



UNIVERSITÀ DEGLI STUDI DI NAPOLI  
**FEDERICO II**

# Machine Learning methods for Nuclear Emulsions experiments

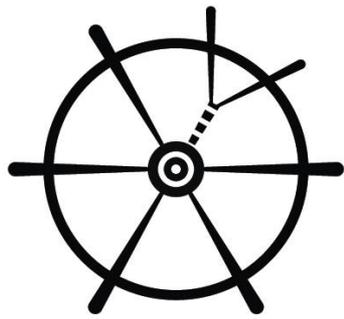
**Artem Golovatiuk**

Universita di Napoli (Italy)

**Principal Investigator:**

**Giovanni De Lellis**

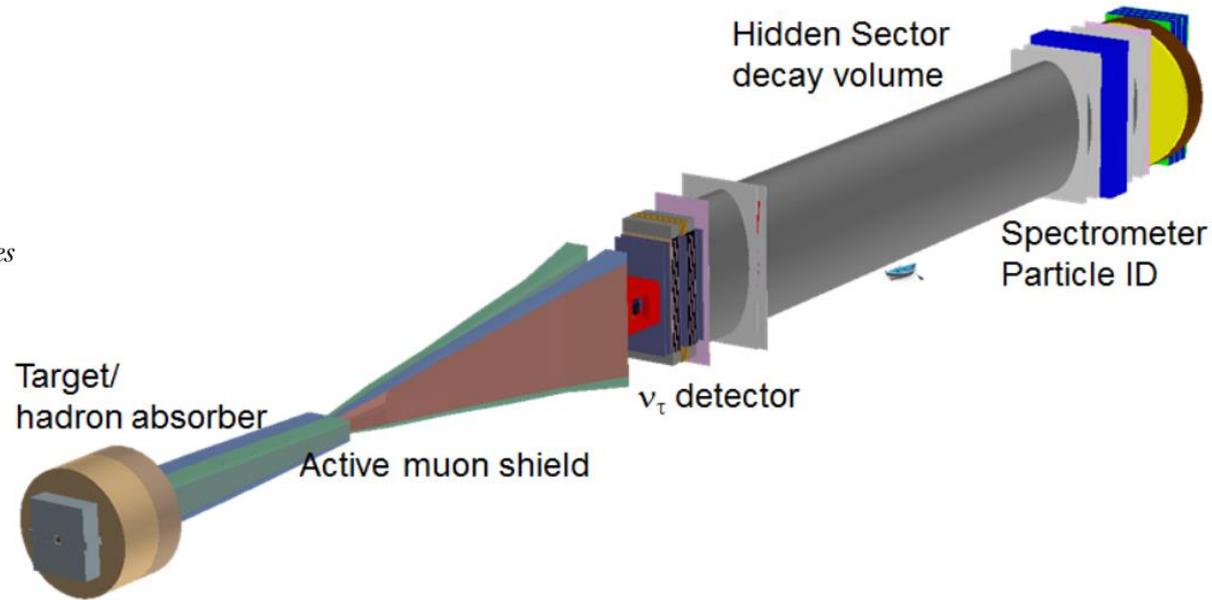
Universita di Napoli (Italy)



**SHiP**

*Search for Hidden Particles*

# The SHiP experiment



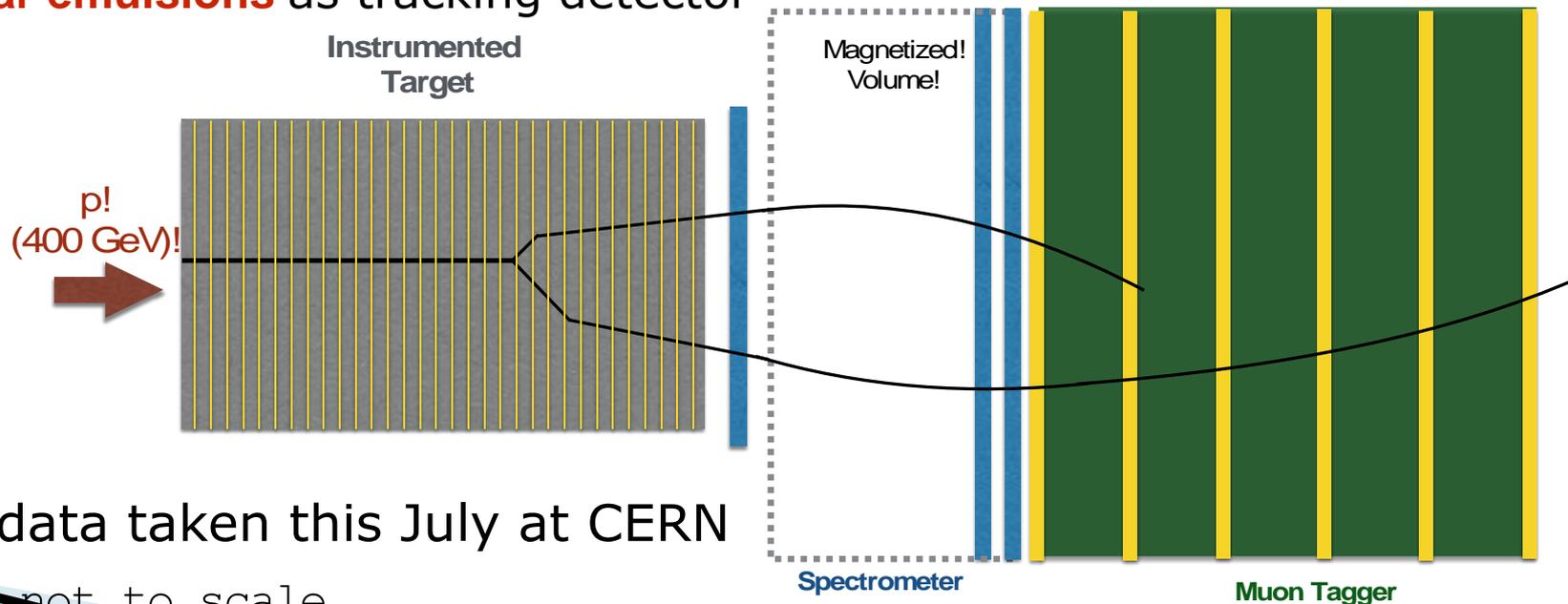
- ▶ Fixed target experiment at the CERN SPS
- ▶ Main goals are:
  - Search for very weakly interacting particles with masses below the Fermi scale
  - Study tau neutrino physics



# The SHiP Charm project

- ▶ Charm production in **proton interactions** and in **hadron cascades** in the SHiP target important for Hidden Sector searches normalization and  $v_T$  cross-section measurement

- ▶ Proton collisions in Lead target instrumented with **nuclear emulsions**
- ▶ **Nuclear emulsions** as tracking detector



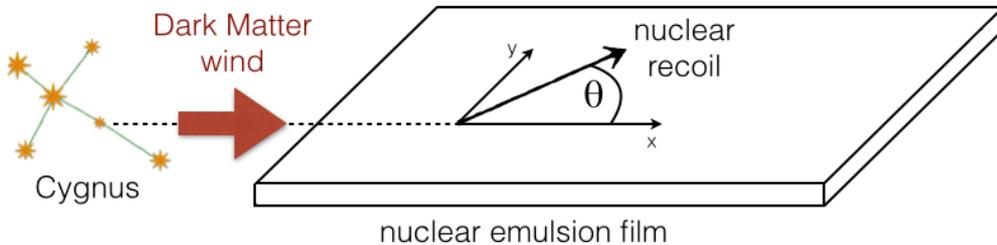
- ▶ First data taken this July at CERN

not to scale

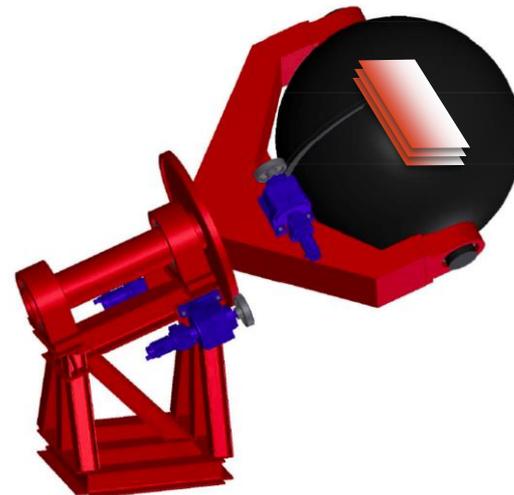


# The NEWSdm experiment

- **Aim:** detect the direction of **nuclear recoils** produced in WIMP interactions
- **Required resolution:** O(100 nm)
- **Target:** nuclear emulsions acting both as target and tracking detector
- **Background reduction:** neutron **shield** surrounding the target
- **Fixed pointing:** target mounted on **equatorial telescope** constantly pointing to the Cygnus Constellation
- **Location:** Underground Gran Sasso Laboratory



Equatorial Telescope

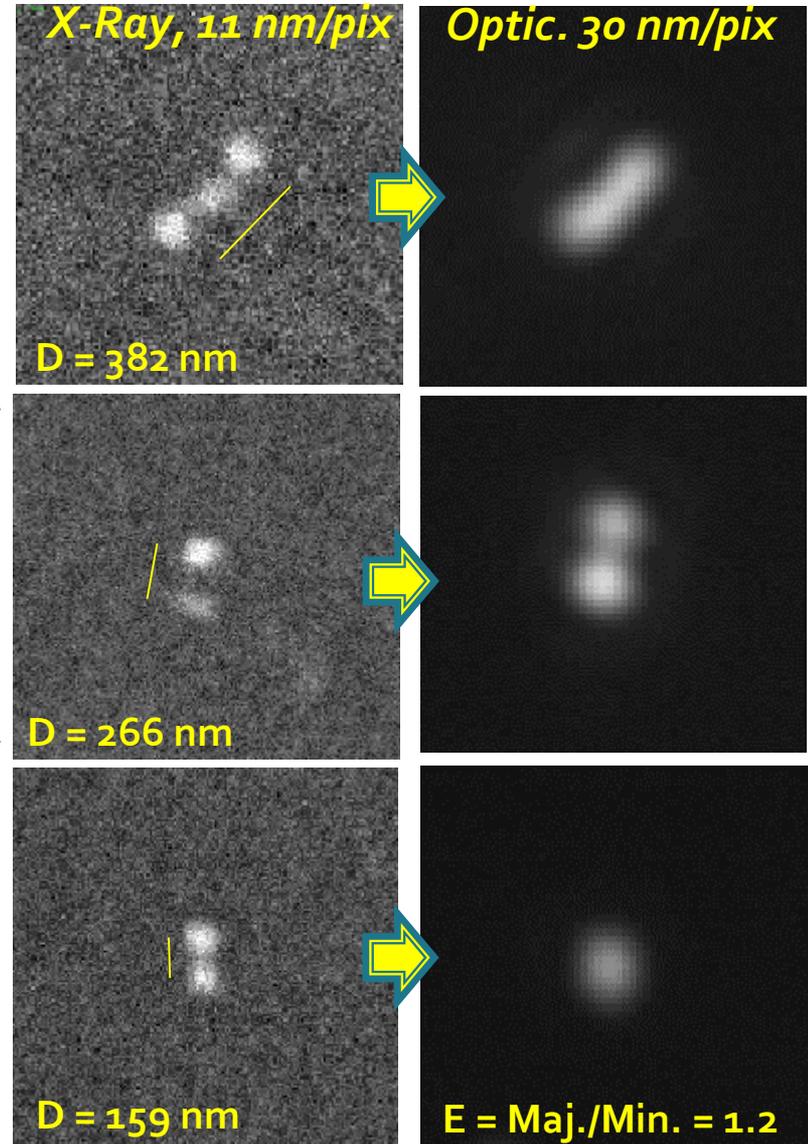
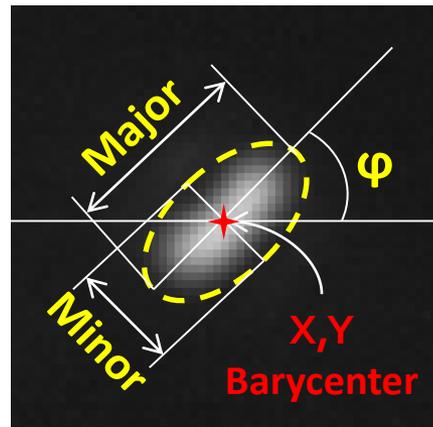




# Track images analysis

- ▶ Track is a sequence of close and aligned grains
- ▶ Resolve a track = distinguish it from a single grain
- ▶ Grains are closer than the mic. resolution (200 nm) → single spot
- ▶ Single grain = round spot  $1.4 \mu\text{m}$

- ▶ Plasmon resonance effect allows probing the internal structure of the cluster with polarized light



# Why do we need ML?

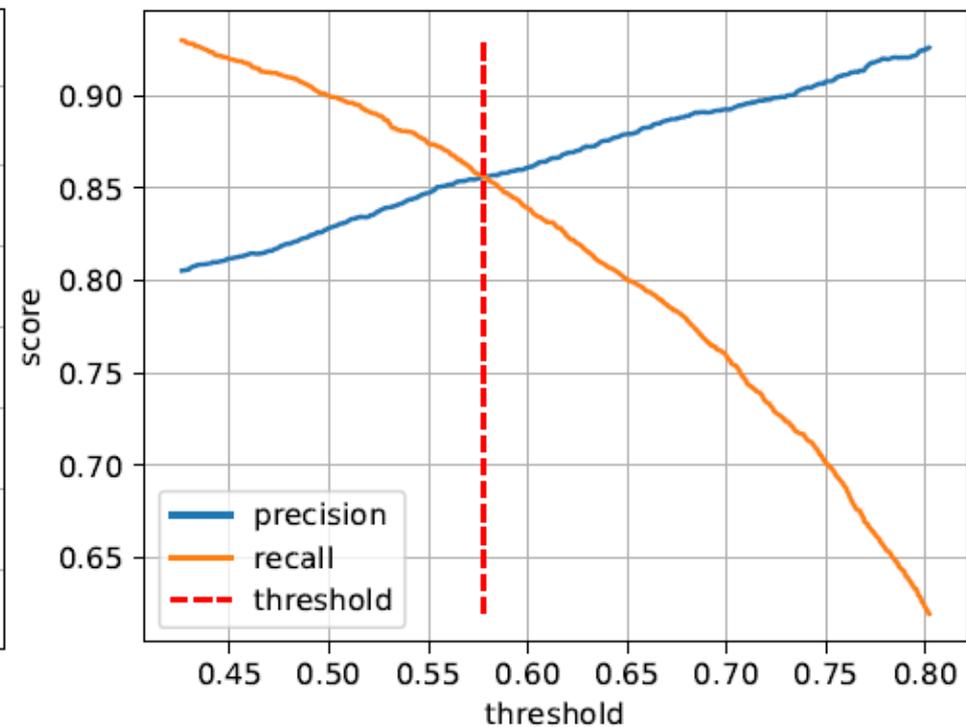
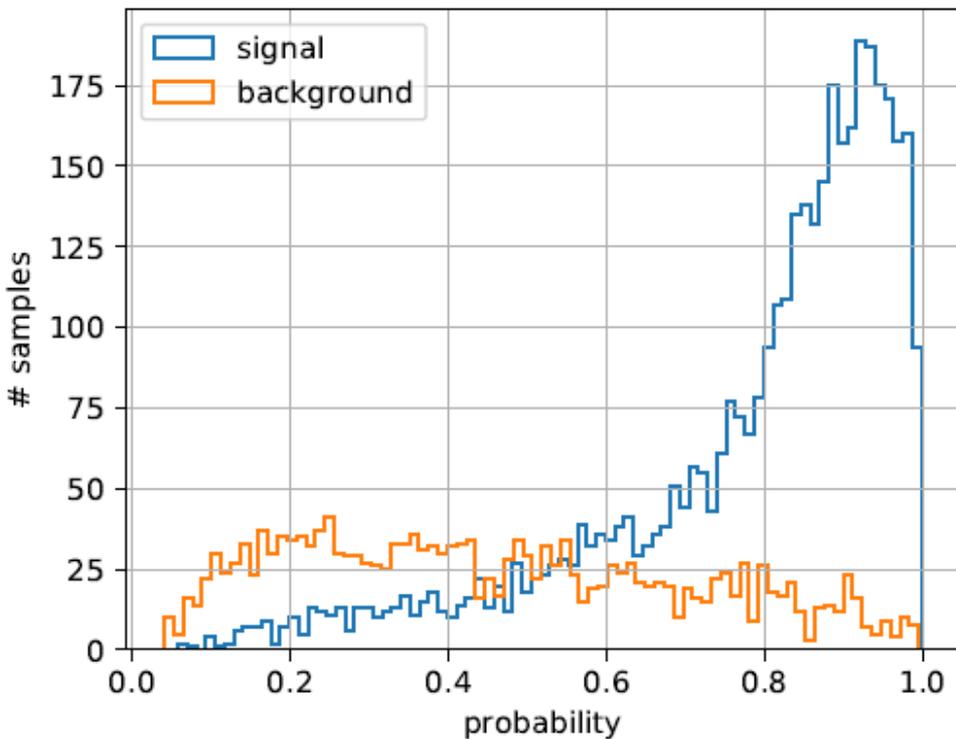
- ▶ Our goal:
  - Reducing the number of background events in potentially signal data
  - Exploring the correlations between event parameters and purification quality.
- ▶ Machine Learning approach:
  - Can discover complex correlations between features
  - Can be robust to insignificant variations in case of high input dimensions.

# Previously tested approaches

- ▶ **Boosted Decision Trees:**
  - Composition of small decision trees, next one improves result of the previous one.
  - Limited possibility to parallelize
- ▶ **Random Forest:**
  - Composition of very deep trees, each one makes its own decision, result is the average of probabilities.
  - Highly parallelizable on CPUs
- ▶ **Trees weakness:**
  - Performance strongly depends
  - on the features choice.



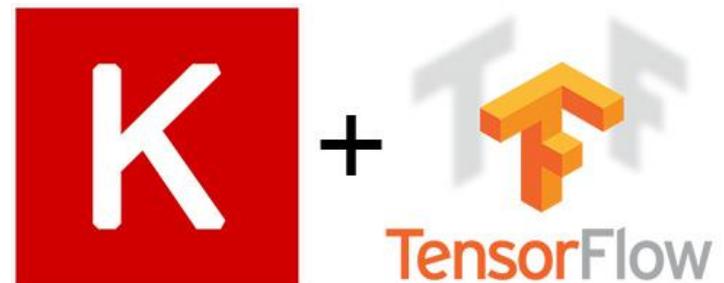
# Preliminary results (Trees)



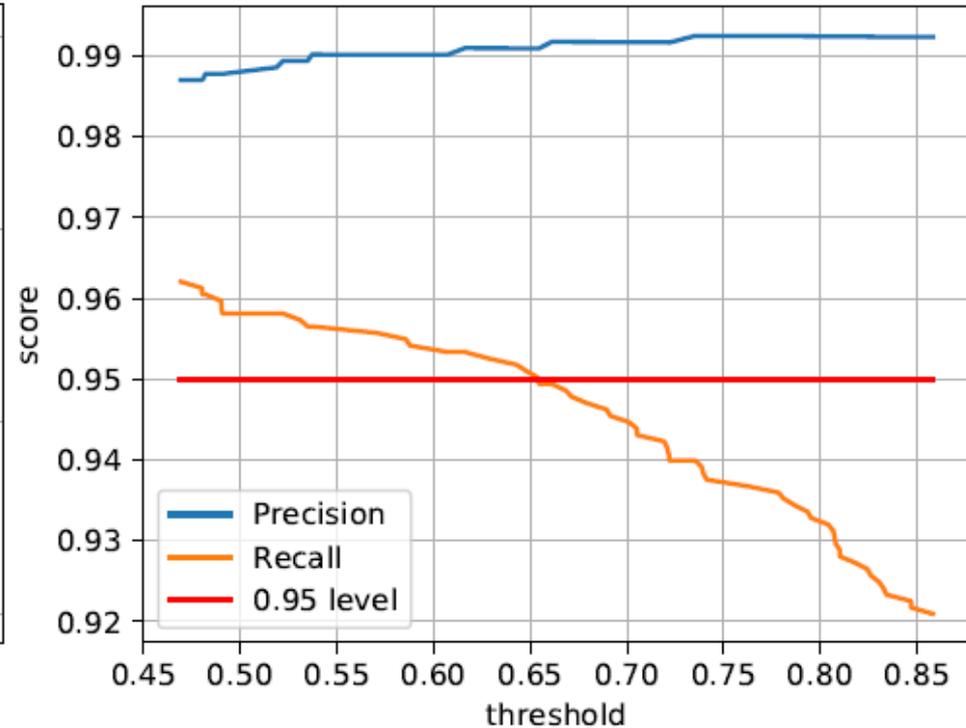
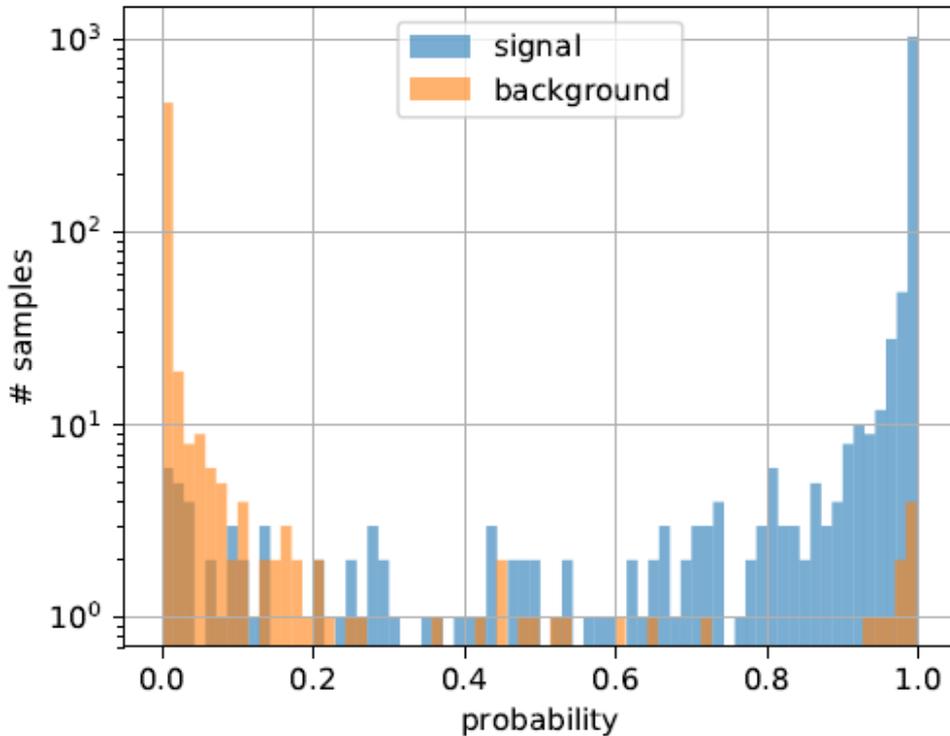
- ▶ Random Forest with  $10^4$  trees test output and physical scores

# Previously tested approaches

- ▶ Convolutional Neural Networks:
  - Compared 2D and 3D architectures
  - Compared Deep and Shallow Networks
  - Working directly with the cluster images
  - Requires large computational power (e.g. GPU)
  - Larger datasets can be highly profitable for performance



# Preliminary results (CNNs)



- ▶ 3D Conv4 test output and physical scores

# Further possible applications of ML

## ▶ SHiP experiment:

- Track and vertex reconstruction in nuclear emulsions in presence of high occupancy (*several hundreds events/ 1000 cm<sup>3</sup>*)
- Event classification

## ▶ NEWSdm experiment:

- Improving the track classification
  - Using color images (scattered wavelength depends on the form of the grains)
  - Enlarging and diversifying the dataset
  - Studying the correlation between classification performance and track direction
- Fast denoising of the images from the microscope to speed-up the scanning

▶ More applications TBA