PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

**The National Higher School Of Artificial Intelligence**

# Internship Report

---

## Artificial Intelligence-Based Unfolding Strategy for Neutron Spectrum Reconstruction

---


| | |
|---|---|
| **Student:** | MAZROU RAYANE |
| **Year / Program:** | 4th Year Student – Artificial Intelligence |
| **Host Institution:** | Nuclear Research Center of Algiers (CRNA) |
| **Supervisor:** | BOUCHAMA LAZHAR |
| **Training Period:** | September 6 – October 4, 2025 |

# Contents

# Acknowledgements

I wish to express my heartfelt gratitude to my supervisor, **Mr. Bouchama Lazhar**, for his valuable guidance and support throughout this internship. His expertise in the domaine played a vital role in the completion of this work. I would like also to extend my gratitude to **Mr. Mazrou Hakim** who was kindly working as a second supervisor. His suggestions and guidance were of great significance for the advancement of my internship and the creation of this report.

I would also like to thank the **Nuclear Research Center of Algiers (CRNA)**, and in particular the Division of Radiobiology and Dosimetry of Ionizing Radiation for hosting me and offering a pleasant research environment.

Special thanks to the researchers and engineers who took the time to present the center's facilities during my visits. I'm particularly grateful to:

- **Mr. Belhout Khaled**, for his valuable presentation about accelerators, their different types, and how they operate.
- **Mr. Toumert Idir**, for his insightful explanation on characterization techniques, the machines involved, and how AI can be utilized in that field.
- **Mr. Medjadj Toufik**, for presenting the different activities belonging to the Secondary Standard Dosimetry Laboratory.

Finally, I would like to thank all the staff members and colleagues that contributed to enrich this internship experience.

# Introduction

## 1.1 Algiers Nuclear Research Center Description

Algiers Nuclear Research Center (CRNA) is a government institution dedicated to research and development in nuclear science and technology. Its mission covers nuclear physics, radiological physics, nuclear applications, environmental studies, safety, and radioactive waste management, with strong emphasis placed on operational radioprotection. Training and capacity building in radioprotection, nuclear safety, and related sciences are also core activities of the center.

CRNA is equipped with many facilities such as a Van de Graaff accelerator, a neutron generator, an ion implantation system, and a gamma irradiator. It also possesses advanced laboratories for dosimetry, isotopic hydrology, and radioisotopic analysis providing a strong technical foundation for basic research as well as applied projects.

For me, I was part of the laboratory of ionizing radiation dosimetry which is considered one of the most fondamental infrastructure existed at the CRNA, its principal missions consist of the development of external dosimetry measurement and calculation techniques for radiation protection purposes, developpemnt and qualification of dosimetric system, student supervising, etc.

## 1.2 Context of the Training

The use of Artificial Intelligence (AI) in scientific disciplines extends far beyond traditional applications such as computer vision or natural language processing. In the discipline of nuclear science, a number of challenges take the form of inverse problems, where the goal is to retrieve the physical quantities from available experimental measurements. One of these problems is unfolding of neutron energy spectra from detector counts, which is crucial to dosimetry of radiation, radiological protection, and reactor physics. For me, as an AI student, training within a nuclear research center was an excellent opportunity to explore data science and machine learning abilities on Nuclear physics challenges. Being exposed to real experimental configurations, better measurement infrastructures, and multidisciplinary expertise in the Nuclear Research Center of Algiers (CRNA) was insightful.

## 1.3 Objectives of the Internship

The objectives of this internship are both pedagogical and technical:

- **Pedagogical Objectives:**

- Understand how AI methods can be applied to solve inverse problems in physics.
  - Develop problem-solving skills by applying supervised learning techniques to scientific data.
  - Strengthen my knowledge in data preprocessing, model training, evaluation, and result interpretation.

- **Technical Objectives:**
  - Design and train machine learning models (boosting,stacking,trees...) to solve the dilemma of inverse problem.
  - Apply data preprocessing techniques such as normalization,cleaning, visualization, and interpolation.
  - Test model performance on real test cases to compare the results with related literature.
  - Document the methodology and results in a scientific-style report and presentation.

Through this training, I acquired both theoretical and practical experience, while contributing to the exploration of innovative AI-based solutions for neutron spectrum unfolding.

# Project Implementation

## 2.1 Problem Statement

Inverse problems are common in the physical sciences. In nuclear science, one such inverse problem is the reconstruction of neutron energy spectra from detector count data, a process known as **neutron spectrum unfolding** [1]. This is fundamentally a regression problem in which the goal is to map measured data (detector counts) to the physical quantity (neutron spectrum).

The neutron spectrum unfolding from Bonner Sphere Spectrometer (BSS) counts [2] is described by a system Fredholm integral equations of the first kind, written as:

$$C_i = \int R_i(E)\, \Phi(E)\, dE, \quad i = 1, 2, \ldots, M \tag{2.1}$$

where:

- $C_i$ is the count vector measured by the Bonner sphere number i;
- $R_i(E)$ is the function response of the ith sphere;
- $\Phi(E)$ is the unknown neutron fluence vector

Mathematically, Equation (2.1) has not a unique solution since a finite number of discrete measurements cannot define a continuous function. In order to allow a numerical solution The equation (2.1) has to be transformed at least approximately into the discretised linear matrix equation as follows [3]:

$$C_i = \sum_{j=1}^{N} R_{ij}(E_j)\, \Phi_j(E_j)$$

where:

- $R_{ij}(E_j)$ is the response function of the $i^{\text{th}}$ sphere to neutrons of the energy that corresponds to the $j^{th}$ energy bin;
- N is the number of energy bins;
- $\Phi_j(E_j)$ is the fluence in the $j^{th}$ energy bin.

Nowaday, a lot of unfoding codes, based on variuos mathematical approach, are developed and widely used [4], however, they have some drawbacks that make them less efficient as the complexity in their use, the need of realistic prior informations such as a "default spectrum", as close as possible to the spectrum to be obtained, an analysed and quantified response function, etc [5, 6]

In recent years, the use of the AI-based unfolding methods have been proposed to solve neutron spectrum unfolding problems by researchers. A lot of works were published using different methods such as: Artificial Neural Network, Genetic Algorithms, etc [7, 8, 9]

AI-based regression methods provide an alternative by learning an approximation function that maps data counts $C_i$ to neutron fluence $\Phi(E)$ in a more reliable way. This highlights both the ill-posedness of the unfolding problem and the motivation for using modern machine learning techniques.

## 2.2 Methodology

### 2.2.1 Data Preprocessing

#### 1.Data Collection

The dataset used in this work was constructed from neutron spectrum vectors provided in the IAEA[1] Technical Report Series TRS-403 [10]. From this technical report, around 114 representative spectra covering a wide range of neutron energy distributions were collected.

Since the raw spectra were not uniformly distributed in energy, an interpolation was applied to obtain a uniform discretizations with 10 energy bins per decade. This made the spectra uniform for all spectra. Furthermore, the spectra distribution in the TRS-403 are expressed in lethargy units; to normalize them into standard differential spectra, each bin content was multiplied by $\ln\left(\frac{E_{i+1}}{E_i}\right)$, where $E_i$ and $E_{i+1}$ are the $i^{th}$ bin boundaries.

Once the lethergy unit is removed, the detector count vectors which are the input to the unfolding model were computed. These counts were obtained by multiplying each spectrum with a BSS response function matrix, delivered by EURADOS [11].

At the end of this process, the dataset was composed of:

- **Input:** detector count vectors of length 12
- **Output:** neutron spectrum vectors of length 104

#### Data Cleaning

After the construction of the raw dataset (114 spectra extracted from TRS-403), the cleaning process was performed. The goal of this step was to identify and retain the most representative spectra and exclude noisy, irrelevant, and redundant ones.

To perform this, the entire dataset was visualized by plotting all spectra simultaneously. After careful examination, 54 spectra were non-representative of practical neutron fields and thus removed. As a result, the dataset was reduced from 114 spectra to 60 representative spectra, which form the core basis for the unfolding study.

[1]IAEA: International Atomic Energy Agency

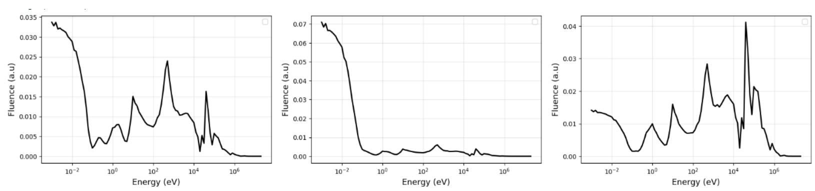



Figure 2.1: Examples of spectra removed during the cleaning process.

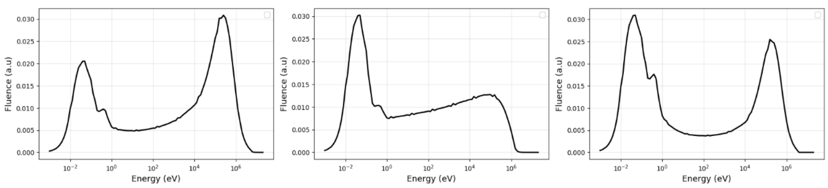



Figure 2.2: Examples of representative spectra retained in the final dataset.

#### Data Augmentation

After cleaning, the dataset contained only 60 representative spectra, which was insufficient for training. To encounter this problem data augmentation was applied, generating five new samples from each original spectrum. The following techniques were used simultaneously:

- **Gaussian Noise:** Adds random fluctuations to simulate detector noise.
- **Smoothing:** Reduces sharp variations to mimic lower-resolution spectra.
- **Peak Scaling:** Adjusts peak heights to reflect changes in intensity.
- **Multiplicative Noise:** Applies proportional variations to simulate count fluctuations.

This process extended the dataset while maintaining realistic spectral characteristics.

#### Dataset Expansion

In addition to augmentation, the dataset was further expanded by adding more spectra from the IAEA TRS-403. Initially, only accelerator and power reactor spectra were selected, as they were the primary focus of this study. Later, 40 more representative spectra from other source types were included to increase data quality and the training efficiency.

Thus, the dataset size increased from 60 to 100 representative spectra, which will introduce more variability to the dataset.

### 2.2.2 Testing Dataset

The model was validated against five test cases, which are realistic and diverse neutron fields relevant to radiation protection and dosimetry, used during the 2017 international comparison exercise on neutron spectra unfolding in Bonner spheres spectrometry, organized by EURADOS [11]. These are two points in an medical linear accelerator (LINAC) field (one at the entrance of the maze, and the other near the treatment head), an 241Am-Be based workplace field, a calibration room with moderated 241Am-Be source, and a Skyshine field around a nuclear plant.

Together, the five test cases are important in so far as they span thermal, epithermal, and fast neutron regimes, and are both controlled reference fields and complex workplace settings. This diversity ensures that the unfolding model testing does not take place in a small range of energy but does show its stability under different practical conditions.

### 2.2.3 Model Architecture and Evaluation Metrics

The unfolding task was modeled as a multi-output regression problem, where the input is a count vector of 12 detector channels and the output is a neutron spectrum vector of 104 energy bins. Each spectrum was unit normalized before training so that the model can perform well on inputs with different units.

#### Model Architecture

We applied gradient boosting regression trees using XGBoost [12], Since XGBoost natively handles single-output regression, this approach implements a model for each of the 104 energy bins then we combine the result in the output.

To ensure optimal performance, the model parameters were tuned using `Optuna` [13], an automatic hyperparameter optimization framework. The search space included the number of estimators, maximum depth, learning rate, subsample ratio, and column sampling ratio. The final selected parameters were:

- Number of estimators: 300
- Maximum depth: 6
- Learning rate: 0.05
- Subsample: 0.8
- Column sampling by tree: 0.8
- Tree method: histogram-based growth

#### Evaluation Metrics

Model performance was primarily assessed using the coefficient of determination $R^2$ between predicted and true spectra. To provide domain-relevant insights, evaluation was also performed over physically meaningful energy regions:

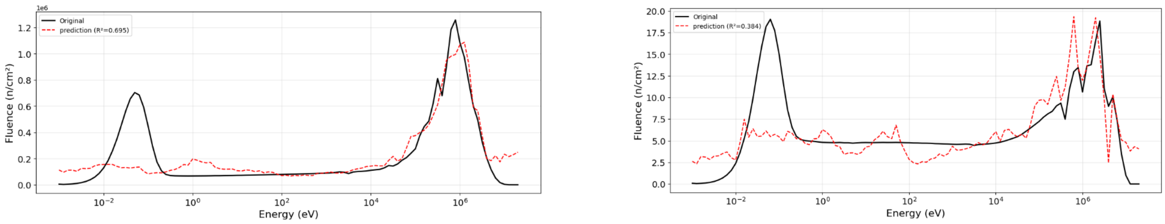
- **Thermal region:** $E < 0.4$ eV
- **Epithermal region:** $0.4 \leq E < 10^5$ eV
- **Fast region:** $E \geq 10^5$ eV
- **Total spectrum:** full 104-bin range

For each spectral region (thermal, epithermal, and fast), the sum of spectrum values of the predicted spectra was compared with the one of the reference . The relative difference (in %) was then calculated. This regional analysis provides a more detailed assessment of the model's accuracy, highlighting its performance not only on the overall spectrum but also within the most physically significant energy domains.

## 2.3 Results and Discussion

The performance of the proposed approach was evaluated across four preprocessing stages: raw datset, cleaned dataset, augmented dataset, and expanded dataset. For each stage, two representative test cases were selected and plotted side by side to illustrate the model's predictive ability in different scenarios. The corresponding plots (Figures 2.3–2.6) highlight how the predictions improves as preprocessing advances. Finally, a summary table (Table 2.1) reports the performance on two important test cases.

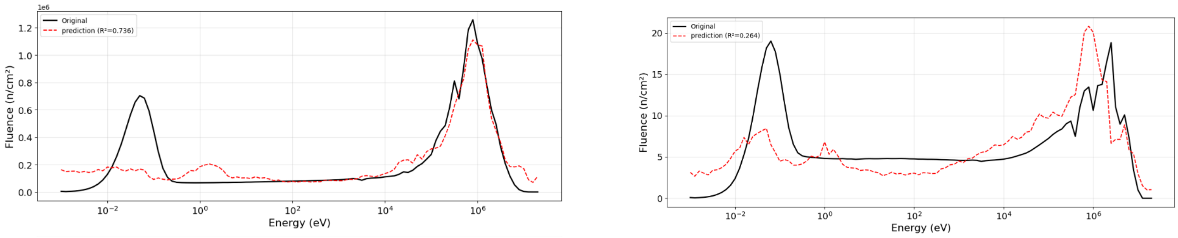
According to the raw dataset, the model performs poorly in prediction with $R^2$ values of 0.69 and 0.38 for the two reference cases, which can be seen from Figure 2.3 While the overall spectral shape is roughly captured, the similarity with the reference is low, with significant deviations in both the thermal and epithermal regions (lower and medium energies). we can see that the peaks of the fast region(higher energies) are recreated to some extent, but the peaks of the thermal region are completely missed. Such failures point towards noisy and non-representative spectra included in the raw data.




(a) (b)

Figure 2.3: Comparison of the predicted spectra (dotted red line), gotten using raw dataset, with the corresponding test cases (continuous black line) : the LINAC (a) and Skyshine (b) scenarios

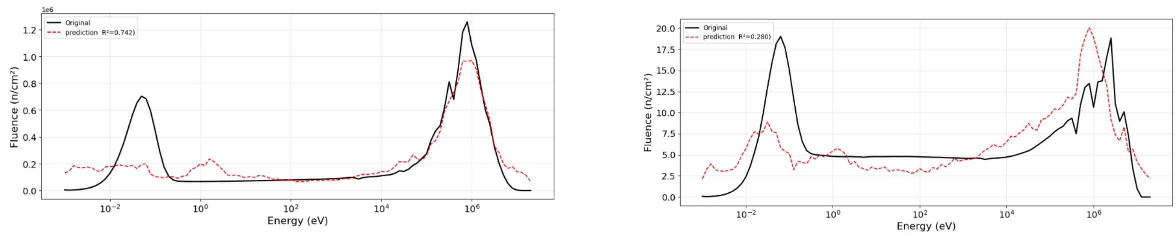
After data cleaning, model performance shows some improvements in certain aspects, which can be seen from Figure 2.4. The removal of inconsistent spectra allows for slightly better matching especially in the epithermal region(medium energies) where we can see more stable prediction compared to the fluctuations seen in the previous stage. however,the predictions in the fast region are worst in some cases compared to the previous step ,while the thermal region(low energies) didn't saw any improvements. This indicates that although data cleaning reduces noise, the reduced dataset size restricts generalizability of the model. So, further augmentation is required to establish balance between data quantity and quality.




(a) (b)

Figure 2.4: Comparison of the predicted spectra (dotted red line), gotten using cleaned dataset, with the corresponding test cases (continuous black line) : the LINAC (a) and Skyshine (b) scenarios

From the augmented dataset, the model has a better diversity in training but no big changes where observed as we can see in Figure 2.5, because augmentation introduces artificial variability not always reflective of the true spectral distribution, thereby explaining the poor performance on more challenging spectra (e.g., $R^2 \approx 0.29$). This implies that while augmentation helps in mitigating the limitations of a small dataset, its success is greatly dependent on the quality of generated samples which emphasizes the necessity to collect more representative samples.

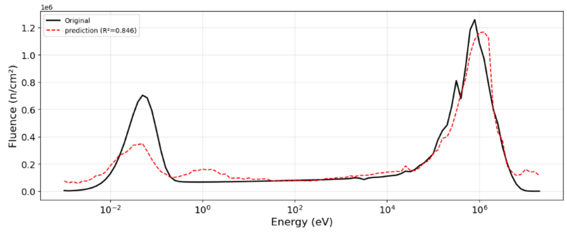



(a) (b)

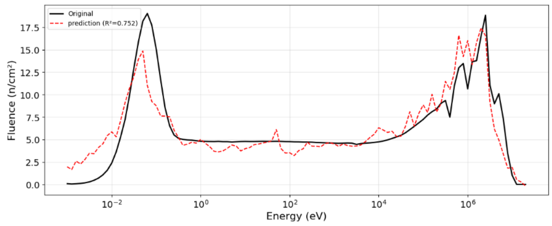
Figure 2.5: Comparison of the predicted spectra (dotted red line), gotten using augmented dataset, with the corresponding test cases (continuous black line) : the LINAC (a) and Skyshine (b) scenarios

With the expanded dataset, the model performs best compared to previous stages, as can be seen from Figure 2.6. The predictions match very well with the original spectra in both the thermal and fast energy ranges with $R^2$ values of 0.84 and 0.75 for the two reference cases. This expanded dataset provides a more dense and well-balanced training base as it contains more qualitative samples that represent the general form of a clean spectrum leading to precise predictions particularly in regions of strong energy as well as a big improvement in the thermal region(low energies) where it captured the peaks significantly according to other stages.

Table 2.1 provides a quantitative summary of the model's performance on the linear medical acceleartor case and skyshine nuclear reactor case across the four stages . The progression from raw to expanded datasets shows a consistent reductions in relative errors, confirming the effectiveness of preprocessing, augmentation, and dataset expansion.



(a)



(b)

Figure 2.6: Comparison of the predicted spectra (dotted red line), gotten using expanded dataset, with the corresponding test cases (continuous black line) : the LINAC (a) and Skyshine (b) scenarios

| **Method** | **Rel. Error Thermal (%)** | | **Rel. Error Epithermal (%)** | | **Rel. Error Fast (%)** | | **Rel. Error Total (%)** | |
|---|---|---|---|---|---|---|---|---|
| | **LINAC** | **SKYSHINE** | **LINAC** | **SKYSHINE** | **LINAC** | **SKYSHINE** | **LINAC** | **SKYSHINE** |
| Raw Data | 46.46 | -29.17 | 35.97 | -4.99 | 9.44 | 18.27 | 0.33 | -3.94 |
| Cleaned Data | -37.34 | -22.87 | 38.29 | -0.023 | -0.009 | 17.23 | -1.59 | -0.57 |
| Augmented Data | -29.35 | -21.95 | 37.48 | 1.27 | -1.03 | 19.23 | -0.16 | 0.84 |
| Expanded Data | -24.79 | 2.47 | 24.07 | -3.08 | 3.15 | 3.04 | 0.27 | 0.42 |

Table 2.1: Relative difference (%) between the thermal, epithermal and fast fluence values provided by the model and the corresponding reference values for the LINAC and Skyshine scenarios.

## 2.4 Web Deployment of the Model

To make the trained model accessible and easy to use, it was deployed as a web application using the Flask framework and hosted on Render at: `https://spectrum-unfolding-app-3.onrender.com/`. This deployment allows users to upload detector count data in CSV format, run predictions directly through the browser, visualize the reconstructed spectra, and download results for further analysis.

### 2.4.1 Homepage and Upload Interface

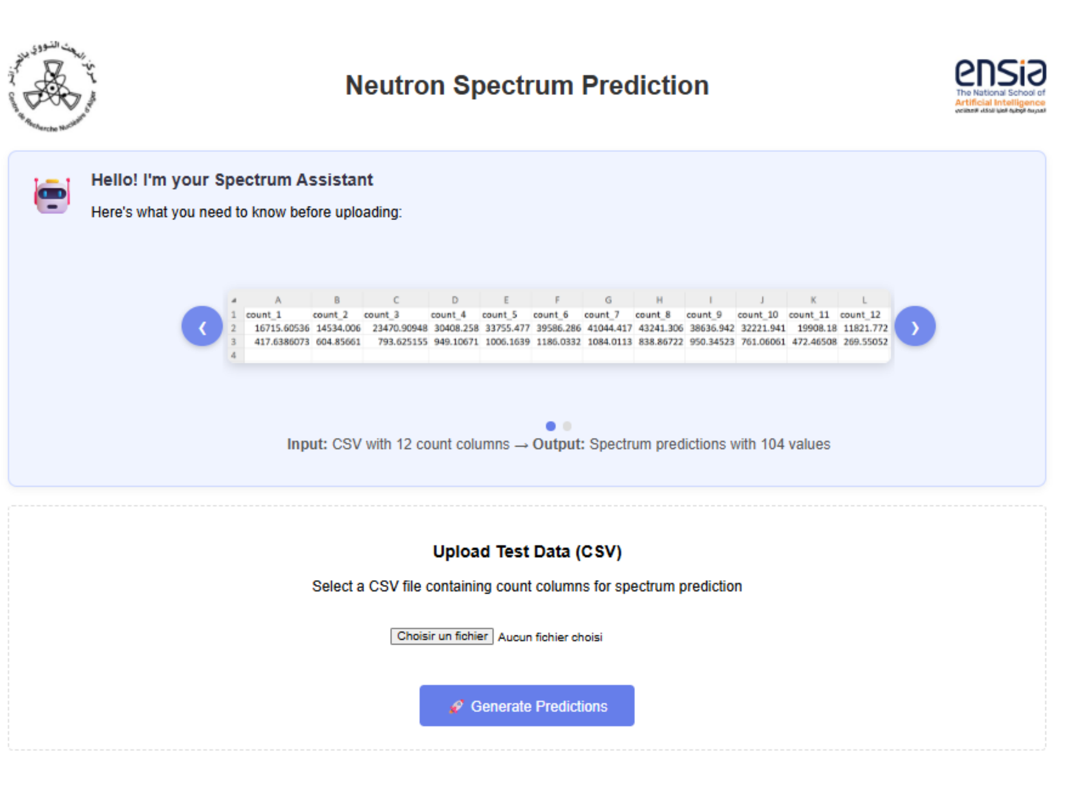
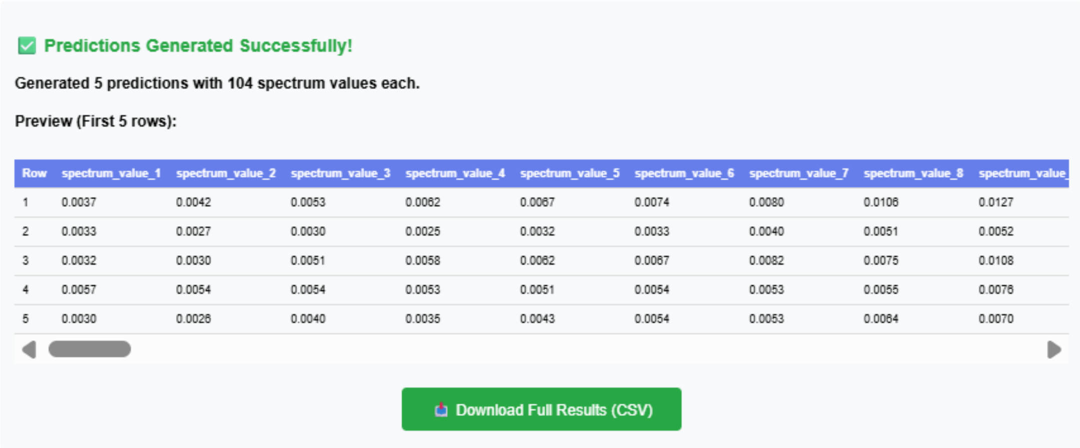



Figure 2.7: Homepage of the web application where users can upload CSV files containing detector counts.

When the user opens the application, they are greeted with a simple homepage containing an upload form. The required input format is clearly indicated: a CSV file containing detector counts. Once the file is selected, the user clicks the *"Generate Predictions"* button to initiate spectrum unfolding.

### 2.4.2 Predictions and Preview



✅ Predictions Generated Successfully!

Generated 5 predictions with 104 spectrum values each.

Preview (First 5 rows):

| Row | spectrum_value_1 | spectrum_value_2 | spectrum_value_3 | spectrum_value_4 | spectrum_value_5 | spectrum_value_6 | spectrum_value_7 | spectrum_value_8 | spectrum_value_ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.0037 | 0.0042 | 0.0053 | 0.0062 | 0.0067 | 0.0074 | 0.0080 | 0.0106 | 0.0127 |
| 2 | 0.0033 | 0.0027 | 0.0030 | 0.0025 | 0.0032 | 0.0033 | 0.0040 | 0.0051 | 0.0052 |
| 3 | 0.0032 | 0.0030 | 0.0051 | 0.0058 | 0.0062 | 0.0067 | 0.0082 | 0.0075 | 0.0108 |
| 4 | 0.0057 | 0.0054 | 0.0054 | 0.0053 | 0.0051 | 0.0054 | 0.0053 | 0.0055 | 0.0076 |
| 5 | 0.0030 | 0.0026 | 0.0040 | 0.0035 | 0.0043 | 0.0054 | 0.0053 | 0.0064 | 0.0070 |

Download Full Results (CSV)

Figure 2.8: Prediction results with preview of the unfolded spectrum values (first rows of the CSV).

After uploading the dataset, the backend normalizes the detector counts, passes them through the ensemble of trained XGBoost models, and generates the predicted neutron spectra. A preview of the output table is displayed directly on the webpage with the possibility to download the generated output file as csv.

### 2.4.3 Spectrum Visualization

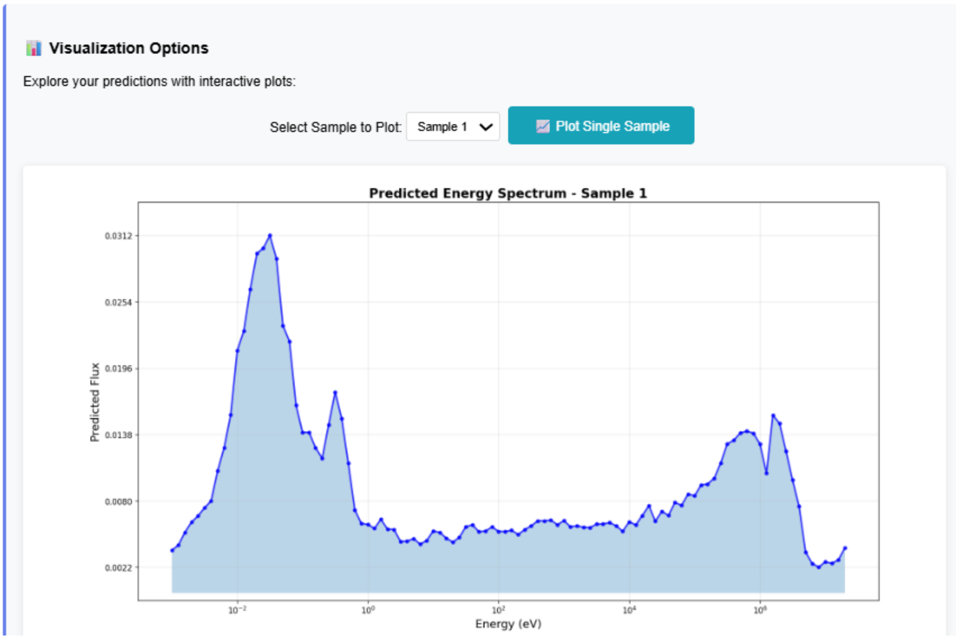



Figure 2.9: Predicted energy spectrum for a single sample.

The interface provides interactive links to visualize the unfolded spectra. Users can generate plots for each single sample. These figures are dynamically created with Matplotlib on the server side and displayed inline in the browser.

# Challenges and Learning Outcomes

## 3.1 Challenges Encountered

Throughout the course of this internship, several challenges were faced, both technical and organizational:

- **Technical Challenges:** Handling nuclear physics datasets was very different from the datasets I had encountered before. The small dataset size, noise in measurements, and the complexity of inverse problems made model design more difficult. Preprocessing steps such as cleaning, augmentation, and normalization were also non-trivial and required significant discussion with domain experts.
- **Knowledge Gaps:** Initially, I lacked in-depth knowledge of neutron physics, spectrum unfolding methods, and dosimetry principles. Bridging this gap required an intensive learning effort, through reading reference materials, and constant interactions with my supervisors.

## 3.2 Skills and Knowledge Acquired

Despite the difficulties, this internship was an enriching experience that allowed me to develop both technical and soft skills:

- **Technical Skills:**
  - Data preprocessing (cleaning, augmentation, interpolation, normalization).
  - Implementation and optimization of machine learning models (XGBoost, stacking, hyperparameter tuning).
  - Model evaluation using physics-oriented error metrics ($R^2$, relative errors by energy range).
  - Deployment of an AI model into a web interface using Flask.
- **Soft Skills:**
  - Teamwork in a multidisciplinary environment combining AI and nuclear physics.
  - Scientific writing and reporting in a formal style.
  - Communication skills when discussing technical concepts with experts outside my core domain.
  - Time management and problem-solving under tight deadlines.

# Conclusion and Perspectives

## 4.1 Conclusion

This internship provided a unique opportunity to apply Artificial Intelligence techniques to a challenging problem in nuclear science: neutron spectrum unfolding. The project covered the entire pipeline, from dataset preparation (cleaning, augmentation, and expansion), through model design and optimization, to deployment as a user-friendly web application. The results obtained demonstrate that machine learning models, especially gradient boosting ensembles, can achieve promising accuracy and robustness across multiple test cases.

Beyond technical outcomes, this internship allowed me to deepen my understanding of nuclear applications, strengthen my research and teamwork skills, and gain valuable exposure to interdisciplinary scientific collaboration.

## 4.2 Perspectives

The work conducted here opens the way for several future directions:

- **AI in Neutron Dosimetry:** Developing hybrid models that combine AI predictions with physical constraints for higher accuracy and interpretability.
- **Nuclear Safety Applications:** Extending unfolding models for real-time monitoring of workplace neutron fields and accident scenarios.
- **Data-Driven Modeling:** Building larger datasets through international collaborations (e.g., EURADOS, IAEA) to train more generalizable models.
- **Predictive Maintenance:** Using AI to analyze signals from radiation detection instruments and anticipate equipment failures.
- **Medical Applications:** Applying similar unfolding strategies to optimize radiation protection in accelerator-based therapies.

Overall, this internship confirmed the importance of AI for advancing nuclear research and safety, and it encouraged me to pursue further exploration of interdisciplinary applications where artificial intelligence meets physical sciences.

# References


[1] V. Blobel, "Unfolding methods in particle physics," 2011.

[2] R. L. Bramblett, R. I. Ewing, and T. Bonner, "A new type of neutron spectrometer," *Nuclear instruments and Methods*, vol. 9, no. 1, pp. 1–12, 1960.

[3] M. Matzke, "Unfolding of particle spectra," in *International Conference Neutrons in Research and Industry*, vol. 2867, pp. 598–607, 1997.

[4] M. Reginatto, "Overview of spectral unfolding techniques and uncertainty estimation," *Radiation Measurements*, vol. 45, no. 10, pp. 1323–1329, 2010.

[5] S. Barros, V. Mares, R. Bedogni, M. Reginatto, A. Esposito, I. F. Gonçalves, P. Vaz, and W. Rühm, "Comparison of unfolding codes for neutron spectrometry with bonner spheres," *Radiation protection dosimetry*, vol. 161, no. 1-4, pp. 46–52, 2014.

[6] O.-R. et al, "A neutron spectrum unfolding computer code based on artificial neural networks," *Radiation physics and chemistry*, vol. 95, pp. 428–431, 2014.

[7] Q.-J. Zhu, L.-C. Tian, X.-H. Yang, L.-F. Gan, N. Zhao, and Y.-Y. Ma, "Advantages of artificial neural network in neutron spectra unfolding," *Chinese Physics Letters*, vol. 31, no. 7, p. 072901, 2014.

[8] J. Wang et al, "Neutron spectrum unfolding using three artificial intelligence optimization methods," *Applied Radiation and Isotopes*, vol. 147, pp. 136–143, 2019.

[9] C. Cao, Q. Gan, J. Song, P. Long, B. Wu, and Y. Wu, "A two-step neutron spectrum unfolding method for fission reactors based on artificial neural network," *Annals of Nuclear Energy*, vol. 139, p. 107219, 2020.

[10] International Atomic Energy Agency, "Compendium of Neutron Spectra and Detector Responses for Radiation Protection Purposes," Tech. Rep. 403, International Atomic Energy Agency, Vienna, 2001.

[11] J. Gómez-Ros, R. Bedogni, C. Domingo, J. Eakins, N. Roberts, and R. Tanner, "Results of the eurados international comparison exercise on neutron spectra unfolding in bonner spheres spectrometry," *Radiation Measurements*, vol. 153, p. 106755.

[12] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

[13] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631, 2019.