LABORATÓRIO DE INSTRUMENTAÇÃO E FÍSICA EXPERIMENTAL DE PARTÍCULAS

partículas e tecnologia

Big Data and Simulation Competence Centre: Machine Learning

Jornadas Científicas do LIP, Braga, 2020

Miguel Crispim Romão

mcromao@lip.pt

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

Fundação para a Ciência e a Tecnologia

UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

ML@LIP

Current Machine Learning Activity Across Groups

- Minho: ATLAS, Pheno, Advanced Computing
- Coimbra: LZ/LUX, Gamma Cameras
- Lisboa: ATLAS, Pheno, CMS, SHiP, COMPASS, HADES, LATTES, Distributed Computing, LIP-Nielsen

ML@LIP

Dedicated ML Resources

Thanks to the computing groups in Lisboa e Minho for installing and managing these resources!

- Minho
  - NVIDIA Titan XP
  - NVIDIA 2080 ti
- Lisboa
  - NVIDIA Tesla P100

LIP Computing

Infraestrutura Nacional de Computação Distribuída

Numbers and Metrics of Growth

BigDataHEP Project Example

Type	Total	2018	2019
Publications	6	2	4
Articles in international journals (with direct contribution from team)	6	2	4
Presentations	12		12
Oral presentations in international conferences	3		3
Seminars	1		1
Outreach seminars	2		2
Presentations in national conferences	4		4
Oral presentations in international meetings	1		1
Poster presentations in national conferences	1		1
Teses	4		4
Master	4		4
Events	2		2
Workshops	1		1
Collaboration Meetings	1		1

Organizing Committee:

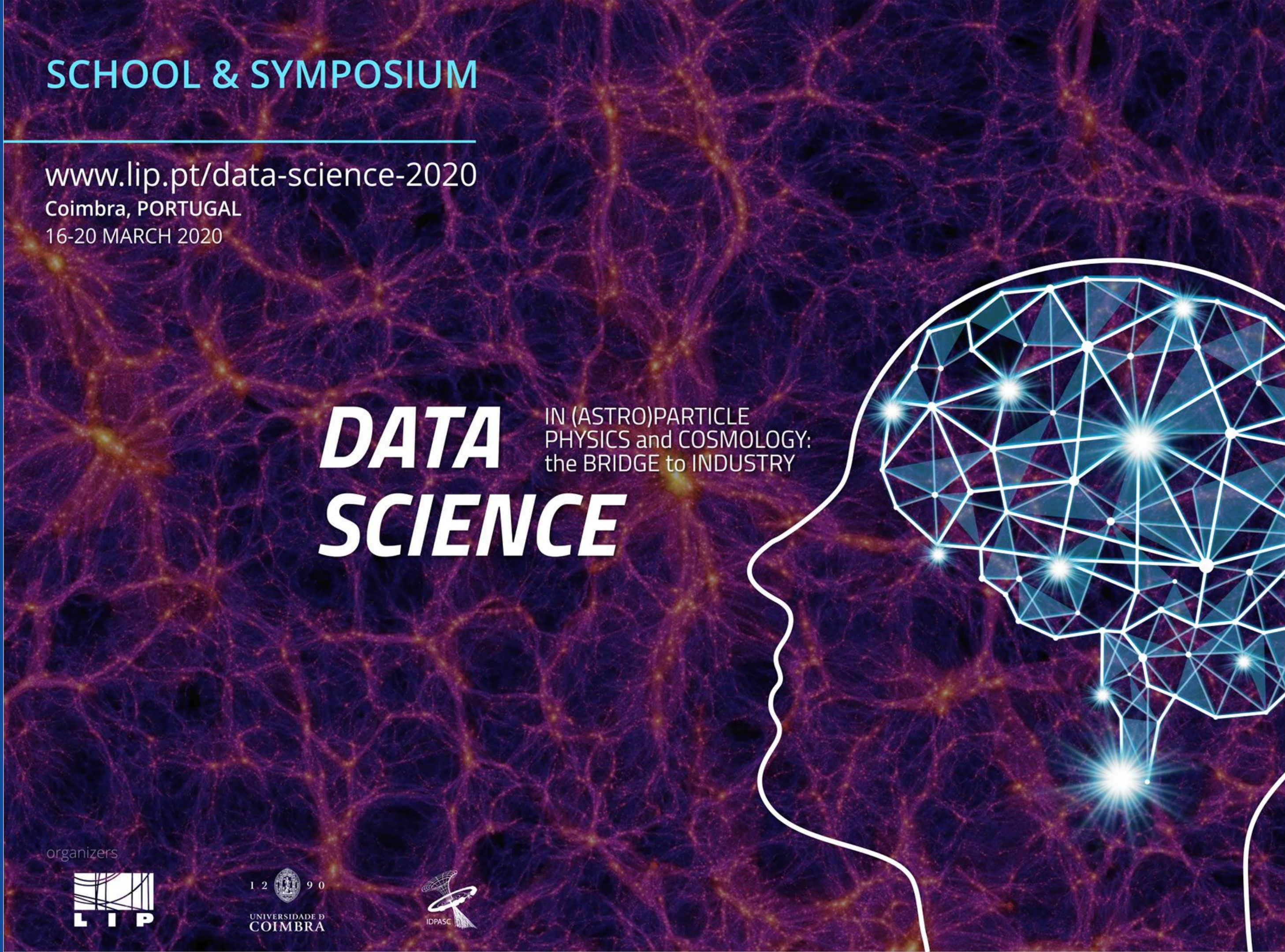
Liliana Apolinário
Gonzalo Parente Bermudez
Nuno Castro (co-chair)
Lorenzo Cazon
Ruben Conceição
Rui Ferreira Marques
Ricardo Gonçalo (chair)
Alexandre Lindote
Isabel Lopes
Valentina Lozza
Andrei Morozov
Francisco Neves
Vladimir Solovov
Bernardo Tomé
Filipe Veloso

Design and Development:

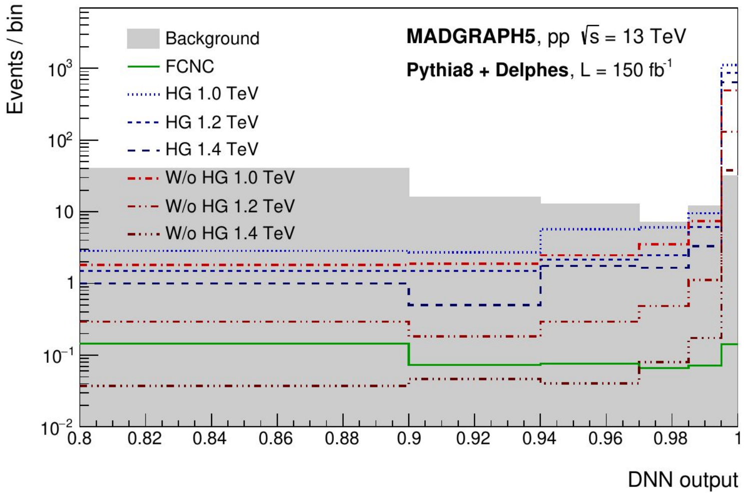
Henrique Carvalho
Carlos Manuel

Secretariat:

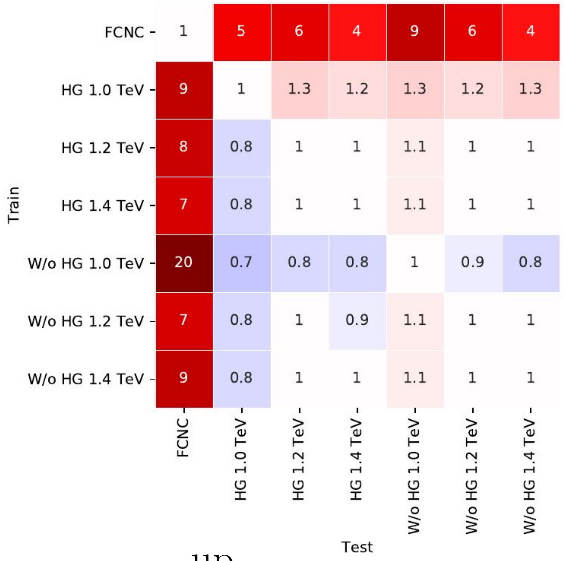
Natália Antunes



Transferability of Deep Learning Models in Searches for New Physics at Colliders (1912.04220, PRD Accepted)



Normalised μ

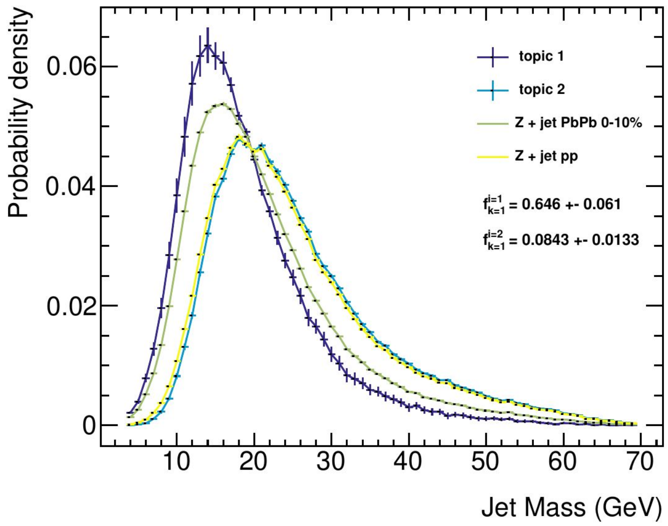


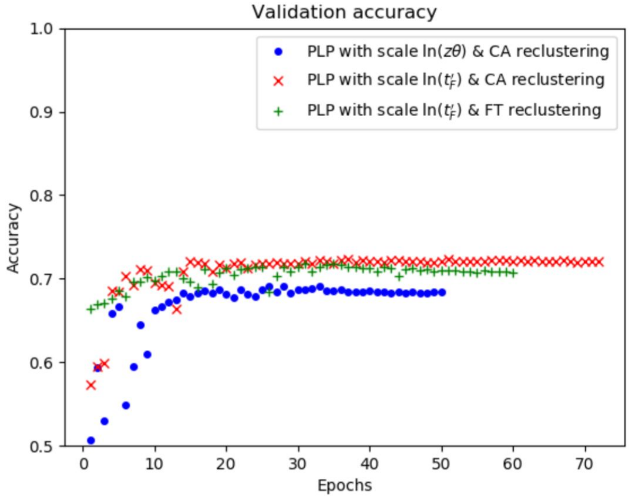
$$\mu = \frac{\sigma_{exp}^{\mathrm{up}}}{\sigma_{th}}$$

ATLAS Pheno

Big Data HEP

Completed Master Theses on Data Driven and ML Methods for New Physics Observables





Classifying Heavy Ion Jets

João Gonçalves (IST)

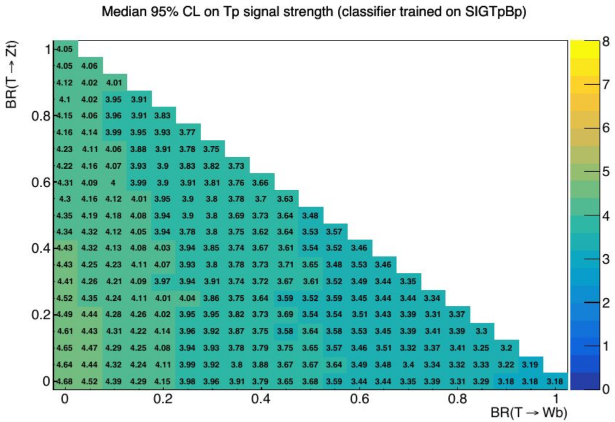
New observables and techniques for the study of jets in hadron collisions
Filipa Peres (U.Minho)

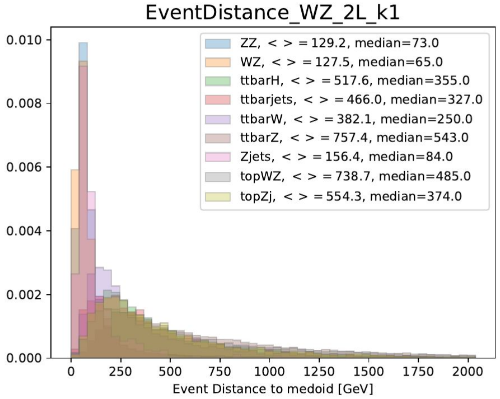
Pheno Big Data HEP

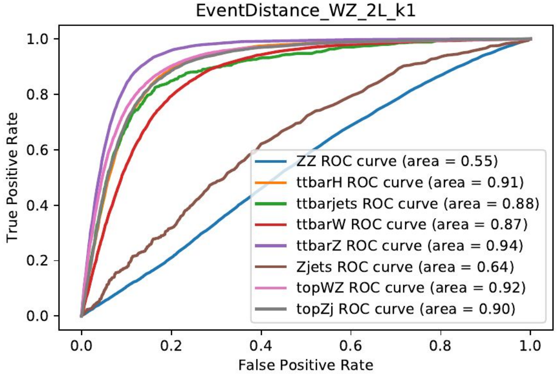
On-going Work

ML in the Search for New Physics

See poster by Rute Pedro







Generic VLQ search strategies

Earth Moving Distance based new observables

ATLAS Big Data HEP

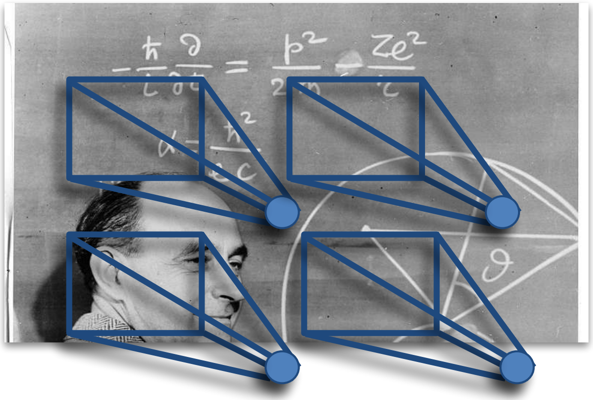
- Exploring relations between low and high level variables
- Using ML to reduce the bias on New Physics assumptions

See talk by Liliana Apolinário

On-going Work

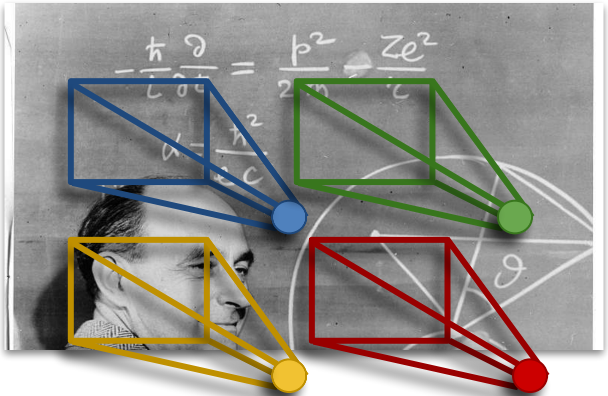
Data-driven and ML Methods for QGP studies

Convolutional Neural Networks

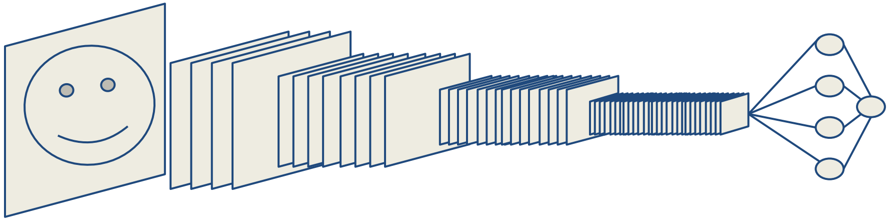


Receptive field moves along the picture => Position Invariant Features

Locally Connected Neural Networks



Receptive field fixed for each region of the picture => Position Dependent Features

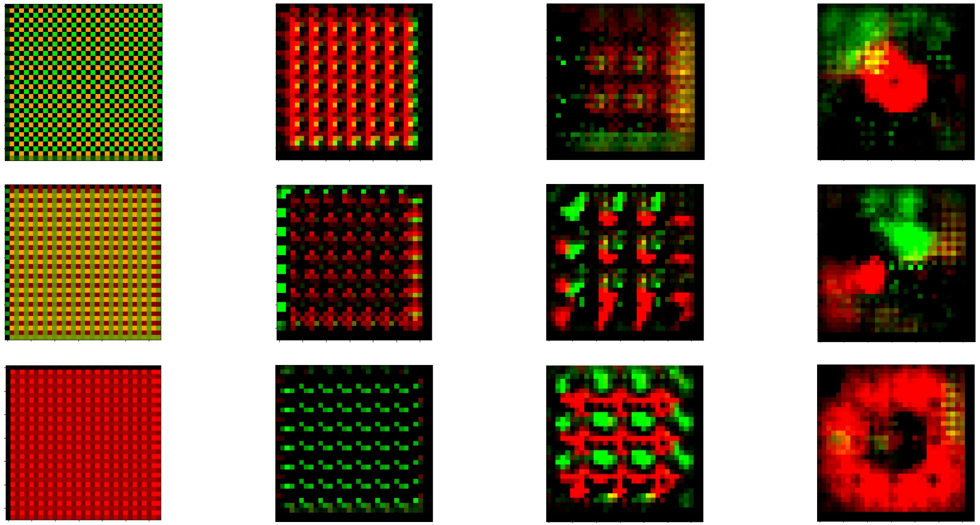


On-going Work

Data-driven and ML Methods for QGP studies

CNN (receptive field composed of 3x3 filters with stride of 2)

Layer 1 | Layer 2 | Layer 3 | Layer 4



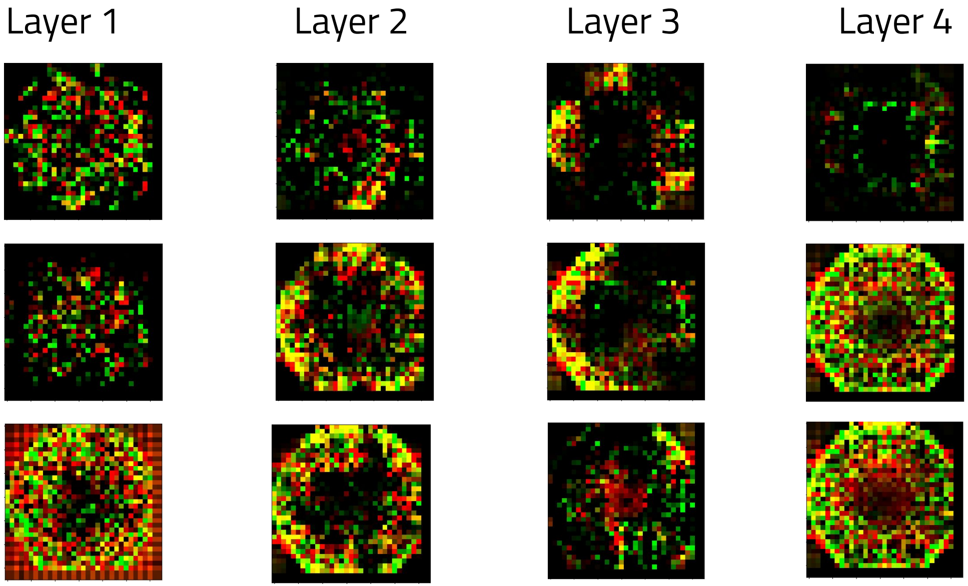
```
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 33, 33, 2)]       0
_________________________________________________________________
batch_normalization (BatchNo (None, 33, 33, 2)         8
_________________________________________________________________
conv2d (Conv2D)              (None, 16, 16, 128)       2432
_________________________________________________________________
leaky_re_lu (LeakyReLU)      (None, 16, 16, 128)       0
_________________________________________________________________
spatial_dropout2d (SpatialDr (None, 16, 16, 128)       0
_________________________________________________________________
batch_normalization_1 (Batch (None, 16, 16, 128)       512
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 7, 7, 256)         295168
_________________________________________________________________
leaky_re_lu_1 (LeakyReLU)    (None, 7, 7, 256)         0
_________________________________________________________________
spatial_dropout2d_1 (Spatial (None, 7, 7, 256)         0
_________________________________________________________________
batch_normalization_2 (Batch (None, 7, 7, 256)         1024
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 3, 3, 384)         885120
_________________________________________________________________
leaky_re_lu_2 (LeakyReLU)    (None, 3, 3, 384)         0
_________________________________________________________________
spatial_dropout2d_2 (Spatial (None, 3, 3, 384)         0
_________________________________________________________________
batch_normalization_3 (Batch (None, 3, 3, 384)         1536
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 1, 1, 512)         1769984
_________________________________________________________________
leaky_re_lu_3 (LeakyReLU)    (None, 1, 1, 512)         0
_________________________________________________________________
spatial_dropout2d_3 (Spatial (None, 1, 1, 512)         0
_________________________________________________________________
batch_normalization_4 (Batch (None, 1, 1, 512)         2048
_________________________________________________________________
flatten (Flatten)            (None, 512)               0
_________________________________________________________________
dropout (Dropout)            (None, 512)               0
_________________________________________________________________
dense (Dense)                (None, 1)                 513
=================================================================
Total params: 2,958,345
Trainable params: 2,955,781
Non-trainable params: 2,564
```

Pheno Big Data HEP

On-going Work

Data-driven and ML Methods for QGP studies

LCNN (receptive field composed of 3x3 filters with stride of 2)



Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 33, 33, 2)]	0
batch_normalization (BatchNo	(None, 33, 33, 2)	8
locally_connected2d (Locally	(None, 16, 16, 128)	622592
leaky_re_lu (LeakyReLU)	(None, 16, 16, 128)	0
spatial_dropout2d (SpatialDr	(None, 16, 16, 128)	0
batch_normalization_1 (Batch	(None, 16, 16, 128)	512
locally_connected2d_1 (Local	(None, 7, 7, 256)	14463232
leaky_re_lu_1 (LeakyReLU)	(None, 7, 7, 256)	0
spatial_dropout2d_1 (Spatial	(None, 7, 7, 256)	0
batch_normalization_2 (Batch	(None, 7, 7, 256)	1024
locally_connected2d_2 (Local	(None, 3, 3, 384)	7966080
leaky_re_lu_2 (LeakyReLU)	(None, 3, 3, 384)	0
spatial_dropout2d_2 (Spatial	(None, 3, 3, 384)	0
batch_normalization_3 (Batch	(None, 3, 3, 384)	1536
locally_connected2d_3 (Local	(None, 1, 1, 512)	1769984
leaky_re_lu_3 (LeakyReLU)	(None, 1, 1, 512)	0
spatial_dropout2d_3 (Spatial	(None, 1, 1, 512)	0
batch_normalization_4 (Batch	(None, 1, 1, 512)	2048
flatten (Flatten)	(None, 512)	0
dropout (Dropout)	(None, 512)	0
dense (Dense)	(None, 1)	513

Total params: 24,827,529
Trainable params: 24,824,965
Non-trainable params: 2,564

Opened Avenues of Research

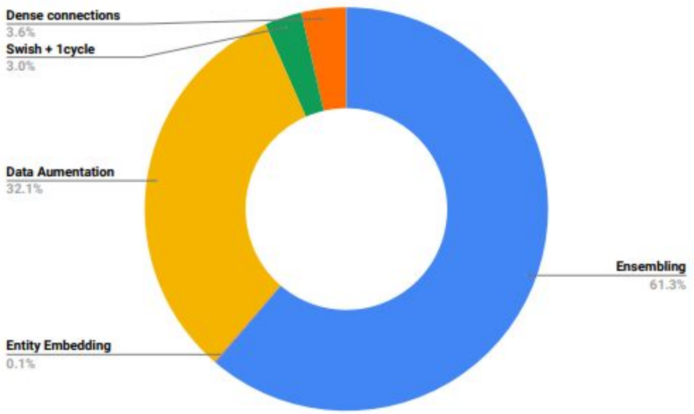
- Consolidated expertised lead to new ideas and questions on the usage and potential of Machine Learning and Deep Learning in HEP
- Submitted a proposal for UTAustin-Portugal call on Exploratory Research Projects (waiting results)
  - Generative and Interpretable Deep Learning in HEP

UTAustin Portugal

ATLAS Pheno Big Data HEP

CMS

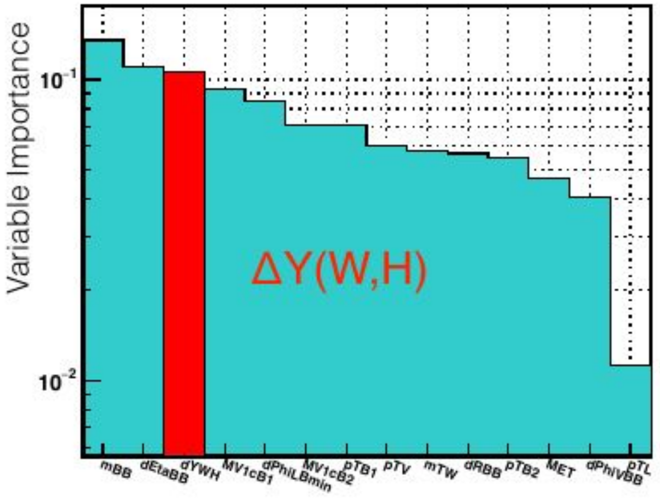
- Giles Strong PhD focused on Deep Learning applications in HEP
  - Crucial contribution to ML training and skill development in LIP with many tutorials and workshops
  - Studies on advanced DNN methods using the Higgs ML dataset (related publication soon)
  - LUMIN: a PyTorch wrapper for deep learning in HEP

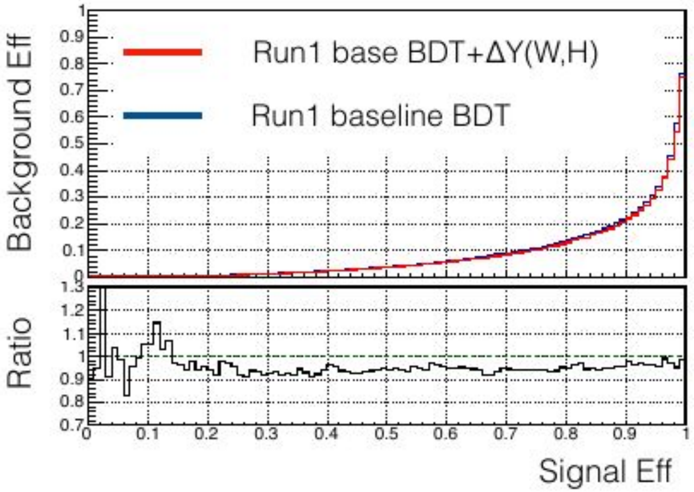


2002.01427

ATLAS
Example: Measurement of VH, H->bb

See talk by Emanuel Gouveia and poster by Ana Peixoto





Two new discriminants increased sensitivity up to 12%

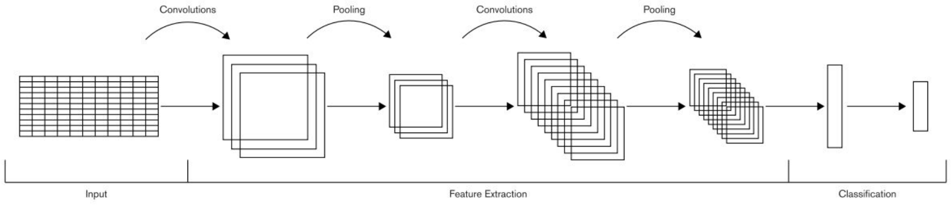
Cosmic Rays

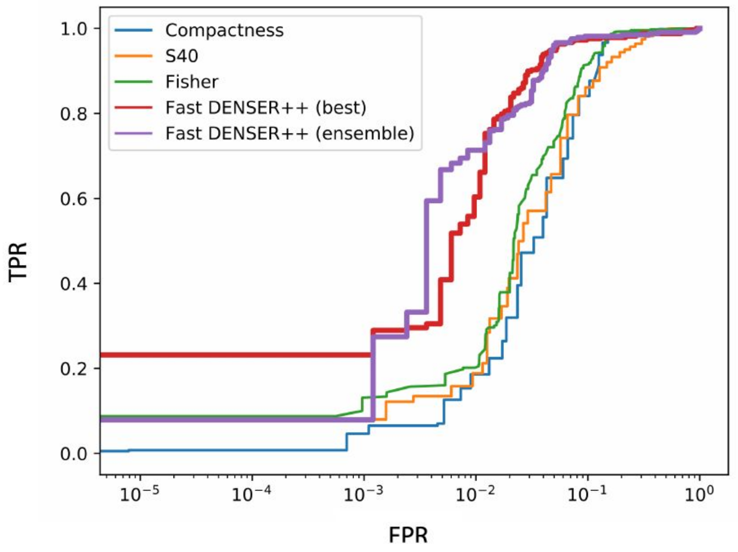
- LATTES

10.1109/ACCESS.2019.2933947

Automatic Design of Artificial Neural Networks for Gamma-Ray Detection

FILIPE ASSUNÇÃO[1], JOÃO CORREIA[1], RÚBEN CONCEIÇÃO[2], MÁRIO JOÃO MARTINS PIMENTA[2], BERNARDO TOMÉ[2], NUNO LOURENÇO[1], AND PENOUSAL MACHADO[1]
[1]CISUC, Department of Informatics Engineering, University of Coimbra, 3030-290 Coimbra, Portugal
[2]LIP/IST, 1600-078 Lisbon, Portugal





See talk by Rúben Conceição

Structure of Matter

ARTICLES
https://doi.org/10.1038/s41567-019-0583-8
nature physics

Probing dense baryon-rich matter with virtual photons

The HADES Collaboration*

- HADES
    - NN for e+ and e- identification
- COMPASS: ML for background vs signal discrimination

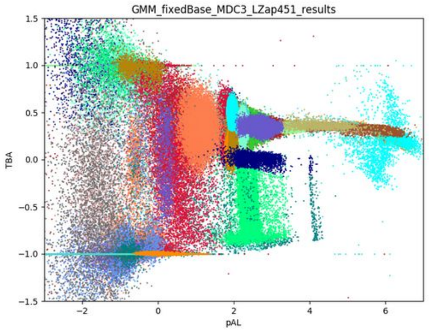
See talks by Alberto Blanco and Márcia Quaresma

Neutrinos and Dark Matter

See talks by Celso Franco, Ricardo Cabrita, and Paulo Brás

- SHiP: Guilherme Soares (MSc Student) ML to increase hidden particle selection efficiency, IST
- LUX/LZ: Paulo Brás (PhD student), A. Solovov (MSc student), U. Coimbra

Example of the parameter space after clustering with
Gaussian Mixture Model (GMM)



Pulse Classification (Paulo Brás)

Paper (in preparation): "Machine Learning for Pulse Classification in LZ"

PhD Thesis (in preparation): "New Physics Phenomenology and Development of Data Processing Tools for the LZ Dark Matter Direct Search Experiment"

N,1,1 RBF SVM

	predicted label		
actual label	91%	9%	1e
	19%	81%	b2b
	1e	b2b	

Y,1,0 kNN

	predicted label		
actual label	75%	25%	1e
	45%	55%	b2b
	1e	b2b	

N,2,0 G. proc.

	predicted label		
actual label	82%	18%	1e
	10%	90%	b2b
	1e	b2b	

N,1,0 R. frst.

	predicted label		
actual label	77%	23%	1e
	26%	74%	b2b
	1e	b2b	

Rare Event Identification (A. Solovov)

Masters Thesis (in preparation): "Exploration of machine learning techniques for discrimination of neutrino less double beta decay of 136Xe"

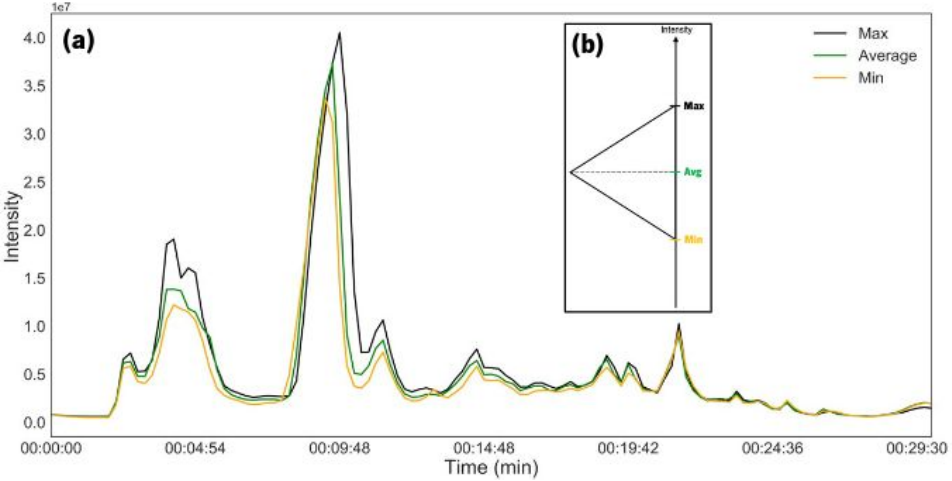
Big Data HEP

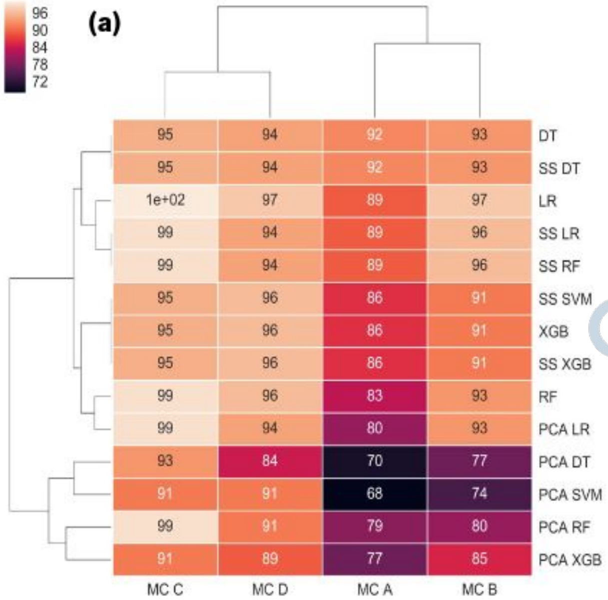
Nielsen-LIP Collaboration

R. Conceição, L. Cazon, G. Strong, Felix Riehn (LIP)
Aynur Kocak, Rita Lima, Carla Silva (Nielsen)

Machine Learning in Analytical Chemistry

- MSc Thesis by Diogo Gonçalves (see his poster), U. Minho



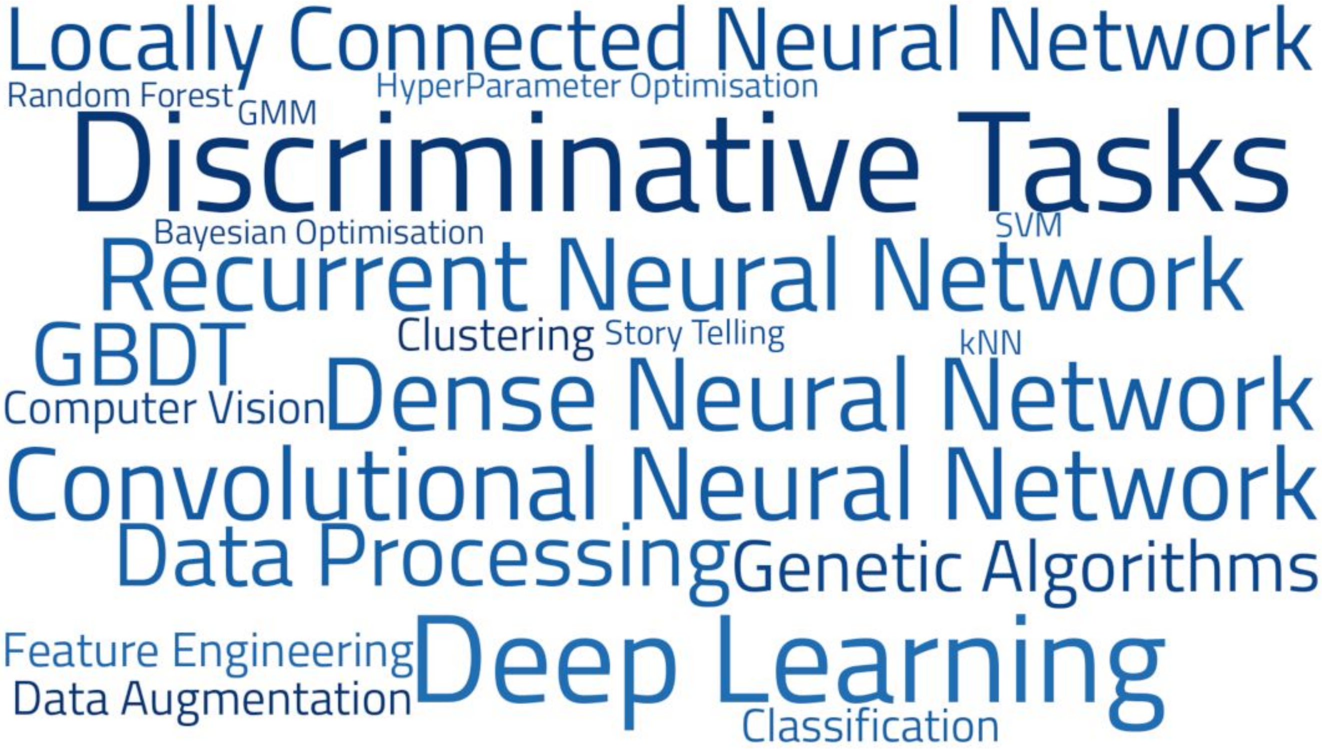


CENTRO DE QUÍMICA
UNIVERSIDADE DO MINHO

Big Data HEP

Where we are?

Assessed ML Expertises Found in LIP

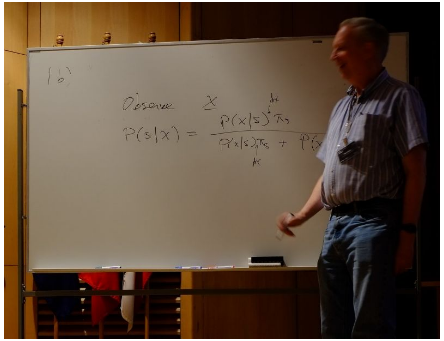


What’s next?

Advanced Training in ML@LIP

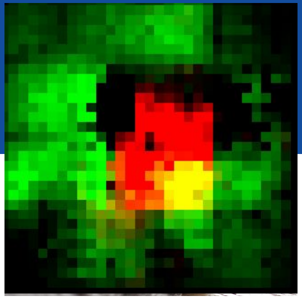
See talk by N. Leonardo (ECO3)

- Comprehensive programme in several topics at LIP, but we need to increase the ML focus
  - We identified a lot of specialised experts
  - The challenge is to profit from each other and teach the next generation
- Outwards training might be a service in itself that LIP can provide
  - Need to identify the key topics and demand
  - Very competitive field, we need to be clever to identify our strong points



Next Steps

- After the first two years exploring the potential of ML@LIP, we are now focusing towards specific strategic points
    - Interpretability and Transferability of DL
    - Anomaly Detection
    - Generative Methods
    - Long term sustainability
        - Data heavy endeavours
        - Reproducibility of increasingly complex methods (udocker)



Thanks!

You can find me around or mcromao@lip.pt