Developing an on-site inspection system using machine learning

Author: Vladislav Mitsyk[1], [1]National Research Nuclear University MEPhI

Annotation

An automated system for the analysis of γ spectra for on-site inspections (OSI) within the framework of the CTBT has been proposed and implemented. The system combines physical modeling in Geant4 to generate realistic spectra and machine learning algorithms (CatBoost) to quickly identify radionuclides and estimate the time of a nuclear event. The solution provides high accuracy and speed of analysis, and is integrated into the IDC CTBTO analytical infrastructure.

Keywords: OSI, γ spectrometry, Geant4, machine learning, CatBoost, radionuclide identification.

Introduction

Traditional analysis of γ spectra an on-site inspections (OSI) for nuclear test verification is time-consuming. This work offers an automated solution that combines physical modeling of the decay and interaction of γ quanta in Geant4 with modern machine learning methods for nuclide classification and regression of the "age" of the radioactive mixture. This reduces analysis time, improves identity reliability, and ensures system scalability.

I. Spectrum Modeling in Geant4

The purpose of the simulation is to generate realistic γ spectra from a sample (gauze contaminated with radionuclides) registered by a germanium detector.

Model geometry: includes a detector (germanium crystal in an aluminum case with a beryllium window), a sample (gauze 30×30×0.5 cm) and a stand. Geometry ensures correct particle tracking.

Physical processes: a custom Geant4 physics list is used, including electromagnetic processes (G4EmStandardPhysics), radioactive decay (G4RadioactiveDecayPhysics), and hadron interactions for completeness of the simulation.

Generation of primary particles: a special generator is implemented that randomly selects the nuclide (^{131}I, ^{132}Te, ^{140}Ba, ^{137}Cs) and the decay point inside the gauze volume γ.

Spectrum Shaping: the total energy released in the detector crystal is recorded. The resulting spectrum is subjected to Gaussian smearing to account for the final energy resolution of the real detector.

II. Data Analysis Using Machine Learning

The program processes the experimental spectrum, represented in the form of "channel → number of sample" pairs.

Loading and preprocessing: spectrum data is loaded from a file.

Peak search: photopeaks are detected by the find_peaks algorithm (SciPy).

Energy calibration: based on the peaks found and the reference base of γ-lines, a linear calibration dependence $E = a \cdot n + b$ is constructed using the least squares method.

Radionuclide identification: the trained CatBoostClassifier model matches calibrated peak energies with radionuclides from the database. The results are displayed in the form of a table.

Estimation of the time of the event: the age of the mixture is estimated by the ratio of nuclide activities (e.g. ^{131}I and ^{137}Cs). On the synthetic data generated, the regressor CatBoost is trained, which predicts the time that has elapsed since the formation of the mixture based on the measured ratio of activities.

Conclusion

The developed system demonstrates high efficiency for OSI tasks. The combination of accurate physical modeling in Geant4 and machine learning (CatBoost) allows you to:

1. Automatically identify radionuclides.
2. To reliably estimate the time of a nuclear event using regression.
3. Easy to integrate into your existing monitoring infrastructure (CTBTO).

Work prospects include expanding the database of nuclides, taking into account the background and adaptive learning on streaming data. The system makes a significant contribution to the tools for operational radionuclide monitoring and strengthening the nuclear non-proliferation regime.