International Symposium on Grids and Clouds 2022
20 - 25 March 2022, Academia Sinica, Taipei, Taiwan

The ML_INFN initiative

L. Anderlini [1], T. Boccali [2], S. Dal Pra [3], D. C. Duma [3], D. Michelotto [3], D. Spiga [4]
for the ML_INFN initiative

Istituto Nazionale di Fisica Nucleare

1. Sezione di Firenze
2. Sezione di Pisa
3. CNAF
4. Sezione di Perugia

INFN Research and structures

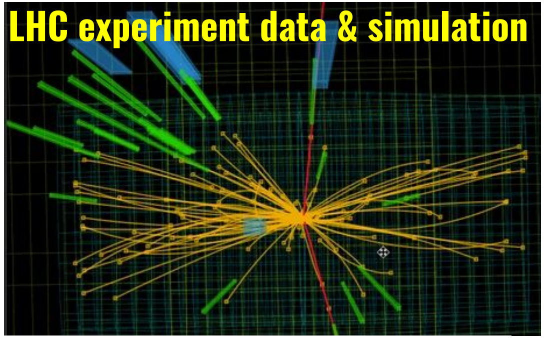
216 activities distributed in 33 structures (labs, groups and divisions)

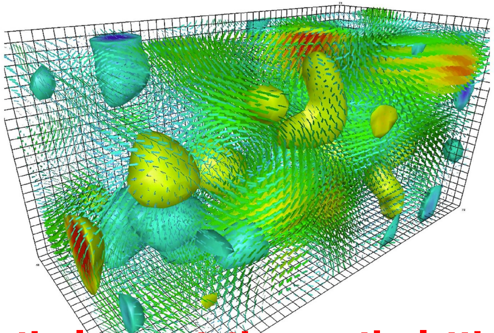
CSN1	**Particle Physics**	17 experiments
CSN2	**Astroparticle Physics**	45 experiments
CSN3	**Nuclear Physics**	23 experiments
CSN4	**Theoretical Physics**	35 initiatives
CSN5	**Technological Research**	96 experiments

Machine Learning Technologies for INFN

Most of the experiments and initiatives produce, analyse or process digital data.

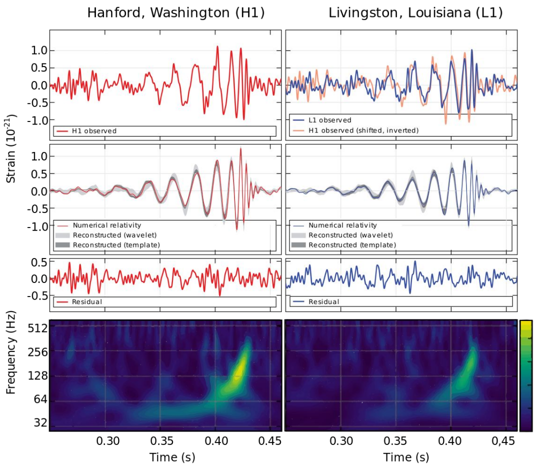
Enthusiasm on the modern data processing technologies!



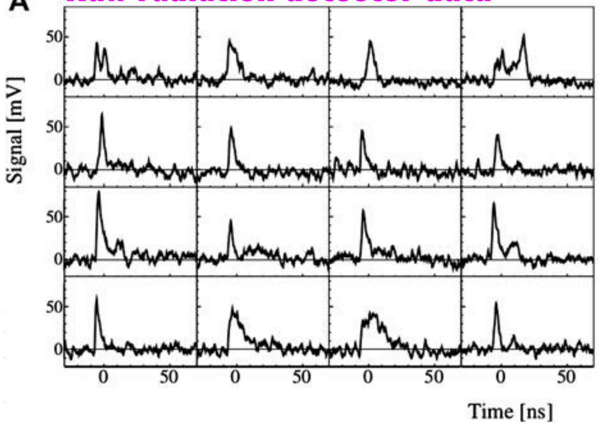


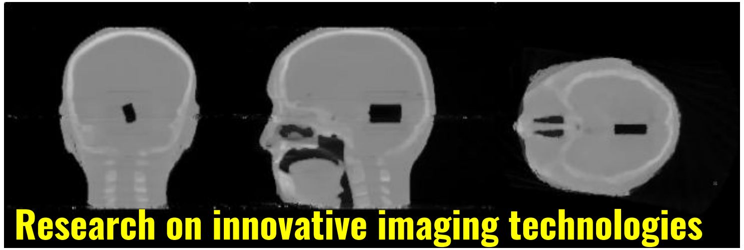
Theoretical computations on the lattice

Gravitational wave detection



Raw radiation detector data





The potential barriers

Employing machine learning techniques often requires:

- specialized hardware and software setup
- specific training to identify tools and learning resources
- a community of experts providing support to research use cases

~~The potential barriers~~ Lowering the potential barriers with ML_INFN

Employing machine learning techniques often requires:

- specialized hardware and software setup

WP1: provide a centrally maintained cloud-based infrastructure for interactive and batch ML fast prototyping, with access to modern GPU hardware and systems tuned for ML performance

- specific training to identify tools and learning resources

WP2: organize national training events for INFN users (Machine Learning hackathons)

- a community of experts providing support to research use cases

WP3: provide and organize example applications in a knowledge base

The numbers of ML_INFN

11 INFN **structures** involved in the preparation of the knowledge base

79 **researchers** devoting a fraction of their time to promote ML techniques for research

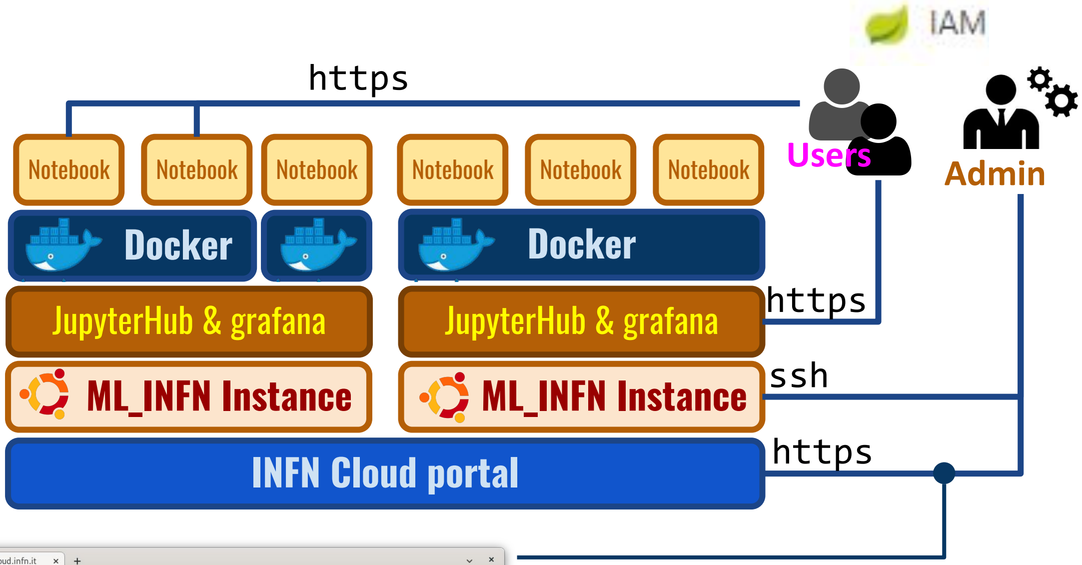
14 professional **GPUs** made available and accessible through the INFN Cloud Interface

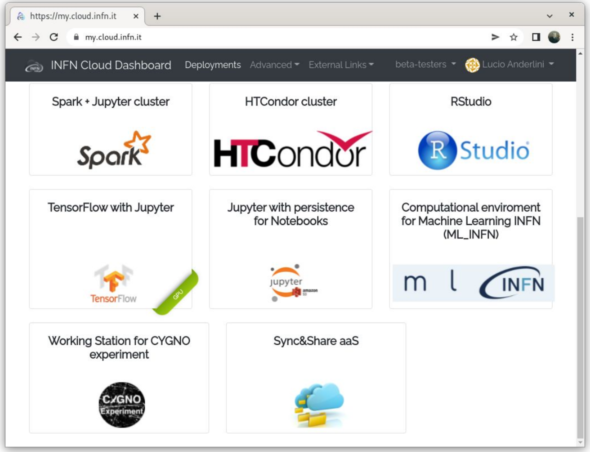
110 **participants** to the **hackathons**, ranging from students to permanent staff members

WP1. The infrastructure

INFN Cloud

ML_INFN is built on top of **INFN Cloud**: a data lake-centric, heterogeneous federated Cloud infrastructure spanning multiple sites across Italy, providing an extensible portfolio of solutions tailored to **multidisciplinary scientific communities**.





See also…

Davide's Talk on Friday on **INFN Cloud**

Igor's Talk on Tuesday on **CYGNO in INFN Cloud**

Daniele's Talk on Friday on **INFN Cloud for clinical data**

Federated baremetal resources

1× SuperMicro + 1× E4 servers:

- 1 TB RAM
- 64-128 CPU cores
- 36 TB local storage (NVMe)
- 8× **Tesla T4** GPUs
- 5× **RTX 5000** GPUs
- 1× **A30** GPU
- 10 GbE connection to CNAF resources

Federated to CNAF OpenStack and INFN Cloud

Storage solutions

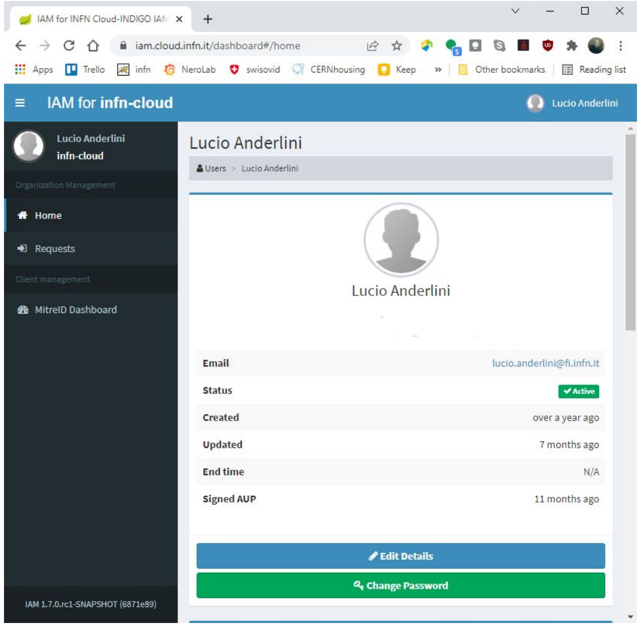
Storage from CERN experiments can be mounted with NFS from the Tier-1 storage

Hypervisors integrated to Ceph to manage persistent virtual volumes accessed from the VM with POSIX

Authentication & authorization mechanism

Identity and Access Management is handled with **INDIGO IAM,** same tool as used by the WLCG community:

- The official INFN Cloud instance instance
- A secondary instance providing **temporary access** for demonstrations, hackathons and tutorials.



Authentication is demanded to the central INFN service (AAI) based on Kerberos and LDAP.

Authorization in **INDIGO IAM** is based on the use of "**groups**" that define the level of access (e.g. admin or user) to the solutions and resources.

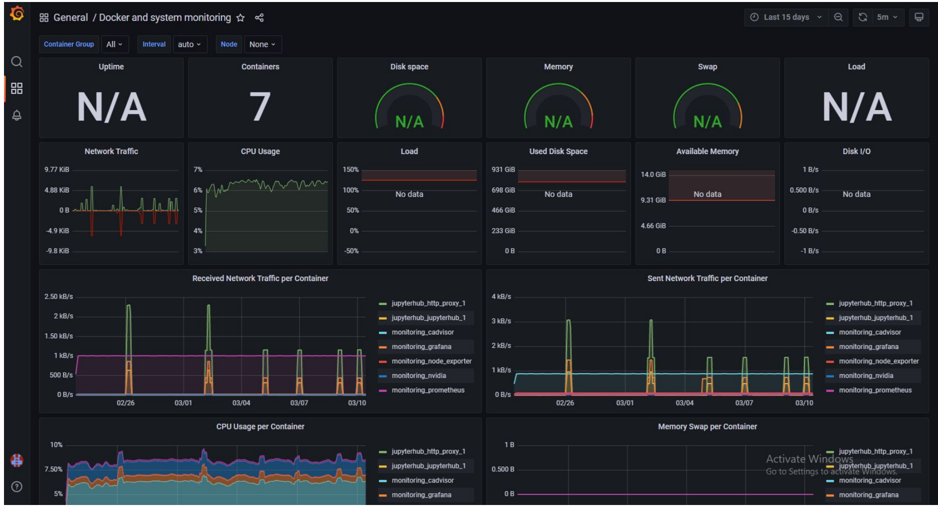
INFN
CCR - AAI

Virtualization stack

- CentOS7 hypervisor
- OpenStack virtualization
- Ubuntu 18 and Ubuntu 20 images with CUDA installed
- cvmfs available, if requested
- JupyterHub integrated with IAM
- Default docker image for the Jupyter Notebook provided with a series of widely used ML tools
  - can be customized /enriched with other tools requested by the users, made available through DockerHub
- Miniconda used to make ROOT available as a Jupyter Kernel

Public repository
ML-INFN

Monitoring and arbitration



Grafana monitoring dashboard deployed with each VM.

Resource assignment managed at:

- ML_INFN level (projects and sys. admins)
- Project level (research groups and users)

Some users contribute to the financial effort to acquire the resources and gets priority in arbitration.

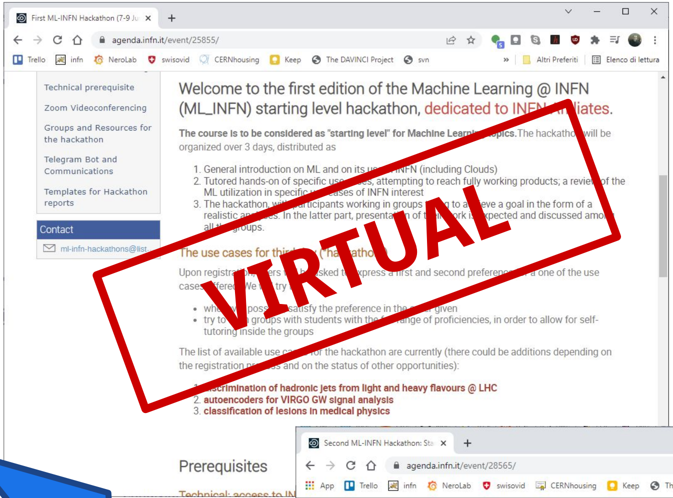
In case of pressure, the resources allocated to privileged users may be reclaimed if monitoring reports low usage.

WP2. Stewardship

Hackathons in Covid-19 Era

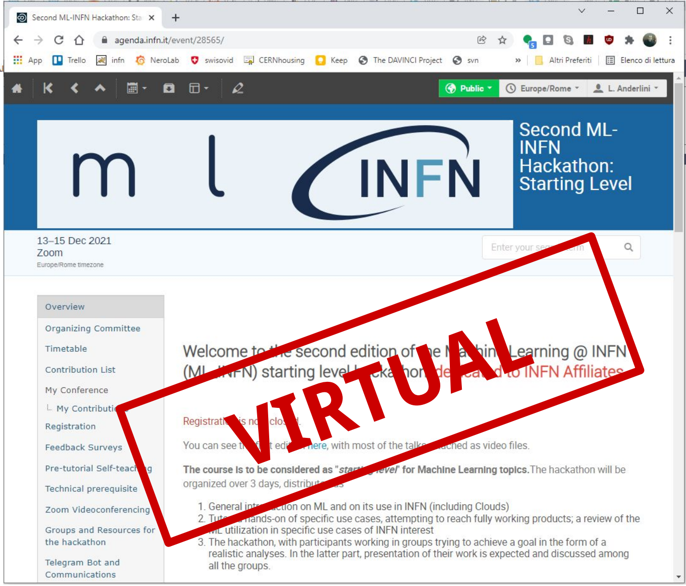
Originally planned as satellite events of scientific workshops, canceled due to pandemic, hackathons have been transformed in virtual events.

Registrations limited to 60 to guarantee decent tutor-per-student and RAM-per-student ratios.



"sold-out" in 72 hours

Repeated



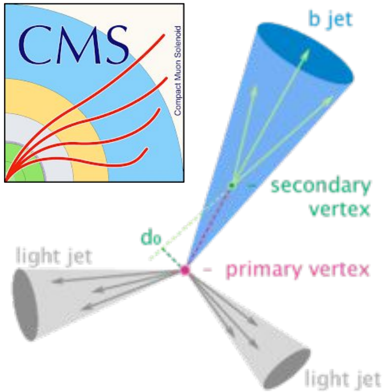
Lecture Program

Day 1	Day 2	Day 3
Lectures **Theoretical introduction to ML**	Hands-on **Neural Networks**	Hackathon **Exercises with tutors *continuous support***
Lectures **Cloud and Cloud Resources**	Seminars **Deep Neural Network Applications to INFN research**	
Lunch break		
Hands-on **Numpy, Pandas and Keras**	Hands-on **Exercises with tutors *on demand***	Closure **Reports from the students**

Hackathon use cases: *10 groups, one tutor per group*

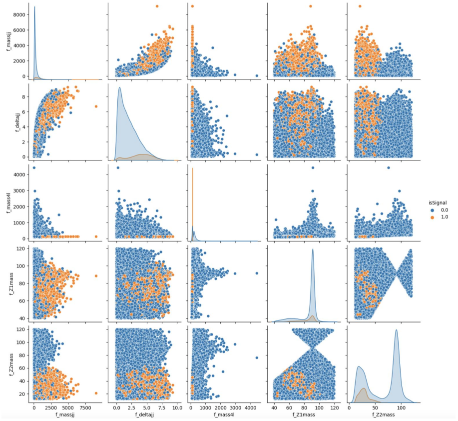
Jet b-tagging at CMS

Recurrent Neural Networks with LSTM



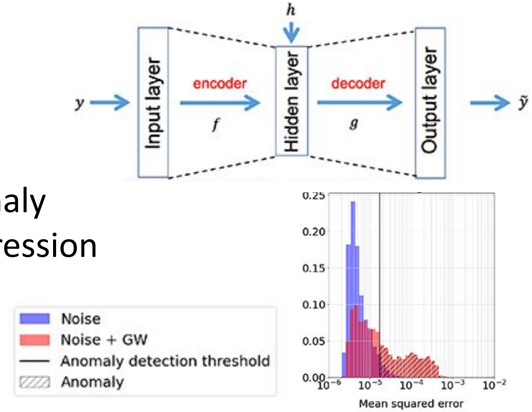
Higgs searches at CMS

Deep Neural Networks and Advanced Keras



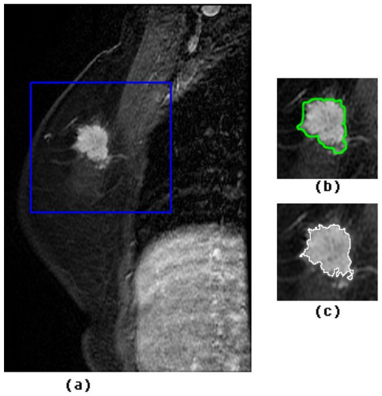
Gravitational Waves with Virgo

Autoencoders, anomaly detection and compression



Segmentation of CT scans

Convolutional Neural Networks
Handling 2D and 3D datasets



Final survey

A satisfaction questionnaire was submitted to the participant at the end of each event.

Generous feedback on:

- Level of difficulty
- Relevance and interest
- Technical setup

About a half of the participant responded they would have been willing to spend **more time** on the hands on and/or the hackathon.

The Cloud / Jupyter setup

Did you find the technical setup using Cloud + Jupyter reasonable for ML oriented analyses?

A. Yes, worked for me: 19 (95.00%)

B. Yes, it generally worked for me (please add comment below): 1 (5.00%)

C. No (please explain below why): 0 (0.00%)

On the difficulty level:

On the choice of topic:

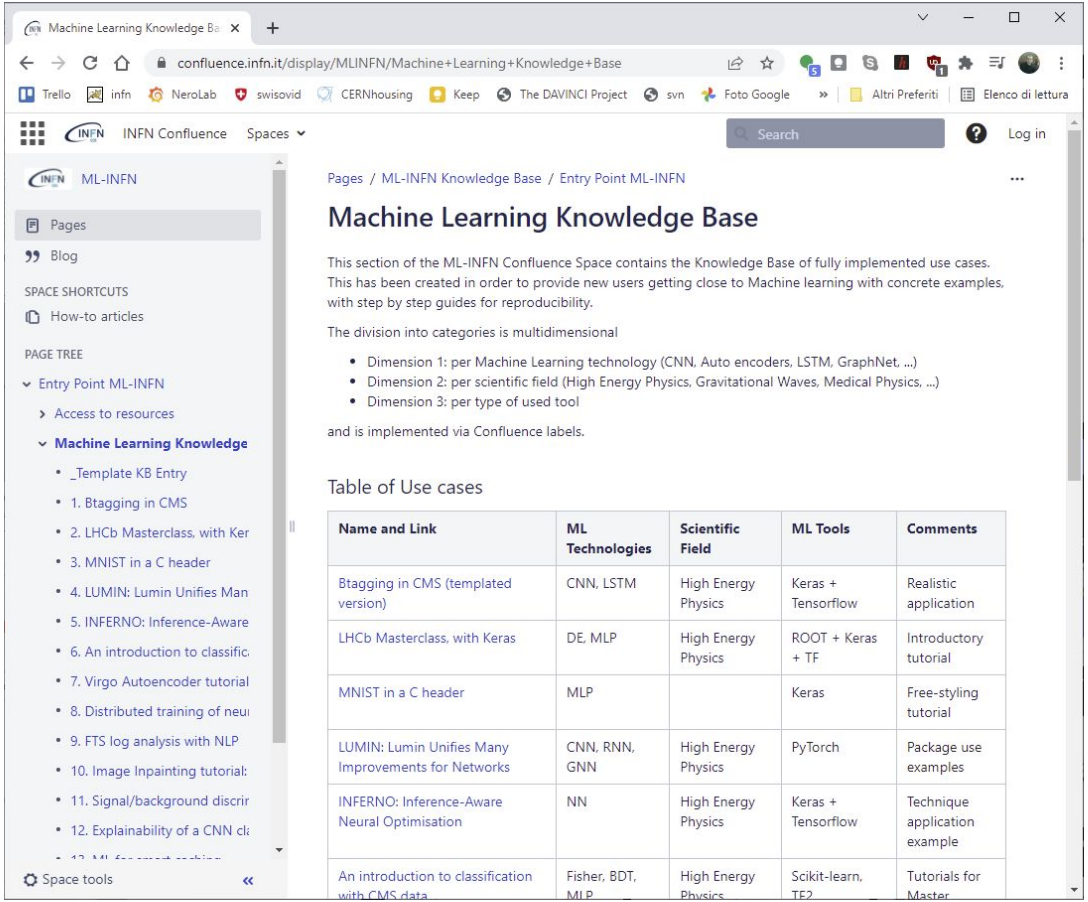
WP3. Network and Knowledge Base

Confluence Knowledge Base

Atlassian Confluence was used to build a **Knowledge Base** reporting several machine-learning use cases, including those discussed at the hackathon.

Each entry includes:

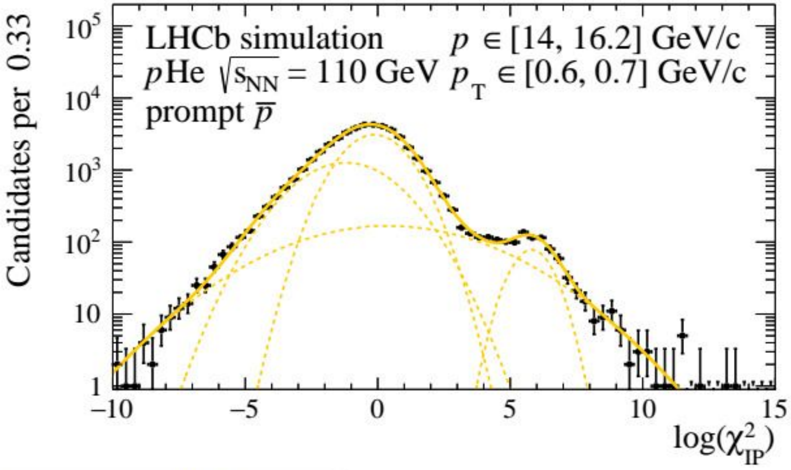
- Runnable **example** as a jupyter notebook or a git repository
- **Contact information** of one or more experts

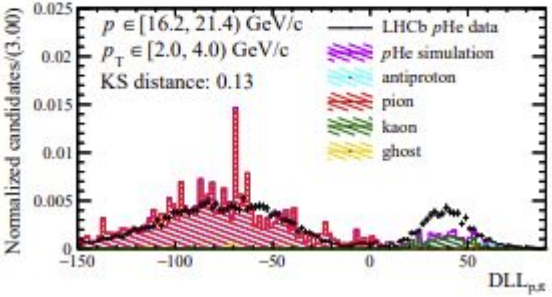


Stories of success [1]: *building template models for LHCb*

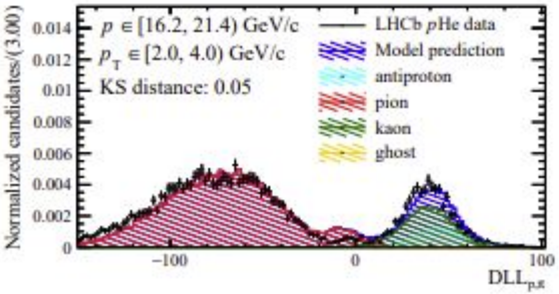
ML_INFN infrastructure was used to develop a model for the Particle Identification response of the LHCb detector as a Gaussian-Mixture model.

With Gaussian parameters inferred with a Deep Neural Network.





Traditional method based on reweighted MC



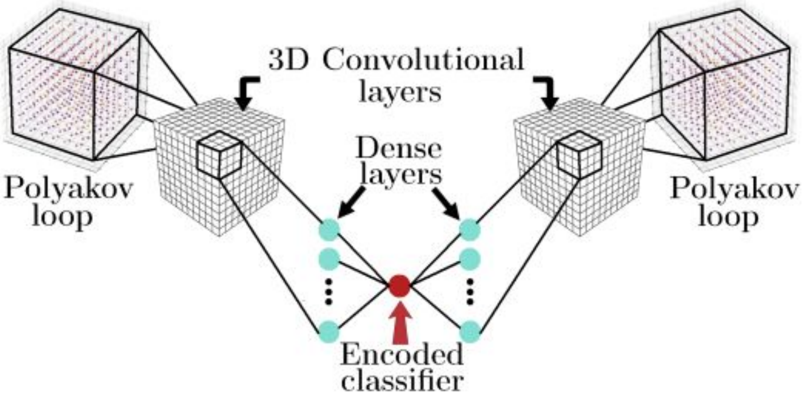
Deep Learning model

S. Mariani *et al*, *"A Neural-Network-defined Gaussian Mixture Model for particle identification applied to the LHCb fixed-target programme"*, JINST 17 (2022) P02018

Public repository **PID4SMOG**

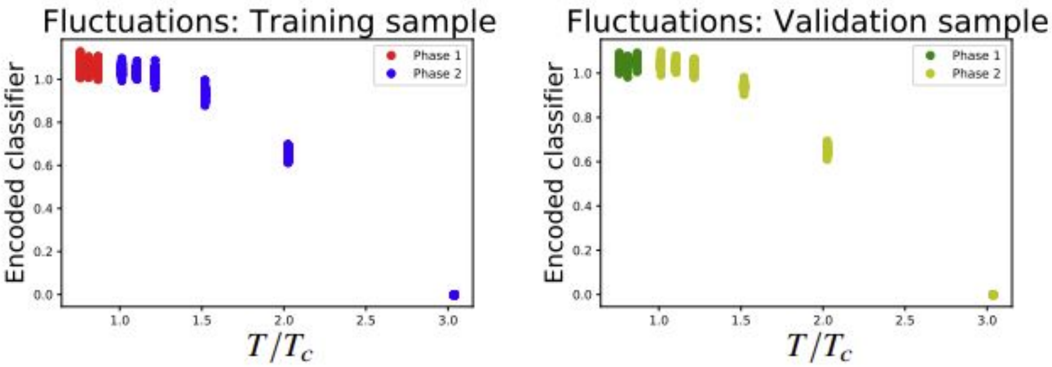
Stories of success [2]: *studying LQCD with CNNs*

A Deep Neural Network is trained in a semi-supervised manner to define an effective order-parameter for Gauge theory where a real order-parameter is not defined.



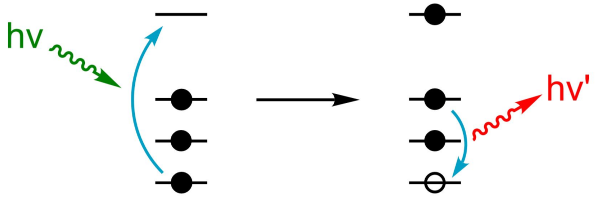
The study was made possible thanks to the GPUs provided by the ML_INFN initiative.

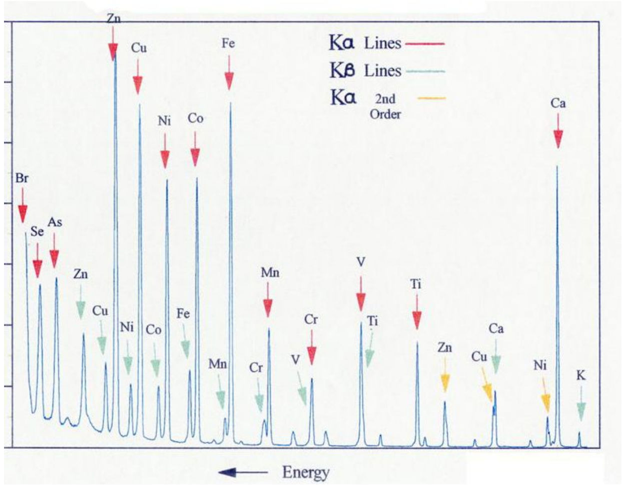
A. Palermo, M.P. Lombardo *et al.*
"Machine learning approaches to the QCD transition", Proceeding of LATTICE21



Stories of success [3]: *X-rays to visible colors for CH*

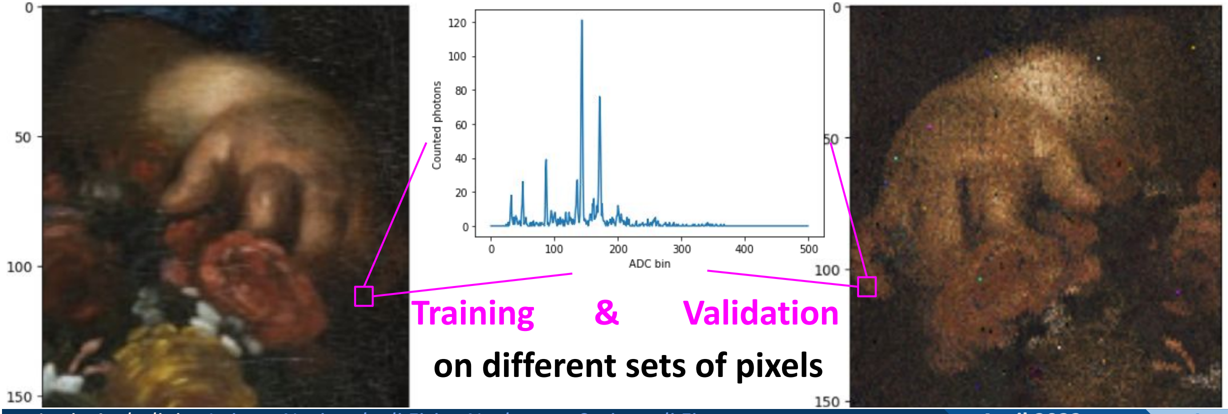
X-ray fluorescence spectroscopy widely used for Heritage Conservation and non-invasive probe of pictorial artworks.





INFN CHNet
Istituto Nazionale di Fisica Nucleare
Cultural Heritage Network

EOSC-Pillar
Coordination and Harmonisation of National &Thematic Initiatives to support EOSC



Deep Neural Network models are trained to reconstruct the image from the XRF scan of the pixels.

Public repository
Hackathon tutorial

Summary and conclusions

Summary

The ML_INFN initiative has been providing many INFN experiments with the hardware and the knowledge base to assess the potential **benefit of machine learning to their research** for three years.

The **ML_INFN** project relies on **INFN Cloud** solutions and it federates resources optimized for ML performance in interactive and batch-like usage patterns (high-end professional GPUs, NVMe disks, many-core high-RAM systems)

A series of national training events (***machine learning hackathons***) and a collection of tutorials and real applications within the INFN community (***knowledge base***) contribute to building **a network of experienced and enthusiast machine learning practitioners**, lowering the skill gap to benefit from machine learning developments.

Outlook

Machine Learning is here to stay. In the next future:

- We will organize ***Advanced Training Course(s)*** on Deep Learning
- We will provide Cloud-based access to **FPGAs** as **Machine Learning accelerators**
  - Two U50 and a U250 Xilinx FPGAs recently federated to the cloud

See Marco's Talk on Thursday on **FPGA in the Cloud for ML**

- We will keep supporting students and researchers employing **Machine Learning technologies** in their daily activities.

See also…

Davide's Talk on Friday on **INFN Cloud**

Igor's Talk on Tuesday on **CYGNO in INFN Cloud**

Daniele's Talk on Friday on **INFN Cloud for clinical data**