Neural Architectures and Data Processing Pipelines for Irradiation Experiments: from the Automatic Assessment of Proposals to the Monitoring of the Beam Quality

Jaroslaw Szumega, jaroslaw.szumega@cern.ch

On behalf of the team:
Jaroslaw Szumega, Lamine Bougueroua, Blerina Gkotse, Pierre Jouvelot, Federico Ravotti

ABSTRACT

Irradiation facilities, infrastructures for assessing devices and material radiation-hardness, face a variety of challenges, from the management of the experiment-selection process to the monitoring of the beam quality they need. While addressing vastly different issues, the answers may be found in carefully engineered **Machine Learning** and **Artificial Intelligence** (AI) solutions.

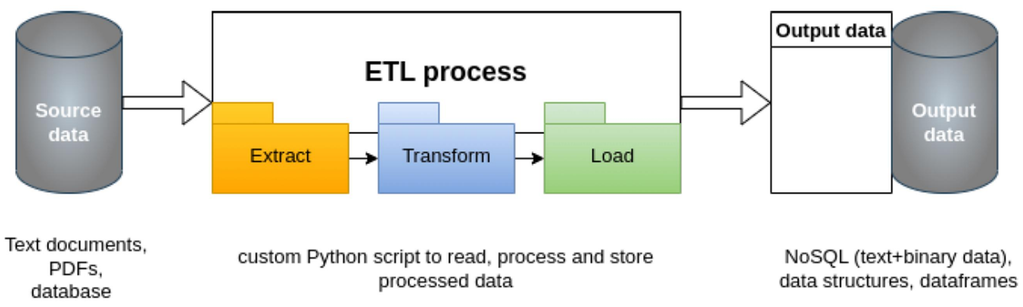
The applications of AI models in **High-Energy-Physics** (HEP) data analysis are well-established, in particular with neural networks and deep-learning algorithms. We suggest that the recent advances in **Natural-Language-Processing** techniques such as transformer architectures may be used for the experiments' proposals assessment and the development of new attention-based monitoring and anomaly detection tools used during their execution. We provide supporting evidence for our approach by describing 1) how we help assess HEP-related scientific proposals within the **RADNEXT** EU-project and 2) how we monitor and evaluate the transverse beam profile quality in real-time at the CERN IRRAD facility in the **EURO-LABS** EU-project.

ETL PROCESS

Extract, Transform, Load (ETL): A standardized workflow of data pre-processing, starting from its raw form up to a structure suitable for processing. This process is broken up into three steps that:

- **extract** the raw data from the initial storage;
- **transform** it into a useful format;
- **load** it into memory.

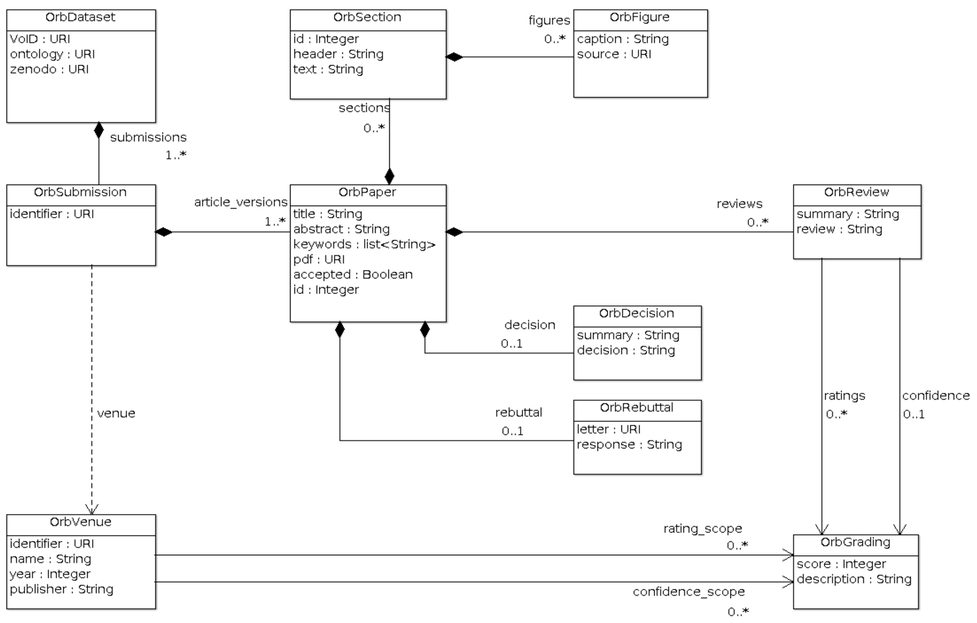
This modular architecture allows to process the multimodal data coming from heterogenous sources.



OPEN PEER REVIEW DATA

Open Peer Review (OPR) is a new peer-review model that provides transparency in the process of scientific publication assessment.

The aim is to build upon this data to create a knowledge representation that will support RADNEXT TA in various ways, including support to users at the time when they are submitting their experimental proposal up to the entire evaluation process.



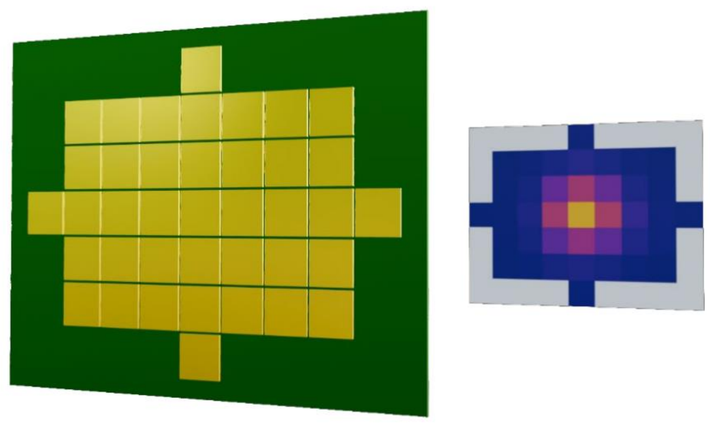
A new Open Review-Based dataset called ORB was created in order to facilitate future work in the domain of automatic assessment of scientific experiment proposals.

This highly structured data is meant to provide a reusable resource that will be accessible and usable for various tasks in the rapidly evolving field of NLP.

BEAM PROFILE MONITOR

The IRRAD beam profile is monitored by a BPM sensors with 40 channels are connected to an upgraded DAQ system.

This new readout electronics measures **currents in the 1-100 nA range** and provides a sampling **rate of up to 1 kHz** that allows the signal processing of data coming from four BPMs in parallel, totalling in **160 channels per DAQ unit**.

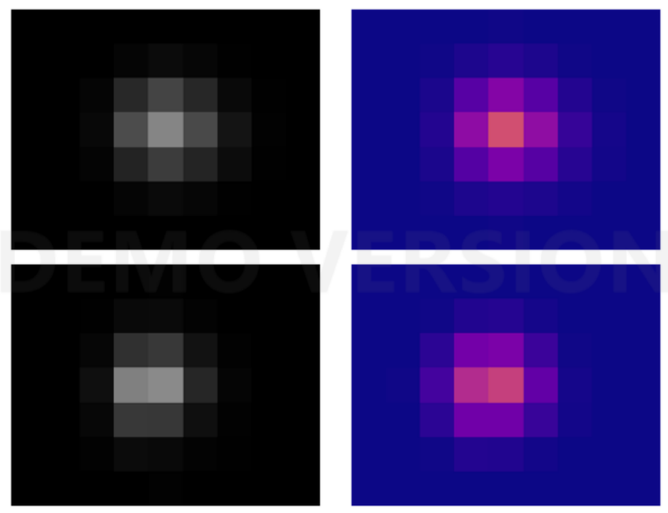


A model of the BPM-sensor printed circuit board (PCB) on the left-hand side, confronted with an example of measurement data represented as an image (right-hand side).

DATASET OF BEAM PROFILES

The dataset is composed of images representing centred- and off-centred beam profiles. They are a representation of **beam monitoring data taken during IRRAD run in year 2024**.

Currently, this dataset consists of around **6,700 images of desirable, good-quality beam profiles** (i.e., aligned on the central BPM pad) and more than **2,000 examples of off-centred ones**. This set of samples is clearly imbalanced, due to the nature of operational beam quality, which (thankfully) stays within acceptable ranges most of the time.

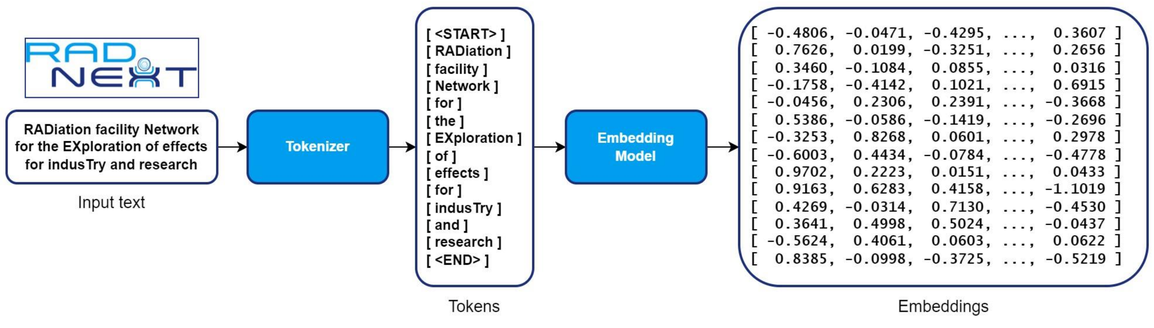


The BPM recorded data are visualized in the form of low-resolution images, where each pixel and its greyscale brightness correspond to the value measured at each individual pad.
To provide a better visualisation, a colour-map may be assigned (right-hand side). Top half of the figure shows a well-centered beam profile, while the bottom - an off-centred beam.

NATURAL LANGUAGE PROCESSING TECHNIQUES

Natural Language Processing (NLP) is a subfield of Machine Learning (ML) and linguistics. It provides the tools that enable the processing of natural language – human-readable texts.

One of the tasks is to work with such documents as scientific texts of experiment proposals. To provide assistance to users and support to experts, semantic relationships need to be properly represented and processed. NLP techniques are the core of such processing.

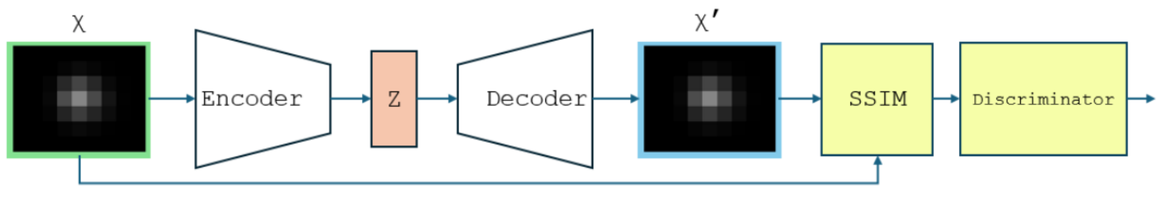


The first stage of the ML pipeline uses NLP to transform texts into "embeddings" - vectors of numbers. Once computed, they are processed further, as as only the numerical representations are suitable for ML tasks.

ANOMALY DETECTION

The Convolutional Autoencoder (CAE) is a specific type of autoencoder capable of learning patterns and specific features that are present in a given set of images. Forthe learned features, image reconstruction is possible as well.

The original data, based on 40 BPM channels, are treated as images represented by matrices of 7×9 pads. Structural Similarity Index Measure (SSIM) is a perception-based metric designed to capture the change in structural information, luminance and contrast. It compares the local patterns of pixel intensities.



The architecture of the anomaly-detection solution. Both the input (X) and reconstructed (X') images are used to calculate the similarity metric. Then the discriminator outputs the decision, based on the metric value and a threshold that was experimentally established.

www.radnext.web.cch

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101008126.

SMARTHEP
Edge Machine Learning School
23.09.2024

https://web.infn.it/EURO-LABS/

This project has received funding from the European Union's Horizon Europe Research and Innovation programme under grant agreement No 101057511.