Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Machine Learning Techniques for Space Calorimeter Experiments

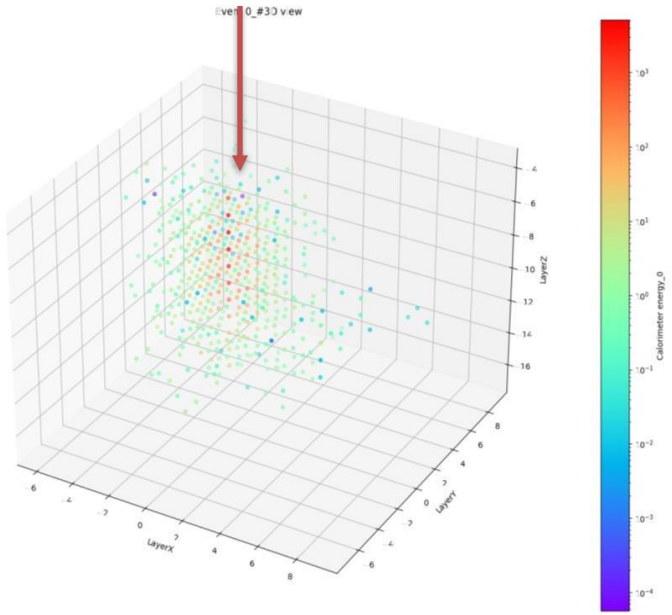
Maria Bossa, Federica Cuna, Fabio Gargano

Referee: Gianpaolo Carlino (INFN Sezione di Napoli), Pasquale Lubrano (INFN Sezione di Perugia)

Spoke 3 Technical Meeting, Bologna 17–19 /12, 2024

Looking Back (Elba meeting)

We simulated a pixelated spatial calorimeter using a toy Monte Carlo simulation, using linear monoenergetic primary particles propagated. Then we applied XGBoost algorithm for electron and proton discrimination.



On a sample of 20k events, the results are :

- **Accuracy** XGB Classifier: **99.85%**
- **Recall** XGB Classifier: **99,90%**
- **Precision** XGB Classifier**: 99,80%**

Results presented at ML4Astro in Catania

Planned next steps

- Conduct tests with various algorithms, exploring their performance across different particle types and energies. ✓
- Explore the possibility of training a Convolutional Neural Network (CNN) using GPU acceleration. (CNN->Transformer) ✓
- Perform classification tests using more sophisticated and robust simulations. ✓
- Develop an integrated framework to facilitate spatial experiments, incorporating both tracker (see Federica Cuna's talk) and calorimeter functionalities.(Ongoing)

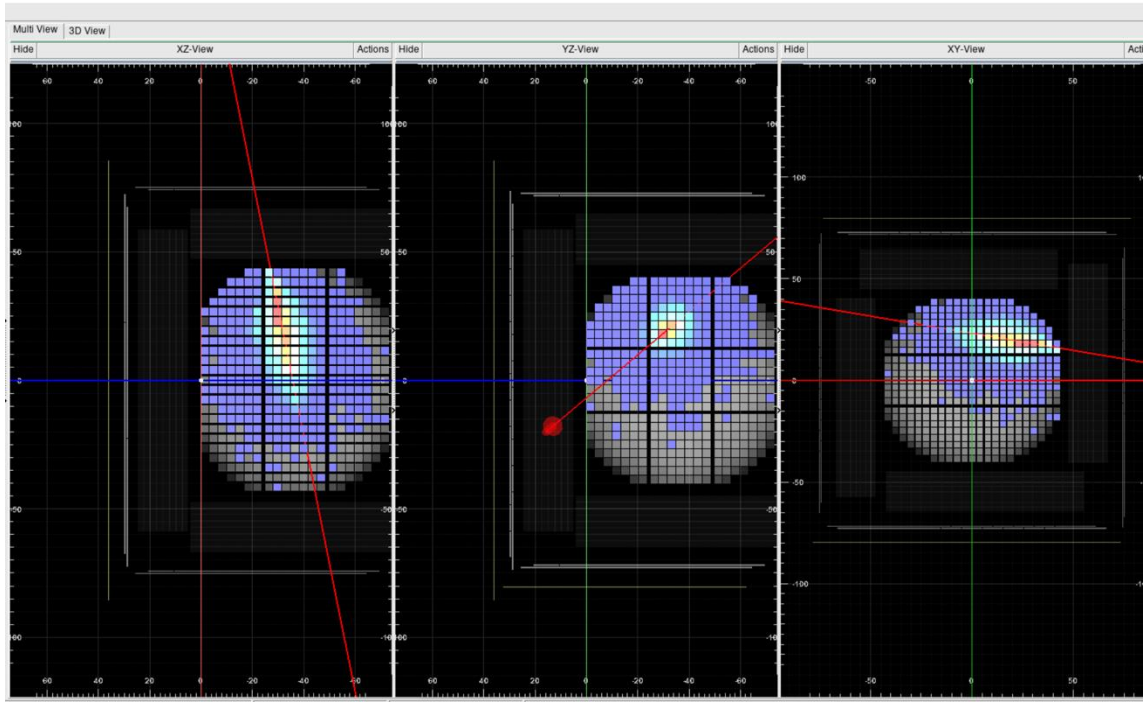
Progress so far

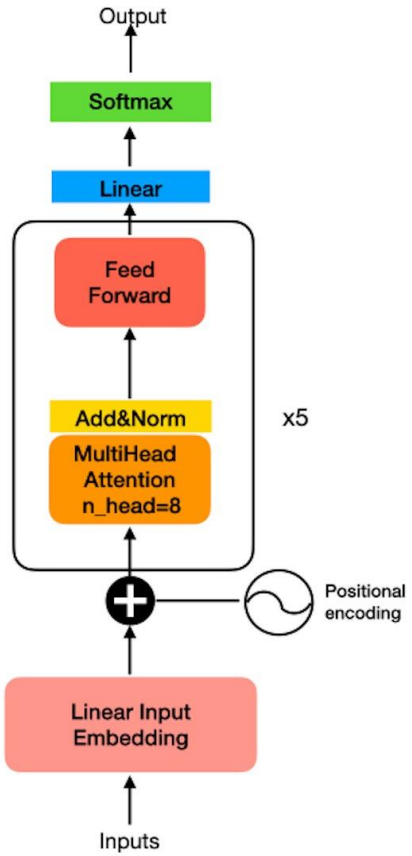
- We have worked with a new set of simulation data from the HERD detector. HERD has been chosen as test bench for this application, since the simulation software, based on Geant4 toolkit, provides complete and accurate description of the detector.

- We simulated 2150000 events for electrons and 2150000 events for protons. They were generated with an energy range between 100 GeV and 1 TeV, following a power-law spectrum of E^{-1}

- After data preprocessing we moved to use transformers for the classification task: we choose them for their versatility in handling sequential and high-dimensional data. Moreover they are largely used in particle physics (jet tagging in CMS/ATLAS)

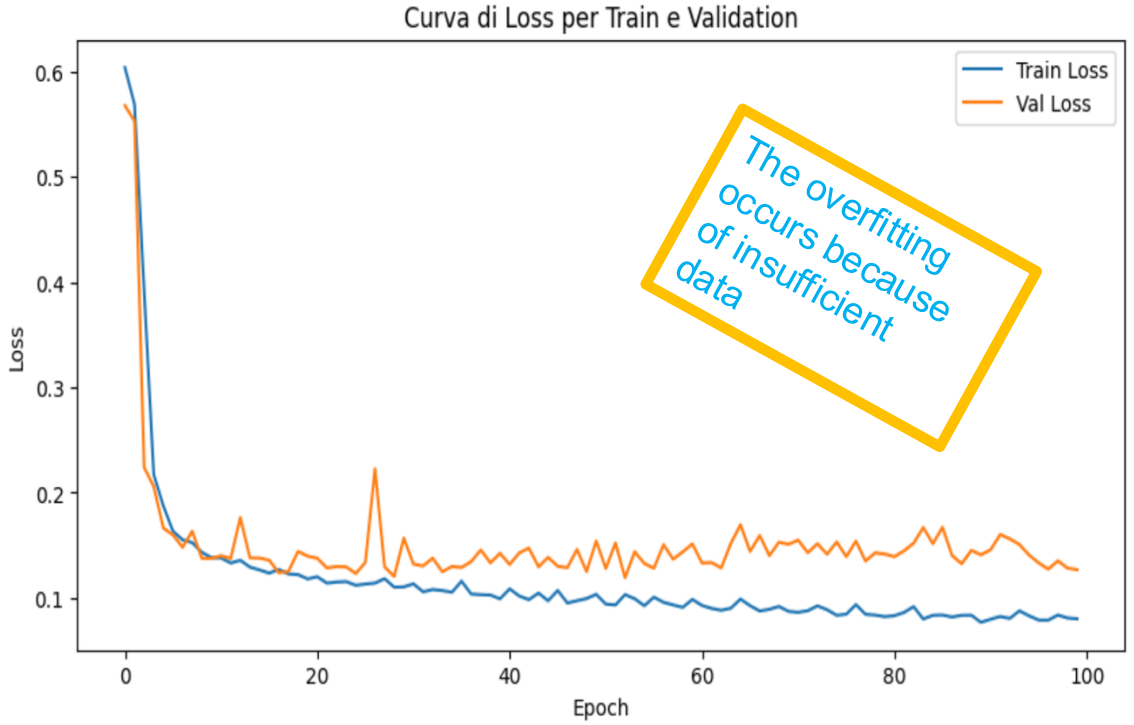
Simulations, choice of features and transformer's architecture





For each event, we record the following information: energy deposition and the X, Y, Z coordinates of the activated pixel

Transformer's results



Test Accuracy: 0.9455
Test Precision: 0.9258
Test Recall: 0.9681

The results have been reported in PDP paper

We are working on the parallelization of the codes on several GPUs and exploring the possibility of using ensemble learning.

Next steps

•**Increase statistics**: collect a larger amount of simulated data to improve the statistical significance and reliability of the results.

•**Extend the dynamic range of the simulation**: broaden the range of energy range simulated to ensure the model captures a wider variety of scenarios and events.

•**Integration with tracking**: combine the simulation data with tracking algorithms to enhance the precision and completeness of the analysis.

Cascade fundings

Moreover, we are also involved in the cascade funding projects under the supervision of Fabio Gargano:

- AI-Legs in collaboration with *Università Parthenope*
- LEGIMAC in collaboration with *Nuclear Instruments*
- GRAIL in collaboration with *UniMarconi*

All projects focus on the identification of gamma rays in space applications. Each of them addresses a different issue and seeks to solve it with advanced Artificial Intelligence techniques. All projects include in the final stages the porting of the codes to FPGAs or low-power GPUs.

Thank you for your attention