Universidade do Minho
Escola de Ciências

LIP

LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

Big Data HEP

Data Science in High Energy Physics

Nuno Castro, Rute Pedro,
Miguel Romão, Tiago Vale

9th Ibergrid
25th September 2019
Santiago de Compostela

POCI/01-0145-FEDER-029147
PTDC/FIS-PAR/29147/2017

FCT Fundação para a Ciência e a Tecnologia

Lisb@20[20]

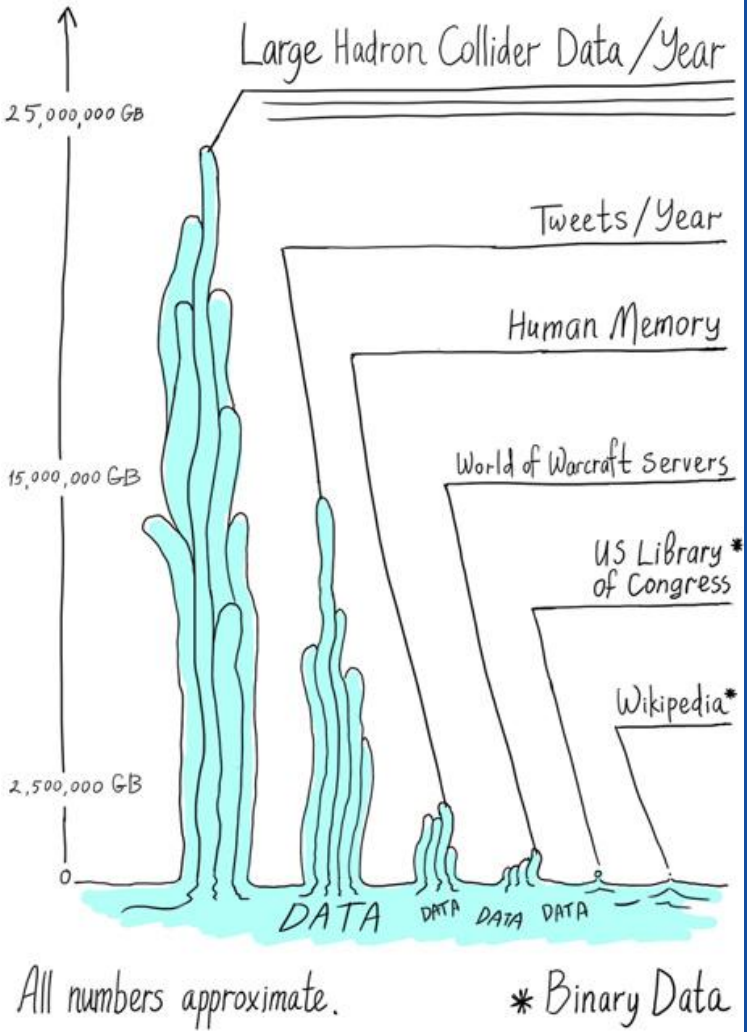
COMPETE 2020

PORTUGAL 2020

UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional

Data Science
In HEP

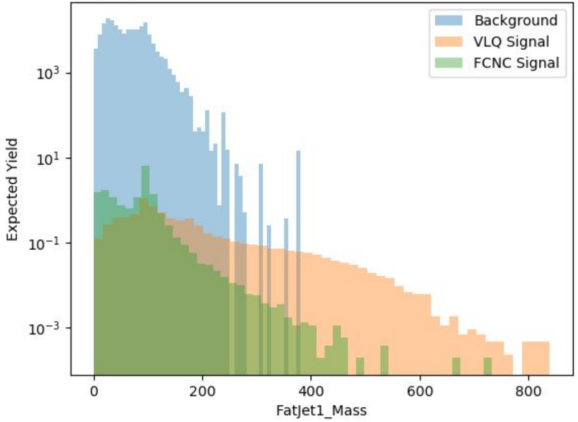
- In High Energy Physics (HEP) we produce large amounts of data

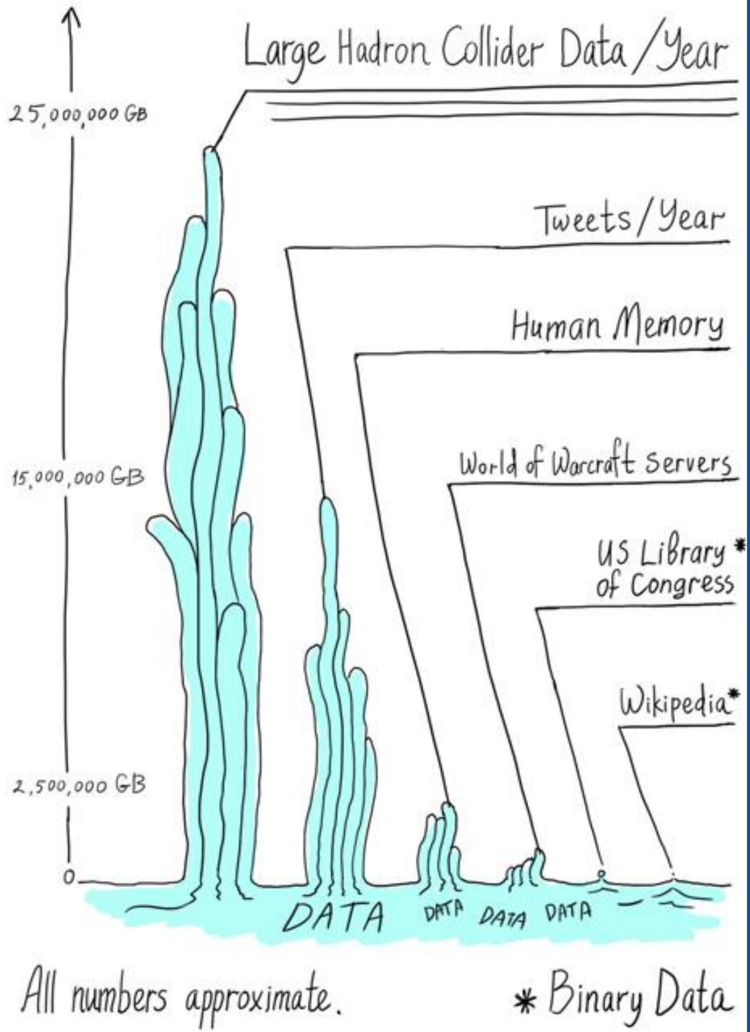


Data Science
In HEP

- In High Energy Physics (HEP) we produce large amounts of data

- The relevant processes are usually several orders of magnitude less frequent than the irreducible backgrounds

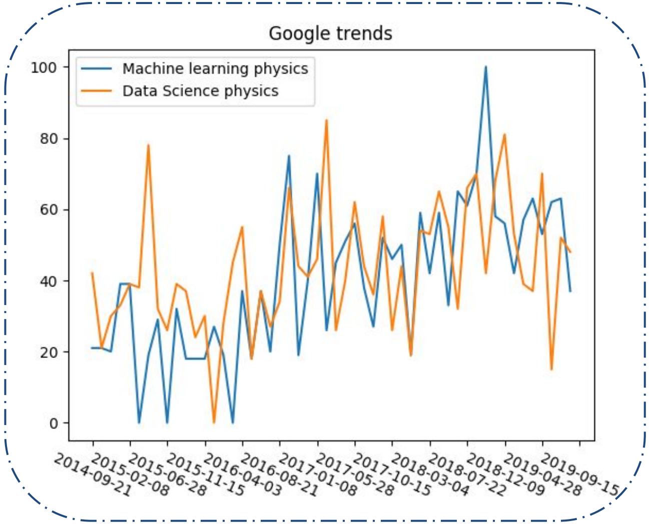




Data Science
In HEP

- Due to these challenges Machine Learning (ML) is becoming increasingly more popular

- Mostly used for classification
  - Other uses are generally:
    - Generative models for Monte Carlo related tasks
    - Convolutional Neural Networks for image based discrimination
    - Adversarial Neural Networks to fight systematic errors
    - Recurrent Neural Networks for time sensitive subjects



Data Science
at LIP for LHC

- Search for new physics using Neural Networks (NN)
- Use the discriminating power of NN to avoid limiting our studies to very narrow physics scenarios
- Best of both worlds:
  - High discrimination
  - Broad signal search

BB →
Z(ll)b Z(ll)b
W(lv)t Z(ll)b
W(lv)t W(lv)t
Z(ll)b W(qq)t
Z(ll)b Z(qq)b
Z(ll)b H(bb)b
W(lv)t W(qq)t
W(lv)t Z(qq)b
W(lv)t H(bb)b
W(qq)t W(qq)t
W(qq)t Z(qq)b
Z(qq)b W(qq)t
Z(qq)b Z(qq)b
W(qq)t H(bb)b
Z(qq)b H(bb)b
H(bb)b H(bb)b

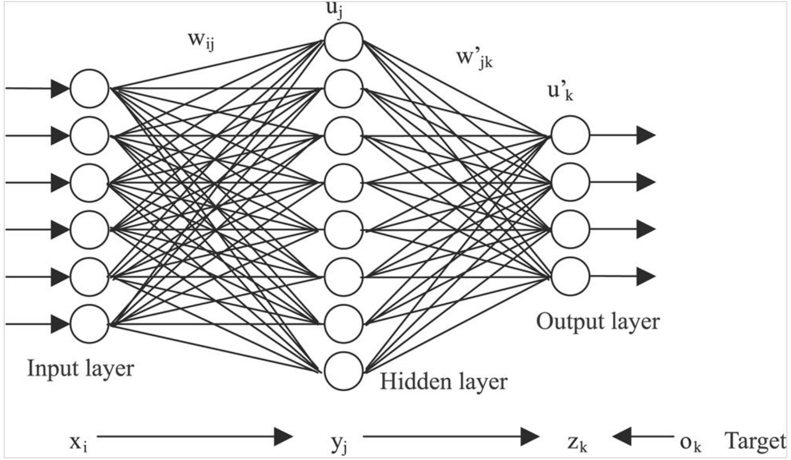
TT →
Z(ll)t Z(ll)t
W(lv)b Z(ll)t
W(lv)b W(lv)b
Z(ll)t W(qq)b
Z(ll)t Z(qq)t
Z(ll)t H(bb)t
W(lv)b W(qq)b
W(lv)b Z(qq)t
W(lv)b H(bb)t
W(qq)b W(qq)b
W(qq)b Z(qq)t
Z(qq)t W(qq)b
Z(qq)t Z(qq)t
W(qq)b H(bb)t
Z(qq)t H(bb)t
H(bb)t H(bb)t

Single VLQ
(B,T,X,Y,B',q*)
→
Z(ll)b
Z(ll)t
W(lv)b
W(lv)t
V(qq)b
V(qq)t
H(bb)b
H(bb)t

×2 for decays to light quarks

Also:
tttt → 0, 1, 2, 3, 4 leptons + 4b + jets

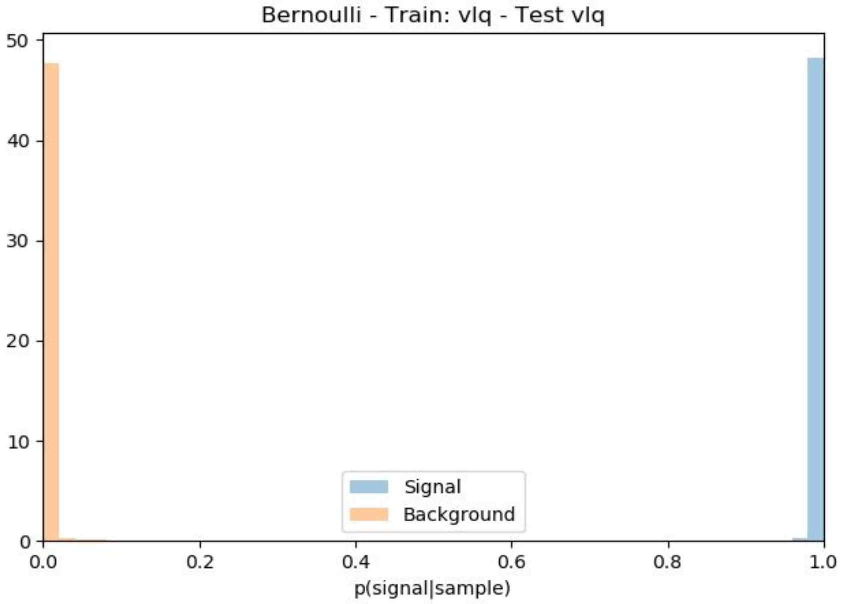
Data Science
at LIP for LHC

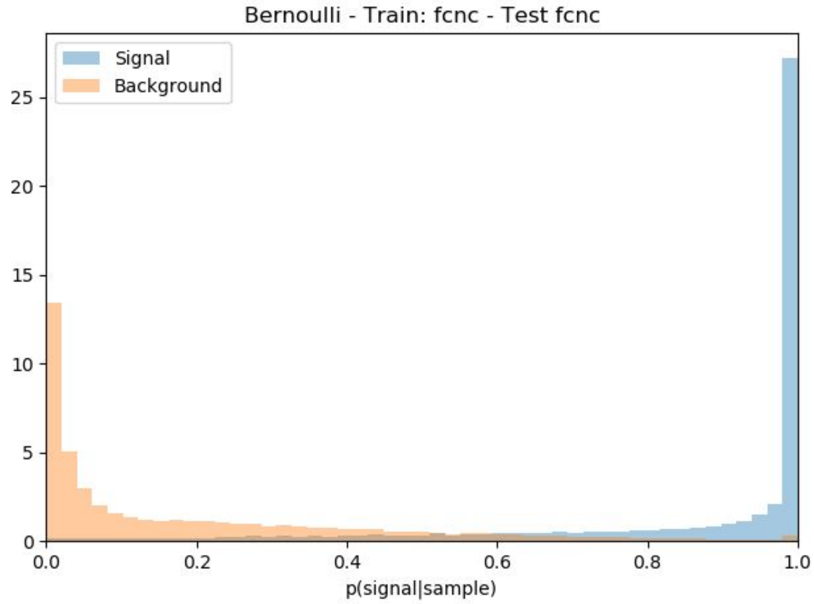


- NN built with Keras, Scikit-learn for data pre-processing
- 2 dense hidden layers of 128 nodes with 0.25 dropout
- Low level information as features:
  - Leptons and jets kinematic variables
  - Missing transverse energy
- Trained on 2 GPUs
  - Titan XP and GeForce RTX 2080 TI

Data Science
at LIP for LHC

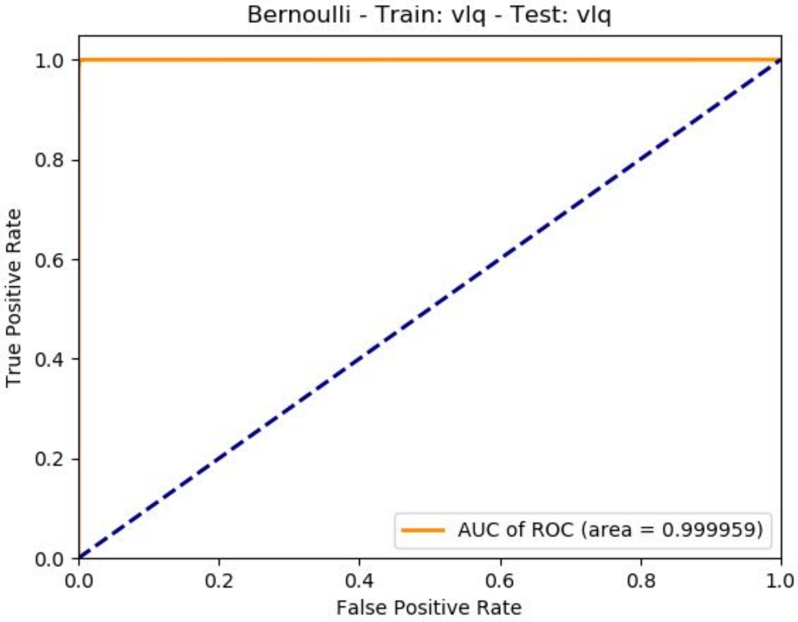
- Training NN for two different new physics models

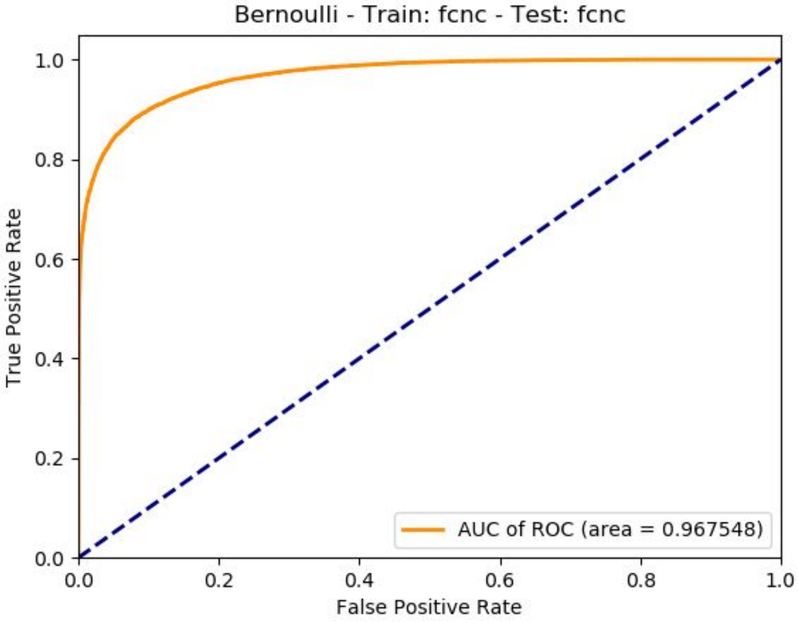




Data Science
at LIP for LHC

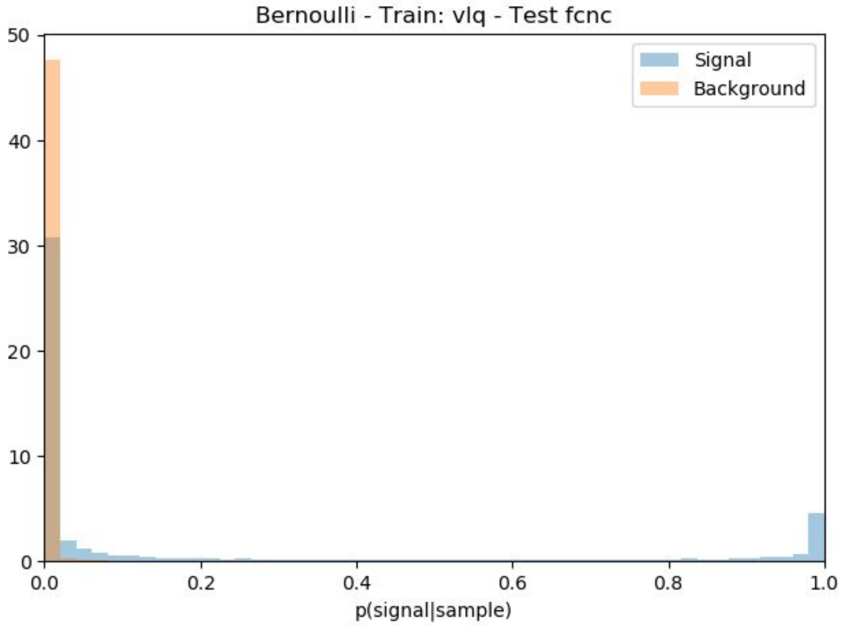
- Training NN for two different new physics models

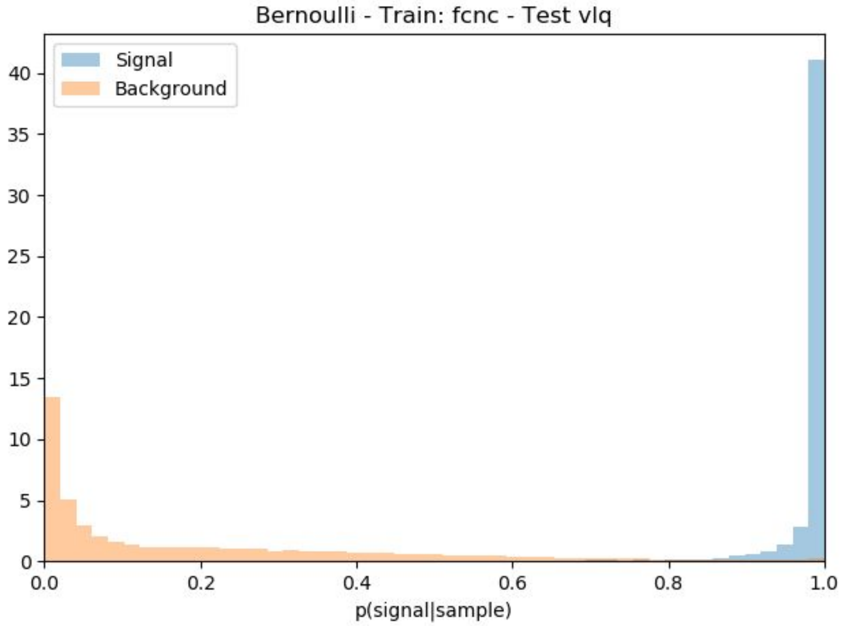




Data Science
at LIP for LHC

- And testing with a class not seen during training

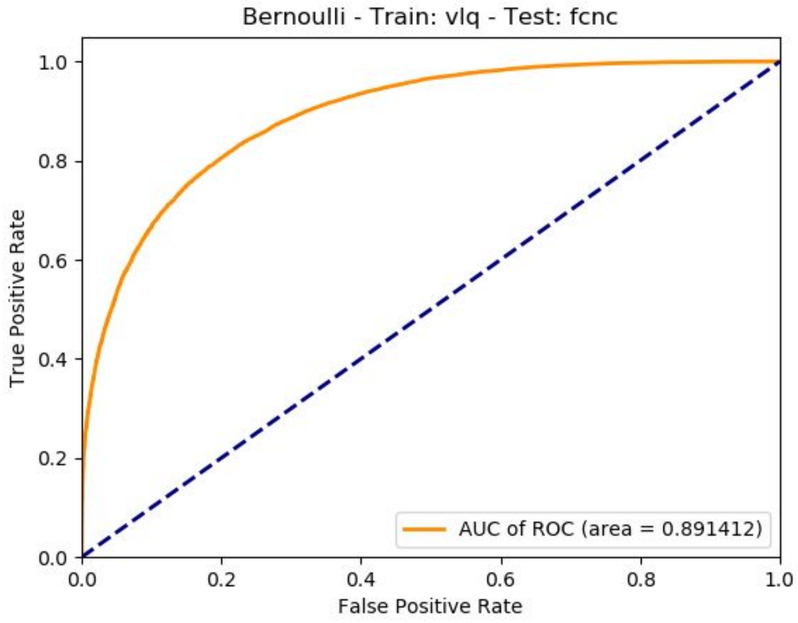


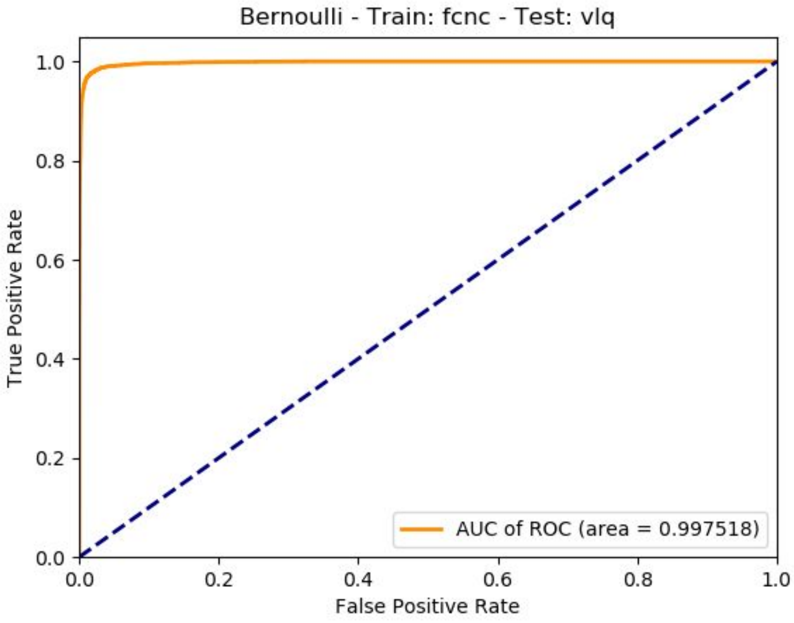


Data Science
at LIP for LHC

Paper in preparation

- And testing with a class not seen during training





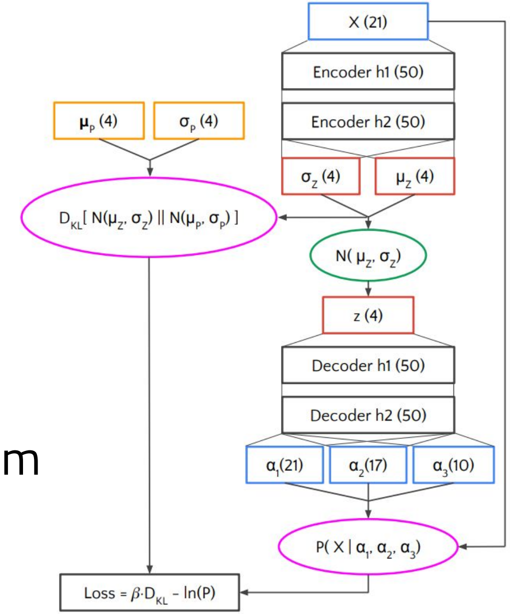
Data Science
at LIP for LHC

- Expertise in machine learning for discrimination are consolidated
- There are very promising applications of NN in HEP besides classification
- Generative models are an area of interest to the group
- Great use cases already developed in the community

https://arxiv.org/abs/1811.10276

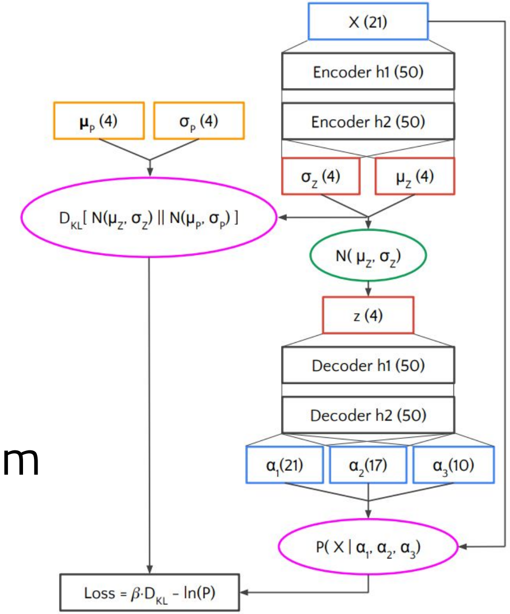
Generative models in HEP

- Variational Autoencoders to learn all background
- Use reconstruction error as a discriminant
  - New signal processes will be triggered by the algorithm
- Semi-supervised method that does not depend on the training signal class
- We do not want to risk missing new physics because it was not foreseen by theory



https://arxiv.org/abs/1811.10276

Generative models
in HEP

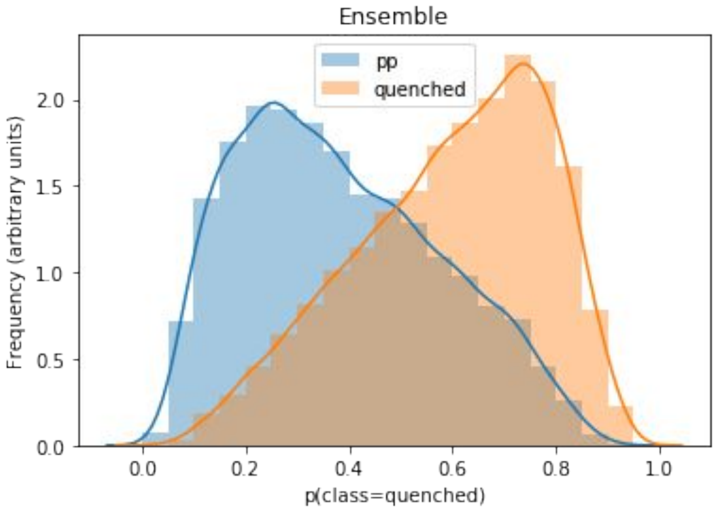


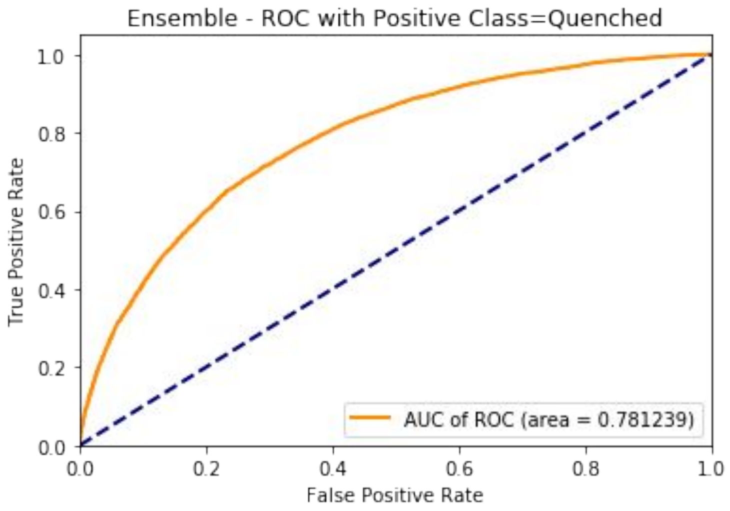
- Variational Autoencoders to learn all background
- Use reconstruction error as a discriminant
  - New signal processes will be triggered by the algorithm
- Semi-supervised method that does not depend on the training signal class
- We do not want to risk missing new physics because it was not foreseen by theory
- The group is developing a similar approach to a general anomaly detector for HEP

Studying jets at the LHC
using ML to tag jets passing through a dense medium

- Ensamble of a Convolutional NN, a Recurrent NN and a locally connected network

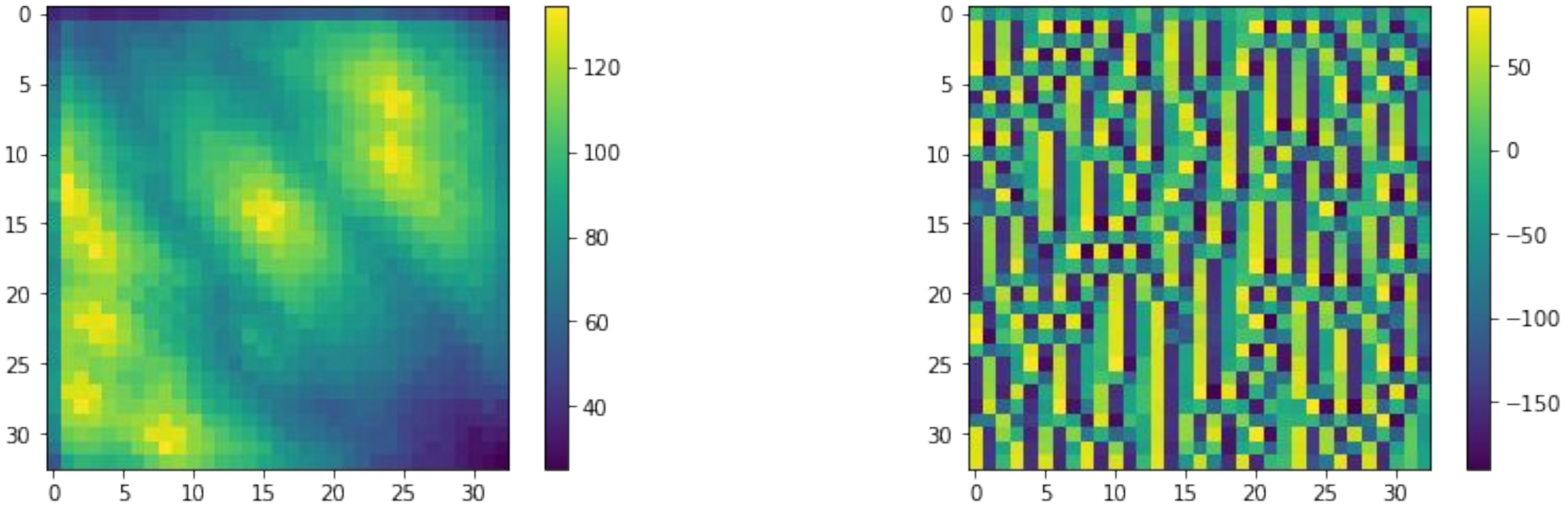
- Classifying QGP quenched jets from vacuum





Studying jets at the LHC
using ML to tag jets passing through a dense medium

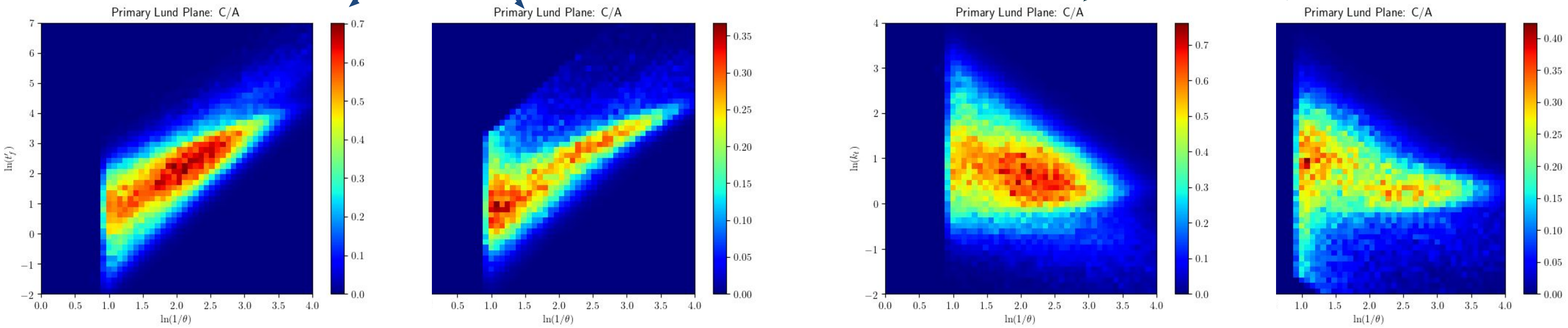
- Use CNN filters to learn physics



Work done by Filipa Peres,
University of Minho Msc Student

Studying jets at the LHC
using ML to tag jets passing through a dense medium

- distinction of quenched and unquenched jets using Lund planes
    - using $t_f = 1 / (p_T\, z\, \Theta^2)$ instead of the traditional k_T splitting



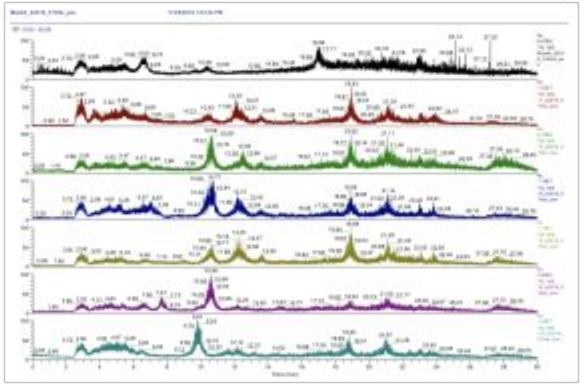
Vacuum QGP Vacuum QGP

Work done by Diogo Barros, University of Minho Msc student

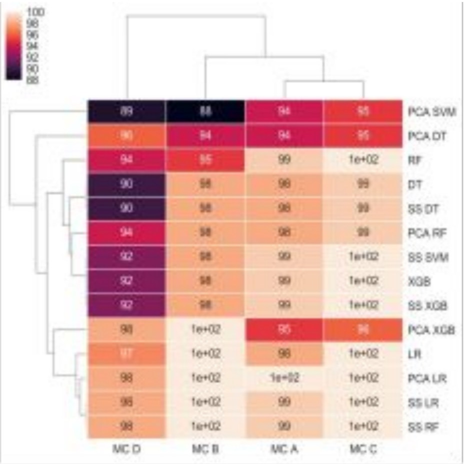
Machine Learning in Analytical Chemistry
collaborating with UMinho colleagues

- Train a model to predict PCB's manufacturing conditions (MC) using chemical data using HPLC-MS.
- Great potential to be used in quality control, forgery detection and pharma industry.

Printed Circuit Board (PCB)



chemical data from HPLC-MS



model performance (accuracy)

Final Remarks

- Machine learning expertises in being consolidated by the group
- Many different fronts of ML application
- Novel approaches to NN use for generative models are currently under study
  - The goal is to have a general purpose anomaly detector for HEP
- Applications beyond HEP are being developed
  - ML is proving to be a very useful tool for analytical chemistry

Thanks

tiago.vale@cern.ch