WG-4 – Analysis techniques

Machine learning in Alice and Atlas

Alexandre Suaide

Organização

- Algoritmos e técnicas
- Infraestrutura (cluster e grid)
- Cronograma de atividades

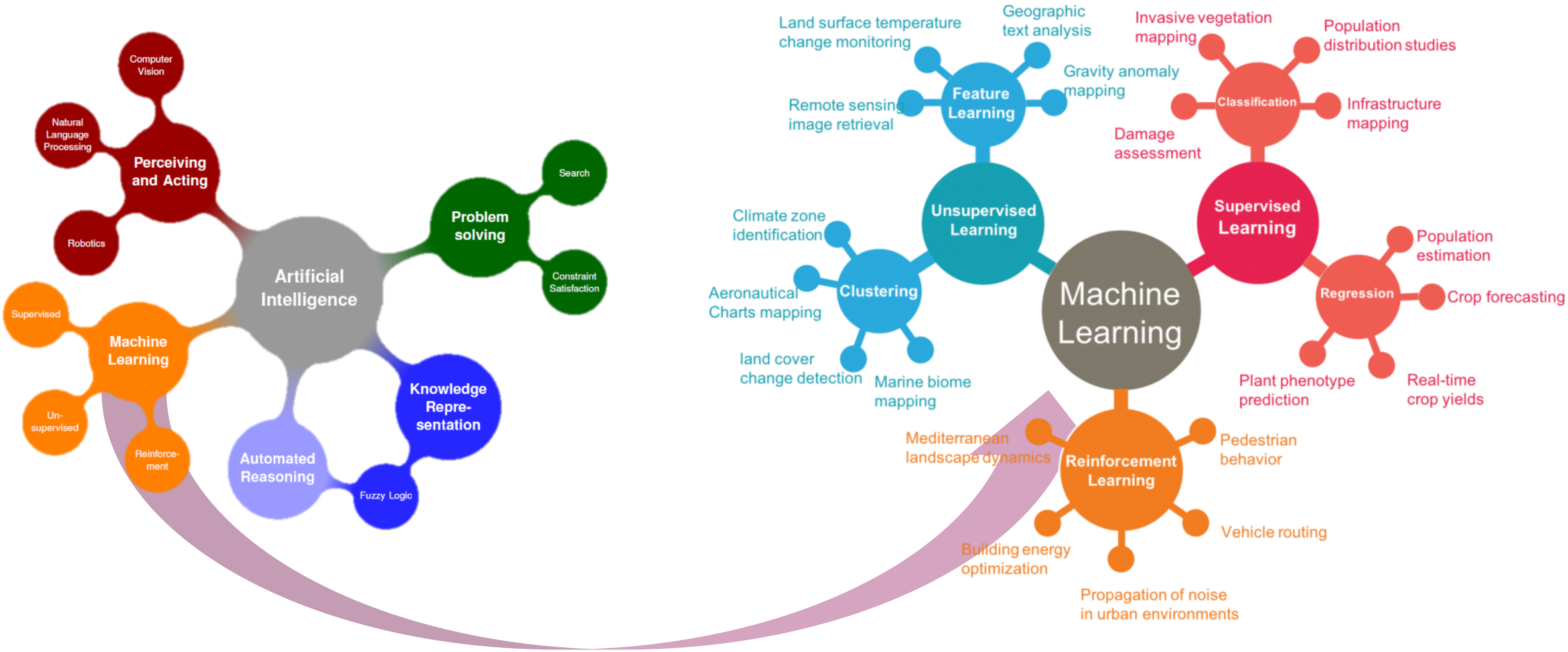
Objetivos do WG-4

- Desenvolver técnicas de análise e simulações envolvendo ML
  - Incorporar essas técnicas ao dia-a-dia das análises no Alice e Atlas
  - Grande interface com demais grupos
  - Grande envolvimento dos estudantes do projeto

- Infraestrutura computacional do projeto

Ciência de dados

- Tecnologia da informação
  - Software + hardware
- Dados
  - Pré-processamento e estruturação
  - Visualização
- Matemática e estatística
- Aprendizado de máquina
  - Um tópico em Inteligência Artificial



Artificial Intelligence
Perceiving and Acting
Computer Vision
Natural Language Processing
Robotics
Problem solving
Search
Constraint Satisfaction
Machine Learning
Supervised
Un-supervised
Reinforce-ment
Knowledge Repre-sentation
Automated Reasoning
Fuzzy Logic
Machine Learning
Unsupervised Learning
Feature Learning
Land surface temperature change monitoring
Geographic text analysis
Gravity anomaly mapping
Remote sensing image retrieval
Clustering
Climate zone identification
Aeronautical Charts mapping
land cover change detection
Marine biome mapping
Supervised Learning
Classification
Invasive vegetation mapping
Population distribution studies
Infrastructure mapping
Damage assessment
Regression
Population estimation
Crop forecasting
Plant phenotype prediction
Real-time crop yields
Reinforcement Learning
Mediterranean landscape dynamics
Pedestrian behavior
Vehicle routing
Building energy optimization
Propagation of noise in urban environments

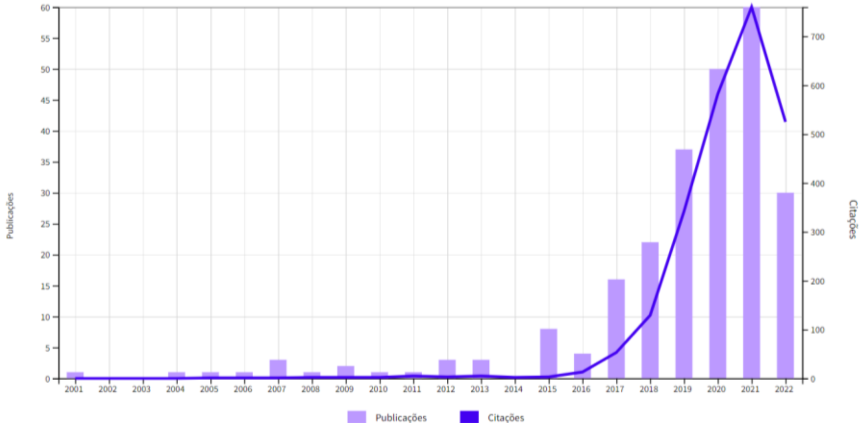
High Energy Physics and Instrumentation Center at IFUSP

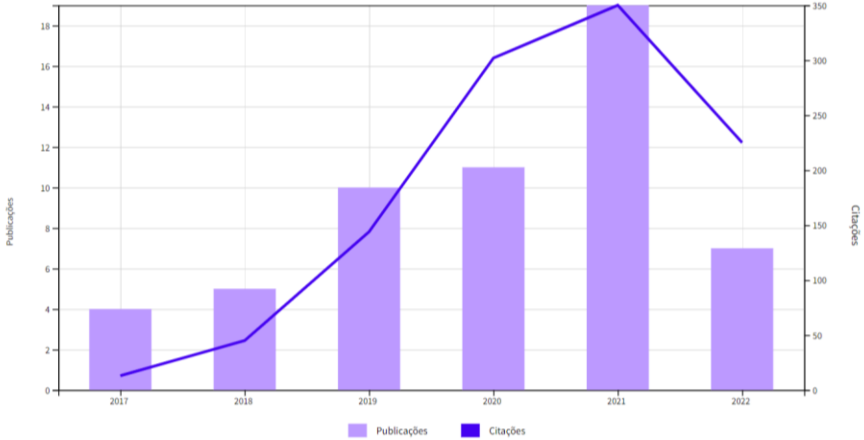
Aprendizado de máquina em HEP

Número de artigos e citações envolvendo “Machine learning” e “LHC” tem aumentado consideravelmente nos últimos anos

Uma fração considerável desses artigos está relacionada a jatos

- ~ 1/3 das publicações e ~ 1/2 das citações em 2021





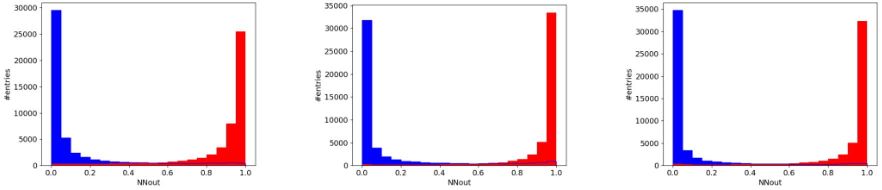
Identificação de partículas

Identifying D_s decays at the LHC

E. Bakos[1*], N. de Groot[2], N.Vranjes[1]

1 Institute of Physics, University of Belgrade, Pregrevica 118, 11080, Belgrade, Serbia
2 Institute for Mathematics, Astrophysics and Particle Physics (IMAPP), Radboud University Nijmegen, Heyendaalseweg 135, 6525 AJ Nijmegen, Netherlands and Nikhef, Netherlands
evelin.bakos@cern.ch

July 28, 2022



(a) DNN (b) CNN (c) Combined network

Figure 4: Output of the different networks for signal (blue) and background (red).

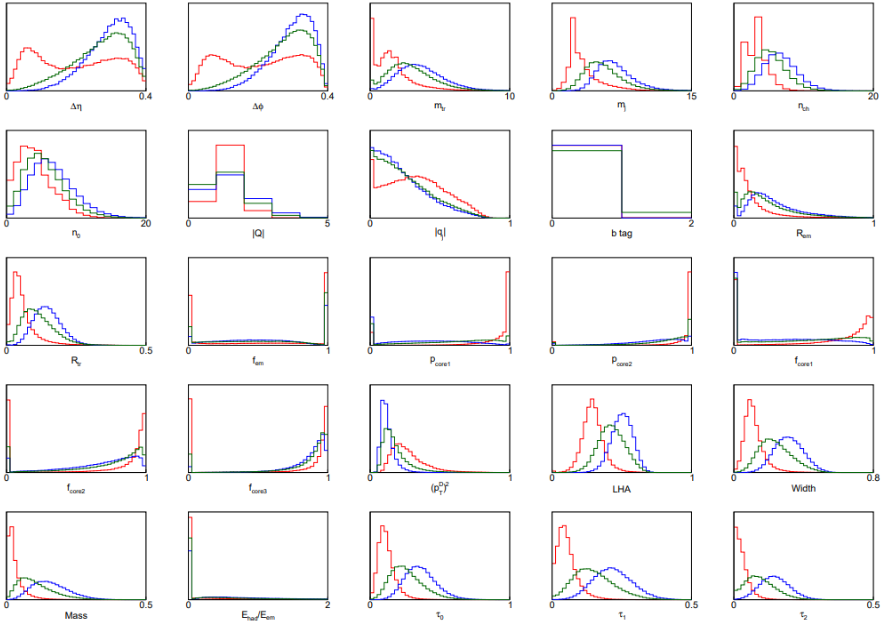


Figure 1: Distributions of the variables used for D_s identification, using DNN. The signal is presented with a red line, while the gg and qq backgrounds are drawn with blue and green colours respectively.

arXiv:2207.13587

Machine learning e jatos no LHC

Identifying quenched jets in heavy ion collisions with machine learning

Lihan Liu,[a] Julia Velkovska,[a] Marta Verweij[b]

[a] *Department of Physics and Astronomy*
Vanderbilt University
PMB 401807, 2301 Vanderbilt Place, Nashville, TN 37235, USA

[b] *Department of Physics*
Utrecht University
Heidelberglaan 8 3584 CS Utrecht, The Netherlands

arXiv:2206.01628

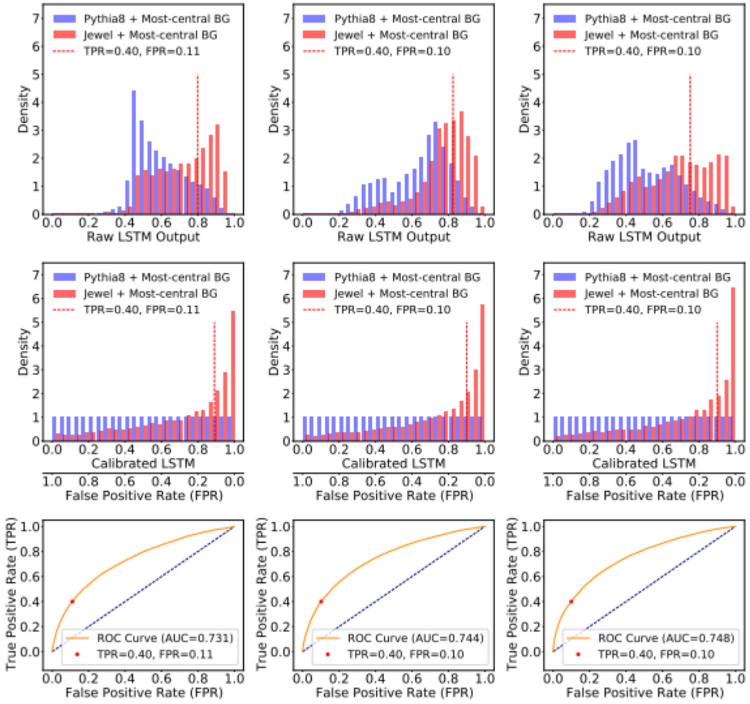


Figure 7. Distributions of the raw LSTM output from top three best-trained neural networks (top panels). Distribution of calibrated LSTM output (middle panels), and the corresponding ROC curve (bottom panels). The red lines/points corresponds to thresholds determined with TPR=0.4.

Machine learning e jatos no LHC

Quantum Machine Learning for b-jet identification

Alessio Gianelle[1], Patrick Koppenburg[2], Donatella Lucchesi[1,3], Davide Nicotra[3,4], Eduardo Rodrigues[5], Lorenzo Sestini[1], Jacco de Vries[4], Davide Zuliani[1,3,6]

[1] *INFN Sezione di Padova, Padova, Italy*
[2] *Nikhef National Institute for Subatomic Physics, Amsterdam, Netherlands*
[3] *Università degli Studi di Padova, Padova, Italy*
[4] *Universiteit Maastricht, Maastricht, Netherlands*
[5] *University of Liverpool, Liverpool, United Kingdom*
[6] *European Organization for Nuclear Research (CERN), Geneva, Switzerland*

arXiv:2202.13943

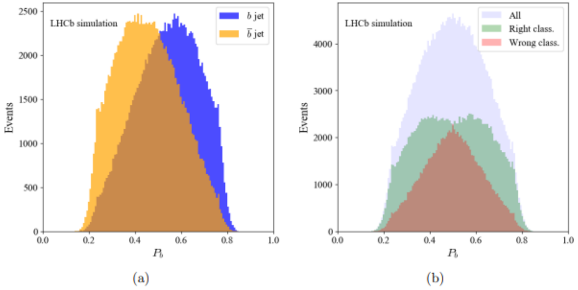


Figure 4: Probability distributions for jet tagged to b (blue) and $\bar{b}$ quarks (yellow), showing separation around 0.5 (a). Probability distribution for the Angle Embedding circuit: jet correctly (wrongly) tagged are plotted in green (red), showing around 0.5 worse classification. The probability distribution for all jets is shown in grey (b).

(AUC) for the DNN and the quantum classifiers for the *muon dataset* and the *complete dataset*.

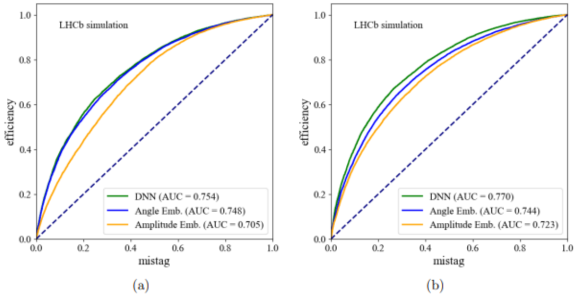


Figure 5: ROC distributions and AUC score for DNN (green), Angle Embedding (blue) and Amplitude Embedding circuits (yellow) for the *muon dataset* (a) and the *complete dataset* (b). The dashed line represents a random classifier.

High Energy Physics and Instrumentation Center at IFUSP

Machine learning e jatos no LHC

Deep Learning Jet Image as a Probe of Light Higgsino Dark Matter at the LHC

Huifang Lv,[1,*] Daohan Wang,[2,3,†] and Lei Wu[1,‡]

[1]*Department of Physics and Institute of Theoretical Physics, Nanjing Normal University, Nanjing, China*

[2]*CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

[3]*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

arXiv:2203.14569

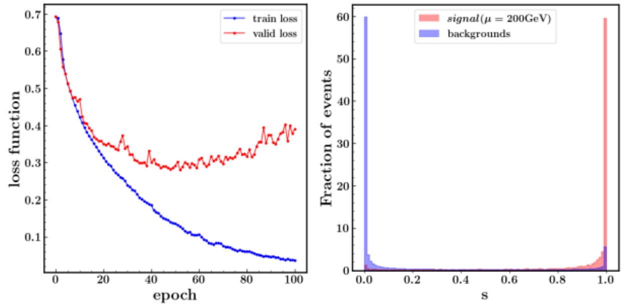


FIG. 8. The dependence of the loss function on the epoch for training samples and validation samples (left panel); The classification effect diagram of the test set of signals and backgrounds (right panel).

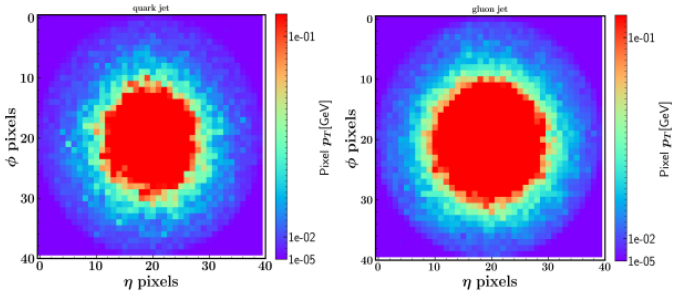
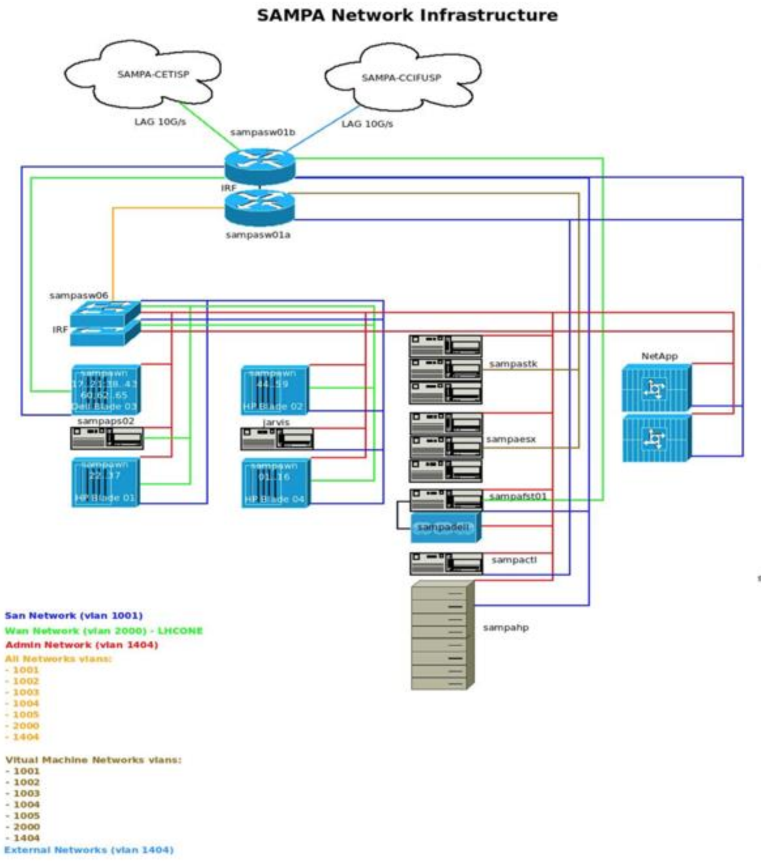


FIG. 6. The jet images of the quark jet (left panel) and the gluon jet (right panel) in the signal after the translation and rotation. The higgsino mass is taken as $\mu = 200$ GeV.

Infraestrutura
Cluster SAMPA

- Configuração atual
  - 2408 cores
    - 18.7 kHS06
  - 0.85 TB de armazenamento
  - 2 GPU Tesla K20
- Uso
  - Processamento em GRID do ALICE e ATLAS
  - Análise de dados e simulações
  - Processamento em geral
  - Desenvolvimento



Cronograma de upgrades

	Requested Processing Power (kHS06)	Extra Compute Modules for the SAMPA cluster	Processing Power per Compute Module (kHS06)
2022	20,7	4	0.62
2023	23,8	6	0.62
2024	27,4	5	0.75
2025	31,5	6	0.75
2026	36,3	6	0.90
TOTAL		27	

Table 2: *Expect grown of the processing power of the computer cluster assuming 15% per year increase in the demand and 20% evolution every 2 years of the processing power of the compute modules*

	Requested Disk Storage (PB)	Number of Extra Disks for the SAMPA Cluster	Single Disk Capacity (TB)
2022	2,12	91	14
2023	2,44	23	14
2024	2,80	23	16
2025	3,22	27	16
2026	3,71	27	18
TOTAL		191	

Table 3*: Expect grown of the storage power of the computer cluster assuming 15% per year increase in the demand and 20% evolution every 2 years of single-disk capacity*

Atividades planejadas

- Aprendizado e formação em ciência de dados
- Análises de dados e desenvolvimento
- Gerenciamento de infraestrutura
- Service work com viés computacional no ALICE