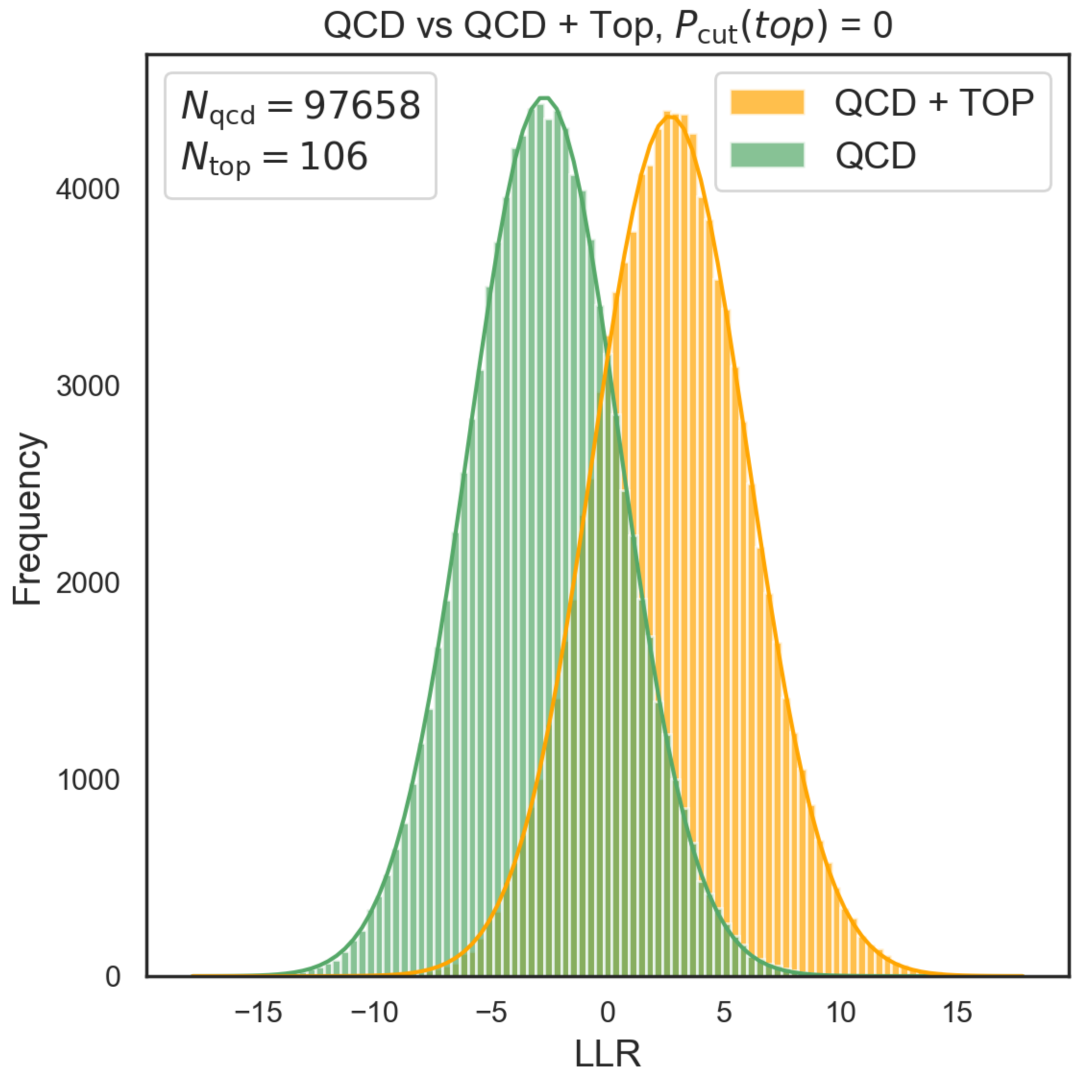
H_0 (Null Hypothesis) H_1 (Alternative Hypothesis)

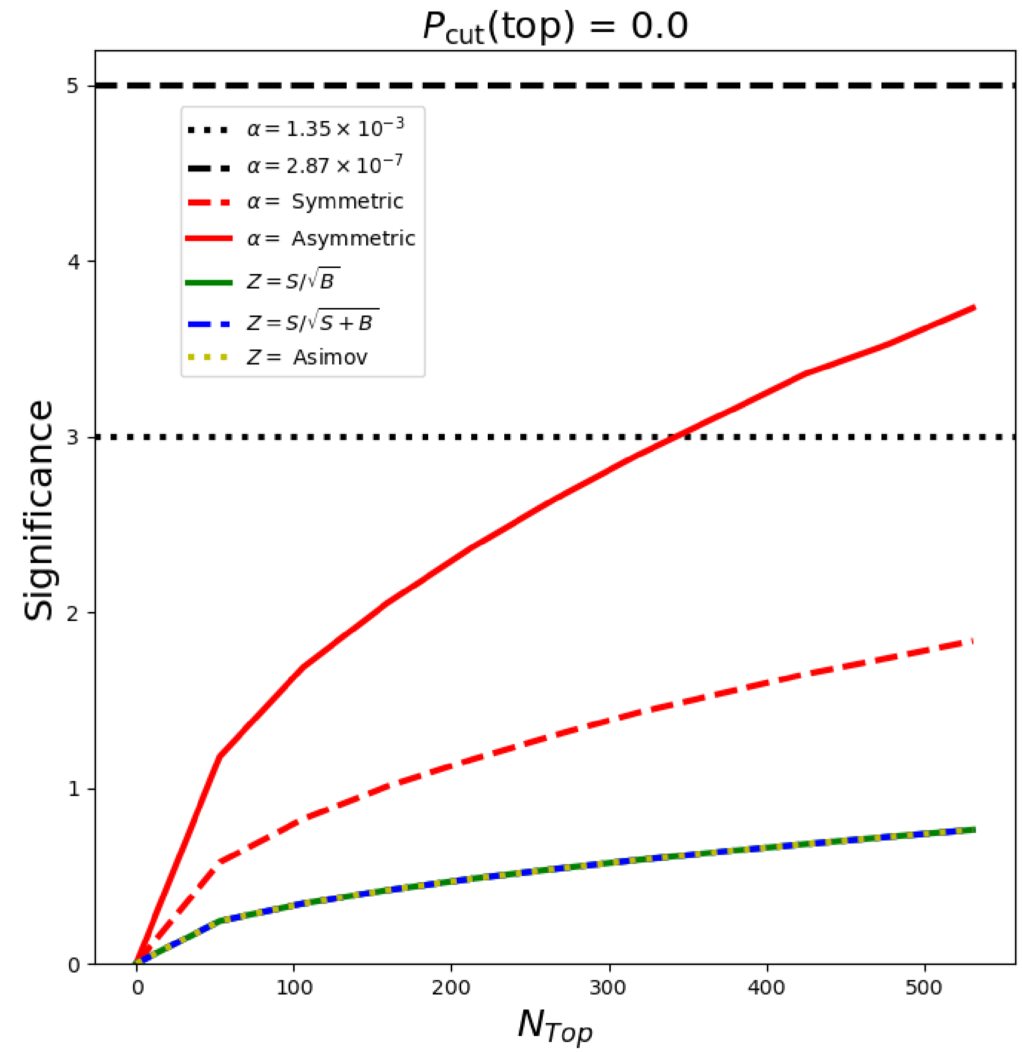


Type II Error Type I Error

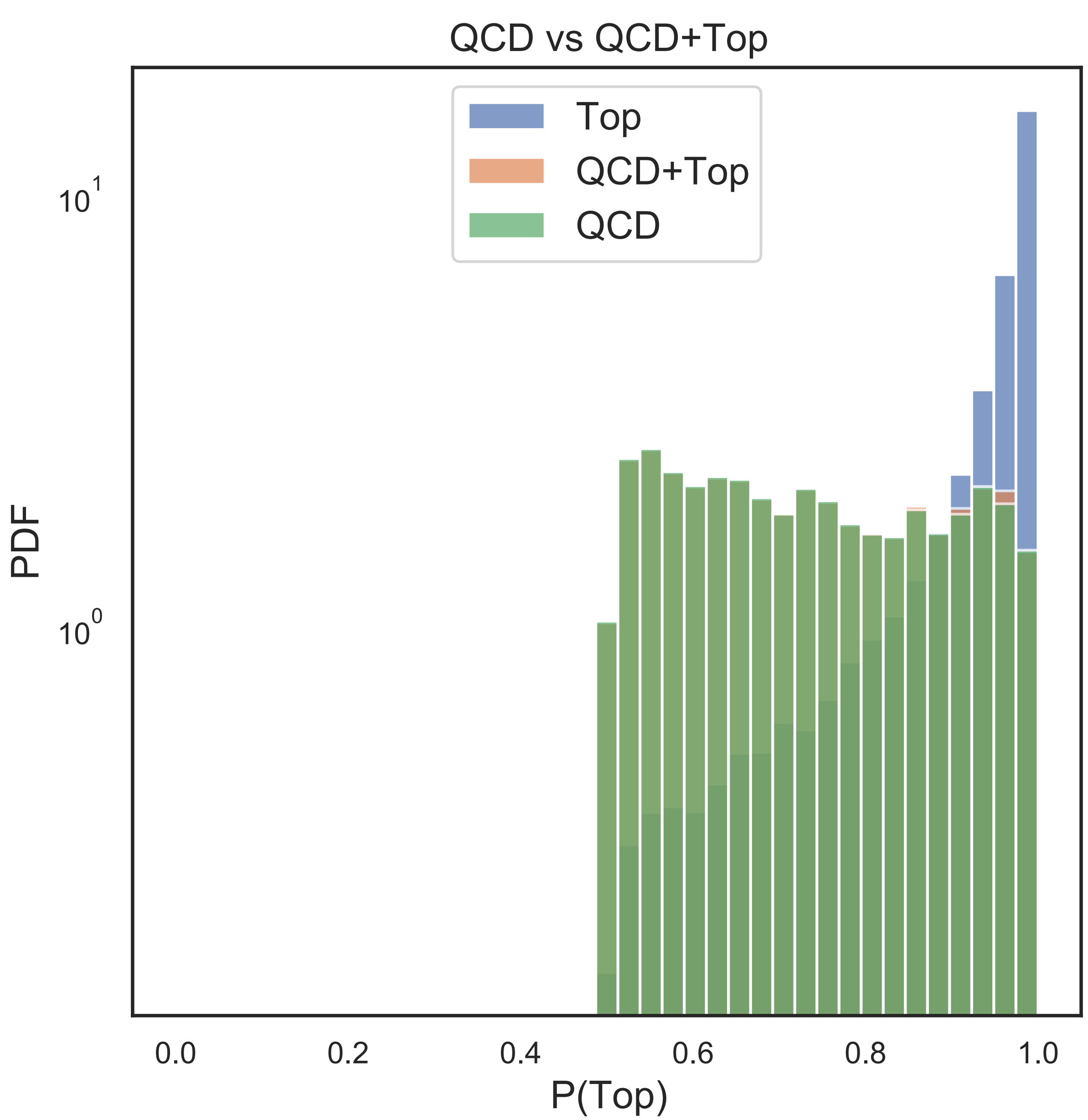
$$\alpha = \frac{\int_{\Lambda_{cut}}^{\infty} f_0(\Lambda)d\Lambda}{\int_{-\infty}^{\infty} f_0(\Lambda)d\Lambda} \qquad \beta = \frac{\int_{-\infty}^{\Lambda_{cut}} f_1(\Lambda)d\Lambda}{\int_{-\infty}^{\infty} f_1(\Lambda)d\Lambda}$$

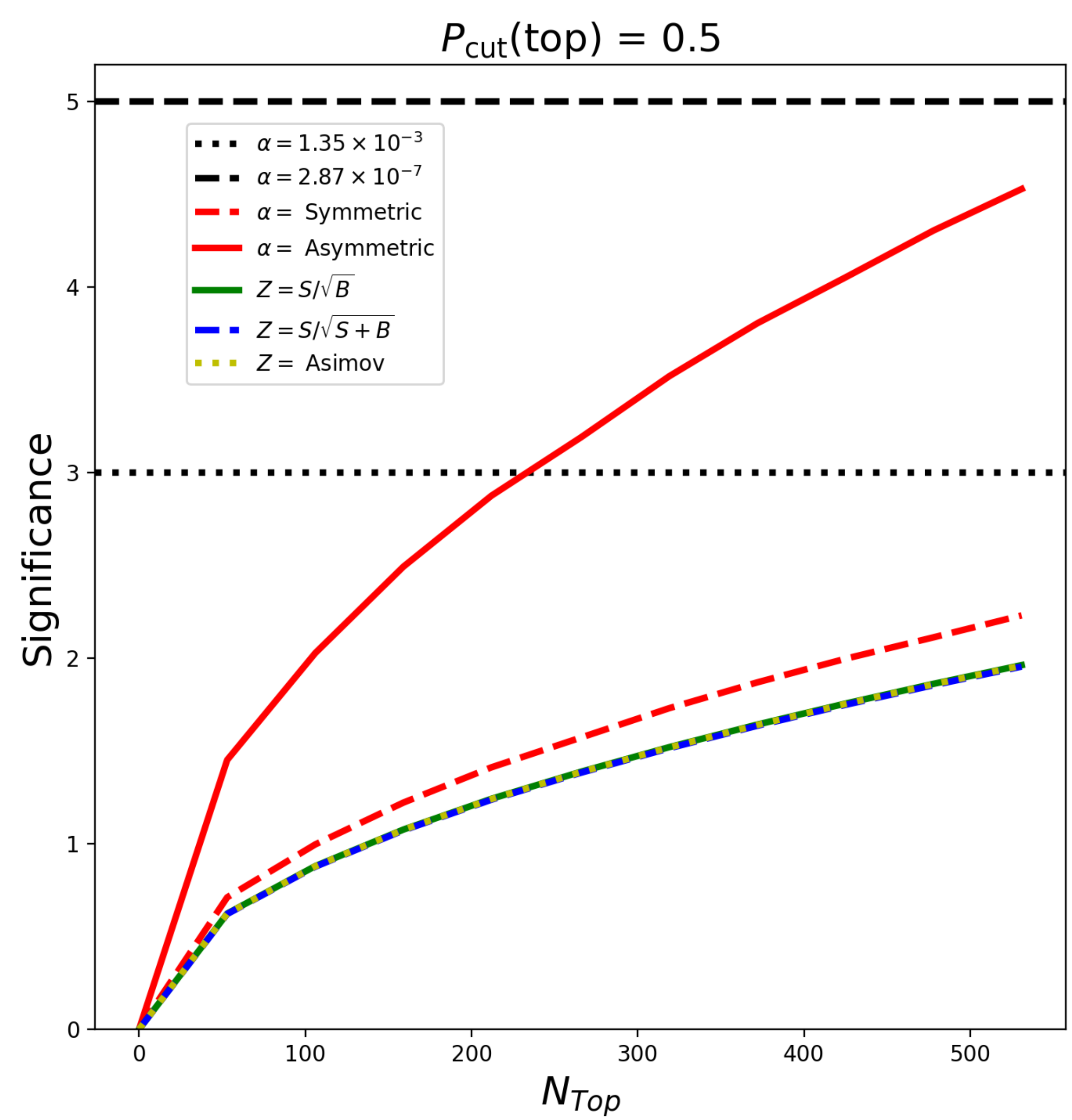
Significance Comparison

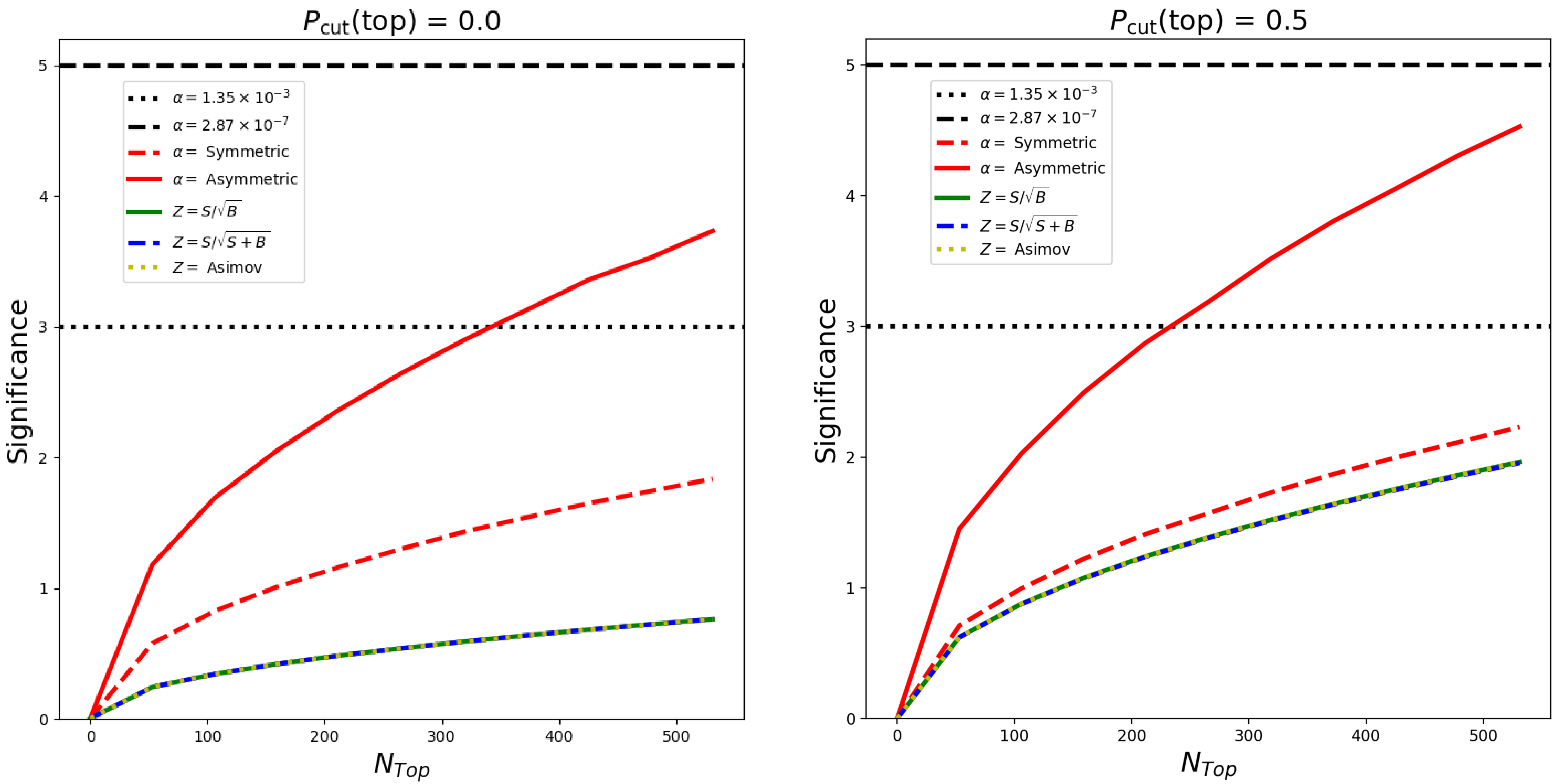




A cut on Classifier Output

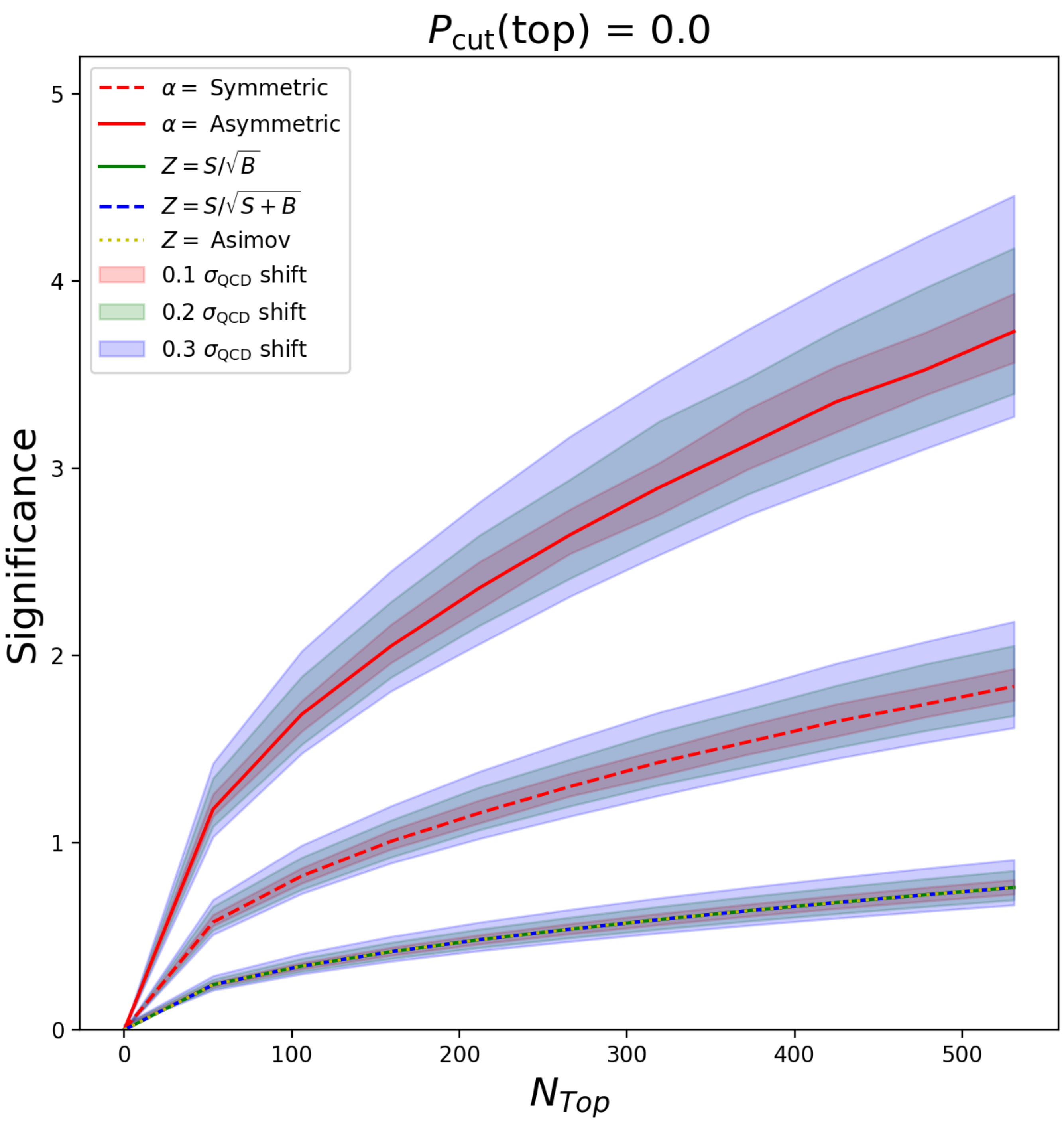




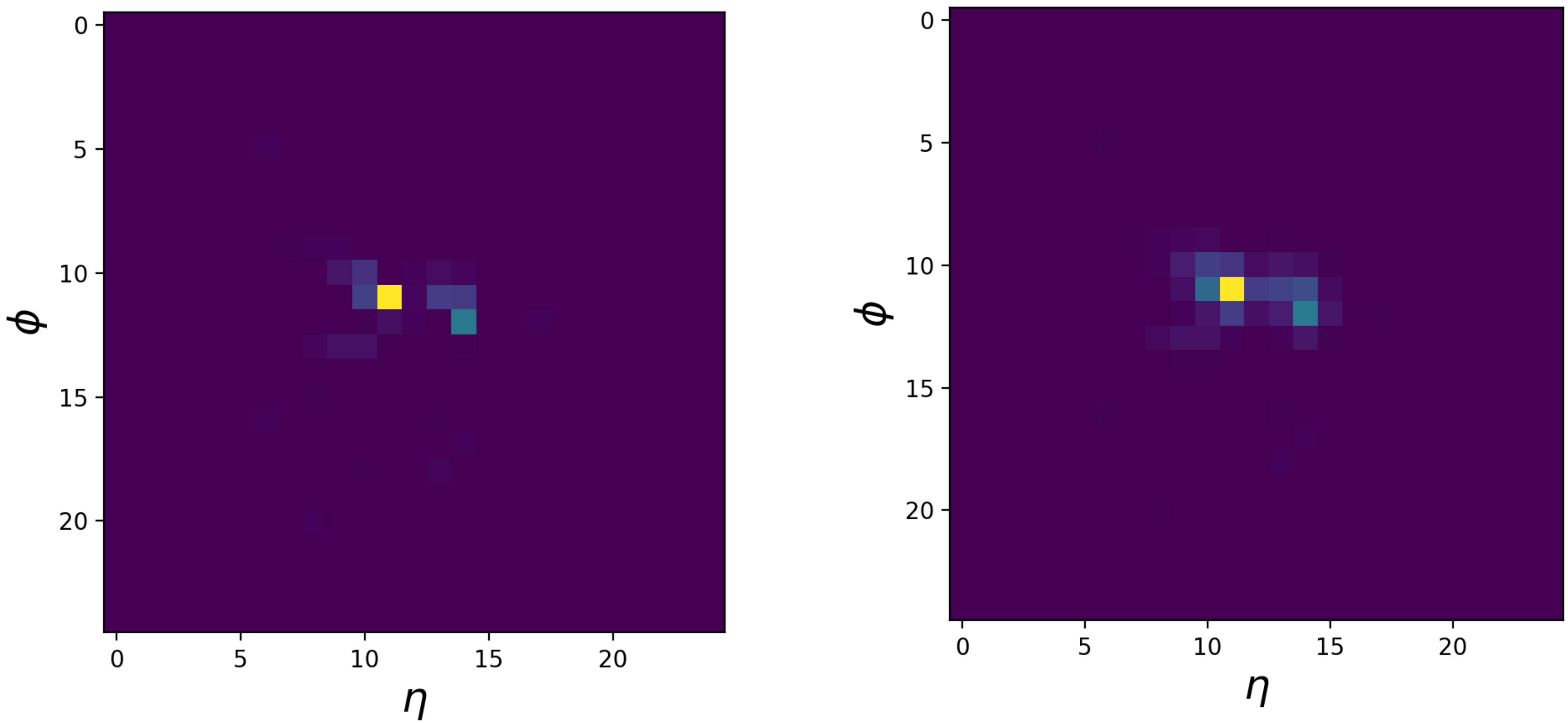


P_{cut}(top) = 0.0
P_{cut}(top) = 0.5
Significance
N_{Top}
$\alpha = 1.35 \times 10^{-3}$
$\alpha = 2.87 \times 10^{-7}$
$\alpha =$ Symmetric
$\alpha =$ Asymmetric
$Z = S/\sqrt{B}$
$Z = S/\sqrt{S+B}$
$Z =$ Asimov

Effect of Uncertainties

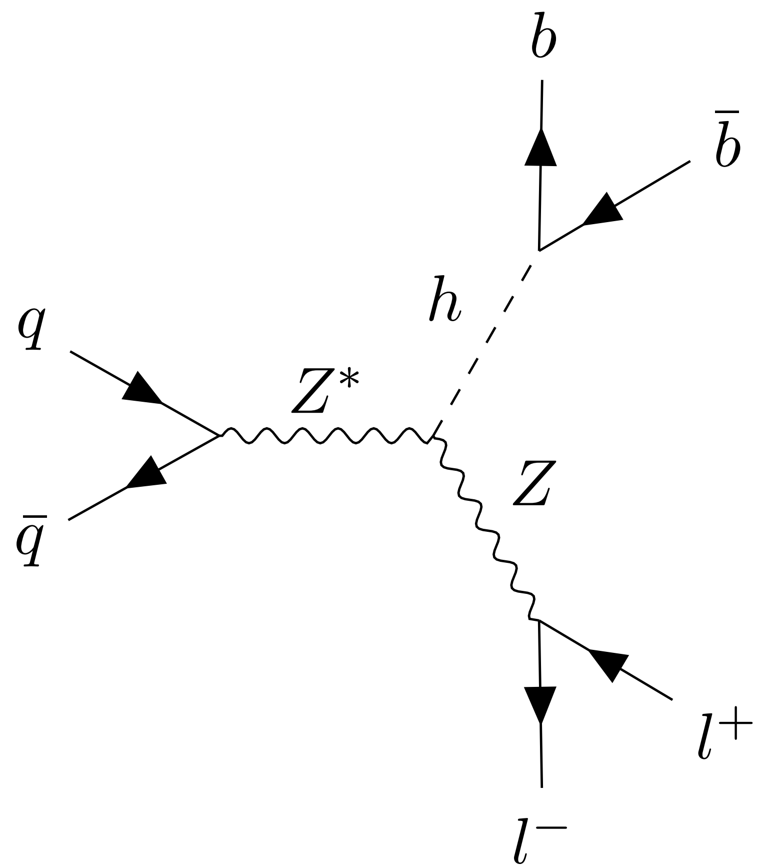


Effect of Uncertainties



QCD jet image before and after blurring.

Higgs EFT Benchmark: Supervised Approach

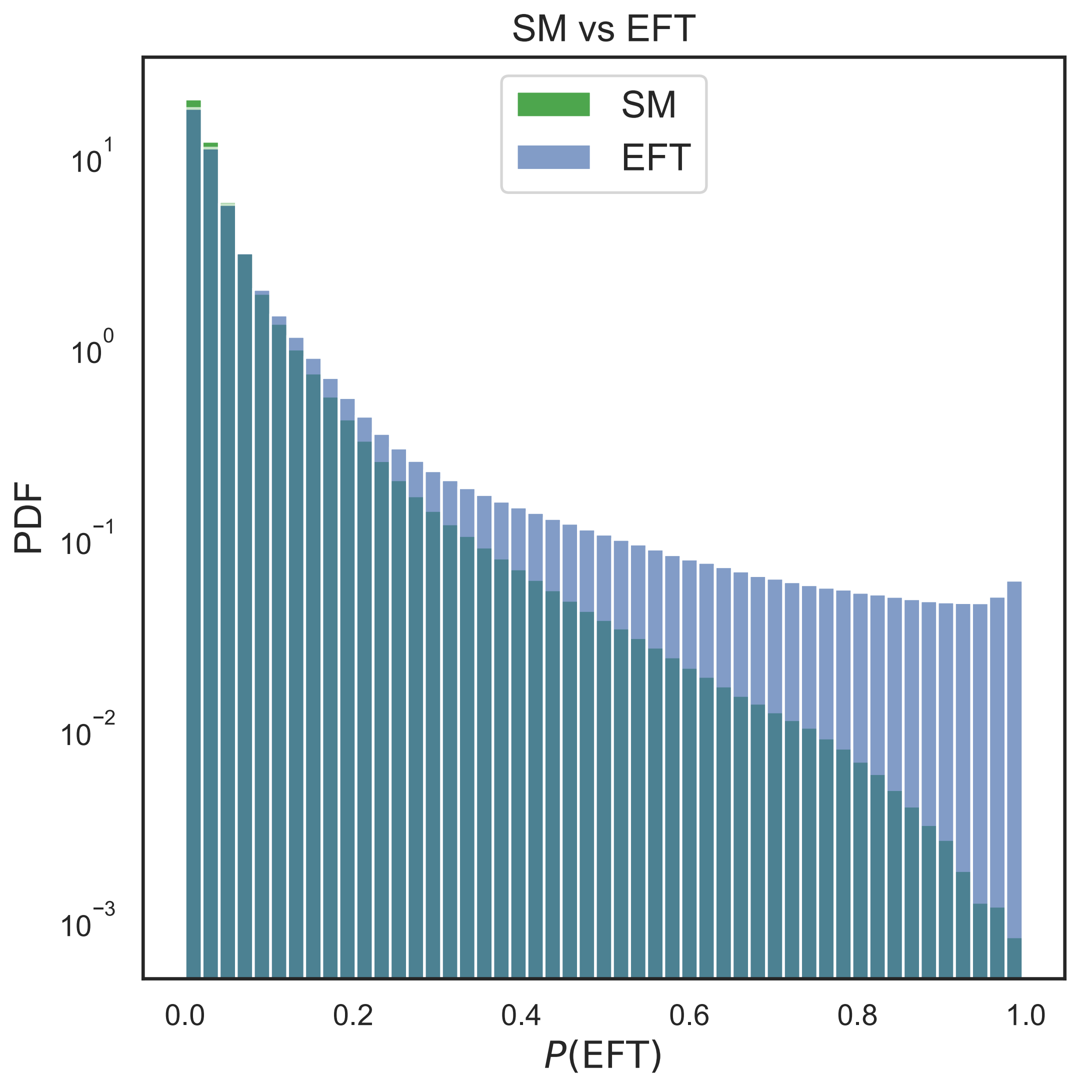


$$\mathscr{L}_{EFT} = \mathscr{L}_{SM} + \mathscr{L}_{BSM} \quad \text{where } \mathscr{L}_{BSM} = \frac{1}{\Lambda^{2n}} \sum_i c_i \mathcal{O}_i$$

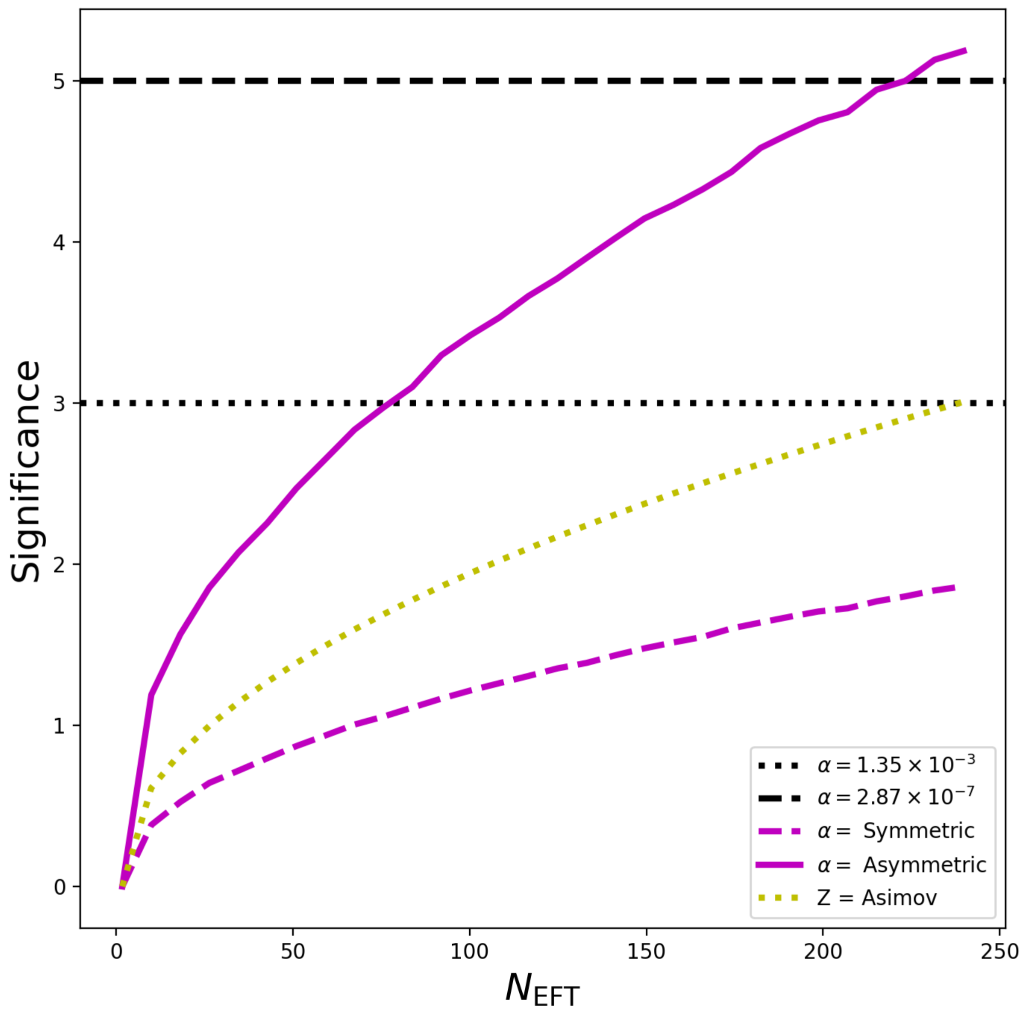
$$\mathscr{L}_{BSM} \supset ig \frac{c_{HW}}{m_W^2} (D^\mu H)^\dagger \sigma_a (D^\nu H) W^a_{\mu\nu}$$

$$p_T^{b_1}, p_T^{b_2}, p_T^{l_1}, p_T^{l2}, p_T^H, \eta_H, \phi_H, \delta R_{l_1 l_2}, \delta R_{b_1 l_1}, M_T^{ZH}, p_T^{ZH}, \delta\phi_{l1b1}, d\phi_{l1b2}$$

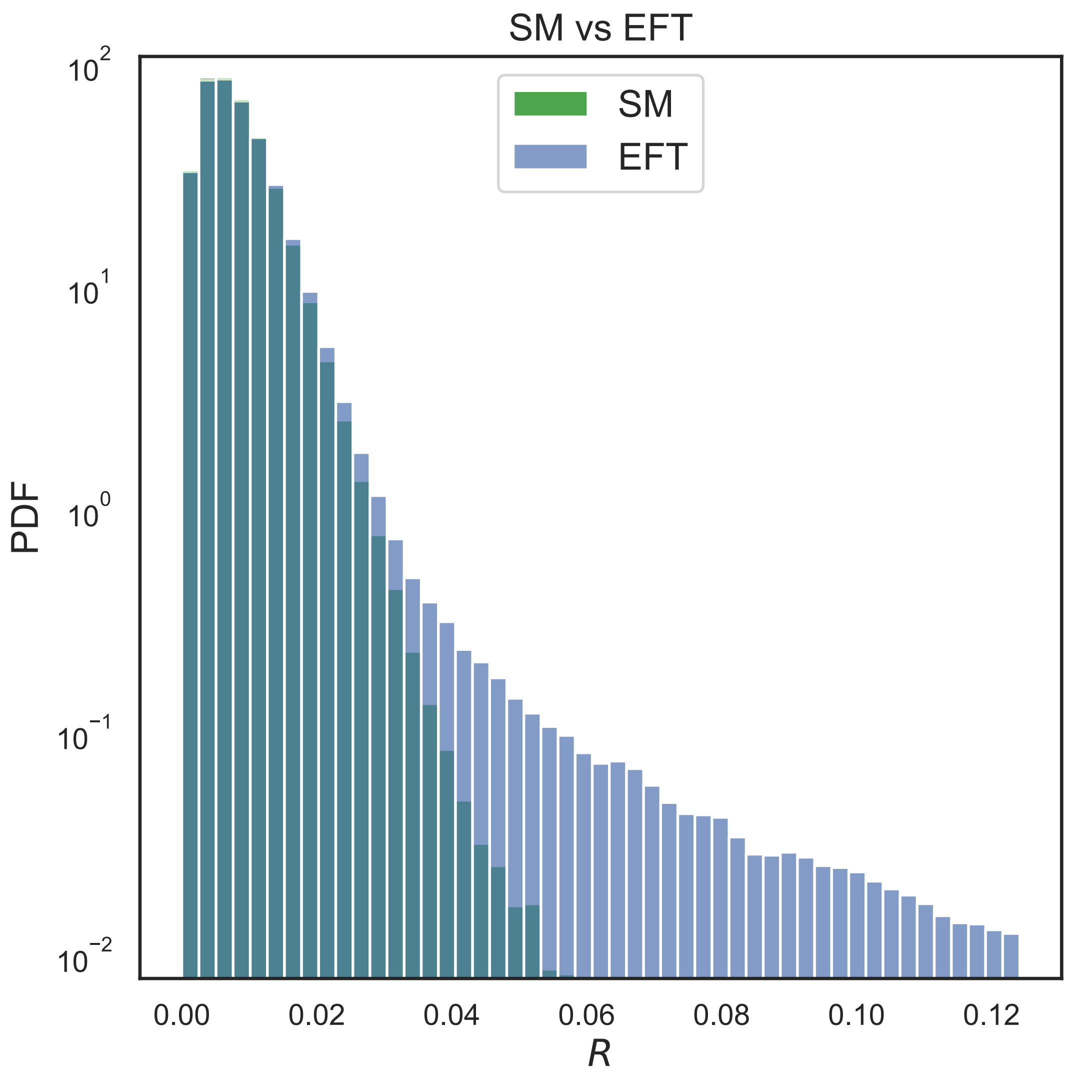
Higgs EFT Benchmark

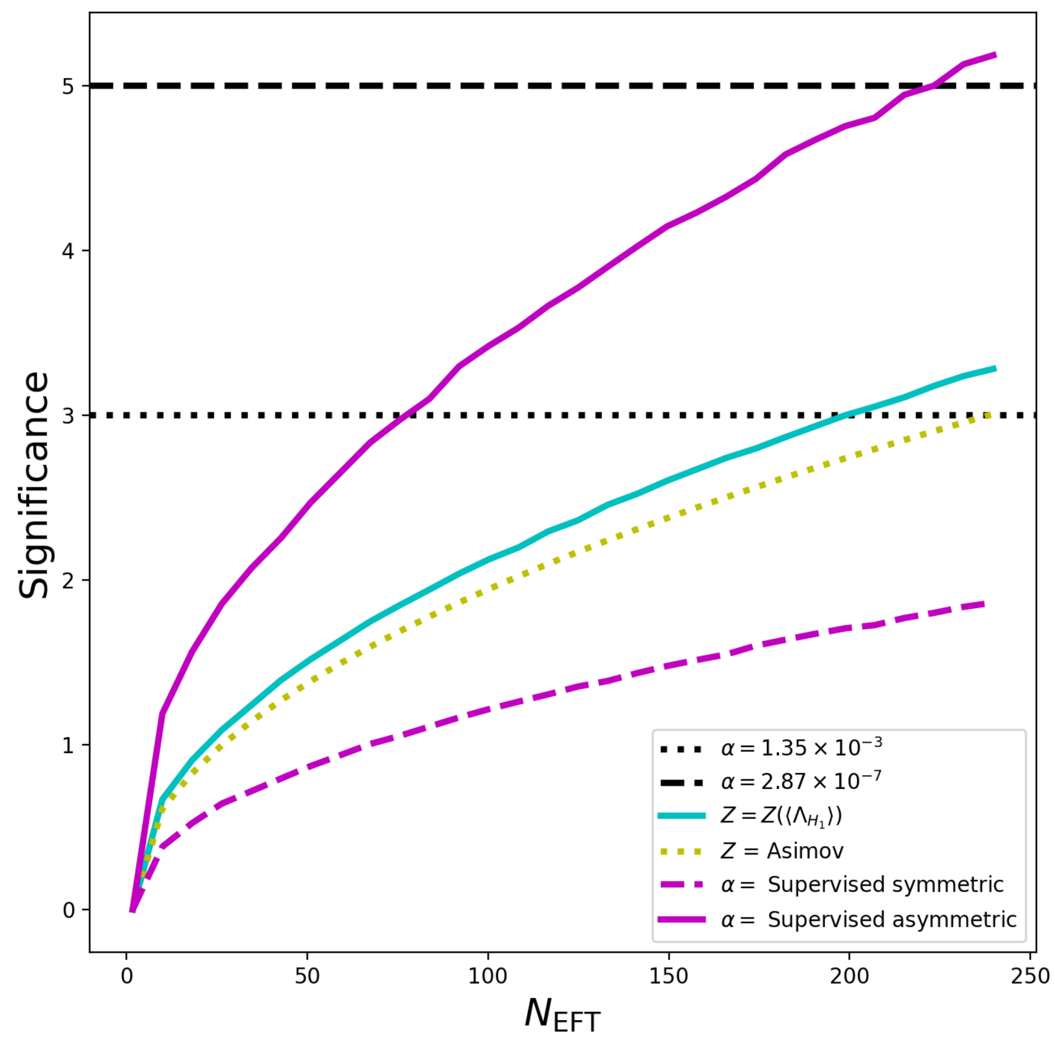


DNN output

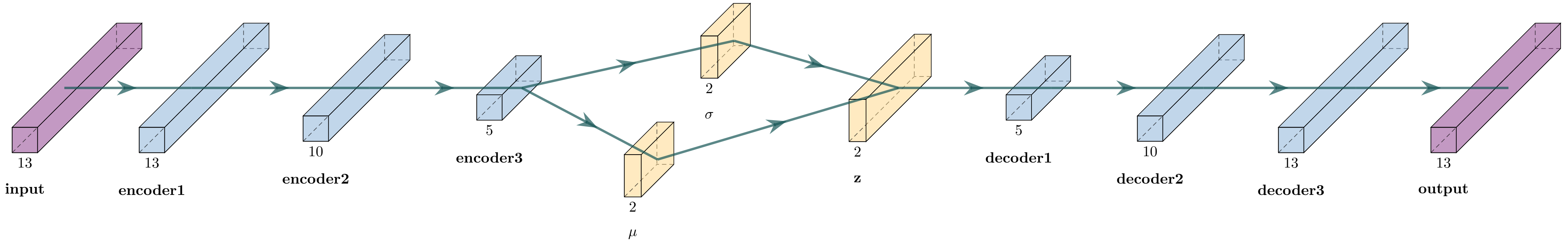


Unsupervised Machine Learning





Reconstruction error using VAE



Other Interesting Proposals

- Madminer and DNNlikelihood (inference from likelihood ratios)

J. Brehmer, F. Kling, I. Espejo and K. Cranmer, Big Sci. 4(1), 3 (2020), 1907.10621. 139 (2021), J. Brehmer, K. Cranmer and F. Kling, Int. J. Mod. Phys. A 35(15n16), 2041008 (2020). K. Cranmer, J. Pavez and G. Louppe, Cla (2015), 1506.02169. K. Cranmer, J. Brehmer and G. Louppe, The frontier of simulation-based inference, Proc. Nat. Acad. Sci. 117(48), 30055 (2020). A. Coccaro, M. Pierini, L. Silvestrini and R. Torre, The DNNLikelihood: enhancing likelihood distribution with Deep Learning, Eur. Phys. J. C 80(7), 664 (2020).

- Anomaly Detection

R. T. D'Agnolo and A. Wulzer, Phys. Rev. D 99 (2019) no.1, 015014), P.De Castro and T. Dorigo, Comput. Phys. Commun. 244 (2019), 170-179.

- And many more (see HEPML-LivingReview for a complete list of references.)

Summary and Outlook

- ML techniques are emerging as a competitive tool to look for new phenomena in complex data and more efficiently identify known objects. It seems like an appropriate tool for particle physics.

- Our community is adapting these techniques for various tasks. I talked about the ways to bridge the gap between the ML outputs and statistical criteria.

- In particular, I talked about how from the ML output one could perform a simple hypothesis testing.

- As a next step, it would be interesting to consider more complicated situations to assess the performance of this set-up.

- Other ways of using a neural network as a test statistic and further uncertainty quantification.

Future public releases of likelihoods from experiments could incorporate information on ML training outputs following the lines of this paper.

SciPost Physics Submission

Publishing statistical models: Getting the most out of particle physics experiments

September 9, 2021

Abstract

The statistical models used to derive the results of experimental analyses are of incredible scientific value and are essential information for analysis preservation and reuse. In this paper, we make the scientific case for systematically publishing the full statistical models and discuss the technical developments that make this practical. By means of a variety of physics cases — including parton distribution functions, Higgs boson measurements, effective field theory interpretations, direct searches for new physics, heavy flavor physics, direct dark matter detection, world averages, and beyond the Standard Model global fits — we illustrate how detailed information on the statistical modelling can enhance the short- and long-term impact of experimental results.