



Fourth ML_INFEN hackathon

Welcome & introduction

Lucio Anderlini
Istituto Nazionale di Fisica Nucleare
Sezione di Firenze



Istituto Nazionale di Fisica Nucleare
SEZIONE DI FIRENZE



INFN Research and structures



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



Particle Physics

17 experiments



Astroparticle Physics

45 experiments



Nuclear Physics

23 experiments



Theoretical Physics

35 initiatives



Technological Research

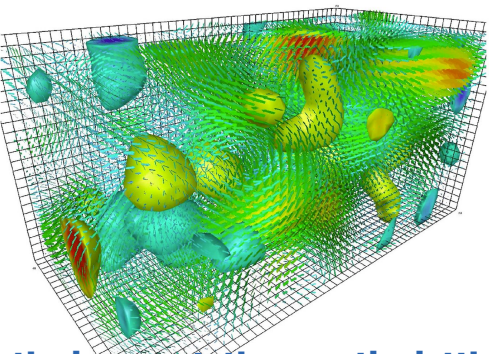
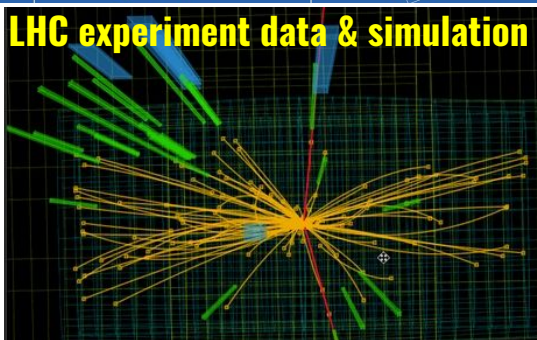
96 experiments



Machine Learning Technologies for INFN

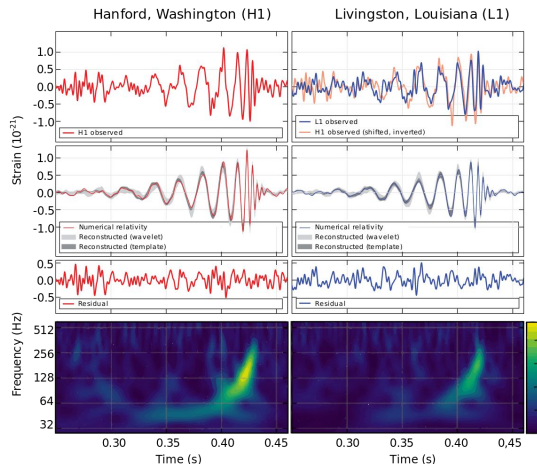
LHC experiment data & simulation

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

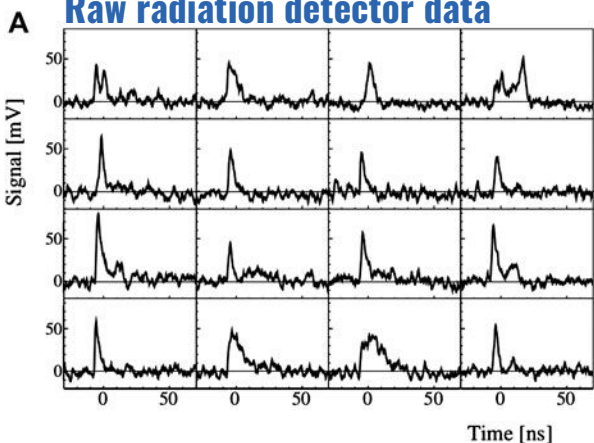


Enthusiasm on the modern data processing technologies!

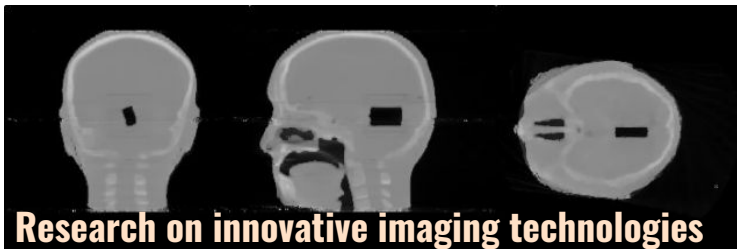
Gravitational wave detection



Raw radiation detector data



Theoretical computations on the lattice

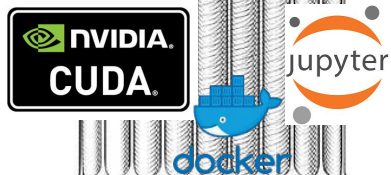


Research on innovative imaging technologies

ML_INFN: The structure of the project

Applications of Machine Learning
HEP, MedPhys, GW detection, Theory...

Infrastructure



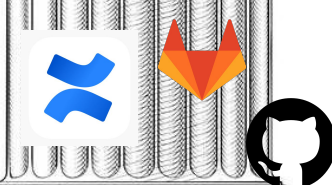
WP 1

Stewardship



WP 2

Knowledge Base



WP 3



Virtualization and orchestration layer
developed and maintained by INFN Cloud



The numbers of ML_INFN

12 INFN **structures** involved in the developments, training activities and hackathons

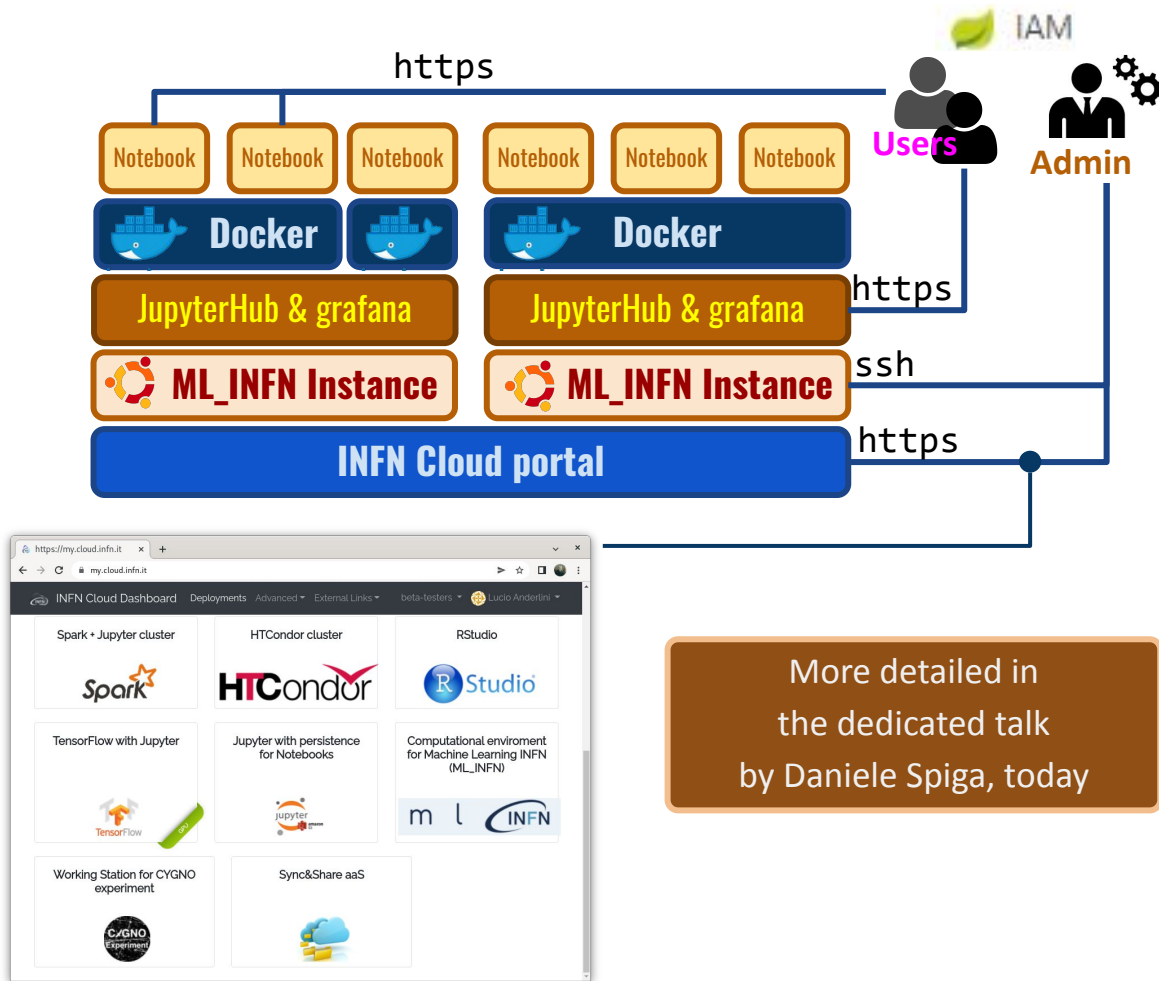
79 **researchers** devoting a fraction of their time to promote ML techniques for research

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

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

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**.



More detailed in
 the dedicated talk
 by Daniele Spiga, today

Federated bare-metal resources

1x SuperMicro + 1x E4 servers:

- 1 TB RAM
- 64-128 CPU cores
- 36 TB local storage (NVMe)
- 8x **Tesla T4** GPUs
- 5x **RTX 5000** GPUs
- 1x **A30** GPU
- 1x **A100** GPU, served as 7 independent MIG slices
- 10 GbE connection to CNAF resources



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

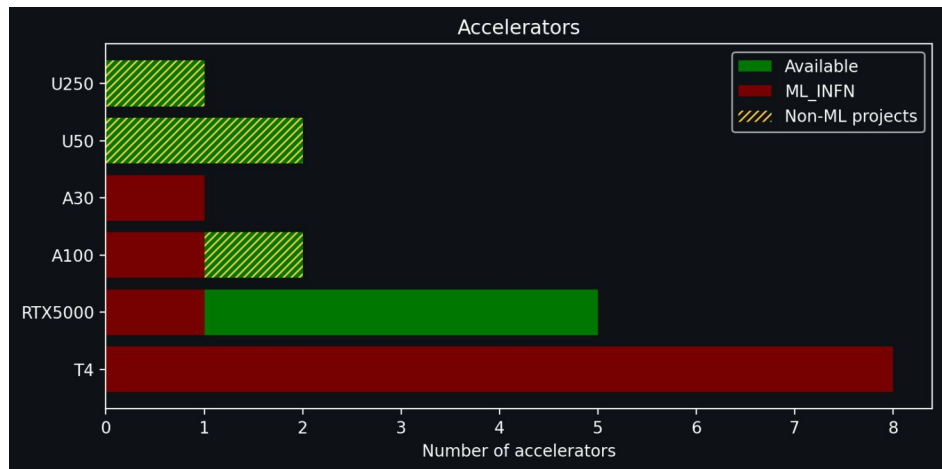
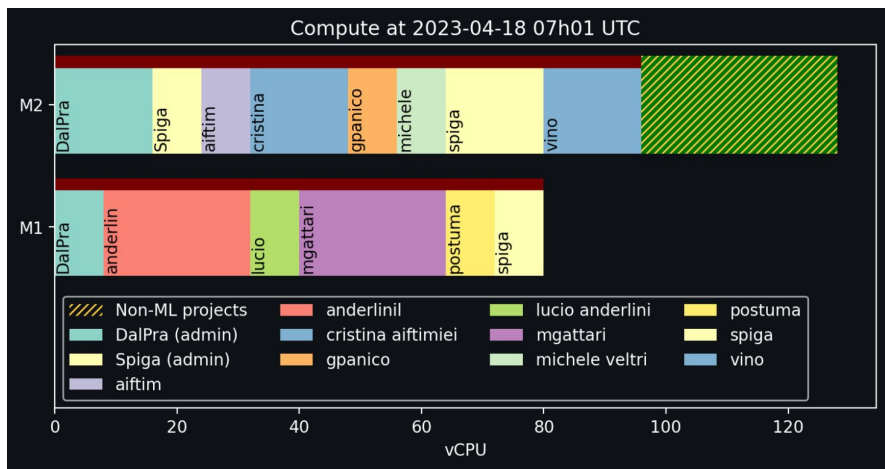
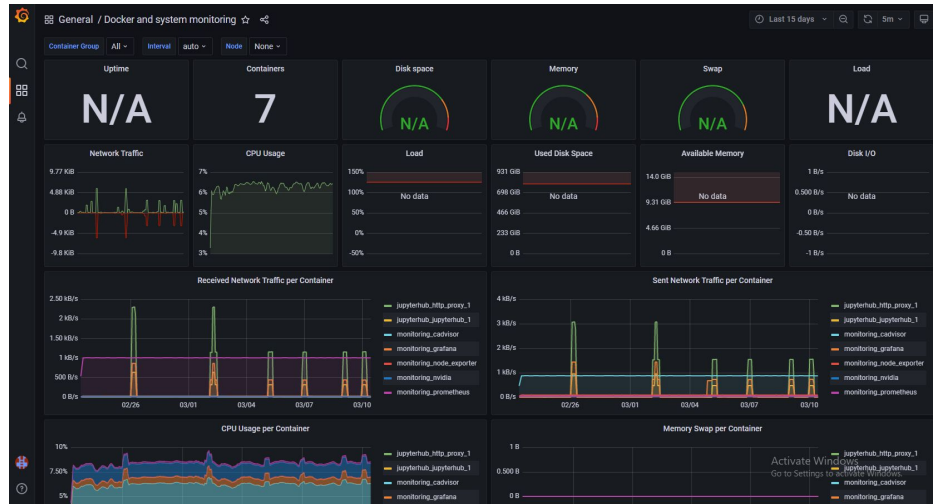
Federated to CNAF OpenStack and INFN Cloud

Monitoring and accounting

On top of INFN Cloud services, special tools tuned for ML workflows.

Streamlit dashboard to visualize share of resources at regular snapshots

Grafana dashboard per-VM for utilization.



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

The screenshot shows the Confluence interface for the ML-INFN Knowledge Base. The left sidebar contains a navigation menu with sections for Pages, Blog, SPACE SHORTCUTS (How-to articles), and PAGE TREE. The main content area displays the title 'Machine Learning Knowledge Base' and a description: 'This section of the ML-INFN Confluence Space contains the Knowledge Base of fully implemented use cases. This has been created in order to provide new users getting close to Machine learning with concrete examples, with step by step guides for reproducibility. The division into categories is multidimensional'.

The division into categories is multidimensional

- Dimension 1: per Machine Learning technology (CNN, Auto encoders, LSTM, GraphNet, ...)
- Dimension 2: per scientific field (High Energy Physics, Gravitational Waves, Medical Physics, ...)
- Dimension 3: per type of used tool

and is implemented via Confluence labels.

Table of Use cases

Name and Link	ML Technologies	Scientific Field	ML Tools	Comments
Tagging in CMS (templated version)	CNN, LSTM	High Energy Physics	Keras + Tensorflow	Realistic application
LHCb Masterclass, with Keras	DE, MLP	High Energy Physics	ROOT + Keras + TF	Introductory tutorial
MNIST in a C header	MLP		Keras	Free-styling tutorial
LUMIN: Lumin Unifies Many Improvements for Networks	CNN, RNN, GNN	High Energy Physics	PyTorch	Package use examples
INFERN0: Inference-Aware Neural Optimisation	NN	High Energy Physics	Keras + Tensorflow	Technique application example
An introduction to classification with CMS data	Fisher, BDT, MLP	High Energy Physics	Scikit-learn, TE2	Tutorials for Master

Machine Learning hackathons: *Base and Advanced level*

To foster the adoption of machine learning tools and techniques in INFN community, we organize events to discuss ML algorithm with the time to look at (and hack) the code.

Starting-level Hackathons

- **online events** with no fee
- up to **60 participants**
- **1 tutor per 5 participants**
- INFN Cloud **CPUs** with shared filesystem

Advanced Hackathons

- **in-person events**
- up to **30 participants**
- **(almost) 1 tutor per participant**
- INFN Cloud **GPUs** with shared filesystem

Day 1

Lectures and introduction
to the hands-on

WEDNESDAY, 21 JUNE

09:00	→ 09:20	Introduction and logistics	🕒 20m	📄
		Speaker: Lucio Anderlini (Istituto Nazionale di Fisica Nucleare)		
09:20	→ 11:15	Introduction: General Introduction to Machine Learning		📄
	09:20	What is Machine Learning and why it is relevant for research?	🕒 50m	📄
		Speaker: Luca Scrucca (University of Perugia, Department of Economics)		
		📄 Scrucca Intro2ML.pdf		
	10:10	Neural Networks: principles and common architectures	🕒 1h 5m	📄
		Speaker: Stefano Giagu (ROMA1)		
		📄 ML_INF_N_Hackatho...		
11:15	→ 11:30	Break	🕒 15m	
11:30	→ 12:30	Introduction: INFN Cloud for ML-INFN		📄
	11:30	INFN-Cloud, and how to use It for ML_INFN	🕒 1h	📄
		Speaker: Daniele Spiga (PG)		
12:30	→ 13:45	Lunch break	🕒 1h 15m	
13:45	→ 16:15	Hands-on: Getting started with python and numpy (+ ...) on the cloud infrastructure		📄
	13:45	Hands-on - Playing with CERN Open Data with numpy, pandas and pyplot	🕒 1h 15m	📄
		Speaker: Lucio Anderlini (FI)		
	15:00	Introduction to keras	🕒 30m	📄
		Speakers: Andrea Rizzi (INFN Pisa), Andrea Rizzi (PI)		
	15:30	Hands-on - Introduction to keras	🕒 45m	📄
		Speakers: Andrea Rizzi (PI), Andrea Rizzi (INFN Pisa)		

Day 2

Seminars on real use-cases of machine learning in INFN research and exercises inspired to research use cases

We will suggest exercises during the morning.

During the afternoon you will have the time to go through (some of) them, getting ready for the Friday hackathon.

We will be around for support and discussion.

Ask questions!

THURSDAY, 22 JUNE			
09:00	→ 09:40	Machine Learning Applications for Gravitational Wave science Speaker: Elena Cuoco (Istituto Nazionale di Fisica Nucleare)	🕒 40m
09:40	→ 10:15	ML Basics: hands-on on a simple categorization example in HEP Speaker: Lucio Anderlini (FI)	🕒 35m
10:15	→ 10:30	Coffee break	🕒 15m
10:30	→ 11:30	Hands-on: Convolutional Neural Networks Conveners: Andrea Rizzi (INFN Pisa), Andrea Rizzi (Istituto Nazionale di Fisica Nucleare)	
11:30	→ 12:30	Real applications of ML in INFN activities - Image Restoration in heritage Speaker: Alessandro Bombini (FI)	🕒 1h
12:30	→ 13:10	Machine Learning Applications for Medical Physics Speaker: Piernicola Oliva (University of Sassari and INFN Cagliari)	🕒 40m
13:10	→ 14:20	Lunch break	🕒 1h 10m
14:20	→ 16:20	Hands-on: Continuation and finalization of hands on	

FRIDAY, 23 JUNE

Day 3

On Friday, the real hackathon.

You will be split in 10 groups.

Each group is assigned to a tutor who will introduce and support your work.

Each group will access a dedicated machine at CNAF.

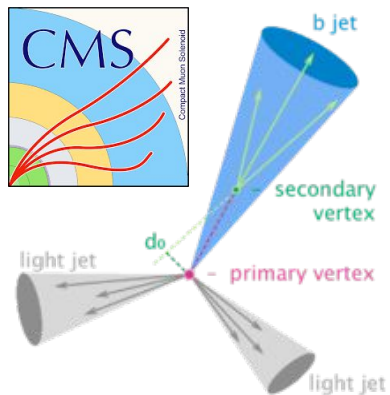
During the afternoon we will reconvene all together to discuss your solutions and close the event

Time	Activity	Duration
09:00 → 09:40	Hackathon Introduction: Description of the exercises	
09:00	Discrimination of hadronic jets from light and heavy flavours @ LHC Speaker: Matteo Malucchi (INFN Pisa)	10m
09:10	Signal/background discrimination for the VBF Higgs four lepton decay channel with the CMS experiment Speaker: Walaa Elmetenawee (Istituto Nazionale di Fisica Nucleare)	10m
09:20	Autoencoders for VIRGO GW signal analysis Speakers: Lucia Papalini (Istituto Nazionale di Fisica Nucleare), Michele Vacatello (Istituto Nazionale di Fisica Nucleare)	10m
09:30	Classification of lesions In medical physics Speaker: Francesca Lizzi (Istituto Nazionale di Fisica Nucleare)	10m
09:40 → 13:00	Hands-on: Hackathon	
09:40	b-Tagging at CMS with Recurrent Neural Networks Speakers: Andrea Rizzi (INFN Pisa), Andrea Rizzi (Istituto Nazionale di Fisica Nucleare), Francesco Vaselli (INFN Pisa), Matteo Malucchi (INFN Pisa)	3h 20m
09:40	Classification of lesions In medical physics Speakers: Alessandra Retico (Istituto Nazionale di Fisica Nucleare), Andrea Berti (Istituto Nazionale di Fisica Nucleare), Camilla Scapicchio (Istituto Nazionale di Fisica Nucleare), Francesca Brero (Istituto Nazionale di Fisica Nucleare), Ian Postuma (Istituto Nazionale di Fisica Nucleare), Dr Lorenzo Marini (Istituto Nazionale di Fisica Nucleare), Raffaella Cabini (Istituto Nazionale di Fisica Nucleare), Sara Saponaro (Istituto Nazionale di Fisica Nucleare)	3h 20m
09:40	Processing Gravitational Waves with Machine Learning Speakers: Luca Rei (Istituto Nazionale di Fisica Nucleare), Lucia Papalini (Istituto Nazionale di Fisica Nucleare), Marco Serra (Istituto Nazionale di Fisica Nucleare), Massimiliano Razzano (University of Pisa and INFN-Pisa), Michele Vacatello (Istituto Nazionale di Fisica Nucleare)	3h 20m
09:40	Selecting Higgs Candidates with Deep Neural Networks Speakers: Ms Brunella D'Anzi (INFN - Bari), Giorgia Miniello (Istituto Nazionale di Fisica Nucleare), Nicola De Filippis (BA), Walaa Elmetenawee (Istituto Nazionale di Fisica Nucleare)	3h 20m
13:00 → 14:30	Lunch break	1h 30m
14:30 → 16:30	Hackathon Introduction: Discussion on exercises	

Hackathon use cases: 10 groups, one tutor per group

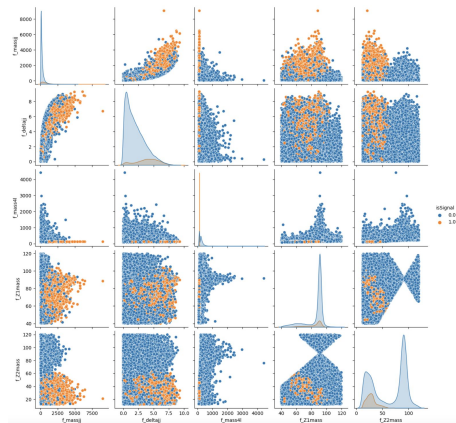
Jet b-tagging at CMS

Recurrent Neural Networks with LSTM



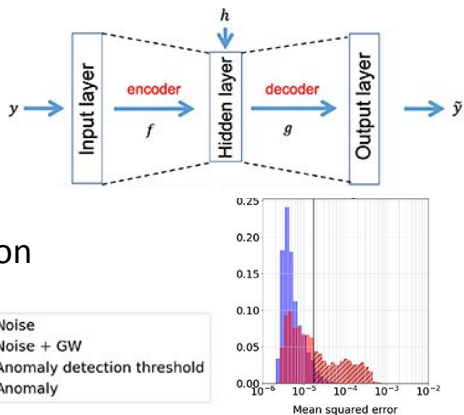
Higgs searches at CMS

Deep Neural Networks and Advanced Keras



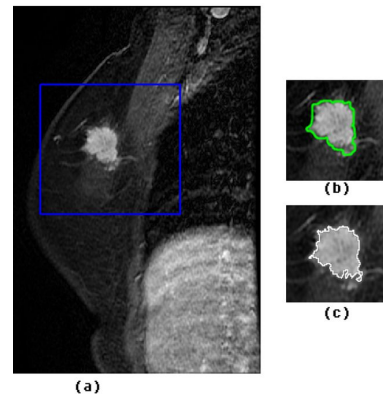
Gravitational Waves with Virgo

Autoencoders, anomaly detection and compression



Segmentation of CT scans

Convolutional Neural Networks
Handling 2D and 3D datasets



Very important:

Indico page:

agenda.infn.it/event/35607


If you haven't yet, get your account to access the machines

Check the group (and the machine) assigned to you.

You can use your machine also for the hands-on exercises of day 1 and day 2.

Fourth ML-INFN Hackathon: Starting Level

21–23 Jun 2023
Zoom
Europe/Rome timezone

Overview
Organizing committee
Timetable
Registration
Participant List
Get a computing account
Pre-tutorial self-teaching
Groups and computing resources
Contact
 ml-infn-hackathons@list...

Welcome to the fourth edition of the Machine Learning @ INFN (ML-INFN) starting level hackathon, **dedicated to INFN Affiliates**.

If you are looking for the previous editions, check the following links:

- [First edition](#) (starting level)
- [Second edition](#) (starting level)
- [Third edition](#) (advanced level)

The course is to be considered as "**starting level**" for Machine Learning topics. The hackathon will be organized over 3 days, distributed as

1. General introduction on ML and on its use in INFN (including Clouds)
2. Tutored hands-on of specific use cases, attempting to reach fully working products; a review of the ML utilization in specific use cases of INFN interest
3. The hackathon, with participants working in groups trying to achieve a goal in the form of a realistic analyses. In the latter part, presentation of their work is expected and discussed among all the groups.

The use cases for third day ("hackathon")

Upon registration, users will be asked to express a first and second preference for a one of the use cases offered. We will try to

- whenever possible, satisfy the preference in the order given
- try to form groups with students with the full range of proficiencies, in order to allow for self-tutoring inside the groups

The list of available use cases for the hackathon are currently (there could be additions depending on the registration process and on the status of other opportunities):



Summary

Welcome to the Fourth ML_INF N Hackathon!

The event is online to lower the costs and enable younger students to get involved.

Let's make an effort to make it equally effective:

1. ask questions (either on zoom or through the mailing list ml-infn-hackathons@lists.infn.it)
2. pick the exercises that you're most curious about and give them a try!
3. have fun!