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

ESPAI

Enhancing Signal Purity with Artificial Intelligence in X-band Telescopes

R. Anfuso, A. Barca, S. Calì, V. Del Zoppo, L. Naso

Spoke 3 III Technical Workshop, Perugia 26-29 Maggio, 2025

Team Koexai

https://koexai.com

Outline

- Brief introduction about ESPAI astrophysical landscape
- Project overview
- What has been done so far
- Next steps
- Conclusions

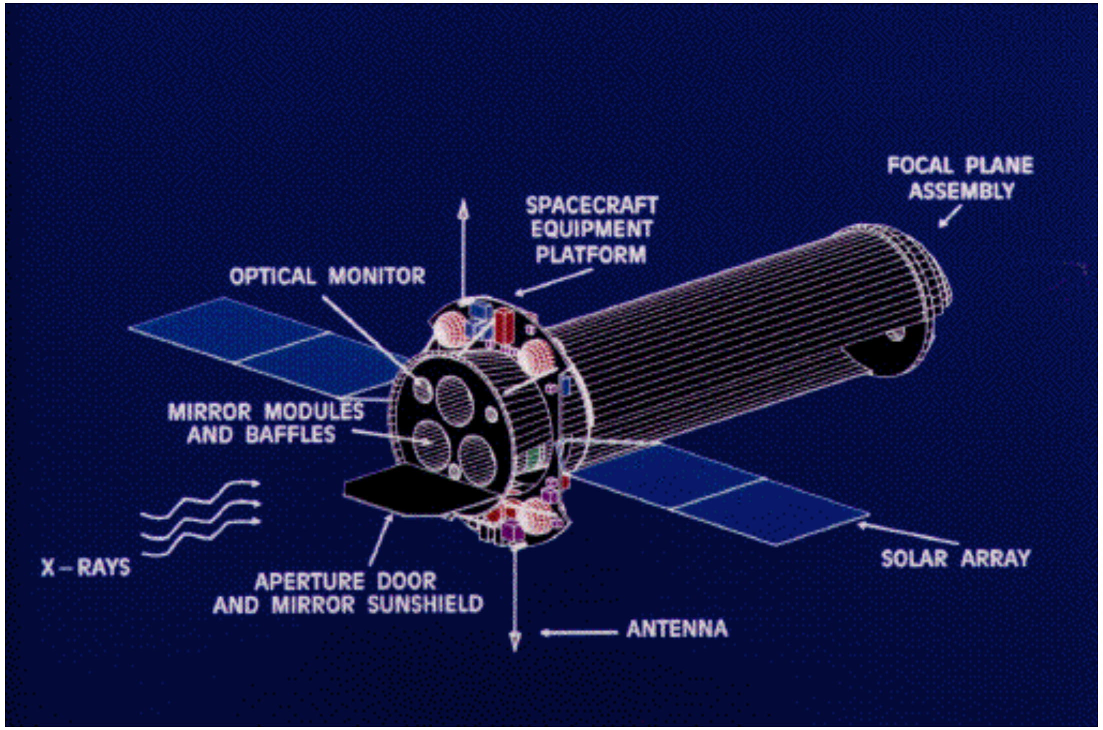
Astrophysical Framework

- X-rays allow to observe a vast set of astrophysical objects exploring the high-energy processes that describe them
- Nowadays, several space missions study this interesting field like the Chandra X-ray Observatory, Swift and XMM-Newton
- Observations can be contaminated by various kinds of noise
- Among the various sources of astrophysical background affecting X-ray astronomy, solar flares are one of the most relevant
- Data taken during solar flare activity are usually not usable for physical analysis and therefore discarded
- Attempt to recover a significant part of the discarded data, by removing the solar flare contamination

Project Overview

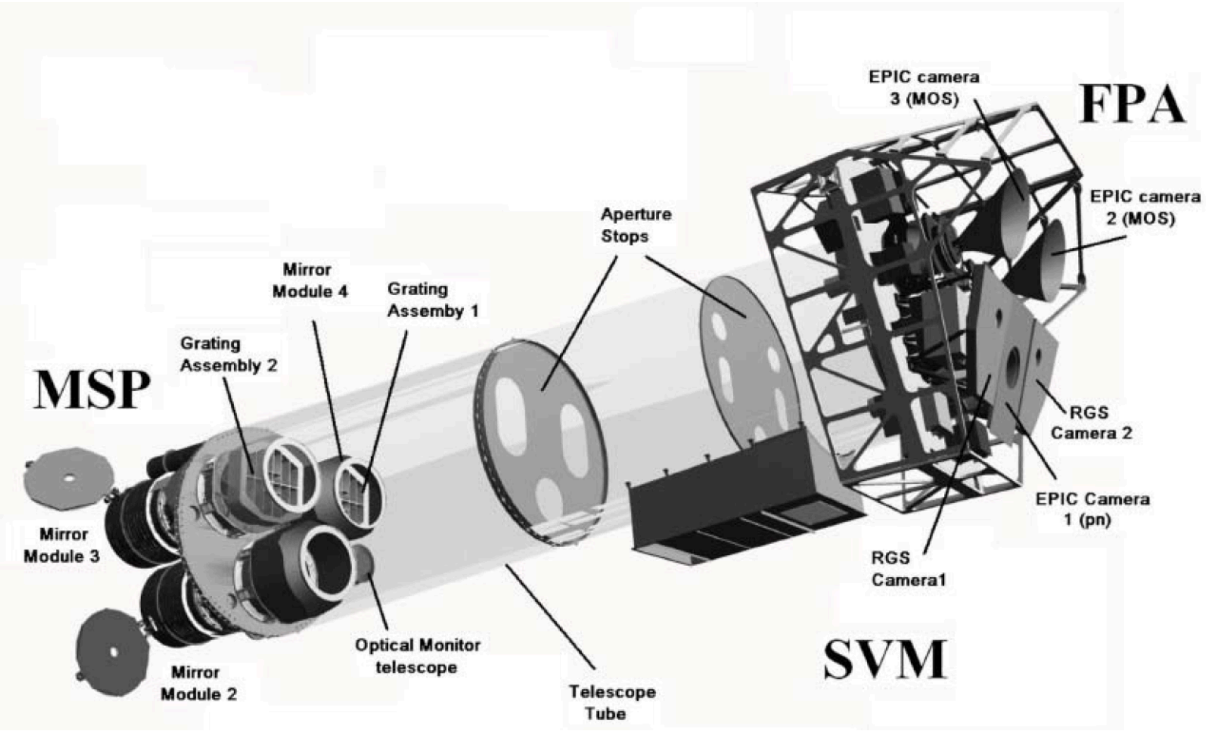
The ESPAI project aims to develop Deep Learning and Generative AI solutions to remove solar flares contamination from signal observations

- Classification models to identify solar flare photons
- Gen AI models to generate synthetic solar flare datasets
- Focus on removing the solar flare contamination from the data from the XMM-Newton satellite
- Project lifespan: from Dec 2024 to Nov 2025



XMM-Newton Dataset

- Focus on the data from the EPIC cameras
- **Key variables:** energy, time of arrival, X and Y coordinates of the photon, pattern and quality flags (FLAG, PATTERN)
- Extensive EDA performed on the dataset
- SAS analysis pipeline to calibrate data and define the Good Time Intervals (GTIs)
  - Count rate based method to define time intervals non affected by solar flare
  - Bad Time Intervals (BTIs) are flare-dominated but contain small fractions of signal



Classification Model

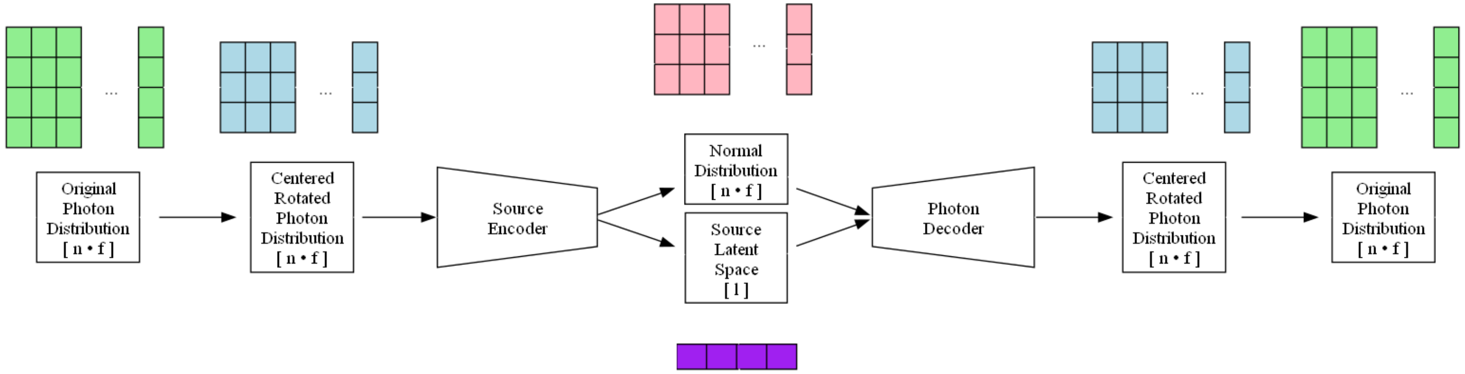
Balanced dataset in terms of clean vs flare-contaminated signal ⇒ classification task

- Architecture based on Multi Layer Perceptron (MLP) trained on a small sample dataset
  - Very simple model providing benchmark performances
- Signal dataset assumed as data inside GTIs, flare dataset as data inside BTIs
  - Non negligible astronomical contamination of the flare dataset
- Accuracy ~ 62% and F1 score ~ 68% on train dataset ⇒ Preliminary but encouraging results:
  - Very limited dataset
  - Contaminated flare dataset ⇒ get cleaner solar flare datasets with Gen AI
- More sophisticated models will be implemented and tested (Gradient Boosting Machines,Transformers, CNNs, …)

Gen AI Model

Generate high-purity solar flare dataset with Gen AI models

- Defined the architecture of the DeepLearning model
  - PointNet architecture ideal to deal with the generation of several features for each photon
- Several models have been considered; AutoEncoders (AE) are the first choice
  - Encode key dependencies of the input features into a lower dimensionality latent space
  - Deterministic latent space where each input is mapped to a fixed point in latent space



Next Steps

- Set up the SAS analysis pipeline on the ISCRA-C cluster and run it over a large part of the XMM-Newton dataset
- Improve MLP model performances and implement more sophisticated ones
- Obtain a high-purity sample of solar flare to train our Gen AI model
  - Datasets with high-intensity solar flare contamination
  - Datasets from sky regions with only few known point-like sources
- Fully implement Gen AI model to produce a high-purity solar flare dataset and train the classification model on it

Conclusions

- Studied sample dataset to well understand the domain
  - Dedicated dashboards created to visualize most important features of the dataset
- Analyzed the SAS pipeline to apply to raw data
- Requested and just obtained ISCRA-C computational resources to train our Deep Learning algorithms
- Implemented a MLP to attempt the classification of signal vs solar flare in a sample XMM-Newton observation
- Fully designed architecture of Gen AI models to create a synthetic, high-purity solar flare dataset

Thanks

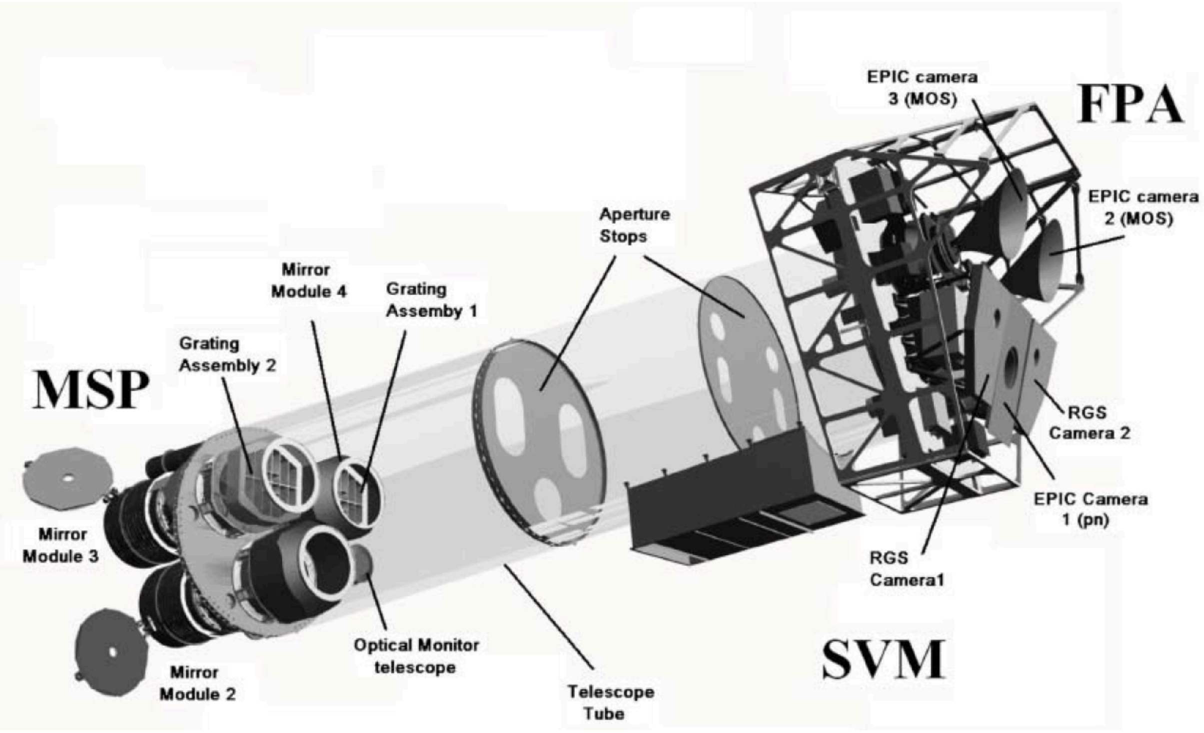
Contact: stefano.cali@koexai.com

Backup

The XMM-Newton Telescope

The XMM-Newton spacecraft was launched in 1999 by the European Space Agency (ESA)

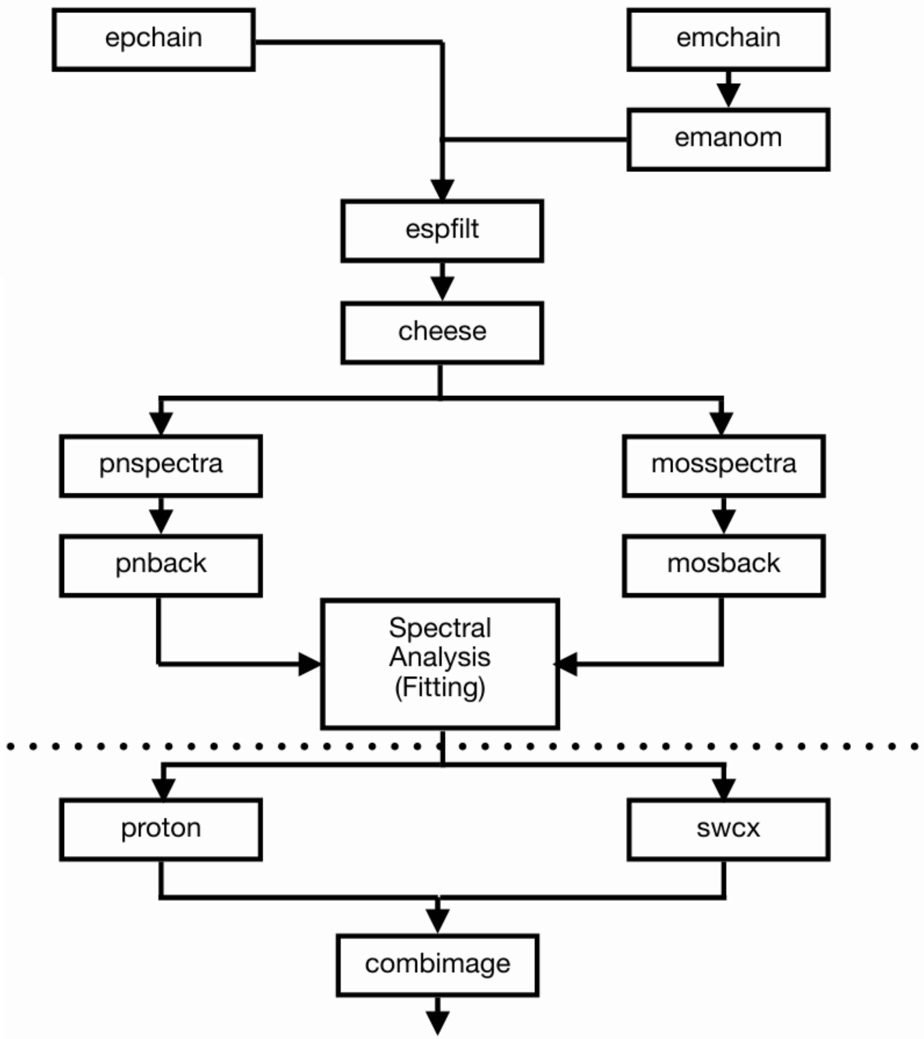
- Equipped with six instruments capable of working at the same time:
  - 3 EPIC (European Photon Imaging Camera) cameras capturing X-rays images and spectra
  - 2 RGS (Reflection Grating Spectrometer) providing high-resolution X-ray spectroscopy
  - 1 OM (Optical Monitor) performing multi-wavelength observations (170-650 nm)



The XMM-Newton Telescope

XMM-Newton analysis pipeline uses SAS to process the Observation Data Files (Raw Data)

- Basic tasks: EpChain/EmChain (calibration), Espfilt (time filtering outside of GTIs), Cheese (point-like sources removal), PnSpectra + PnBack (Quiescent particle background removal)
  - Interested in running only the first tasks of this pipeline



Multi Layer Perceptron

MLP architecture optimized for binary classification on a single sample observation

- Based on three fully connected layers with batch normalization and dropout
- Leaky ReLU used as activation function to avoid the dead ReLU issue
- Trained with a standard BCEWithLogits loss function for binary classification tasks
- Metric used: Accuracy, F1 score, precision and recall