

# **ICALEPCS 2023: 3rd Data Science and Machine Learning Workshop**

Saturday 7 October 2023 - Saturday 7 October 2023

CENTURY CITY CONVENTION CENTRE 4 Energy Lane, Century  
City, Cape Town

## **Book of Abstracts**



# Contents

Workshop Introduction . . . . .	1
Lunch Break . . . . .	1
Wrap-up . . . . .	1
Common Problems in Early Stage Projects at the ISIS Neutron and Muon Source . . . . .	1
Neural Networks for Anomaly Detection in LINACs, Injectors, and Transfer Lines . . . . .	1
A Potential of Use of Language Processing in Accelerator Control Systems . . . . .	2
Addressing protein serial crystallography 36 GB/s data-rate challenge with FPGAs and GPUs . . . . .	2
Use of Machine learning for Denoising Beam Profile Measurements . . . . .	3
Tutorial I: Linear and Logistic Regression . . . . .	3
Coffee Break . . . . .	3
Tutorial II: Neural Networks, Unsupervised Learning and Advanced Topics . . . . .	3
Discussion Sesssion (only if enough interest) . . . . .	4
Additional presentations and/or discussions . . . . .	4



1

## Workshop Introduction

**Author:** Manuel Gonzalez Berges<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Author:** manuel.gonzalez@cern.ch

2

## Lunch Break

3

## Wrap-up

**Project presentations/demos / 4**

## Common Problems in Early Stage Projects at the ISIS Neutron and Muon Source

**Author:** Kathryn Baker<sup>None</sup>

**Corresponding Author:** k.baker@stfc.ac.uk

At the ISIS Neutron and Muon Source, we are still relatively early on in our pursuit to integrate machine learning into the operations of the accelerator. Consultation with various teams across the accelerator has highlighted three key areas where machine learning can be leveraged most effectively, namely fault diagnosis and prediction, the use of virtual diagnostics and intelligent control of the machine. However, in the case of each of these themes we have encountered complications that may limit their development or practical use that we are keen to discuss with other facilities who may have more knowledge and experience mitigating against these issues. Some of these items to be considered include high dimensional feature selection, dealing with highly correlated outputs and how to match models trained on physics simulation with live behavior of the machine.

**Project presentations/demos / 5**

## Neural Networks for Anomaly Detection in LINACs, Injectors, and Transfer Lines

**Author:** Jon Edelen<sup>1</sup>

<sup>1</sup> *Radiasoft*

**Corresponding Author:** jedelen@radiasoft.net

Maximizing up-time of accelerators relies heavily on the ability to detect and diagnose changes in the machine. The application of machine learning for anomaly detection remains a rich area of research.

RadiaSoft has been developing methods for anomaly detection in collaboration with Jefferson Lab, Brookhaven National Lab, and SLAC. Here we provide a survey of recent innovations in anomaly detection for particle accelerators and present results from our recent work. Our studies are focused on the low energy injector at CEBAF, the AGS to RHIC transfer line at BNL, and industrial accelerators for radiotherapy and imaging. We focused on the use of two neural network architectures, inverse models and variational autoencoders. This talk will provide high level context for how these methods are utilized for anomaly detection and results from our studies using both simulation and measurement data.

**Project presentations/demos / 6**

## **A Potential of Use of Language Processing in Accelerator Control Systems**

**Author:** Antonin Sulc<sup>1</sup>

<sup>1</sup> DESY

**Corresponding Author:** antonin.sulc@desy.de

Particle accelerators rely on complex control systems for their operation. As accelerators grow in scale and complexity, developing and maintaining effective control systems becomes increasingly challenging. In this presentation, we will explore the potential for applying natural language processing (NLP) techniques to improve accelerator operations by closely examining the use of textual data.

We will present our applications of NLP algorithms to logbook data from DESY and BESSY. Initial results demonstrate feasibility for using NLP to automatically parse log entries, categorize events, detect problems, and surface important information.

However, challenges remain in handling physics terminology, noisy data, and model generalization. This presentation will provide an overview of how natural language processing can be applied to accelerate logbooks in field of accelerator controls.

**Project presentations/demos / 7**

## **Addressing protein serial crystallography 36 GB/s data-rate challenge with FPGAs and GPUs**

**Author:** Filip Leonarski<sup>None</sup>

**Corresponding Author:** filip.leonarski@psi.ch

Serial crystallography [1] is a technique used at synchrotrons and X-ray free electron lasers to solve protein structures from random still diffraction images of thousands of small crystals. The technique is one of the most data intensive techniques at X-ray facilities. With novel detectors, like the 9 MPixel JUNGFRÄU [2] currently commissioned at the Paul Scherrer Institute, it is possible to acquire a continuous stream of images at 36 GB/s.

Such large data rates challenge the current way images are handled in crystallography, i.e., it is no longer possible to save to disk storage every image and every pixel irrespective of their value for the scientific question [3]. On-the-fly data analysis and compression become a key to sustainable operations of high data rate detectors. Given a very high data throughput, such analysis requires computing accelerators, like field programmable gate arrays (FPGAs) and general-purpose graphical processing units (GPUs).

In this presentation, I will talk about our practical experience from implementing data science methods for on-the-fly analysis on computing accelerators [4]. I will give a practical example of spot finding algorithms that we implemented on GPUs and FPGAs (with high-level synthesis), highlighting differences in both approaches. I will also give an outlook of our early-stage developments in image analysis with machine learning methods.

[1] T. Weinert et al. (2019). *Science*, 365, 61-65.

[2] F. Leonarski et al. (2018). *Nat. Methods*, 15, 799-804.

[3] F. Leonarski et al. (2020). *Struct. Dyn.*, 7, 014305.

[4] F. Leonarski et al. (2023). *J. Synchrotron Rad.*, 30, 227.

## Project presentations/demos / 8

### Use of Machine learning for Denoising Beam Profile Measurements

**Authors:** Javier Martinez Samblas<sup>None</sup>; Manuel Gonzalez Berges<sup>1</sup>

<sup>1</sup> *CERN*

**Corresponding Authors:** javier.martinez.samblas@cern.ch, manuel.gonzalez@cern.ch

Several CERN accelerators are being equipped with Beam Gas Ionization (BGI) profile monitors using high resolution Timepix3 detectors resulting in very powerful and not destructive measurements [1]

The images produced by these detectors contain the signal from ionization electrons as well as noise coming from different sources (mainly beam losses) and other artifacts like noisy pixels or the RF shield.

Several approaches are being studied to remove the noise and the artifacts from the images to improve the beam measurement. The presentation will give an overview of these approaches.

[1] <https://bgi.web.cern.ch/introduction/>

9

### Tutorial I: Linear and Logistic Regression

**Author:** Gianluca Valentino<sup>1</sup>

<sup>1</sup> *University of Malta (MT)*

**Corresponding Author:** gianluca.valentino@cern.ch

10

### Coffee Break

11

### Tutorial II: Neural Networks, Unsupervised Learning and Advanced Topics

**Author:** Gianluca Valentino<sup>1</sup>

<sup>1</sup> *University of Malta (MT)*

**Corresponding Author:** gianluca.valentino@cern.ch

12

## **Discussion Session (only if enough interest)**

**Project presentations/demos / 13**

## **Additional presentations and/or discussions**