

A PHYSICS-GUIDED DIGITAL TWIN FRAMEWORK FOR SYNTHETIC GAMMA-RAY SPECTRUM GENERATION AND VALIDATION IN RADIATION MONITORING APPLICATIONS

Abstract: The limited availability of labelled gamma-radiation data constrains the development and evaluation of machine-learning methods for radiation-monitoring applications, particularly under shielded or otherwise difficult-to-reproduce conditions. This study presents a physics-guided Digital Twin framework integrating Geant4-based Monte Carlo radiation transport, automated scenario orchestration, detector-response aggregation, structured synthetic-data generation, and multiple explicitly separated levels of assessment. Eight radionuclide-labelled source classes were represented by characteristic monoenergetic gamma emissions ranging from 0.051 to 2.610 MeV. Simulations covered four shielding materials—lead, stainless steel, concrete, and polyethylene—at thicknesses of 0, 1, 2, 5, and 10 cm, resulting in 160 source–material–thickness configurations. The simulated detector responses reproduced the expected dependence of photon transmission on energy, shielding material, and absorber thickness. After incorporating counting-statistical uncertainties, the fitted linear attenuation coefficient for 1.25-MeV ^{60}Co -labelled photons in lead was $0.618 \pm 0.042 \text{ cm}^{-1}$, with a 95% confidence interval of $0.535 - 0.700 \text{ cm}^{-1}$. The corresponding NIST XCOM reference value of 0.680 cm^{-1} lies within this interval. A limited detector-level experimental assessment was performed using independently acquired NaI(Tl) spectra. A multi-peak blind spectrum was assigned to material containing the ^{232}Th decay chain before disclosure of the source identity, and the assignment was subsequently confirmed by the data provider. This result is reported as a proof-of-concept case rather than a statistical estimate of identification accuracy. An independent UAV radiation-survey dataset containing 3,722 observations across ten flight transects was additionally analysed. It demonstrated a strong and spatially stable relationship between integral gamma activity and estimated ^{137}Cs activity $r=0.990$, $R^2=0.980$. These UAV results provide an application-level reference but are not interpreted as direct validation of Geant4-generated spectra. The proposed framework provides a transparent and extensible basis for physics-guided synthetic-data generation while explicitly distinguishing attenuation verification, limited detector-level spectral assessment, and application-level consistency analysis.

Keywords: Digital Twin; Geant4; Monte Carlo simulation; synthetic data generation; gamma-ray spectra; radiation monitoring; verification and validation; physics-guided modelling.

1. INTRODUCTION

1.1. Context

Automated identification of radioactive sources is a critical component of modern nuclear security systems, ranging from border control monitoring to the decommissioning of nuclear facilities (Ivanov et al., 2022). Standard hand-held or drone-mounted spectrometers generate complex, noisy gamma-ray spectra that are difficult to interpret using traditional peak-fitting algorithms, especially in scenarios with heavy shielding or low signal-to-noise ratios.

Machine Learning (ML) and Deep Learning (DL) have emerged as effective approaches for Radioisotope Identification (RIID), offering improved performance in complex non-linear environments. However, the development of robust data-driven models is fundamentally constrained by the "data scarcity" problem. Unlike computer vision, where millions of labelled images are publicly available (e.g., ImageNet), the nuclear domain lacks large-scale, open-access datasets of real-world gamma-ray spectral signatures due to safety regulations, restricted access to radioactive materials, and the rarity of illicit trafficking events. This limitation motivates the development of high-fidelity simulation frameworks capable of

generating synthetic data that accurately reproduce the physical characteristics of real measurement scenarios.

Recent advances in Digital Twin technologies provide new opportunities for creating physically consistent virtual environments that support synthetic data generation, simulation-based analysis, and validation of radiation monitoring systems. By integrating detailed physical models with configurable computational workflows, digital twins enable the systematic generation of representative gamma-ray spectra under diverse source, shielding, and detector configurations while reducing the dependence on expensive and difficult-to-reproduce experimental measurements. Such frameworks provide a reproducible infrastructure for the development, testing, and benchmarking of future intelligent radiation monitoring applications.

1.2. Problem Statement: The Data Bottleneck in Radiation Monitoring and Synthetic Dataset Construction

Despite significant advances in neural network architectures, their practical deployment in radiation safety applications remains constrained by a critical challenge commonly referred to as the data bottleneck. Modern deep learning algorithms are inherently data-hungry: achieving robust generalization performance and avoiding overfitting typically requires training datasets comprising thousands or even tens of thousands of labelled examples.

In conventional computer vision tasks, this challenge is mitigated through large-scale publicly available datasets such as ImageNet or COCO. However, in the domain of radiation monitoring, the construction of comparable datasets faces fundamental obstacles. Acquiring real-world data for rare or hazardous scenarios - such as radioactive material leakage from damaged containment - is infeasible due to ethical and safety considerations. The deliberate release of radioactive aerosols for the purpose of “training” autonomous platforms is clearly unacceptable. Moreover, access to sources of special nuclear materials, such as ^{239}Pu or ^{232}Th , is strictly regulated under international non-proliferation protocols. As a result, the collection of representative gamma-ray spectra for these isotopes remains inaccessible to most researchers and developers of intelligent radiation monitoring systems. Existing open spectral databases are typically obtained under idealized laboratory conditions and fail to account for shielding effects, detector variability, and scattered radiation, which are inevitably present in real-world accident or field scenarios.

The limited availability and restricted diversity of experimental measurements not only affect the training of data-driven algorithms but also complicate the verification, benchmarking, and comparative evaluation of radiation monitoring technologies. Models and decision-support systems developed using small and highly specialized datasets often exhibit a significant reduction in performance when environmental conditions differ from those represented during data acquisition. Consequently, there is an urgent need for computational tools capable of generating high-fidelity synthetic datasets that realistically reproduce complex physical interactions and compensate for the scarcity of real measurement data.

From a Data-Centric AI perspective, improving dataset quality, coverage, and physical consistency is often more beneficial than increasing model complexity. In this context, validated simulation frameworks and digital twin technologies provide a promising approach for the systematic generation of physically consistent synthetic gamma-ray spectra under controlled and reproducible conditions. Such frameworks can support not only the training of future intelligent algorithms but also the verification, testing, and benchmarking of radiation monitoring applications where comprehensive experimental datasets are unavailable.

1.3. Existing Solutions and Their Limitations

Traditional physics-based modeling methods. A substantial body of research (Knoll, 2010; Saheli et al., 2021; Shorten and Khoshgoftaar, 2019) focuses on the use of Monte Carlo codes (e.g., MCNP, Geant4) for radiation shielding calculations. While these methods provide

high physical accuracy, they are rarely employed as systematic synthetic data generation frameworks and are typically limited to static dose estimation or attenuation analysis.

Machine learning applications in radiometry. Several studies (García-Esteban et al., 2023; Rathkopf, 2026) demonstrate the successful application of neural networks to gamma-ray spectrometry. However, most of these works encounter a severe shortage of labelled training data, relying on relatively small datasets obtained from laboratory experiments. This limitation significantly reduces the generalization capability and practical applicability of the resulting models.

Synthetic data generation using generative models (GAN/VAE). To address data scarcity, some authors (Agostinelli et al., 2003; Prokhorets, 2024; Shadura, 2019) propose the use of generative adversarial networks (GANs) and related generative models. A key limitation of this approach is the risk of producing physically inconsistent spectra (“hallucinations”) that do not necessarily satisfy the fundamental laws governing radiation–matter interactions.

Despite these advances, there is currently no comprehensive framework that combines the physical fidelity of Monte Carlo simulation tools such as Geant4 with a configurable and reproducible workflow for large-scale synthetic gamma-ray spectrum generation and validation. Existing approaches typically address either accurate physical modelling or data-driven algorithm development, while the integration of these capabilities into a unified digital twin environment remains limited. The development of such a physics-guided digital twin framework, capable of generating physically consistent synthetic datasets for the development, testing, and benchmarking of radiation monitoring applications, constitutes the primary objective of this work.

1.4. Digital Twins for Physics-Guided Data Generation

The concept of a Digital Twin has evolved from industrial manufacturing toward a general paradigm for representing, monitoring, and optimizing complex cyber-physical systems. Jones et al. (2020) emphasized that a Digital Twin integrates computational models, observational data, and domain knowledge to reproduce the behavior of a physical system, while Schweiger and Barth (2023) identified adaptability, data integration, and decision support as its defining characteristics.

In the nuclear domain, Digital Twin technologies are increasingly being investigated for monitoring and predictive analysis. Herrera et al. (2020) proposed a gamma-background radiation digital twin combined with machine learning algorithms, while Mengyan et al. (2024) reviewed emerging Digital Twin architectures for nuclear-energy applications. More recently, Hossain et al. (2025) demonstrated a virtual-sensing Digital Twin framework for real-time monitoring of nuclear systems using deep neural operators.

At the same time, synthetic data generation has become an important strategy for addressing the scarcity of representative datasets required for the development and evaluation of modern intelligent monitoring systems. Reviews by Goyal and Mahmoud (2024) highlight the growing use of GANs, VAEs, and other generative approaches for synthetic data creation. However, such methods may generate physically inconsistent outputs and often lack guarantees of compliance with the governing laws of radiation transport and interaction. Similar concerns regarding reliability and hallucination effects in generative AI systems have also been reported by Rathkopf (2026).

Recent advances in Physics-Informed Machine Learning demonstrate the benefits of integrating scientific knowledge with data-driven modelling (Karniadakis et al., 2021). In contrast to approaches that embed physical constraints directly into neural-network architectures, the present work incorporates physics at the data-generation stage. The proposed framework employs Monte Carlo radiation transport simulation as the computational core of a Digital Twin, ensuring that the generated datasets remain consistent with the underlying

radiation–matter interaction mechanisms while preserving the flexibility required for large-scale scenario generation.

Although Digital Twins, synthetic data generation, and AI-assisted radiation monitoring have each received considerable attention, they are typically investigated as separate research directions. There remains a lack of unified frameworks that integrate physics-based simulation, automated synthetic data generation, verification and validation procedures, and reproducible dataset construction within a single computational workflow. Addressing this gap constitutes the primary motivation of the present study.

1.5. Proposed Physics-Guided Digital Twin Framework

To address the identified limitations, this study proposes a physics-guided Digital Twin framework for configurable radiation-transport simulation and structured synthetic-data generation. The framework integrates scenario definition, Python-based simulation orchestration, Monte Carlo radiation transport using Geant4, detector-response aggregation, structured dataset construction, and explicitly separated procedures for physical verification, limited detector-level spectral assessment, and application-level consistency analysis.

The computational core is implemented in C++ using the Geant4 toolkit. It models photon transport and energy deposition under user-defined source, shielding, and detector configurations. In the present implementation, each radionuclide-labelled source class is represented by a characteristic monoenergetic gamma emission rather than by a complete radioactive-decay scheme. Accordingly, the model is intended to investigate energy-dependent radiation transport, attenuation, scattering, and buildup effects and does not claim to reproduce the complete measured spectrum of each radionuclide.

Simulation control is provided by an external Python-based orchestration layer. This component manages the systematic variation of source energy, shielding material, and shielding thickness; initiates batch simulations; parses Geant4 outputs; applies detector-event filtering; calculates transmission; and stores the resulting observations and configuration metadata in a structured CSV dataset. This separation between the simulation core and the orchestration layer enables additional scenarios to be generated without modifying or recompiling the principal Geant4 implementation.

The resulting workflow extends the use of Geant4 beyond individual radiation-transport calculations by supporting reproducible scenario generation and automated detector-response aggregation. Physical verification is performed by analysing attenuation behaviour and comparing fitted attenuation coefficients with NIST XCOM reference data while accounting for counting-statistical uncertainties. A limited detector-level spectral assessment is conducted separately using laboratory NaI(Tl) measurements, including a blind proof-of-concept identification of a multi-peak spectrum associated with the ^{232}Th decay chain.

An independent UAV radiation-survey dataset from the Chernobyl Exclusion Zone is used exclusively for application-level consistency assessment. The UAV analysis examines the strength and cross-transect stability of the relationship between integral gamma activity and estimated ^{137}Cs activity under operational environmental conditions. Because these integral survey variables are not directly equivalent to the Geant4-generated energy-deposition data, the UAV measurements are not used to calibrate the simulation, update its parameters, or validate full spectral fidelity.

The proposed framework should therefore be interpreted as a configurable and reproducible computational environment that connects physics-guided synthetic-data generation with three distinct levels of assessment: attenuation verification against reference physical data, limited detector-level spectral assessment using laboratory measurements, and independent application-level consistency analysis using UAV observations. This separation prevents application-level environmental measurements from being treated as direct evidence of simulated spectral fidelity.

1.6. Research Gap and Contributions

Despite significant progress in Monte Carlo radiation transport simulation and machine learning for radiation monitoring, an important research gap remains. Existing studies typically focus either on physics-based simulation environments or on the development of AI algorithms. Comparatively little attention has been devoted to integrated Digital Twin frameworks that systematically combine synthetic data generation, physics-informed validation, and structured dataset construction within a unified and reproducible computational workflow.

A major challenge limiting the practical deployment of intelligent radiation monitoring systems is the Sim-to-Real gap. Data-driven methods developed using laboratory measurements or synthetic datasets often exhibit degraded performance when applied to operational monitoring scenarios characterized by environmental variability, detector uncertainty, and complex radiation–matter interactions. Consequently, ensuring the physical realism and practical relevance of synthetic datasets remains a critical requirement for the trustworthy development and evaluation of computational methods in nuclear safety and environmental monitoring applications.

To address this challenge, this study proposes a physics-guided Digital Twin framework that combines Geant4-based synthetic data generation, attenuation verification against reference physical data, and an independent application-level consistency assessment using UAV radiation-survey measurements from the Chernobyl Exclusion Zone. The UAV component provides an operational environmental context for examining radiation-related relationships but is not used to validate the Geant4-generated spectra or detector response.

The main contributions of this work are as follows:

1. Development of a configurable Geant4-based Digital Twin workflow for physics-guided synthetic radiation-data generation.
2. Automated construction of a structured dataset covering eight source classes, four shielding materials, and five shielding thicknesses.
3. Verification of attenuation behaviour against NIST XCOM reference data with propagated counting uncertainties.
4. Limited detector-level spectral assessment using laboratory NaI(Tl) measurements and blind proof-of-concept identification of a decay-chain spectrum.
5. Independent application-level consistency assessment using UAV radiation-survey data from the Chernobyl Exclusion Zone.

Beyond its role as a synthetic data generation platform, the proposed Digital Twin framework provides a practical computational infrastructure for the development, testing, and benchmarking of radiation monitoring technologies. The framework enables systematic evaluation of shielding scenarios, detector configurations, and radionuclide combinations that are difficult or impossible to reproduce experimentally, thereby supporting both operational system design and future intelligent radiation monitoring applications.

2. RELATED WORK AND RESEARCH SCOPE

2.1. Radiation Transport Modeling and Research Gap

Monte Carlo radiation-transport codes, particularly MCNP and Geant4, are widely used to model the interaction of ionizing radiation with matter. Geant4 provides a flexible open-source environment for defining radiation sources, material compositions, detector geometries, and relevant interaction processes (Agostinelli et al., 2003; Allison et al., 2016). This flexibility also supports its integration into configurable Digital Twin architectures for radiation monitoring, virtual sensing, and nuclear-system analysis (Herrera et al., 2020; Mengyan et al., 2024; Hossain et al., 2025).

Machine-learning methods have increasingly been applied to radionuclide identification, source localization, and radiation-field analysis (de Oliveira et al., 2023; Pant,

2025; Singh et al., 2025). However, the development and comparative evaluation of such methods remain constrained by the limited availability of labelled radiation data, particularly for shielded, hazardous, or rarely observed source configurations. Existing studies commonly address either radiation-transport simulation or machine-learning analysis, whereas fewer provide reproducible workflows connecting configurable Monte Carlo simulations, automated scenario generation, structured dataset construction, physical verification, and subsequent data-driven evaluation.

A further limitation is that data-generation scripts, simulation parameters, and preprocessing procedures are not always sufficiently documented to support independent reproduction or extension. Consequently, there remains a need for transparent computational workflows that generate physically interpretable synthetic observations under user-defined radiation-monitoring scenarios and explicitly distinguish physics verification, experimental spectral assessment, and application-level consistency evaluation.

2.2. Research Objective and Scope

The objective of this study is to develop and assess a physics-guided Digital Twin framework for automated synthetic data generation in radiation-monitoring applications. The framework integrates Geant4-based Monte Carlo radiation transport, Python-based scenario orchestration, structured detector-response aggregation, physical verification, and machine-learning evaluation within a reproducible computational workflow.

The present implementation represents eight radionuclide-labelled source classes using characteristic monoenergetic gamma emissions between 51 keV and 2.61 MeV. It evaluates unshielded and shielded configurations involving lead, stainless steel, concrete, and polyethylene. The radionuclide labels denote representative photon-energy classes rather than complete radioactive-decay schemes.

The specific objectives are to:

1. implement a configurable Geant4-based model combining an isotropic point source, variable shielding, and a sensitive detector volume;
2. automate the generation of source–material–thickness scenarios and aggregate detector-response data into a structured synthetic dataset;
3. quantify attenuation and photon-buildup behaviour across shielding materials and compare fitted attenuation coefficients with NIST XCOM reference data, including propagated counting uncertainties;
4. perform a limited experimental spectral assessment using independently measured NaI(Tl) spectra and demonstrate a blind proof-of-concept identification of a multi-peak ^{232}Th decay-chain spectrum;
5. evaluate the predictive utility of the generated variables using machine-learning models;
6. conduct an independent application-level consistency assessment using UAV radiation-survey data from the Chornobyl Exclusion Zone, without treating these integral measurements as validation of full spectral fidelity.

This scope positions the proposed framework as a reproducible computational workflow for physics-guided data generation and evaluation rather than as a complete metrological model of radionuclide decay, detector electronics, or absolute spectrometric efficiency.

3. MATERIALS AND METHODS

3.1. System Architecture

To achieve the stated objectives, a hybrid Digital Twin architecture was developed for automated synthetic data generation and validation in radiation-monitoring applications. The framework combines high-fidelity Monte Carlo simulation, automated experiment

orchestration, data management, and machine-learning-oriented processing within a unified reproducible workflow. The architecture is organized into two logical layers and three functional components, balancing the requirements of physical accuracy, computational efficiency, and experimental flexibility.

The overall architecture of the proposed framework is illustrated in Fig. 1. The system consists of the following components:

1. **Simulation Core (C++/Geant4).** The simulation core is implemented using the Geant4 toolkit (version 11.0) and serves as the physics engine of the Digital Twin. It is responsible for detector geometry construction, primary particle generation, radiation–matter interaction modeling, and event-level tracking of gamma photons. The module is compiled into a standalone executable optimized for multithreaded execution, enabling efficient large-scale Monte Carlo simulations.

2. **Orchestrator (Python 3.9+).** The orchestration layer controls the execution of virtual experiments. It automatically generates Geant4 macro files, configures simulation parameters, launches simulation runs, monitors execution status, and manages error handling. This component enables dynamic reconfiguration of isotopes, shielding materials, detector settings, and experimental conditions without recompilation of the simulation core.

3. **Data Pipeline.** The data pipeline is responsible for collecting simulation outputs, parsing generated spectra, aggregating results, and storing data in structured formats suitable for subsequent machine learning applications. The subsystem also performs dataset organization, metadata management, and preparation of AI-ready training data.

Unlike conventional Geant4 workflows, where simulations are typically executed individually and manually configured, the proposed architecture enables fully automated generation of large numbers of experimental scenarios. As a result, the framework transforms Geant4 from a stand-alone radiation simulation tool into a scalable Digital Twin platform for synthetic dataset generation, physics-informed validation, and AI-oriented analysis.

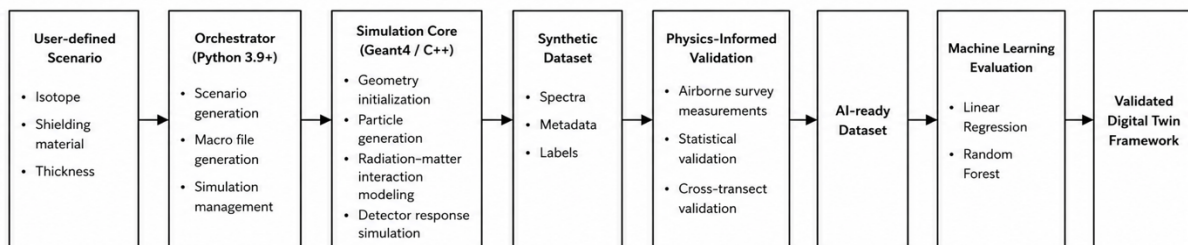


Fig. 1. Digital Twin architecture integrating the Geant4 physics engine, automated experiment orchestration, and AI-oriented data processing.

3.2. Automated Data Generation Workflow

The synthetic data generation process is fully automated and follows a cyclic execution model. Initially, the orchestrator reads user-defined scenario parameters, including radionuclide type, shielding material, shielding thickness, detector configuration, and experimental conditions. Based on these parameters, configuration files and Geant4 macro commands are generated automatically.

For each scenario, the orchestrator launches the simulation core and executes a Monte Carlo experiment. During execution, energy-deposition events and detector-response information are recorded and exported to intermediate output files. Upon completion of the simulation, the generated outputs are parsed, aggregated, and transformed into structured records suitable for machine learning workflows.

The resulting dataset contains both radiation-response information and metadata describing the simulated experimental conditions. This approach enables systematic exploration of large parameter spaces while ensuring complete reproducibility of the simulation process.

The generated datasets subsequently undergo a multi-stage evaluation procedure. First, attenuation behaviour is verified through comparison with theoretical reference data derived from radiation-transport theory. Second, limited detector-level spectral assessment is performed using independently acquired laboratory NaI(Tl) measurements. The UAV radiation-survey dataset is analysed separately to characterize an application-level radiation relationship under operational environmental conditions and is not directly compared with the Geant4 outputs. Finally, the structured synthetic dataset is prepared for subsequent data-driven evaluation.

The proposed workflow implements a Physics-Informed Digital Twin paradigm that combines automated simulation, structured synthetic-data generation, attenuation verification, and explicitly separated detector- and application-level assessment procedures within a unified computational environment. This capability is particularly important for developing machine learning models in domains where acquisition of real radiation measurements is expensive, restricted, hazardous, or impractical.

3.3. Digital Twin Implementation

The development of the detector Digital Twin was based on the object-oriented architecture of Geant4 (version 11.0). The software implementation consists of three key components that define the spatial configuration, material composition, and physical interaction models. The reliability of Geant4-based frameworks for radiation shielding calculations has been confirmed in a number of laboratory studies (MacFadden et al., 2018).

The geometric model was constructed using the method of constructive solid geometry. A schematic representation of the simulated scene is shown in Fig. 2.

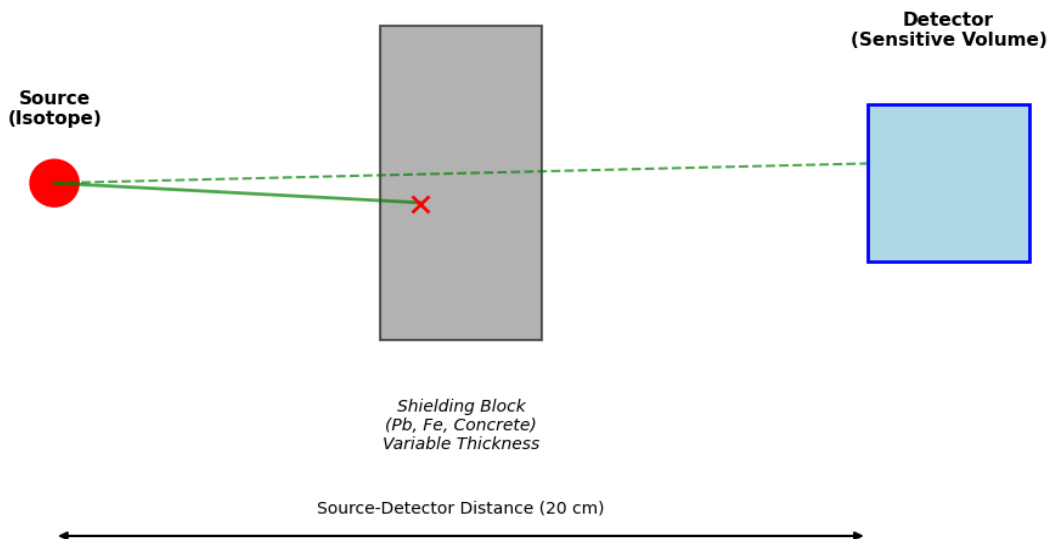


Fig. 2. Schematic representation of the experimental setup modeled in Geant4. The simulation geometry consists of a point isotropic gamma source (^{137}Cs), a variable shielding block (Lead, Steel, Concrete, Polyethylene) with thickness $t \in [0, 10]$ cm, and a sensitive detector volume located at a fixed source-to-detector distance of 20 cm.

The simulation setup includes a point isotropic gamma source placed within a defined geometry. The experimental domain consists of three primary volumes:

1. A cubic volume with dimensions of $100 \times 100 \times 100$ cm filled with air (G4_AIR).
2. A variable volume (G4Box) positioned between the radiation source and the detector. To ensure physical fidelity of the material properties, the NIST material database is utilized via the G4NistManager class. This guarantees that the modeled materials (e.g., G4_Pb, G4_STAINLESS-STEEL, G4_CONCRETE, and G4_POLYETHYLENE) accurately correspond to their real-world counterparts.
3. A sensitive volume representing a scintillation detector located at a fixed distance from the shielding block.

A critical feature of the implementation is the incorporation of dynamic geometry control. Instead of hard-coding geometric parameters, the DetectorConstruction class is equipped with the G4GenericMessenger interface. This allows the external Python-based orchestrator to modify shielding thickness and material type at runtime without requiring recompilation of the simulation core.

The selection of an appropriate physics model is critical to the accuracy of radiation transport simulations. Since the investigated energy range extends from 51 keV to 2.61 MeV, the framework requires accurate modeling of low-energy electromagnetic interactions. Standard physics packages optimized primarily for high-energy particle physics are therefore insufficient for the objectives of this study.

In this work, the G4EmStandardPhysics_option4 physics constructor is employed. This physics list provides enhanced electromagnetic interaction models, including photoelectric absorption, Compton scattering, pair production, and secondary particle generation processes relevant to gamma-ray transport and shielding analysis. The selected configuration incorporates advanced low-energy interaction models derived from the Livermore and Penelope electromagnetic packages and is widely used in radiation physics applications requiring high simulation fidelity.

The resulting Digital Twin implementation serves as the computational foundation for automated synthetic-dataset generation. Its attenuation behaviour is verified through comparison with theoretical reference data, while detector-level spectral interpretability is assessed separately using limited laboratory NaI(Tl) measurements. The independent UAV dataset is used only for application-level consistency assessment and is not treated as validation of the Geant4-generated spectra.

Figure 3 presents an illustrative energy-deposition spectrum generated for the ^{137}Cs -labelled source configuration. Its purpose is to demonstrate the form of the Geant4 output used by the processing pipeline; it is not used as evidence of experimental spectral validation.

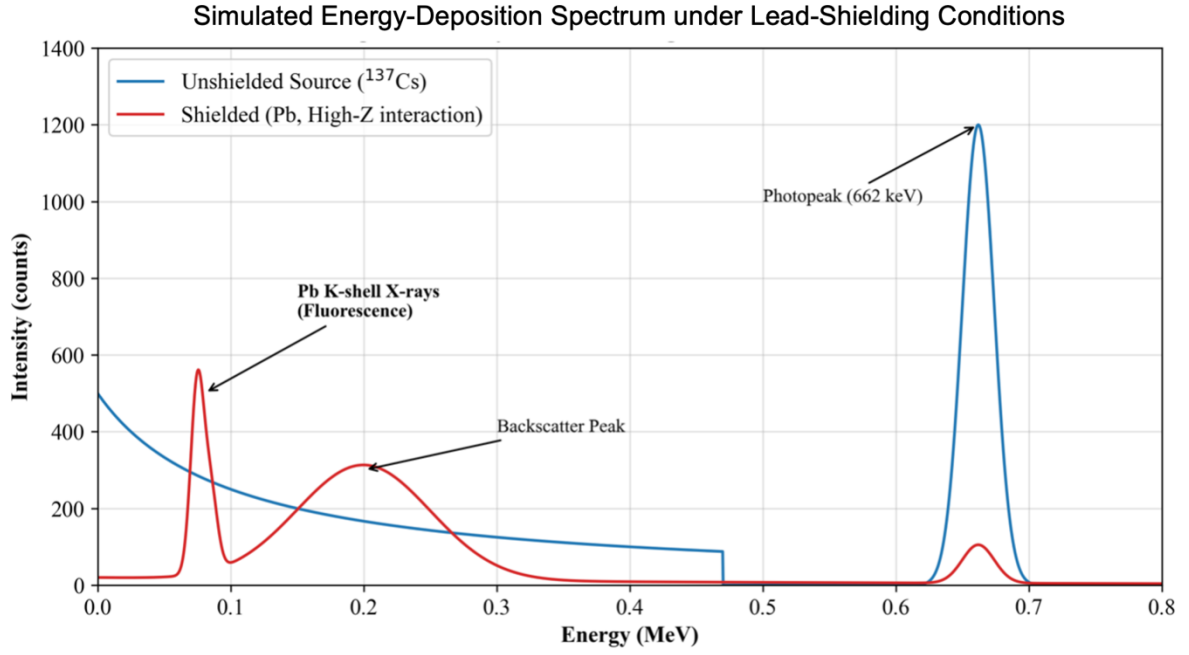


Fig. 3. Simulated energy-deposition spectrum for a monoenergetic ^{137}Cs -labelled gamma source under lead-shielding conditions. The spectrum illustrates the detector response generated by the Geant4 model, including the full-energy deposition region and lower-energy contributions associated with photon scattering and secondary interactions. The figure represents simulation output and is not presented as experimental validation of spectral fidelity.

The simulated spectrum contains a full-energy deposition region near the prescribed primary photon energy and a lower-energy continuum resulting from partial energy deposition and scattering processes. Detector-specific broadening and electronic-response effects were not explicitly modelled in this simulation.

3.4 Orchestration and Data Processing Algorithm

To ensure full reproducibility of the experiment and eliminate human intervention, a deterministic automation algorithm was developed and implemented in Python. The algorithm performs an exhaustive exploration of the experimental state space, defined as the Cartesian product of the sets of isotopes (I), materials (M), and shielding thicknesses (T). The logical structure of the data generation process is presented in Algorithm 1.

Algorithm 1: Synthetic Radiation Data Generation Pipeline

Input:

- I: Dictionary of isotopes and their energies {Name: E_{MeV} }
- M: List of NIST material codes
- T: List of shielding thicknesses $[t_0, t_1, \dots, t_n]$
- N: Number of primary events (history count)

Output:

- D: Structured CSV dataset

- 1: Initialize dataset D with column headers
- 2: FOR each isotope i with energy E in I DO
- 3: FOR each material m in M DO
- 4: BaseCounts $\leftarrow 0$ // Counts without shielding
- 5:

```

6:   FOR each shielding thickness t in T DO
7:       // Step 1: Macro generation
8:       MacroFile ← CreateMacro(Energy = E,
9:                               Material = m,
10:                              Thickness = t,
11:                              Events = N)
12:
13:       // Step 2: Simulation execution
14:       ExecuteSimulation("./exampleB1", MacroFile)
15:
16:       // Step 3: Raw data parsing
17:       RawHits ← ReadCSV("B1_output.csv")
18:       FilteredCounts ← 0
19:
20:       FOR each hit_energy in RawHits DO
21:           // Software low-level discriminator (LLD)
22:           IF hit_energy >= 0.001 MeV THEN
23:               FilteredCounts ← FilteredCounts + 1
24:           END IF
25:       END FOR
26:
27:       // Step 4: Feature computation
28:       IF t = 0 THEN
29:           BaseCounts ← FilteredCounts
30:           Transmission ← 1.0
31:       ELSE
32:           Transmission ← FilteredCounts / BaseCounts
33:       END IF
34:
35:       // Step 5: Dataset aggregation
36:       AppendToDataset(D,
37:                     [i, E, m, t,
38:                      FilteredCounts,
39:                      Transmission])
40:   END FOR
41: END FOR
42: END FOR

```

Particular attention is devoted to the preprocessing stage of the raw simulation logs (Steps 3–4). The Geant4 output files contain information about every particle interaction registered within the detector volume. To approximate realistic measurement conditions, a software-based discrimination threshold, referred to as a low-level discriminator (LLD), is applied. Events with deposited energy $E < 1$ keV are filtered out as electronic noise or insignificant background fluctuations. The selected threshold is consistent with typical electronic discrimination settings used in gamma-spectrometry systems and prevents contamination of the dataset with non-informative events.

It should be noted that the parameter N in Algorithm 1 specifies the number of simulated primary particle histories used in the Monte Carlo experiment, whereas Detected_Counts represents the number of detector events recorded after application of the low-level

discriminator. Consequently, Detected_Counts is a derived measurement quantity and should not be interpreted as the total number of simulated primary particles.

In addition to the absolute number of detected events (Detected_Counts), the algorithm automatically computes the normalized transmission coefficient:

$$T_r(t) = \frac{N(t)}{N(0)}, \quad (1)$$

where $N(t)$ denotes the number of detected events measured in the presence of shielding with thickness t , and $N(0)$ represents the corresponding unshielded reference measurement for the same radionuclide and detector configuration. This parameter serves as an informative feature for both physics-informed validation and subsequent machine-learning analysis.

Instead of storing large volumes of raw particle-track information, the framework performs automated aggregation into a compact tabular dataset. Each row corresponds to a completed simulation scenario and contains both measurement results and metadata describing the experimental configuration. This representation is directly compatible with standard data-analysis and machine-learning libraries such as Pandas and Scikit-learn.

The generated dataset includes detector-response information together with radionuclide, shielding-material, and shielding-thickness metadata. Transmission coefficients slightly exceeding unity may occur for low- Z materials due to photon buildup and multiple-scattering effects, which may increase the number of detected events relative to the corresponding unshielded reference configuration.

The framework was benchmarked on a workstation powered by an Apple M4 processor (8-core CPU, 16 GB Unified Memory). The orchestration pipeline utilizes Python multiprocessing to distribute simulation tasks across available processor cores, enabling efficient execution of large numbers of independent Monte Carlo experiments. This approach significantly reduces dataset-generation time while maintaining full reproducibility of the simulation workflow.

The resulting dataset forms the basis of the proposed Digital Twin framework. By systematically varying radionuclide type, shielding material, and shielding thickness, the system automatically produces labeled synthetic observations that can subsequently be employed for machine-learning training, benchmarking, uncertainty analysis, and physics-informed validation without requiring access to hazardous radioactive sources.

3.4.1. Dataset Augmentation for Machine Learning Evaluation

To evaluate the AI-oriented applicability of the proposed framework, the original synthetic dataset was expanded through Poisson-based statistical augmentation. The augmentation procedure was designed to emulate counting-statistics variability that naturally occurs in repeated radiation measurements.

For each original simulation configuration, 50 additional realizations of the Detected_Counts variable were generated using a Poisson distribution parameterized by the original count value. This procedure preserves the statistical properties of radiation detection while introducing realistic measurement variability into the dataset.

The original dataset containing 160 unique simulation configurations was expanded to 8,160 samples by generating 50 additional Poisson realizations for each original configuration.

The augmentation process enables evaluation of machine-learning algorithms under more realistic statistical conditions while preserving the underlying physical relationships established by the Digital Twin simulation framework.

3.5. Independent Airborne Radiation-Survey Dataset

An independent airborne radiation-survey dataset was used to provide an application-level reference for examining radiation-related relationships under operational environmental

conditions. The measurements were acquired during an unmanned aerial vehicle (UAV) survey over the Pishchane Plateau in the Chernobyl Exclusion Zone at an approximate flight altitude of 10 m above ground level.

Following quality-control procedures, the dataset contained 3,722 georeferenced observations distributed across ten flight transects. Each observation included integral gamma activity, estimated ^{137}Cs activity, spatial coordinates, and a transect identifier. Quality control included coordinate verification, removal of incomplete records, and consistency checks of the radiometric variables.

Integral gamma activity represents an aggregate measure of the recorded radiation field, whereas ^{137}Cs activity is a radionuclide-specific estimate derived from the airborne measurements. These UAV variables are therefore not directly equivalent to the monoenergetic photon energies, detector-event counts, or transmission coefficients generated by the Geant4 model.

The UAV dataset was not used to calibrate the Geant4 simulation, estimate detector-response parameters, or assess agreement between simulated and measured spectral shapes. Instead, it was analysed independently to determine whether a stable and physically interpretable relationship between integral radiation response and ^{137}Cs activity was present under real survey conditions. The resulting analysis is treated as an application-level consistency assessment and not as detector-level or spectral validation of the Digital Twin.

3.6. Application-Level Consistency Assessment

The UAV analysis was designed to characterize the stability of an operational radiation-monitoring relationship rather than to compare synthetic and measured spectra directly. The assessment focused on the association between integral gamma activity and estimated ^{137}Cs activity within the surveyed radiation field.

The procedure consisted of three stages:

1. **Data quality control.** Records with missing values, invalid coordinates, or incomplete radiometric measurements were excluded.
2. **Overall relationship analysis.** The association between integral gamma activity and ^{137}Cs activity was evaluated using Pearson's correlation coefficient, linear regression, the coefficient of determination and residual analysis.
3. **Cross-transect stability assessment.** The same relationship was evaluated separately for each of the ten flight transects to determine whether it remained stable across spatially distinct survey trajectories.

The resulting statistics characterize the internal consistency and spatial robustness of the relationship observed in the UAV dataset. They do not quantify agreement between UAV measurements and Geant4 outputs because the two datasets describe different measurement levels and do not share directly comparable input and response variables.

Accordingly, the UAV analysis was not incorporated into an iterative calibration or closed-loop update of the simulation model. Its role was limited to providing an independent application context for interpreting the potential relevance of physics-guided synthetic data to environmental radiation monitoring. The overall workflow and the separation between the synthetic-data generation and UAV application-level assessment branches are summarized in Fig. 4.

Physics-guided computational workflow

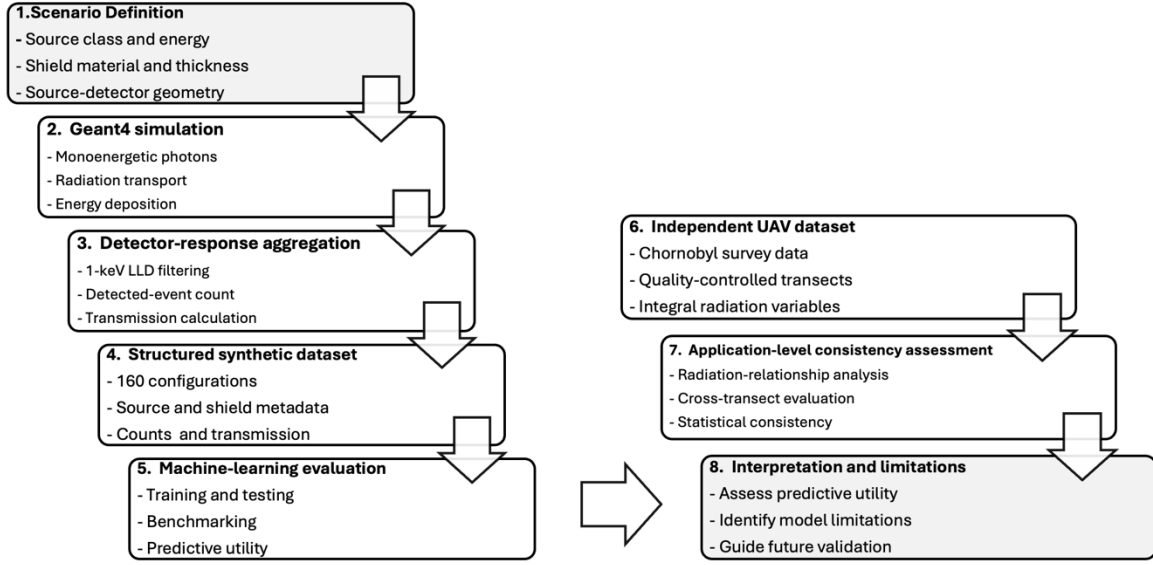


Fig. 4. Architecture of the physics-guided computational workflow integrating scenario definition, Geant4-based radiation-transport simulation, detector-response aggregation, structured synthetic-dataset construction, machine-learning evaluation, and an independent application-level consistency assessment using UAV radiation-survey data from the Chernobyl Exclusion Zone. The UAV dataset was not used to calibrate the Geant4 model or to validate full spectral fidelity.

4. RESULTS AND DISCUSSION

4.1. Simulation Parameters and Feature Space

The following analysis characterizes the parameter space, structure, and physical behaviour of the generated synthetic dataset. The feature space of the generated dataset is defined by three variable parameters: source type (energy), absorber type (material), and shielding thickness.

A. Radiation Sources

To ensure data variability, eight radionuclides were selected, covering an energy range from 51 keV to 2.61 MeV. The dataset includes three classes of sources:

- Special nuclear materials: ^{239}Pu (0.051 MeV) and ^{235}U (0.186 MeV). Their low-energy gamma emissions are easily attenuated, posing a challenging detection problem.
- Medical and industrial isotopes: $^{99\text{m}}\text{Tc}$ (0.140 MeV), ^{131}I (0.364 MeV), ^{137}Cs (0.662 MeV), and ^{60}Co (1.25 MeV). These represent commonly encountered medical and industrial radiation-source classes.
- Naturally occurring radioactive materials: ^{40}K and ^{232}Th , which are used for background calibration purposes.

B. Shielding Materials

To investigate the influence of atomic number (Z) and material density on the shape of the measured gamma-ray spectrum, four shielding materials were considered:

1. Polyethylene (G4_POLYETHYLENE, $\rho = 0.94 \text{ g/cm}^3$), representing lightweight packaging materials and plastics. It is characterized by low photoelectric absorption but significant scattering.

2. Concrete (G4_CONCRETE, $\rho = 2.3 \text{ g/cm}^3$), the most commonly used construction material.

3. Stainless steel (G4_STAINLESS-STEEL, $\rho = 8.0 \text{ g/cm}^3$), typically used for transport containers and equipment enclosures.
4. Lead (G4_Pb, $\rho = 11.35 \text{ g/cm}^3$), the standard material for radiation shielding.

C. Measurement Geometry

The shielding thickness was varied discretely as 0, 1, 2, 5, and 10 cm. The value of “0 cm” corresponds to the baseline condition and is used as a reference for calculating the transmission coefficient. In total, 160 unique experimental configurations were simulated. Automated scenario sweeps were performed with 10^3 primary events, while high-statistics spectral simulations used 10^5 events per configuration to improve statistical reliability. The selected numbers of primary histories define the counting-statistical precision of the present simulations but were not derived from a target experimental energy resolution. Because detector-specific energy broadening was not modelled, the present study does not claim validation at a specified percentage energy resolution. Establishing such a requirement would require a calibrated detector-response function and convergence analysis against experimental spectra acquired under matched conditions.

The number of particle-transport steps was not specified as a fixed input parameter because Geant4 determines step lengths dynamically from the geometry, material boundaries, active physical processes, and interaction sampling. The primary-history count was therefore used as the reproducible measure of simulation depth.

4.2. Characteristics of the Generated Dataset (Benchmark Dataset)

The output of the automated pipeline is a structured dataset stored in CSV format. The dataset contains no missing values and is ready for use in machine learning algorithms without additional preprocessing. The structure of the generated dataset and representative records for different shielding scenarios are presented in **Table 1**.

Table 1. Fragment of the generated dataset demonstrating the dependence of detector response on source, shielding material, and shielding thickness.

Isotope	Energy (MeV)	Material	Thickness (cm)	Detected_Counts	Transmission_Rate
Cs-137	0.662	Lead	0	161	1.000
Cs-137	0.662	Lead	2	19	0.118
Cs-137	0.662	Concrete	5	121	0.752
Cs-137	0.662	Polyethylene	1	163	1.012
Pu-239	0.051	Polyethylene	10	36	1.500

Each record in the dataset is represented by a feature vector with the following attributes:

- Isotope (Categorical): class label of the radiation source (e.g., ^{137}Cs).
- Energy_MeV (Numerical): energy of the primary gamma-ray line (0.051–2.61 MeV).
- Material (Categorical): type of shielding material.
- Thickness_cm (Numerical): thickness of the shielding layer (independent variable).
- Detected_Counts (Integer): number of detector interaction events recorded during the simulation.
- Transmission_Rate (Float): normalized transmission coefficient (Tr), calculated as the ratio between shielded and unshielded detector responses.

It should be noted that Detected_Counts represents the number of detector interaction events recorded within the sensitive volume rather than the number of primary particles emitted by the source. Consequently, the number of detected events may differ from the number of

simulated primary histories due to photon scattering, secondary interactions, and buildup effects occurring within the shielding materials and detector environment.

4.2.1. Statistical Analysis and Correlations

An initial exploratory data analysis (EDA) was performed to evaluate the structure and physical consistency of the generated synthetic dataset. The mean transmission coefficient across the entire dataset was 0.62. However, the distribution was strongly material-dependent. Lead provided the strongest attenuation, with an average transmission coefficient of 0.29, whereas polyethylene exhibited the highest average value of 1.03, reflecting its limited attenuation capability for gamma radiation.

Pearson correlation analysis indicated a moderate negative correlation between shielding thickness and transmission coefficient ($r = -0.398$, $p < 0.001$), which is consistent with the expected attenuation behavior of gamma radiation in matter. Conversely, a positive correlation was observed between photon energy and transmission coefficient ($r = 0.215$, $p = 0.006$), reflecting the higher penetration capability of higher-energy gamma photons.

In 17 cases, corresponding to 10.6% of the dataset, transmission values greater than unity ($T > 1.0$) were observed. The maximum value, ($T = 1.5$), occurred for ^{239}Pu at 0.051 MeV when shielded by 10 cm of polyethylene. This behavior is physically interpretable rather than a simulation artifact. For low-energy photons interacting with low- Z materials, multiple scattering and photon buildup effects may increase the number of detector interaction events relative to the unshielded reference configuration. The presence of such cases is important for AI-oriented analysis because it exposes machine learning models to non-trivial radiation transport scenarios involving scattered and indirect photon contributions.

4.3. Physical Validation of the Digital Twin

The physical validation of the Digital Twin was performed to assess whether the generated synthetic data reproduce established radiation-matter interaction behavior. Theoretical reference values of the linear attenuation coefficients were obtained using the NIST XCOM photon cross-section database (Berger and Hubbell, 1987). The validation procedure consisted of two complementary stages: qualitative validation through analysis of attenuation curve shapes and quantitative verification through comparison with theoretical attenuation coefficients.

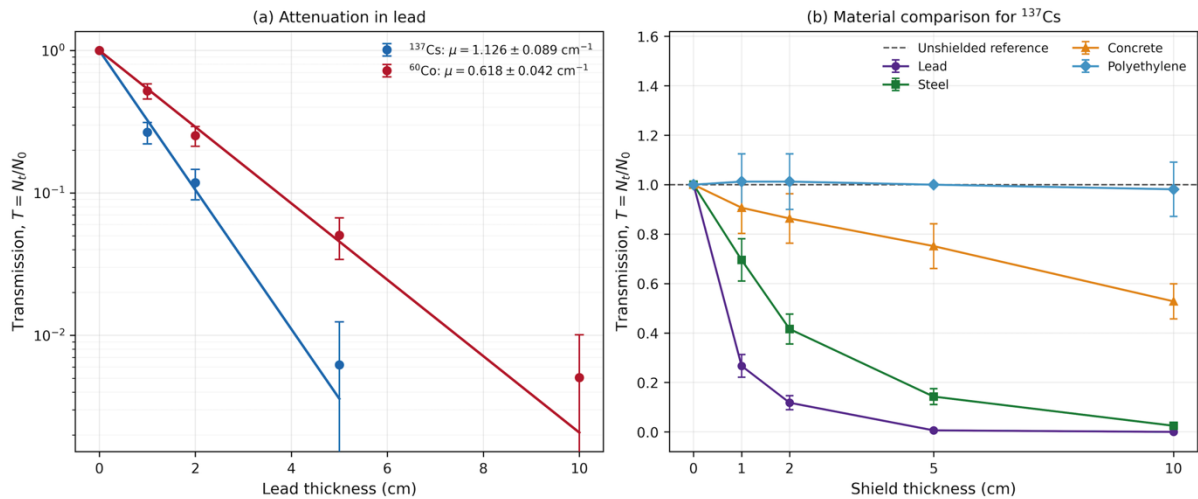


Fig. 5. Gamma-ray transmission through shielding materials obtained from Geant4 simulations: (a) attenuation of ^{137}Cs - and ^{60}Co -labelled radiation in lead, shown on a logarithmic scale together with weighted exponential fits; (b) comparison of transmission through lead, steel, concrete, and polyethylene for a ^{137}Cs -labelled source. Error bars represent propagated Poisson counting uncertainties. The ^{137}Cs data point at 10 cm of lead, for which no detector interactions were recorded ($N_t = 0$), was excluded from the logarithmic plot and attenuation fit.

4.3.1. Analysis of Attenuation Curves

Fig. 5 presents the attenuation behavior of gamma radiation under different shielding conditions. The results demonstrate a clear dependence of shielding effectiveness on material density and effective atomic number, consistent with the fundamental theory of radiation–matter interactions. For lead, a pronounced decrease in registered detector response is observed as shielding thickness increases, reflecting strong attenuation due to photoelectric absorption and Compton scattering. In contrast, polyethylene exhibits a much flatter transmission profile, indicating weaker attenuation of gamma photons in low-Z material.

The left panel of Fig. 5 shows the attenuation of ^{137}Cs and ^{60}Co radiation in lead shielding. The approximately linear behavior on the logarithmic scale supports the expected exponential attenuation trend. The right panel compares transmission coefficients for different materials under ^{137}Cs irradiation. The dashed reference line denotes the unshielded condition ($N/N_0 = 1$), allowing direct identification of attenuation and buildup-dominated regimes.

4.3.2. Buildup-Induced Transmission Increase in Low-Z Material

An important indicator of model fidelity is the ability to reproduce non-monotonic detector-response behavior caused by scattering and buildup effects. For the low-energy source ^{239}Pu ($E = 51 \text{ keV}$) shielded by polyethylene, an increase in the transmission coefficient above unity was observed, reaching ($T = 1.5$) at a shielding thickness of 10 cm.

This effect can be explained by the interaction of low-energy photons with low-Z material. Under these conditions, the shielding medium may act not only as an attenuator but also as a scattering volume, redirecting photons toward the sensitive detector region. Consequently, the number of recorded detector interaction events can exceed that of the unshielded reference configuration. The successful reproduction of this buildup-induced increase supports the physical consistency of the simulation geometry and interaction modeling.

4.3.3. Quantitative Verification against NIST XCOM

For quantitative verification, transmission through lead was calculated as

$$T(x) = \frac{N(x)}{N(0)}, \quad (2)$$

where $N(x)$ is the number of accepted detector-interaction events recorded at shielding thickness x , and $N(0)$ is the corresponding unshielded count. Transmission uncertainties were estimated from independent Poisson counting statistics for the shielded and unshielded configurations.

The attenuation coefficient was determined using the exponential attenuation model

$$T(x) = e^{-\mu x}, \quad (3)$$

After logarithmic transformation, the relationship becomes

$$\ln T(x) = -\mu x, \quad (4)$$

Weighted least-squares regression was applied to $\ln T(x)$, using the propagated counting uncertainties as statistical weights. Configurations with zero accepted detector events were excluded from the logarithmic fit because $\ln(0)$ is undefined.

For ^{137}Cs -labelled photons with an energy of 0.662 MeV, the fitted linear attenuation coefficient in lead was $\mu_{\text{sim}} = 1.126 \pm 0.089 \text{ cm}^{-1}$, with a 95% confidence interval of $0.952 \leq \mu_{\text{sim}} \leq 1.300 \text{ cm}^{-1}$.

For ^{60}Co -labelled photons with a representative energy of 1.25 MeV, the fitted coefficient was $\mu_{\text{sim}} = 0.618 \pm 0.042 \text{ cm}^{-1}$, with a 95% confidence interval of $0.535 \leq \mu_{\text{sim}} \leq 0.700 \text{ cm}^{-1}$.

The corresponding NIST XCOM narrow-beam reference value was $\mu_{\text{XCOM}} = 0.680 \text{ cm}^{-1}$. The relative difference between the fitted coefficient and the NIST XCOM reference value was 9.2%.

Although the central estimates differed by approximately 9.2%, the NIST XCOM reference value lies within the 95% confidence interval of the fitted coefficient. The simulated attenuation behaviour is therefore statistically consistent with the XCOM reference at the uncertainty level supported by the available detector-event counts. This result supports the physical consistency of the simulated attenuation trend but does not constitute complete validation of detector response or spectral fidelity.

4.4. Limited Experimental Spectral Assessment and Blind Identification

A limited experimental assessment was performed using independently acquired spectra measured with a $63 \times 63\text{-mm}$ NaI(Tl) detector. Energy calibration was based on a confirmed reference spectrum, for which the 661.657-keV photopeak was located at channel 400. Because only one confirmed calibration point was available, the measured spectra were used for approximate photopeak localization and qualitative multi-peak interpretation rather than precision energy metrology or absolute-efficiency validation.

One independently measured spectrum, 2026.08.12_05-43-19, was analysed without knowledge of the source identity or activity. Before ground-truth disclosure, its multi-peak structure was assigned to material containing the ^{232}Th decay chain. The data provider subsequently confirmed the source as ^{232}Th , supporting the blind assignment in this proof-of-concept case.

This limited assessment does not establish classification accuracy or complete experimental validation of the Digital Twin. The laboratory detector geometry was not reproduced in the Geant4 model, and only one blind spectrum with disclosed ground truth was available. The result therefore provides supplementary detector-level evidence of qualitative spectral interpretability rather than validation of absolute efficiency, peak areas, or full spectral fidelity.

4.5. Application-Level Consistency Assessment Using Airborne Survey Data

After quality control, the independent UAV dataset contained 3,722 valid observations distributed across ten flight transects. The analysis examined the relationship between integral gamma activity and estimated ^{137}Cs activity within this operational airborne survey dataset. For the complete dataset, Pearson's correlation coefficient was $r=0.990$ ($p<0.001$), and the linear model yielded $R^2 = 0.980$.

These results demonstrate a strong association between the two measured radiation variables within the surveyed area. The coefficient of determination indicates that approximately 98% of the observed variance in the estimated ^{137}Cs activity was associated with the fitted linear relationship under the conditions represented by this dataset.

Figure 6 presents the observed relationship between integral gamma activity and estimated ^{137}Cs activity. The concentration of observations around the fitted line indicates that the association is approximately linear over most of the investigated range. This result characterizes the airborne dataset itself; it does not constitute a direct comparison with Geant4-generated spectra or detector responses.

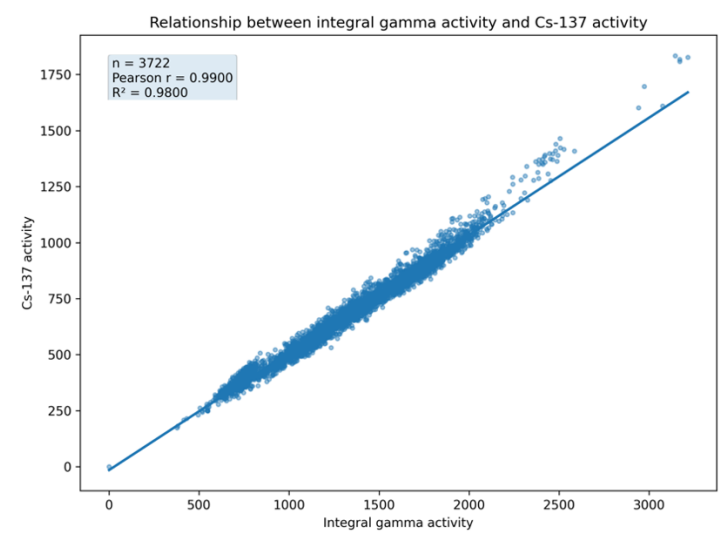


Fig. 6. Relationship between integral gamma activity and estimated ^{137}Cs activity in the independent UAV radiation-survey dataset. The fitted relationship characterizes the airborne measurements and is not a direct comparison with Geant4-generated data.

To evaluate spatial stability, the analysis was repeated separately for each flight transect. Pearson's correlation coefficients ranged from 0.9816 to 0.9943, while R^2 values ranged from 0.9635 to 0.9886. The consistently high coefficients show that the observed association was not restricted to a single flight trajectory and remained stable across spatially distinct transects. The transect-level results are summarized in Figure 7.

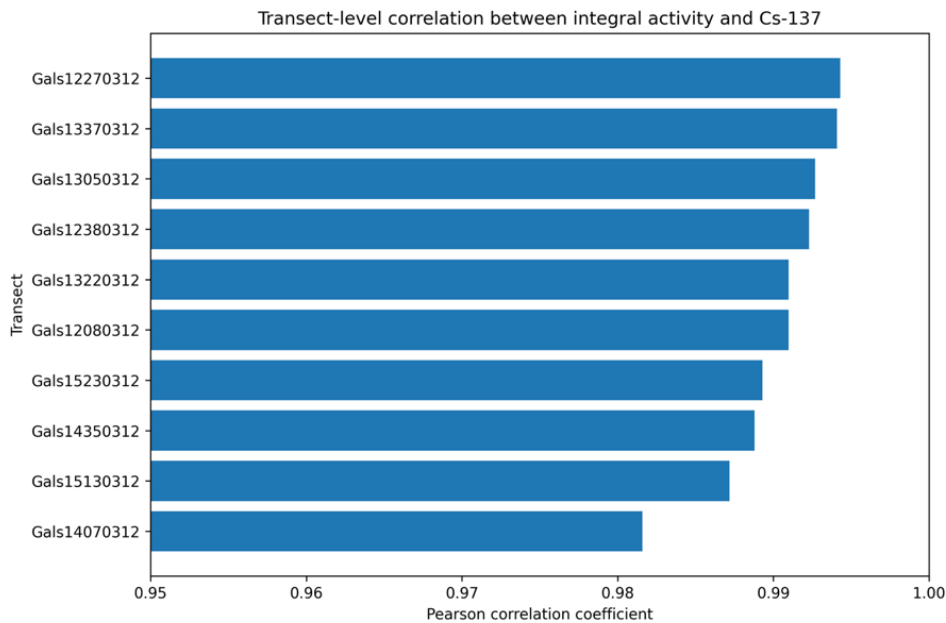


Fig. 7. Transect-level Pearson correlation coefficients between integral gamma activity and estimated ^{137}Cs activity across ten independent UAV flight transects.

Residual analysis is presented in Figure 8. The residuals remained approximately centred around zero, although their dispersion increased at higher activity levels. This heteroscedastic pattern indicates that prediction uncertainty was not constant over the measured

range. It may reflect environmental heterogeneity, spatial variation in contamination, measurement conditions, or counting-statistical effects. Consequently, the residual behaviour should not be described as fully unbiased without qualification.

Overall, the UAV results establish a stable application-level relationship between integral gamma activity and estimated ^{137}Cs activity in the Chornobyl survey dataset. They provide an operational context in which physically interpretable radiation variables exhibit consistent statistical behaviour across multiple flight transects.

However, these results do not validate the spectral fidelity, absolute detector efficiency, radionuclide-decay representation, or shielding response of the Geant4 model. The UAV dataset contains integral and derived activity variables rather than raw spectra acquired under a geometry corresponding to the simulation. It was therefore not used to calibrate the Digital Twin or to estimate simulation parameters. The UAV analysis is interpreted exclusively as an independent application-level consistency assessment.

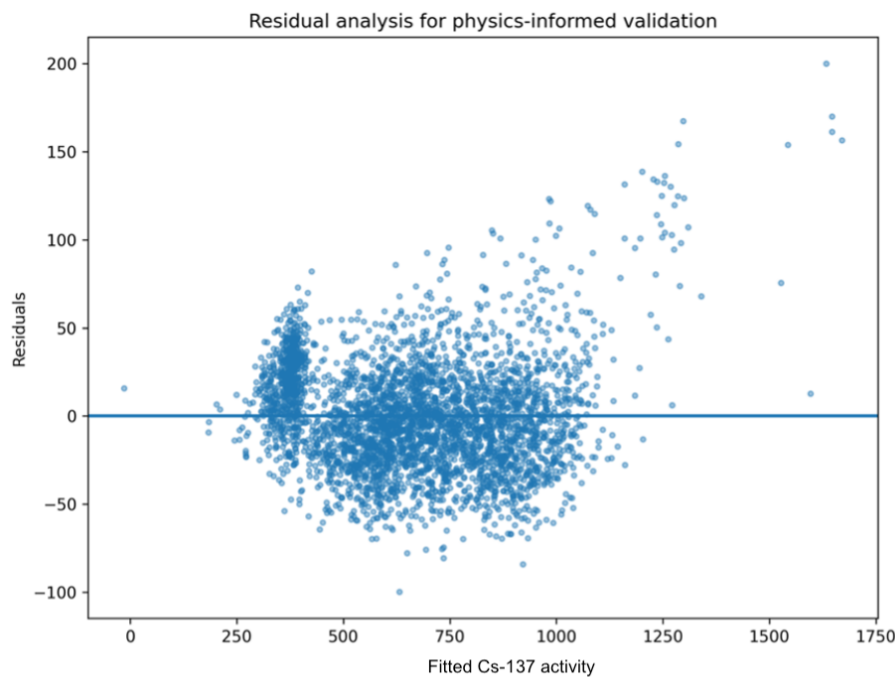


Fig. 8. Residuals of the linear relationship between integral gamma activity and estimated ^{137}Cs activity in the UAV dataset. The increasing residual dispersion at higher activity levels indicates heteroscedasticity.

4.6. Practical Implications of the Proposed Digital Twin Framework

Beyond its role as a synthetic data generation tool, the proposed Digital Twin framework can be viewed as a reusable computational infrastructure for the development, testing, and evaluation of modern radiation monitoring systems. The integration of Monte Carlo simulation, automated scenario orchestration, structured dataset construction, and explicitly separated assessment procedures provides a configurable virtual environment in which a range of controlled radiation-transport conditions can be reproduced.

A key practical advantage of the proposed approach is its ability to systematically generate representative gamma-ray spectra for combinations of radionuclides, shielding materials, detector configurations, and environmental conditions that are difficult or impossible to investigate experimentally. In many real-world applications, including nuclear facility monitoring, emergency response planning, and border security inspection, access to radioactive sources is restricted, while the controlled reproduction of hazardous scenarios is often

infeasible. The Digital Twin provides a virtual testbed that allows such scenarios to be explored safely and reproducibly through computational experiments.

The framework is also relevant to the future development and benchmarking of intelligent radiation-monitoring algorithms. Rather than replacing real measurements, the generated synthetic data can complement available experimental datasets by increasing the coverage of controlled source, energy, shielding-material, and shielding-thickness configurations. The Geant4 results reproduce physically interpretable attenuation and buildup behaviour within the assumptions of the implemented model. The independent UAV analysis provides an application-level environmental context but does not demonstrate synthetic-to-real transfer or validate the spectral fidelity of the generated data.

An additional benefit of the proposed architecture lies in its modular and extensible design. The Python-based orchestration layer enables automated parameter variation, batch execution of Monte Carlo experiments, and structured dataset construction without modifications to the Geant4 simulation core. As a result, the framework can be readily extended to incorporate additional radionuclides, detector types, shielding configurations, or environmental scenarios. This flexibility supports its use not only for the present gamma-ray source identification task but also for a broader range of radiation monitoring and environmental assessment applications.

From a broader perspective, the presented Digital Twin may be regarded as a virtual laboratory for computational experimentation in radiation monitoring. By providing a reproducible workflow that links physics-based simulation, structured synthetic-data generation, attenuation verification, and separate experimental and application-level assessments, the framework facilitates the systematic evaluation of data-processing and decision-support methods under controlled conditions. Such an approach is particularly valuable for safety-critical applications, where the availability of representative experimental datasets is inherently limited and where reliable computational surrogates can substantially accelerate the development and verification of next-generation monitoring technologies.

5. LIMITATIONS AND FUTURE RESEARCH

5.1. Limitations

The results of this study should be interpreted in light of several limitations related to the source representation, detector model, Monte Carlo statistics, experimental assessment, and use of the independent UAV dataset.

First, the simulated radionuclide classes were represented by characteristic monoenergetic primary gamma emissions rather than by complete radioactive-decay schemes. Consequently, the current model does not explicitly reproduce branching ratios, multiple emission lines, correlated gamma cascades, internal-conversion processes, daughter-product evolution, or secular-equilibrium conditions. The radionuclide names used in the synthetic dataset therefore identify representative photon-energy classes and should not be interpreted as complete simulated spectra of the corresponding radionuclides. This simplification is suitable for analysing energy-dependent attenuation and shielding effects but limits direct comparison with experimentally measured multi-peak spectra.

Second, the detector was represented as a sensitive volume recording deposited energy, while several instrument-specific response mechanisms were not modelled explicitly. These include energy-resolution broadening, energy-dependent detection efficiency, electronic noise, gain drift, dead time, pulse pile-up, detector housing, photomultiplier response, and detailed signal-processing electronics. Accordingly, the Geant4 output represents simulated energy-deposition data rather than a complete electronic spectrum produced by a specific measurement instrument. The present study does not claim validation of absolute detection efficiency, absolute source activity, or complete agreement with the shape of an experimental spectrum.

Third, the synthetic dataset contained one Monte Carlo realization for each source–material–thickness configuration. Statistical uncertainties in transmission were therefore estimated from event-counting statistics rather than from variability across independent simulation runs. Configurations with very low or zero accepted detector counts produced comparatively large or asymmetric uncertainties. In particular, zero-count observations could not be included in logarithmic attenuation fitting using conventional Gaussian error propagation. Repeated simulations with independently initialized random-number sequences and larger numbers of primary histories would permit a more comprehensive convergence and uncertainty analysis.

Fourth, the laboratory NaI(Tl) measurements provided only a limited detector-level spectral assessment. Energy calibration was based on a single confirmed ^{137}Cs reference point at 661.657 keV. A zero-intercept channel-to-energy relationship was therefore used for approximate peak localization, but possible calibration offset and non-linearity could not be independently quantified. This limitation is most relevant in the low-energy region, where deviations between observed and reference peak positions were larger.

The experimental geometry and source characteristics were also incompletely specified. Detailed information on the NaI(Tl) crystal housing, inactive layers, optical assembly, detector–container offset, Marinelli-container wall thickness, and exact source-matrix composition was unavailable. In addition, the available background measurement was not acquired with an empty Marinelli container under fully matched conditions. The calibration source activity had a stated uncertainty, and the activities of the other measured sources at the acquisition date were not available. These factors precluded reliable evaluation of absolute efficiency and absolute peak-area agreement between simulated and measured data.

Fifth, the blind radionuclide-identification assessment was restricted to one multi-peak spectrum with subsequently disclosed ground truth. The spectrum was assigned to material containing the ^{232}Th decay chain before disclosure, and this assignment was confirmed by the data provider. Although this result demonstrates the feasibility of blind multi-peak interpretation in a proof-of-concept case, it does not provide a statistically supported estimate of classification accuracy, sensitivity, specificity, or robustness across source types, activities, acquisition times, and background conditions.

Finally, the UAV radiation-survey dataset was not directly comparable with the Geant4-generated energy-deposition data. The airborne dataset contained integral gamma activity and estimated ^{137}Cs activity acquired under spatially variable environmental conditions, whereas the computational model described controlled source, shielding, and detector configurations. The UAV analysis therefore assessed the strength and cross-transect stability of an operational radiation relationship within the independent survey dataset. It was not used to calibrate the Geant4 model, update simulation parameters, evaluate synthetic-to-real transfer, or validate full spectral fidelity. The strong association observed across the ten flight transects should consequently be interpreted as application-level consistency evidence rather than detector-level validation of the Digital Twin.

These limitations define the present framework as a physics-guided and reproducible computational prototype rather than a complete metrological Digital Twin of a specific gamma-spectrometry system. Future validation should use complete decay schemes, experimentally calibrated detector-response functions, repeated Monte Carlo runs, matched source–detector geometries, empty-container background measurements, and a larger independently labelled blind-spectrum dataset.

5.2. Future Research Directions

Future work will focus on extending the physical and experimental fidelity of the proposed framework while preserving its configurable and reproducible architecture.

First, the monoenergetic source representation will be extended to complete radionuclide-decay schemes, including multiple gamma-emission lines, branching ratios, daughter products, and relevant cascade processes. This extension will enable more realistic simulation of multi-peak spectra and improve comparability with experimentally measured radionuclide signatures.

Second, detector-specific response functions will be incorporated into the simulation and post-processing workflow. Planned developments include energy-dependent resolution broadening, absolute and relative detection efficiency, electronic noise, gain drift, dead time, pulse pile-up, detector housing, and experimentally calibrated channel-to-energy conversion. These additions will support detector-level comparison between simulated and measured spectra rather than qualitative peak-position assessment alone.

Third, the Monte Carlo evaluation will be expanded through repeated runs with independently initialized random-number sequences and larger numbers of primary histories. Convergence analysis and variance estimation across independent realizations will provide a stronger basis for uncertainty quantification, particularly for highly attenuated configurations with low or zero detector-event counts.

Fourth, future experimental studies will use matched source–detector geometries, documented source matrices, empty-container background measurements, and multi-point energy calibration. A larger set of independently labelled blind spectra, covering different radionuclides, activities, acquisition times, shielding conditions, and background levels, will allow statistically supported evaluation of identification performance.

Finally, future application-level studies will examine how physics-guided synthetic data can complement operational radiation-monitoring datasets without treating integral field measurements as direct spectral validation. Such studies may evaluate domain adaptation or transfer-learning strategies only when synthetic and real datasets are generated with compatible detector configurations, feature definitions, target variables, and measurement protocols.

6. CONCLUSIONS

This study developed a physics-guided Digital Twin framework for automated synthetic data generation and evaluation in radiation-monitoring applications. The framework integrates Geant4-based Monte Carlo radiation transport, configurable scenario orchestration, detector-response aggregation, structured dataset construction, physical verification, and data-driven evaluation within a unified computational workflow.

The implemented model represented eight radionuclide-labelled source classes using characteristic monoenergetic gamma emissions between 0.051 and 2.610 MeV. Four shielding materials—lead, stainless steel, concrete, and polyethylene—were evaluated at thicknesses of 0, 1, 2, 5, and 10 cm, producing 160 source–material–thickness configurations. The radionuclide labels correspond to representative photon-energy classes rather than complete radioactive-decay schemes; therefore, the current implementation should be interpreted as a configurable radiation-transport model and not as a complete spectrometric representation of each radionuclide.

The simulated detector responses reproduced the expected dependence of photon transmission on energy, shielding material, and absorber thickness. Lead produced the strongest attenuation, whereas polyethylene exhibited substantially weaker attenuation and, in selected low-energy configurations, buildup-related transmission values exceeding unity. After incorporating counting-statistical uncertainties, the fitted linear attenuation coefficient for 1.25-MeV ^{60}Co -labelled photons in lead was $(0.618 \pm 0.042 \text{ cm}^{-1})$, with a 95% interval of 0.535–0.700 cm^{-1} . The corresponding NIST XCOM reference value of 0.680 cm^{-1} lies within this interval. This agreement supports the physical consistency of the simulated attenuation

behaviour while also demonstrating the importance of reporting uncertainty when the number of detected events is limited.

A limited detector-level experimental assessment was additionally performed using independently acquired NaI(Tl) spectra. The calibration spectrum confirmed the principal ^{137}Cs photopeak, while an independently labelled multi-peak spectrum was assigned to material containing the ^{232}Th decay chain before disclosure of the source identity. The subsequent confirmation of ^{232}Th supported the blind assignment in this proof-of-concept case. Because only one blind spectrum with disclosed ground truth was available, this result is not interpreted as a statistical estimate of radionuclide-identification accuracy or as complete experimental validation of the Digital Twin.

The independent UAV dataset demonstrated a strong and spatially stable relationship between integral gamma activity and estimated ^{137}Cs activity across ten flight transects. This result provides an application-level reference for environmental radiation monitoring but does not constitute direct validation of the Geant4-generated spectra. Detector-level spectral assessment was instead addressed separately using laboratory NaI(Tl) measurements and the blind ^{232}Th decay-chain identification.

The principal contribution of the study is therefore a transparent and extensible computational workflow that connects configurable radiation-transport simulations with structured data generation and multiple, explicitly separated levels of assessment. Physics verification was performed using attenuation behaviour and NIST XCOM reference data; limited detector-level spectral assessment was performed using laboratory measurements; and application-level consistency was examined independently using UAV radiation-survey data. Distinguishing these levels avoids treating integral environmental measurements as evidence of full spectral fidelity.

The present implementation remains subject to several limitations. It uses representative monoenergetic photon emissions rather than complete radionuclide-decay schemes and does not explicitly reproduce branching ratios, daughter-product evolution, or decay cascades. Detector-specific effects such as energy-resolution broadening, electronic noise, gain drift, dead time, absolute detection efficiency, and detailed detector housing were not fully modelled. The experimental assessment was limited by single-point energy calibration, the absence of a matched empty-container background measurement, and the small number of blind spectra with disclosed ground truth. In addition, one Monte Carlo realization was available for each simulated configuration, so uncertainty was estimated from event-counting statistics rather than between-run variability.

Acknowledgements

The authors gratefully acknowledge the technical staff of the Institute of Environmental Geochemistry of the National Academy of Sciences of Ukraine for providing the independently acquired NaI(Tl) spectral measurements and associated experimental information used in the detector-level assessment.

REFERENCES

Agostinelli, S., Allison, J., Amako, K.A., Apostolakis, J., Araujo, H., Arce, P., et al. (2003).

Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A*, **506**(3), 250–303.

[https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8)

Alabsy, M.T., Alzahrani, J.S., Sayyed, M.I., Abbas, M.I., Tishkevich, D.I., El-Khatib, A.M. & Elsafi, M. (2021).

Gamma-ray attenuation and exposure buildup factor of novel polymers in shielding using Geant4 simulation. *Materials*, **14**(17), 5051.
<https://doi.org/10.3390/ma14175051>

Allison, J., Amako, K., Apostolakis, J., Arce, P., Asai, M., Aso, T., et al. (2016).
Recent developments in Geant4. *Nuclear Instruments and Methods in Physics Research Section A*, **835**, 186–225.
<https://doi.org/10.1016/j.nima.2016.06.125>

Berger, M.J. & Hubbell, J.H. (1987).
XCOM: Photon cross sections on a personal computer (NBSIR 87-3597). National Bureau of Standards, Washington, DC.

de Oliveira, F.M., Daniel, G. & Limousin, O. (2023).
Artificial gamma-ray spectra simulation using generative adversarial networks and supervised generative networks. *Nuclear Instruments and Methods in Physics Research Section A*, **1047**, 167795.
<https://doi.org/10.1016/j.nima.2022.167795>

García-Esteban, J.J., Cuevas, J.C. & Bravo-Abad, J. (2023).
Generative adversarial networks for data-scarce spectral applications. *arXiv preprint*, arXiv:2307.07454.

Goyal, M., Mahmoud, Q.H. (2024).
A systematic review of synthetic data generation techniques using generative AI. *Electronics*, **13**(17), 3509.
<https://doi.org/10.3390/electronics13173509>

Hayamizu, R., Nakamura, S., Takashima, S., Kataoka, H., Sato, I., Inoue, N. & Yokota, R. (2024).
SIFTer: Self-improving synthetic datasets for pre-training classification models. In *Synthetic Data for Computer Vision Workshop*, CVPR 2024.

Herrera, A., Moore, E.F. & Heifetz, A. (2020).
Development of gamma background radiation digital twin with machine learning algorithms (ANL/NSE-20/64). Argonne National Laboratory.

Hossain, R., Ahmed, F., Kobayashi, K., Koric, S., Abueidda, D. & Alam, S.B. (2025).
Virtual sensing-enabled digital twin framework for real-time monitoring of nuclear systems leveraging deep neural operators. *npj Materials Degradation*, **9**(1), 21.
<https://doi.org/10.1038/s41529-025-00557-y>

Ivanov, A., Muhammad, B., Tonchev, K., Mihovska, A. & Poulkov, V. (2022).
UAV-based volumetric measurements toward radio environment map construction and analysis. *Sensors*, **22**(24), 9705.
<https://doi.org/10.3390/s22249705>

Jones, D., Snider, C., Nassehi, A., Yon, J. & Hicks, B. (2020).
Characterising the digital twin: A systematic literature review. *CIRP Journal of*

Manufacturing Science and Technology, 29, 36–52.
<https://doi.org/10.1016/j.cirpj.2020.02.002>

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S. & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6), 422–440.
<https://doi.org/10.1038/s42254-021-00314-5>

Knoll, G.F. (2010).
Radiation detection and measurement (4th ed.). John Wiley & Sons.

MacFadden, N., Peggs, S. & Gulliford, C. (2018).
Development and validation of a Geant4 radiation shielding simulation framework.
Brookhaven National Laboratory (BNL-211700-2019-TECH).
<https://doi.org/10.2172/1515417>

Mengyan, H., Xueyan, Z., Cuiting, P., Yixuan, Z. & Jun, Y. (2024).
Current status of digital twin architecture and application in nuclear energy field. *Annals of Nuclear Energy*, **202**, 110491.
<https://doi.org/10.1016/j.anucene.2024.110491>

Pant, H.J. (2025).
Applications of radioisotope and NDT techniques in industry. In *Proceedings of ICARST-2025* (IAEA-CN-332/810). IAEA.

Prokhorets, S. (2024).
Mathematical modeling of processes in neutron radiography. In *Proceedings of the International Scientific and Technical Conference “Information Technologies in Metallurgy and Mechanical Engineering”*, 311–315.
<https://doi.org/10.34185/1991-7848.itmm.2024.01.058>

Rathkopf, C. (2026).
Hallucination, reliability, and the role of generative AI in science. *arXiv preprint*, arXiv:2504.08526v2.

Saheli, F., Riazi, Z., Jokar, A., Shahabinejad, H., Vosoughi, N. & Ghasemi, S.A. (2021).
Evaluation of the nonlinear response function of an HPGe detector for 59 keV to 10.7 MeV gamma rays using Monte Carlo simulation. *Journal of Instrumentation*, **16**(07), P07003.
<https://doi.org/10.1088/1748-0221/16/07/P07003>

Schweiger, L. & Barth, L. (2023).
Properties and characteristics of digital twins: Review of industrial definitions. *SN Computer Science*, 4(5), 436.
<https://doi.org/10.1007/s42979-023-01937-4>

Shadura, O.V. (2019).
Stochastic optimization of the performance of a particle transport simulation package in high-energy physics (PhD thesis). Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine.

Shorten, C. & Khoshgoftaar, T.M. (2019).

A survey on image data augmentation for deep learning. *Journal of Big Data*, 6, 60.
<https://doi.org/10.1186/s40537-019-0197-0>

Singh, M., Singh, D.P. & Pandey, K. (2025).

Emerging trends in nuclear medicine: breakthroughs in radiopharmaceuticals, Geant4 simulations, and AI integration. *Isotopes in Environmental and Health Studies*.
<https://doi.org/10.1080/10256016.2025.2584330>

Tiboni, G., Protopapa, A., Tommasi, T. & Averta, G. (2023).

Domain randomization for robust, affordable and effective closed-loop control of soft robots. In *Proceedings of IEEE/RSJ IROS 2023*, 612–619.

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W. & Abbeel, P. (2017).

Domain randomization for transferring deep neural networks from simulation to the real world. In *Proceedings of IEEE/RSJ IROS 2017*, 23–30.

Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., et al. (2018).

Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *CVPR Workshops*, 969–977.

Tremblay, J., Meshry, M., Evans, A., Kautz, J., Keller, A., Khamis, S., et al. (2022).

RTMV: A ray-traced multi-view synthetic dataset for novel view synthesis. *arXiv preprint*, arXiv:2205.07058.