Applications of Machine Learning for Enhancing γ-ray Spectrometry in Nuclear Fusion

Kimberley Samantha Lennon

Sheffield Hallam University

This thesis is submitted in partial fulfilment for the degree of
Doctor of Philosophy

January 2026

I hereby declare that:

1. I have not been enrolled for another award of the University, or other academic or professional organisation, whilst undertaking my research degree.

2. None of the material contained in the thesis has been used in any other submission for an academic award.

3. I certify that this thesis is my own work. The use of all published or other sources of material consulted have been properly and fully acknowledged. I confirm that I have sought and obtained copyright permission for any third party materials included in this thesis.
   I used AI at AITS 2 (AI for Shaping) of the Artificial Intelligence Transparency Scale (AITS). I acknowledge the use of Microsoft Copilot (https:/copilot.microsoft.com/) to check the spelling and grammar in all sections of my thesis.

4. The work undertaken towards the thesis has been conducted in accordance with the SHU Principles of Integrity in Research and the SHU Research Ethics Policy, and ethics approval has been granted for all research studies in the thesis, as shown in the table below.

Ethics review reference number	Title of research study	Approval date	Date of post-approval amendments
ER65629887	Applications of machine learning for improvements in nuclear fusion diagnostics	21/05/2024	n/a

5. The word count of the thesis is 32,670.

Name: Kimberley Samantha Lennon
Award: Doctor of Philosophy
Date: 26th January 2026
College: Business, Technology and Engineering
Director of Studies: Dr Robin Smith

Abstract

Nuclear fusion represents a promising solution to the global energy challenge, offering safe and virtually limitless power. Diagnostics are critical for all life stages of fusion machines, with a core diagnostic being γ-ray spectrometry, used for operation, research, and waste characterisation. This thesis demonstrates machine learning can enhance high-purity germanium γ-ray spectrometry for nuclear fusion applications. Two key applications were identified: reducing the effects of Compton scattering and improving the absolute efficiency calculation process.

To achieve these objectives, two novel algorithms that leverage machine learning were developed. The first, the Machine Learning-based Compton Suppression Algorithm (MLCSA), employs a convolutional neural network to reduce Compton scattering effects through classification of pulses. The MLCSA reduced the Compton continuum by 97%, outperforming a traditional hardware suppression system by 10%. The second algorithm, the Machine Learning-based Efficiency Calculator (MaLBEC), uses a multilayered perceptron to improve the efficiency calculation process. The MaLBEC predicted absolute efficiency values within 6% and with an improved computational time of 99%, when compared to Monte Carlo N-Particle.

I dedicate this work in loving memory of my Grandad, Raymond James Bradley.

Acknowledgements

Professional

I would first like to thank my supervisors: Dr Robin Smith, Dr Chantal Shand, Dr Gemma Wilson, and Prof. Andrew Alderson. Their constant support and guidance made this work possible and I am grateful for all the time and knowledge they provided to support me.

Machine learning was a relatively new area to me prior to embarking on this PhD, and as such I have several people to thank who helped get me up to speed with the techniques in the early days, and who provided guidance throughout the project. First, the UKAEA machine learning group: Dr Rob Akers, Dr Stanislas Pamela, Dr Lorenzo Zanisi, and Vignesh Gopakumar. Second is Dr Ocean Wong, a fellow SHU PhD student who provided me with fundamental knowledge in my first few weeks that gave me a running start. I would also like to thank Dr Dominic Lennon, not only my husband but also a machine learning expert in the quantum field, he acted as an advisor throughout my PhD, guiding my work as any supervisor would, and insisting I learn and apply good coding practices.

Next I would like to like to thank the Applied Radiation Technology (ART) at UKAEA. Here thanks are needed for the general expertise given, mainly in the form of code tips and for helping when my code inevitably, constantly, and often in new and exciting ways, broke: Ross Worrall, Dr Tony Turner, Steve Bradnam, Andy Turner, Tim Eade, Hari Chohan, James Hagues, Dr Callum Grove, to name just a few.

Finally, I would like to extend my gratitude to UKAEA, who granted me the opportunity to undertake this PhD whilst continuing my career. I am so grateful that I was not forced to choose between the career I had started and the desire to follow my dreams to embark upon a PhD.

Personal

I start by thanking my husband, Dr Dominic Lennon. There are not enough words to express my gratitude towards you for your support over our 14 years together, none of this would be possible without you. You make me feel like anything is possible, you push me to excel everyday in everything I do, all the while being my anchor so that I can find my way back when things get too much. You make every single day magical, and when the PhD got tough you whisked me away to my favourite place, Efteling in the Netherlands, to reinforce the magic of life.

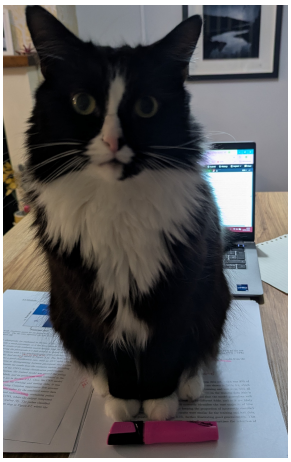
I would like to thank my parents, Paul and Wendy Neels. To my dad, you are my inspiration and I am only where I am today because of your belief in me and your drive to provide me with everything I need to be the best I can be. To my mom, thank you for always believing in me in anything I do, your constant emotional support is my lifeline. Thank you both for everything you have provided me, I will be forever grateful.

To my brother, Sam (Spam) Neels. Thank you for reviewing all of my figures over the years for conference posters and presentations, I am clearly not the artistic one in the family and you have saved many of my posters. Thank you for all of the music recommendations, providing me with a soundtrack to my PhD and to my life.

I thank the fluffy bean that is my cat, Oliver Lennon. I will always cherish the comfort and magic that this lunatic brings to everyday, and for the company on those long thesis writing days.

To my best friend Emma Pritchard, thank you for always being there for me, I am very lucky to have a friend like you as a constant in my life.

To my work bestie Callum, thanks for all the gifs(jifs), the emotional support that many daily gifs can provide should not be underestimated.



Oliver deciding thesis editing is over for the day

Publications

The work presented in this thesis has resulted in the following peer-reviewed publications.

Lennon, K., Shand, C. & Smith, R. Machine Learning Based Compton Suppression for Nuclear Fusion Plasma Diagnostics. J Fusion Energy 43, 17 (2024). DOI:10.1007/s10894-024-00408-9
(Chapter 3)

Lennon, K., Shand, C., Wilson, G. & Smith, R. Machine Learning Based Efficiency Calculator (MaLBEC) for Nuclear Fusion Diagnostics. J Fusion Energy 45, 6 (2026). DOI:10.1007/s10894-025-00546-8
(Chapter 4)

Table of contents

List of figures

List of tables

List of Abbreviations

CO_2	carbon dioxide
UKAEA	UK Atomic Energy Authority
JET	Joint European Torus
JDR	JET Decommissioning and Repurposing
MAST-U	Mega Amp Spherical Tokamak Upgrade
STEP	Spherical Tokamak for Energy Production
DEMO	DEMOnstration Power Plant
RADLab	Radiological Assay and Detection Laboratory
HPGe	high-purity germanium
IAEA	International Atomic Energy Agency
AI	artificial intelligence
MCF	magnetic confinement fusion
ICF	inertial confinement fusion
OOS	out of scope of regulation
LLW	low level waste
ILW	intermediate level waste
MDA	minimum detectable activity
FWHM	full width at half maximum
NaI	sodium iodide
CZT	cadmium zinc telluride
ADC	analogue to digital converter
MCA	multichannel analyser
PE	photoelectric absorption
CS	Compton scattering
PP	pair production
NORM	naturally occurring radioactive materials
MLCSA	Machine Learning-based Compton Suppression Algorithm
MaLBEC	Machine Learning-based Efficiency Calculator

List of Abbreviations

ANN	artificial neural network
CV	cross validation
CNN	convolutional neural network
MLP	multilayer perceptron
CVe	Compton veto ring
BEGe	broad energy germanium
DCSA	Digital Compton Suppression Algorithm
PSD	pulse shape discrimination
RF	random forest
SDGC	stochastic deep graph convolutional model
DT	decision tree
KNN	k-nearest neighbours
LinearSV	linear support vector
SVM	support vector machine
GaussianNB	Gaussian naïve Bayes
LR	logistic regression
SP	single-layer perceptron
XGB	extreme gradient boosting
TPR	true positive rate
ReLU	rectified linear unit
ROC	receiver operating characteristic
FPR	false positive rate
AUC-ROC	area under the ROC curve
SBR	signal-to-background ratio
GPU	graphical processing unit
GV	GammaVision
MCNP	Monte Carlo N-Particle
GEANT4	GEometry ANd Tracking
ISOCS	In Situ Object Counting System
RMSE	root mean squared error
SVR	support vector regression
MSE	mean squared error
nps	number of particle histories
XAI	explainable artificial intelligence

Chapter 1

Introduction

1.1 The fusion landscape

The global energy demand continues to rise, driven by population growth and industrialisation. To meet current and future energy demands, baseload power must be provided by a constant reliable energy source. The baseload is currently provided by a combination of fossil fuels, renewables, nuclear fission, and others, with fossil fuels providing the majority of the baseload. The burning of fossil fuels releases greenhouse gases, such as carbon dioxide (CO_2), which contribute to climate change and must be replaced with alternative methods. Renewables, such as wind and solar, are essential to support the baseload during high demand, but they face intermittency challenges and require large scale storage solutions. Nuclear fusion represents one of the most promising solutions to the global energy challenge. Fusion occurs when two light nuclei combine to form a single heavier one, while releasing vast amounts of energy. The process does not emit greenhouse gases, thereby offering a complementary approach with the potential to replace fossil fuels, offering a pathway to clean, safe (no chain reactions), and virtually limitless (fuel from the sea) power [1]. Achieving controlled fusion on Earth, however, is an immense scientific and engineering challenge.

The UK Atomic Energy Authority (UKAEA) is at the forefront of this effort, leading the UK's national fusion research programme and contributing to and supporting international projects and private fusion companies, as fusion experts. The UKAEA operated the record breaking fusion research facility, the Joint European Torus (JET), which is shown in Figure 1.1, on behalf of EUROfusion for 40 years. The JET facility holds the world record for fusion energy output, where high fusion power was consistently produced for 5 seconds, resulting in a ground-breaking record

of 69 megajoules. The JET facility is now undergoing decommissioning, since 2023, through the JET Decommissioning and Repurposing (JDR) project, which is also a world first and offers exciting opportunities to improve new technologies, provide insight on the impact of years of plasma operations, build the skills required for future fusion power plants, while learning how to minimise waste and maximise fuel recovery [2]. The UKAEA currently operate the Mega Amp Spherical Tokamak Upgrade (MAST-U), which focuses on the development of spherical tokamak technology which the UKAEA has long-supported, with particular focus on the development of divertor technology (which extracts heat and ash). Through initiatives like the Spherical Tokamak for Energy Production (STEP) program, the UKAEA aims to design and build a prototype fusion power plant by the 2040s, positioning the UK as a global leader in fusion technology. At the international level, fusion research is driven by large collaborative projects. The ITER (meaning 'the way' in Latin) machine, currently under construction in France, is designed to demonstrate the feasibility of fusion energy by exploring plasma behaviour at power plant-scale conditions [3]. Its central objective is to achieve a plasma that produces significantly more fusion power than the external power used to heat it, marking the first time a fusion experiment is expected to achieve a net energy gain from a confined and controlled plasma. Building on the scientific and technological advances achieved by ITER, the next part of the European fusion roadmap will be to build the DEMOnstration Power Plant (DEMO), to demonstrate net electricity generation and long-term operation, thereby bridging the gap between experimental fusion devices and commercial fusion power plants [4].

Figure 1.1: The JET fusion machine (with Kimberley Lennon in the foreground), located at the UKAEA's Culham Campus

1.2 Fusion diagnostics

Diagnostics are critical for every stage of the fusion lifecycle, including development, commissioning, operation, maintenance, and decommissioning. Among the suite of diagnostics available for fusion research is γ-ray spectrometry, which is a technique that enables the identification and quantification of radionuclides. There are many applications of γ-ray spectrometry, with one area being radioactive waste disposal and decommissioning, where radioactive waste must be characterised and the resulting activity assessments should be separated by individual radionuclides before disposal, as stipulated in [5]. Another area is in fusion research, where γ-ray spectrometry is a key technique that can be used to determine the purity of materials to be used in nuclear fusion machines, with results used as justification for fusion materials research to minimise the burden and cost of the consignment of waste, as described in [6]. Another example of a research application is the use of γ-ray spectrometry to provide information about the neutron energy spectrum, from neutron activation analysis of materials positioned near to fusion plasmas. The results can be used to indirectly calculate the neutron flux and fusion power, with possible applications in plasma control.

At the UKAEA, γ-ray spectrometry is utilised in both of the aforementioned applications of waste disposal and fusion research, through a state of the art diagnostics lab, the Radiological Assay and Detection Laboratory (RADLab), as shown in Figure 1.2. The experimental parts of this thesis work were conducted utilising some of the RADLab's static and portable γ-ray spectrometry detectors, specifically the high-purity germanium (HPGe) type.

Figure 1.2: The RADLab, at the UKAEA's Culham Campus. The photograph shows a corner of the RADLab, where static HPGe detectors are housed within lead castles (orange, green, blue, and pink)

1.2.1 Machine learning for fusion diagnostics

Machine learning is increasingly recognised as a transformative tool in nuclear fusion diagnostics, particularly in the context of γ-ray spectrometry and spectral data analysis. Machine learning-based algorithms have been applied to enhance γ-ray spectrometry analysis, especially in areas where traditional algorithms face limitations due to noise or spectral complexity. The article [7] provides a comprehensive review of the applications of machine learning in γ-ray spectrometry; the papers reviewed encompass a wide array of applications, from environmental studies to nuclear physics research, highlighting the versatility and impact of machine learning approaches in this field. Challenges in applying machine learning to γ-ray spectrometry include misidentification rates, sensitivity to changes in background radiation, and the need for extensive training datasets. Despite these challenges, the studies demonstrate the potential of machine learning to revolutionise processes such as radioisotope identification, optimisation of detector performance, and environmental monitoring [7].

Although the primary focus of this work is γ-ray spectrometry for fusion waste and research applications, it is important to contextualise this within the broader landscape of machine learning in nuclear fusion. The International Atomic Energy Agency (IAEA) produced a report [8] that outlines a broad range of use cases, including automated data processing, predictive modelling, and enhanced diagnostic capabilities in high-radiation environments such as those encountered in fusion reactors. More generally, strategic frameworks such as the UK's National artificial intelligence (AI) Strategy [9] emphasise the importance of machine learning and AI in scientific innovation, with specific reference to nuclear science and the need for collaborative infrastructure and open data standards. Collectively, these developments and strategic documents highlight the growing relevance of machine learning in fusion diagnostics, particularly in enhancing the capabilities and precision of γ-ray spectrometry systems. Chapters 3 and 4 contain a more comprehensive review of specific γ-ray spectrometry development requirements, which are not met by existing work, identifying a gap for novel improvements.

1.3 Thesis aims

This thesis addresses the gap in novel improvements to γ-ray spectrometry, by developing and evaluating machine learning-based algorithms tailored to the challenges of fusion diagnostics for waste characterisation and research applications. The central research question addressed in this thesis is: *'can machine learning be applied to enhance γ-ray spectrometry analysis, with a focus on nuclear fusion applications?'*. To address this central question, two key challenges of HPGe γ-ray spectrometry were identified: reducing the Compton scattering effect and improving the absolute efficiency calculation process. The primary contributions resulting from this work are two novel, machine learning-based algorithms designed to address these challenges, with further specific research questions addressed in the relevant chapters. This research aims to enhance diagnostic accuracy and efficiency, supporting the advancement of fusion diagnostics for next generation fusion machines. This thesis does not address real time deployment, integration with plasma control systems, application to other detector technologies such as scintillators or diamond detectors, or other nuclear industry applications of γ-ray spectrometry, as this is beyond the scope of the work.

1.4 Thesis outline

This thesis consists of five chapters. Chapter 2 provides the theoretical background relevant to the whole thesis, while the theory relevant to specific chapters is presented within those chapters. Chapter 3 introduces a novel machine learning-based algorithm for reducing the Compton scattering effect on spectra from γ-ray spectrometry, supported by relevant background and experimental validation. Chapter 4 presents a second novel machine learning-based algorithm aimed at improving the absolute efficiency calculation process in γ-ray spectrometry, supported by relevant background and experimental validation. Finally, Chapter 5 summarises the findings from all chapters, discusses the implications and limitations for fusion diagnostics, and outlines recommendations for future work.

Chapter 2

Theory

2.1 Nuclear fusion environments

2.1.1 Fusion reactions

Nuclear fusion is where two light nuclei are given energy to overcome the repulsive forces of the Coulomb potential and collide to form a heavier nucleus, with the release of energy [10]. To give the nuclei enough energy they are heated to a plasma, which is an ionised state of matter, comprising a mixture of positively charged ions and free electrons. To create a plasma in a fusion machine, often multiple different types of auxiliary heating methods are required, and these systems will vary depending on technological approach, for example whether the machine uses magnetic confinement fusion (MCF) or inertial confinement fusion (ICF). At the UKAEA, MCF has been the preferred approach, as it is the most mature technology used on the JET machine. In MCF machines, fusion fuel is heated in excess of 150 million degrees Celsius using ohmic heating, radio-frequency heaters, neutral beam injection [11]. For commercial fusion (i.e. a fusion reactor to produce net energy), three key parameters must be optimised, referred to as the Lawson criteria [10] or triple product: temperature, plasma particle density, and confinement time. These parameters are not maximised independently but instead they must be balanced to maintain plasma stability. For example, confinement time refers to the duration the plasma retains its energy at fusion relevant temperatures before significant losses occur, not simply how long the plasma is physically held. The most commonly used fusion fuels are the isotopes of hydrogen, deuterium (D, ^{2}H) and tritium (T, ^{3}H), due to their high cross section (measure of the effective area a nucleus presents for a fusion reaction to occur) for fusion at the lowest centre of mass energy (and therefore lower plasma temperatures). Figure 2.1 shows the cross sections of several fusion

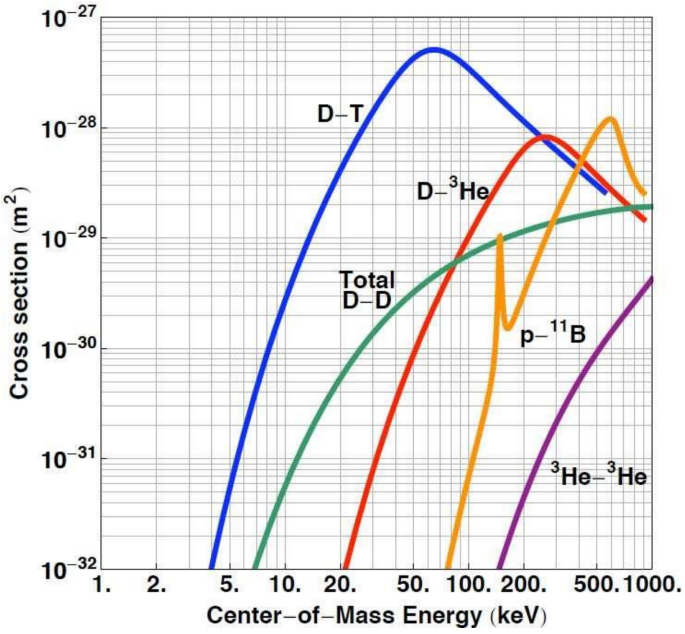


Figure 2.1: The DT reaction (blue) exhibits the highest cross section at relatively low energies (100 keV), making it the most favourable for future commercial fusion machines. The DD reaction (green) has a significantly lower cross section, requiring higher temperatures for comparable reaction rates [12]

fuel options, clearly demonstrating that DT is the optimal fuel mix. The reaction for DT is denoted as [10]

$$^{2}\mathrm{H} + {}^{3}\mathrm{H} = \mathrm{n} + {}^{4}\mathrm{He}, \tag{2.1}$$

where n is a neutron with energy of 14.1 MeV and ^{4}He is helium (an α) with energy 3.5 MeV.

Conversely, DD fusion has two equally probably pathways, namely

$$^{2}\mathrm{H} + {}^{2}\mathrm{H} = \mathrm{n} + {}^{3}\mathrm{He}, \tag{2.2}$$

where n has energy 2.45 MeV and ^{3}He has 0.82 MeV. The second possible DD reaction is

$$^{2}\mathrm{H} + {}^{2}\mathrm{H} = \mathrm{p} + {}^{3}\mathrm{H}, \tag{2.3}$$

where p is a proton with energy 3.02 MeV and T has energy 1.01 MeV. The neutrons released in the reactions from Equation 2.1 and 2.2 are the focus of this work, specifically the 14.1 MeV neutrons as DT will be the primary fuel used in the commercial fusion machines.

2.1.2 Neutron activation

The process of neutron activation in fusion can occur when the neutrons emitted in reactions such as those from Section 2.1.1 interact with nuclei in the surrounding material. Neutron behaviour depends strongly on energy, typically categorised as thermal (<0.5 eV), epithermal (0.5 eV–100 keV), and fast (>100 keV). Fast neutrons first lose energy through both elastic and inelastic scattering. As neutrons slow into the thermal range, inelastic scattering becomes negligible and neutron capture becomes the dominant interaction mechanism. Depending on the energy of the neutron and the reaction cross sections for elements in the material, the nucleus may become excited or transmute into another isotope, emitting prompt γ rays during de-excitation, and delayed γ rays later as the radioactive product decays [13]. The nucleus may also emit β particles, but these are not conventionally measured with γ-ray spectrometry and so have been omitted from this work. There are four primary reactions that can take place:

- **Neutron capture** - (n,γ), this dominates with low-energy (thermal) neutrons. When a neutron is absorbed by the nuclei it collides with, it forms a heavier and often excited nucleus. An example is in cobalt (Co), ^{59}Co(n,γ)^{60}Co, which goes from non-radioactive to radioactive Co.

- **Elastic scattering** - this is the process in which a neutron collides with a nucleus and transfers part of its energy without exciting the nucleus. The interaction leaves the target nucleus in its ground state, and the neutron continues with reduced energy and an altered direction.

- **Inelastic scattering** - this is where the neutron imparts some of its energy leaving the nucleus it collided with in an excited state, where de-excitation then coccus through γ emission.

- **Transmutation** - the changing of one element to another through reactions such as proton emission (n,p), neutron-proton emission (n,np), and alpha particle emission (n,α). These reactions dominate in fast neutron fields (over 0.5 MeV). An additional reaction in this category is the two neutron emission (n,2n) which is also known as a threshold reaction and requires a minimum neutron energy (unique to that interaction) to occur [13]. An example of transmutation is aluminium (Al) to Magnesium (Mg), ^{27}Al(n,p)^{27}Mg. Here fast neutrons can change ^{27}Al into ^{27}Mg, which is radioactive and then beta decays back to aluminium.

Neutron activation and fusion waste

Fusion machines are often constructed out of materials that readily activate, such as stainless steel. Steel is used due to its useful construction properties such as strength, economic viability, and established manufacturing capabilities. The neutron activation of composite materials, such as stainless steel, readily produces complex activation networks that might include multiple reaction types, resulting in the emission of a vast spectrum of γ energies [14]. Table 2.1 provides the dominant reactions and product half-lives (the time required for half of the nuclei in a radioactive sample to decay) for common elements in stainless steel, including iron (Fe), chromium (Cr), Nickel (Ni), manganese (Mn), molybdenum (Mo), and Co.

Table 2.1: The dominant neutron activation reactions in steel, the product half lives, and main γ energies [15]

Element	Reaction	Product nuclide	Half-life	Dominant γ energies (keV)
^{54}Fe	(n,p)	^{54}Mn	312.3 d	835
^{58}Fe	(n,γ)	^{59}Fe	44.5 d	1099, 1292
^{50}Cr	(n,γ)	^{51}Cr	27.7 d	320
^{58}Ni	(n,p)	^{58}Co	70.9 d	811
^{58}Ni	(n,np)	^{57}Co	271.8 d	122
^{55}Mn	(n,γ)	^{56}Mn	2.58 h	847, 1811
^{59}Co	(n,γ)	^{60}Co	5.27 y	1173, 1332
^{98}Mo	(n,γ)	^{99}Mo	65.9 h	740, 181

Material activation can hinder maintenance activities and can impact decommissioning [16]. For example, removing temporary structures from the exterior of the machine (torus hall) such as stairwells and scaffolding which may be in place during fusion pulses, or the removal of components which might have been damaged during operations. Once activated and removed from the fusion machine, items must be handled carefully either through safe storage and shielding until the radiation decays to safe levels, or by being disposed of as radioactive waste. The safe handling of radioactive items is critical because ionising radiation is harmful to humans, therefore monitoring and controlling exposure to radiation is vital. This type of component removal work is ongoing in the JDR project at the UKAEA [16]. Other materials that do not undergo neutron activation so readily are being researched to build future fusion machines [6, 17], selecting those that are the least activated but still have other desirable materials properties, such as high strength. This is an area that is heavily researched already, and is not the focus of this thesis.

When materials are disposed of as radioactive waste, these are consigned to waste storage/disposal facilities based on the activity from γ-ray spectrometry

measurements (discussed in detail in Section 2.2.8). Accurate classification of waste is essential, radioactive waste disposal facilities charge amounts that scale with the activity and therefore it is important that the activity is estimated accurately and not overestimated, as this will greatly increase the costs. Along with the potential financial burden, misclassification can lead to potential legal ramifications (especially when understating the activity of radionuclides), and reputational damage which can impact public perception of the nuclear sector and could influence government funding decisions, which is especially important for nuclear fusion research. The different waste categories and the requirements for each category are defined in UK legislation [5] as per Table 2.2. The table contents been limited to the categories relevant to fusion, as fusion wastes do not produce heat generating waste (unlike nuclear fission), which include: out of scope of regulation (OOS), low level waste (LLW), and intermediate level waste (ILW).

Table 2.2: UK radioactive waste categories and activity limits relevant to fusion [5]

Category	Beta/γ (Bq/g)	Alpha (Bq/g)
OOS	Specific per radionuclide	Specific per radionuclide
LLW	$<$ 12,000	$<$ 4,000
ILW	$>$ 12,000	$>$ 4,000

Neutron activation and fusion research

While neutron activation of materials is problematic for maintenance and decommissioning activities, it can also be beneficial for research activities. Neutron activation of materials can be used as a fusion diagnostic tool, specifically activation foils, such as those used in [18–23]. Activation foils are a form of indirect neutron measurement, which require irradiating materials of known composition and activation pathways, each with specific neutron energy dependent cross sections and threshold energies. Following removal from the source, the resulting γ signatures are analysed via γ-ray spectrometry to determine characteristics of the neutron field. These characteristics include the energy spectrum and neutron fluence, where fluence is the total number of neutrons that have passed through a unit area during a given time period (n cm^{-2} s^{-1}). Through this method fusion power can be inferred, since one 14.1 MeV neutron is released per fusion reaction in DT fusion, counting the neutrons provides a direct measure of the fusion power. The activation foil method also has potential for estimating fuel burn-up [24, 25] and for materials research [17]. Given the importance of the applications of the γ-ray spectrometry analysis results in research, it is vital that the analysis is correct. Incorrectly determining the fuel consumption, through

poor γ-ray spectrometry results, could mean the fusion machine runs out of fuel prematurely, potentially ruining costly and important fusion experiments. This would damage the scientific research, cost money, and could lead to reputational damage (for the company but also for fusion research on a wider scale).

2.2 Radiation detection principles

A common method in the nuclear industry to identify and quantify the presence of γ radiation is γ-ray spectrometry. The method is well developed, with books such as [13, 26, 27] being the standards that provide detailed explanations on all aspects of radiation detection and γ-ray spectrometry. The principles of γ-ray spectrometry are discussed in this section.

2.2.1 Radioactive decay

Radioactive decay is a spontaneous change within the nucleus of an atom which results in the emission of particles or electromagnetic radiation. The modes are α, β, and γ decay. Here, β is the principal mode, which encompasses β^- decay where neutrons change into protons with the emission of an electron and β^+ decay where protons change into neutrons with the emission of a positron. Following α or β decay, γ radiation is commonly emitted, when the daughter nucleus remains in an excited state and subsequently releases surplus energy in the form of γ rays, which typically range from a few keV to several MeV. Thus, γ emission is a mechanism for de-exciting the nucleus to its ground state (which is the lowest energy configuration for that nuclide), without changing its atomic number or mass number [26].

Decay schemes, such as the one illustrated in Figure 2.2 for caesium-137 (^{137}Cs), visually represent the sequence of nuclear transitions that occur during radioactive decay, including the emission of α, β, and γ radiation. They show the energy levels of parent and daughter nuclei, the type of decay involved, and the energy of emitted particles or γ rays. Transitions of γ rays, depicted as a vertical arrows between energy levels, are classified as either prompt or delayed depending on the lifetime of the excited state from which they originate. Prompt γ emissions occur almost instantaneously (femtoseconds), while delayed γ emissions are associated with metastable states (isomers), which can persist for milliseconds to years before de-excitation.

The discrete energy levels associated with nuclear transitions make γ radiation particularly suitable for spectral analysis, as each radionuclide emits γ rays with

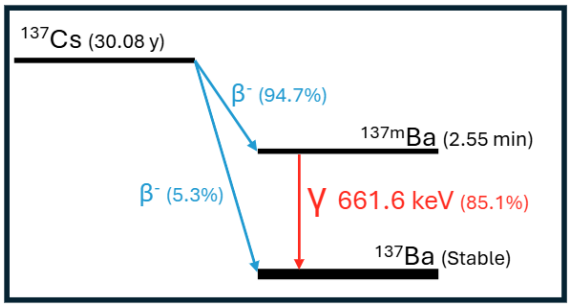


Figure 2.2: A simplified schematic of the ^{137}Cs decay scheme, illustrating the β^- transitions to the ^{137}Ba ground state and the metastable ^{137m}Ba states (blue), as well as the characteristic γ-ray transition from ^{137m}Ba to the ground state (red). Branching ratios and relevant half-lives are indicated

characteristic energies. In γ-ray spectrometry, decay schemes serve as a crucial reference for identifying radionuclides based on their characteristic γ emissions.

2.2.2 The γ-ray spectrometry process

The characteristic γ-rays emitted in decay schemes appear in practice as features within a measured spectrum. A γ-ray spectrum is essentially a histogram of detected γ energies, where discrete photopeaks correspond to the specific transitions of individual radionuclides. By linking each peak to its originating γ-ray energy and relating the peak area to the number of emitted γ rays, the spectrum provides the basis for both radionuclide identification and activity determination. The full γ-ray spectrometry measurement process typically involves several key steps, listed below and shown in Figure 2.3:

1. Sample preparation
    - Identify the item for measurement
    - Select an appropriate container, where relevant
    - Assign unique identifier for tracking
    - Record the sample geometry metadata (e.g. weight, dimensions, materials, material density, fill height, etc.)
    - Define the purpose of the measurement (e.g. waste characterisation, research measurement, etc.)

2. Detector preparation
    - Select the most suitable detector (Section 2.2.4)
    - Ensure the detector is energy calibrated (Section 2.2.5)
    - Perform energy calibration QA checks with reference sources
    - Efficiency calibration to determine an energy dependent efficiency curve for the measurement geometry (Section 2.2.3)
3. Data acquisition
    - Acquire a background spectrum if required (Section 2.2.7)
    - Select an appropriate live counting time, depending on the measurement purpose
    - Set the source-to-detector distance
    - Record all relevant measurement metadata
    - Enable Compton suppression or veto systems, if available (Section 3.1.1)
    - Acquire the spectrum for the predefined live time
4. Spectrum processing
    - Identify photopeaks and assign them to radionuclides
    - Fit photopeak shapes (typically Gaussian)
    - Subtract background, if applicable
    - Determine the net peak area for each identified photopeak (Section 2.2.8)
    - Apply residual corrections that are not already included in the efficiency calibration (e.g. coincidence summing)
    - Activity calculation: calculate activity for each relevant photopeak (Section 2.2.8), decay correction to a reference time if applicable (Section 4.2.1), and uncertainty propagation (Section 2.2.8)
5. Quality assurance
    - Verify sample dimensions and geometry metadata
    - Confirm that radionuclide assignments and activities are physically reasonable, e.g. by way of comparison with historical or expected values, if applicable

6. Reporting

    - Report activities
    - Provide associated uncertainties
    - Include minimum detectable activities (MDA) where relevant (Section 3.3.2)
    - Document all measurement conditions and metadata

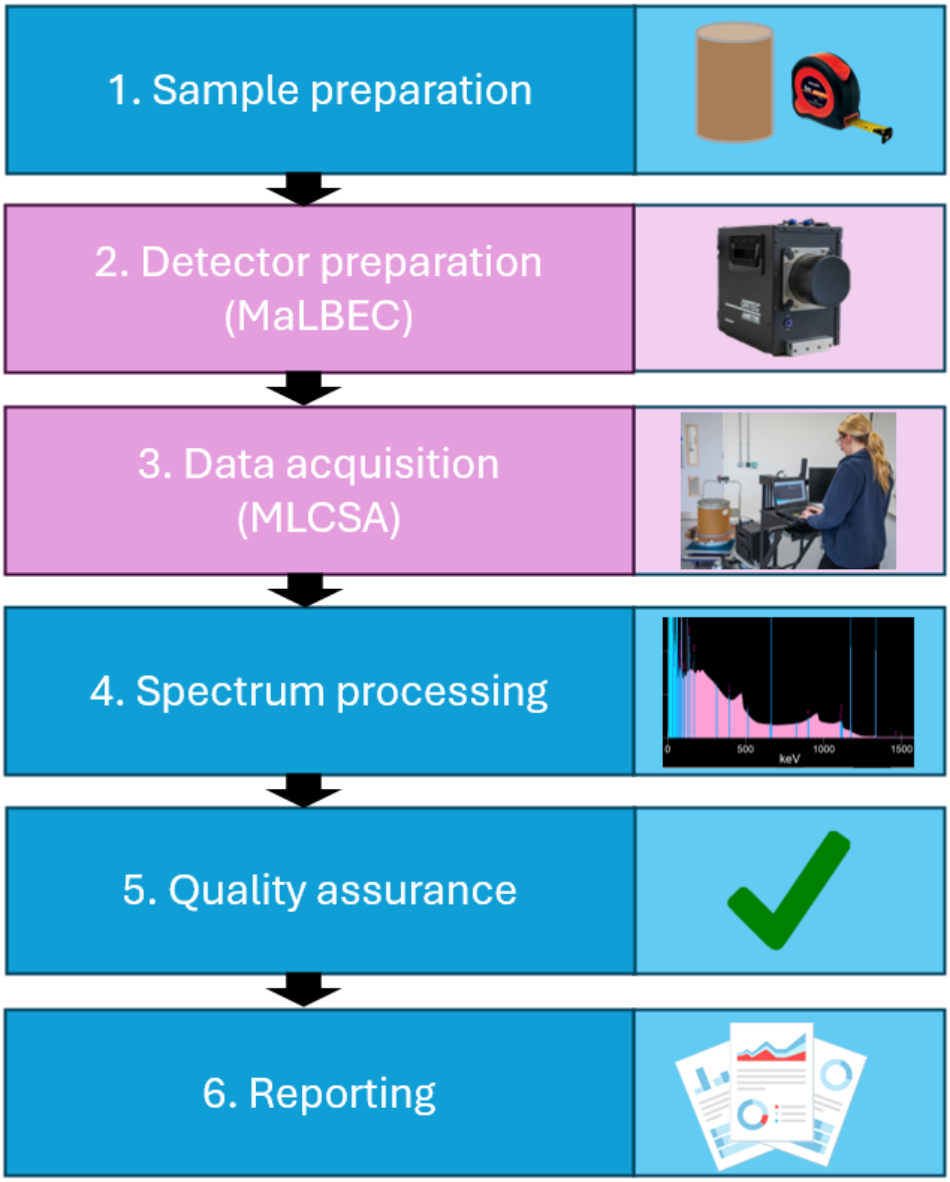


Figure 2.3: A diagram to show the key steps in the γ-ray spectrometry measurement process. The pink coloured boxes show the areas where the algorithms developed in this thesis are directly applicable to

2.2.3 Performance of γ-ray spectrometry detectors

The performance of γ-ray spectrometry is determined by two parameters - energy resolution and detection efficiency. Energy resolution, typically expressed as the full width at half maximum (FWHM) of a photopeak (see Figure 2.4), determines the ability to distinguish closely spaced γ lines and changes as a function of energy. Energy resolution is specified as the FWHM in keV at a given γ-ray energy (e.g. 1 keV at 1332 keV).

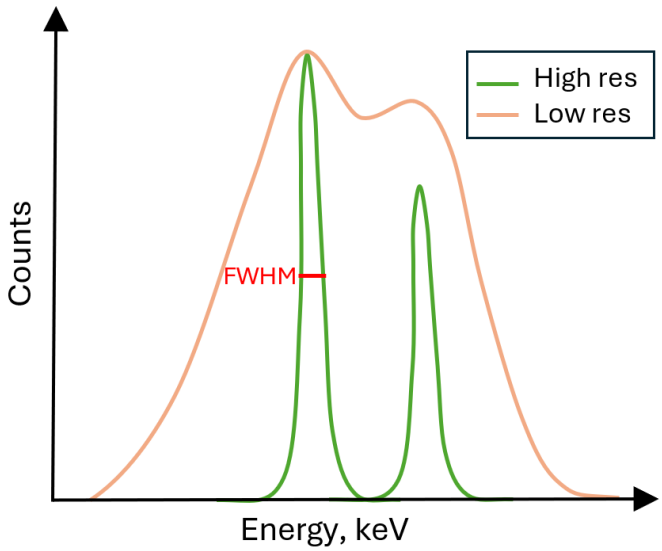


Figure 2.4: An illustration of how high resolution (green) compares in energy/photopeak width to low resolution (orange). The FWHM is shown for one of the high resolution peaks (red)

Detection efficiency quantifies the probability that a γ ray emitted from a source will be detected and correctly registered by the system. It depends on factors such as detector geometry, γ ray energy, source-detector distance, and shielding. Efficiency can be described as either absolute or relative; relative efficiency compares the detector's response to a reference standard - typically a 3"×3" thallium-doped sodium iodide (NaI(Tl)) detector at a source-detector distance of 25 cm, measured at 1332 keV using ^{60}Co - and is more commonly used to compare detectors of different sizes or types, while absolute efficiency considers the total number of γ rays emitted by the source and is used when the true detection probability is required. Absolute efficiency is typically expressed as a function of energy, producing an efficiency curve that characterises the detector's response across the energy range of interest. Efficiency curves exhibit a characteristic shape, rising at low energies before reaching a peak and then decreasing at higher energies, as illustrated in Figure 2.5.

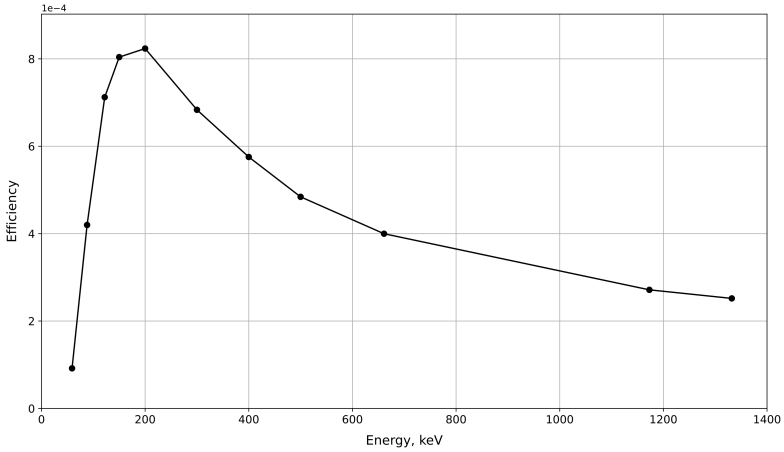


Figure 2.5: An example absolute efficiency curve showing the characteristic shape of detector efficiency as a function of γ-ray energy. Efficiency rises at low energies, reaches a peak at approximately 200 keV, and decreases at higher energies due to the reduced probability of γ-ray interaction within the detector volume.

2.2.4 Types of γ-ray spectrometry detectors

There are two broad families of γ-ray spectrometry detectors, semiconductor detectors and scintillators, and the choice between these two families depends on the budget available, performance requirements (higher resolution vs detection efficiency), and operational feasibility (cooling availability, portability, etc.). Scintillation detectors, such as NaI, convert the energy deposited from incoming γ-rays into flashes of light, which are then measured to determine the γ-ray energy. These offer high efficiency at room temperature, but at the expense of lower energy resolution. Semiconductor detectors, such as HPGe and cadmium zinc telluride (CZT), operate on the principle of converting γ-ray interactions into electrical signals, through charge collection in a crystal medium, and provide superior energy resolution. In this work, resolution is an important factor due to complex spectra arising from extensive activation networks, and so semiconductor detectors were used.

Semiconductors have an electrical conductivity that lies between that of conductors and insulators, meaning they have a relatively small band gap energy between the lower conduction and upper valence band, which readily enables charge carrier generation when ionizing radiation interacts with it [13]. The CZT crystals have a band gap of 1.4 to 2.2 eV, which doesn't require cooling as the gap is significant enough to not allow leakage currents, but provides a medium resolution (better than NaI scintillators but worse than HPGe). Whereas HPGe crystals have a small band gap of 0.67 eV, and due to the small band gap of germanium, thermal excitation at room temperature can produce significant leakage current (the noise level), degrading

performance. To reduce this, HPGe detectors are operated at cryogenic temperatures of ∼77 K, using either liquid nitrogen or electrically cooled cryostats, where the cooling reduces thermal noise and stabilises the detector response. Despite the inconvenience the cooling brings, the high resolution performance of HPGe detectors far outweighs this and so HPGe was selected for use in this work.

High-purity germanium detectors are manufactured in various geometries, each optimised for specific applications. For example, coaxial detectors feature a cylindrical geometry with a central bore and are commonly used for general purpose spectrometry due to their large active volume and consequently high efficiency. Another example is planar detectors, which use flat crystal surfaces and shallow contact structures that reduce electronic noise, providing excellent energy resolution, and enhanced sensitivity to low-energy γ detection. A final example is the Broad energy germanium (BEGe) detector, whose short, flat, non-rounded crystal geometry enhances efficiency below 1 MeV for typical close-geometry samples, while its thin, stable entrance window and wide operational energy range (3 keV – 3 MeV) provide excellent low-energy performance and resolution.

2.2.5 HPGe theory

In its pure form, germanium has an equal number of electrons and holes, which results in limited electrical conductivity. Therefore, to enhance electrical properties, n-type germanium is doped with donor atoms to introduce additional electrons into the conduction band, and p-type germanium is doped with acceptor atoms to create holes in the valence bands by accepting electrons. Where the p-type and n-type regions are brought into contact, they form a p-n junction (Figure 2.6). Here, the electrons from the n-type region diffuse into the p-type region and recombine with holes, while holes from the p-type region diffuse into the n-type region and recombine with electrons. This diffusion process leaves behind immobile ionised donor and acceptor atoms (essentially cancelling each other out, leaving no net charge), resulting in the formation of a depletion region, a zone that has no free charge carriers and is the area where γ-ray detection occurs. This region is extended by the application of a reverse bias voltage across the junction, which strengthens the electric field and sweeps carriers away from the p-n junction, widening the depletion region and effectively increasing the volume of the detector that is sensitive to ionising radiation. In HPGe detectors, the aim is to have a depletion region that covers as much volume of the crystal as possible, to maximise the likelihood of γ rays being detected [13]. A feature to consider is the dead layer, which is an inactive region on the surface of

the crystal, where radiation interactions do not produce measurable signal due to recombination or insufficient electrical field strength.

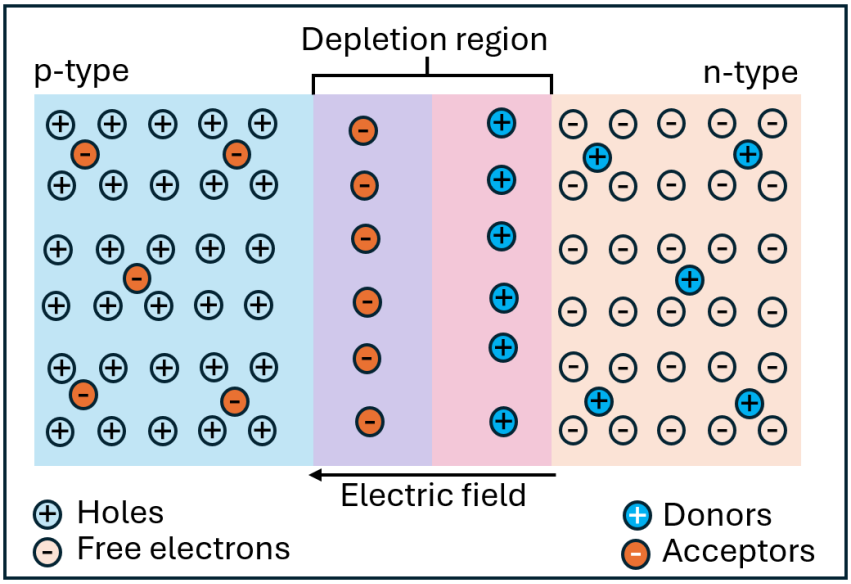


Figure 2.6: A simplified schematic of a p-n junction in a HPGe semiconductor detector, illustrating the formation of a depletion region when p- and n-type germanium are brought into contact

For detector fabrication, the bulk of the crystal can be doped to be either p-type or n-type, depending on the intended application. In the case of a p-type BEGe detector, as used in Chapter 3 of this thesis and described in detail in [28], this has a p+ point contact at the base (boron implantation) and a n+ external contact at the surface (lithium diffusion). A bias is applied to the external n+ contact (+5000 V in this work) and the p+ contact is grounded, therefore the n+ contact collects electrons and the p+ contact collects holes. Due to the large area of the external contact it has a high capacitance and is therefore noisy, hence, the collected signals are not used, meaning the collected signals come from the point contact. The BEGe detectors are discussed in more detail in Chapter 3.

When a γ ray interacts with the crystal within the depletion zone, the result is the liberation of an electron from the atoms within the depletion region, which goes on to generate electron-hole pairs through further interactions. The applied electric field across the depletion region causes the electrons and holes to drift toward their respective electrodes: electrons move towards the positive electrode and the holes move to the negative electrode. The movements of the charge carriers induce a current on the contacts, where a charge is collected over time and processed by the detector electronics, where the time taken for electron-hole pairs to reach the contacts is the charge collection time, the pulse integral is proportionate to the energy

deposited, and the measured pulse amplitude corresponds to the collection of both of the charge carriers, according to the Shockley-Ramo theorem [29]. The charge carrier drift and resulting induced signals for a BEGe detector are discussed in detail in [28]. The total charge collected is proportional to the energy deposited in the detector by the incoming γ ray [30]. During this processing interval the system cannot accept another pulse, creating a short period known as the dead time. Any γ-ray interaction that occurs during this interval will not be recorded. It is important to distinguish that this dead time does not arise from the detector or the charge collection process itself, multiple interactions may occur during charge drift, and preamplifier pulses can overlap without preventing further pulses from being generated. Instead, the dead time is introduced by the pulse processing electronics, particularly the shaping and digitisation stages. In typical modern digital systems, dead time is determined by the pulse processing electronics and is tracked continuously by the acquisition system, which distinguishes between real time and live time (the period during which the system is able to record events). Depending on the system behaviour, dead time losses can follow either a paralysable or non-paralysable model. In the non-paralysable model, events occurring during the dead time interval are lost but do not extend the dead time. In the paralysable model, any event arriving during the dead time interval will extend it further, meaning high rates can effectively shut the system down.

After charge collection, the signal is integrated and amplified by the preamplifier, then digitised by an analogue to digital converter (ADC). The digital processor applies baseline restoration, pole-zero correction, and pile-up rejection before performing digital shaping, typically with a trapezoidal filter, to reduce noise and stabilise the pulse. The flat-top amplitude of the shaped pulse is extracted as the event energy, and this value is sent to the multichannel analyser (MCA), which assigns the pulse to a channel based on its amplitude, which is directly proportional to the energy deposited by the γ ray in the detector. Over time, as many events are recorded, the MCA builds a histogram of counts versus energy, producing the γ-ray spectrum.

Since the channel numbers from the MCA are arbitrary and do not correspond to γ-ray energy, an energy calibration must be performed to enable quantitative analysis. This process establishes a relationship between channel number and energy, ensuring peaks in the spectrum correspond to their respective energies. This typically involves measuring known γ-ray energies from reference sources and fitting a calibration curve to link channel number to energy. Accurate calibration ensures that photopeaks in the spectrum correspond to their true energies, enabling reliable radionuclide identification and quantitative analysis.

2.2.6 Gamma-ray interactions

When a γ ray interacts within the depletion region of an HPGe crystal, it deposits energy through one of three energy-dependent processes; photoelectric absorption (PE), Compton scattering (CS), and pair production (PP) [26]. Figure 2.7 shows these three interactions within a HPGe crystal.

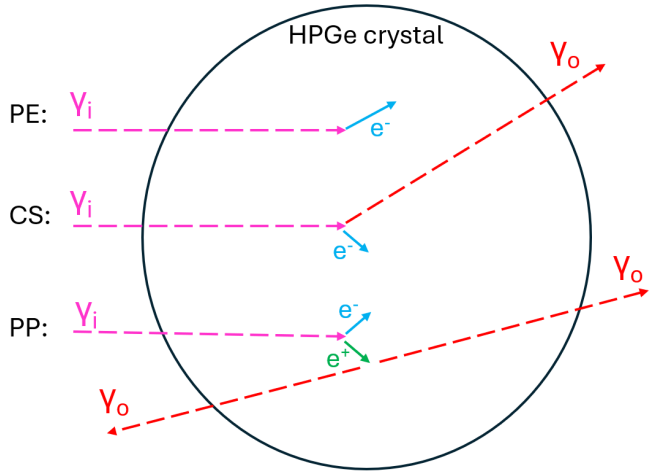


Figure 2.7: A simplified schematic of the three γ interactions within a HPGe crystal. Incoming γ rays are shown by γ_i (pink), escaping γ rays are γ_o (red), the electrons are e^- (light blue) and the positron is e^+ (green). The CS interaction shows a single scatter event. This figure is adapted from [13]

Photoelectric absorption

The PE interaction is dominant at energies < 150 keV and occurs when a γ ray interacts with a tightly bound electron. The γ ray is absorbed in an atom in the germanium crystal and ejects the electron from its shell, becoming a photoelectron, and the γ ray transfers nearly all its energy to the electron with kinetic energy equal to

$$E_e = E_\gamma - E_b, \tag{2.4}$$

where E_e is the energy of the ejected photoelectron, E_γ is the energy of the incident γ ray, and E_b is the energy binding the electron in its shell (E_b is negligible in this work since E_γ is typically thousands of times larger [26]). The photoelectron then travels through the depletion region, undergoing multiple ionizing collisions with surrounding atoms. Each collision liberates additional electrons from the valence band, creating electron-hole pairs. This continues until the photoelectron loses all its kinetic energy. The total number of electron-hole pairs generated is proportional to the initial energy of the γ ray, and these charges are collected by the detector's

electrodes under the influence of an electric field, producing a current pulse whose amplitude reflects the energy deposited.

Compton scattering

The CS effect, dominant at 150 keV – 4500 keV, occurs when an incoming γ ray collides with a loosely bound atomic electron, transferring part of its energy. The electron is ejected from the atom and is referred to as the recoil electron. This recoil electron behaves similarly to the photoelectron in PE, travelling through the depletion region of the crystal, ionizing atoms and creating electron-hole pairs, which are collected by the electrodes and contribute to the current pulse. If the scattered γ ray is eventually fully absorbed via PE absorption (i.e. it does not leave the detector), then this will be a pulse representing the full energy of the incoming γ ray. If the γ ray leaves the detector after a single scatter (as shown in Figure 2.7), then the resulting pulse corresponds to the energy deposited by the singular recoil electron, producing a low-energy event (where the energy is lower than the full γ ray energy) in the spectrum that does not reflect the energy of the γ ray. If the γ ray scatters multiple times before escaping, each scattered γ ray deposits partial energy and contributes to the pulse; the pulse will be larger than the pulse reflecting γ ray escape after a single scatter, but still will not reflect the full energy of the γ ray. The kinetic energy transferred to the recoil electron, E_r, is given by

$$E_r = E_\gamma - E'_\gamma, \tag{2.5}$$

where E'_γ is the energy of the scattered γ ray [26], which is defined as

$$E'_\gamma = \frac{E_\gamma}{1 + \frac{E_\gamma}{m_e c^2}(1 - \cos\theta)}, \tag{2.6}$$

where $m_e c^2$ is the rest mass energy of the electron which equals 511 keV (where m_e is the electron mass and c is the speed of light in vacuum) and θ is the scattering angle of the γ ray [13]. The probability of CS occurring in a material increases approximately linearly with its atomic number Z, since a higher Z implies a greater number of orbital electrons that serve as scattering targets. The Klein-Nishina formula describes the probability that an incident γ ray scatters through an angle θ into a solid angle $d\Omega$. The resulting angular distribution is strongly energy dependent: at lower energies (e.g. 10 keV), γ rays scatter over a broad range of angles with a slight preference for backscattering, whereas at higher energies (e.g. 10 MeV) the distribution becomes increasingly forward peaked. The increased probability

of forward scattering at higher energies implies that higher-energy γ rays are more likely to deposit their energy deeper into the bulk of a detector crystal [13].

The CS effect creates the Compton continuum in γ-ray spectra. With each collision the γ ray is scattered at an angle relative to its original direction, in the case that the γ ray is scattered 180° (backscatter) this gives the electron the maximum possible energy transfer in a single scatter and corresponds to the upper limit of the continuum, called the Compton edge [13]. The gap between the Compton edge and the photopeak is calculated with

$$E_{\gamma}^{\max} = E_{\gamma}\left(1 - \frac{1}{1 + \frac{2E_{\gamma}}{m_e c^2}}\right). \tag{2.7}$$

Pair production

The PP effect occurs when a γ ray with energy exceeding 1022 keV interacts with the electromagnetic field of a nucleus within the germanium crystal. This interaction can result in the conversion of the γ ray into an electron-positron pair (complete absorption of the incoming γ ray). The kinetic energy of the pair is given by the excess energy above the pair creation threshold, and both particles travel through the depletion region, ionizing atoms and generating electron-hole pairs [13]. These charges are collected by the electrodes and contribute to the current pulse. The positron eventually annihilates with an electron, producing two back-to-back 511 keV photons. If both annihilation photons are absorbed within the detector, the total energy deposited corresponds to the full energy of the original incident γ ray. However, if one or both (as shown in Figure 2.7) escape the detector volume, the resulting pulse will be reduced by 511 keV or 1022 keV, respectively, producing a single or double escape peak in the spectrum. The energy deposited by the electron-positron pair is given by

$$E_{\text{dep}} = E_{\gamma} - E_{\text{esc}}, \tag{2.8}$$

where E_{dep} is the energy deposited in the detector and E_{esc} is the energy carried away by escaping annihilation γ rays (either 511 keV or 1022 keV) [26].

γ-ray attenuation

When γ rays pass through matter, some are absorbed or scattered as a result of interactions such as the PE, CS, and PP effects, reducing the number that reach the detector. The extent of attenuation depends strongly depend on γ energy and

the materials involved, meaning the effect varies with sample matrices, container materials, and detector components. Although attenuation is not a major factor in this thesis work, a basic understanding of this behaviour is useful when considering how many γ rays from a sample ultimately contribute to the recorded spectrum.

2.2.7 Spectral features

There are several features in γ spectra that are important to consider during analysis, which are shown in Figure 2.8 and are described in the list below, where the first list is due to interactions within the detector, and the second list originate externally to the detector (e.g. shielding materials) [13, 26].

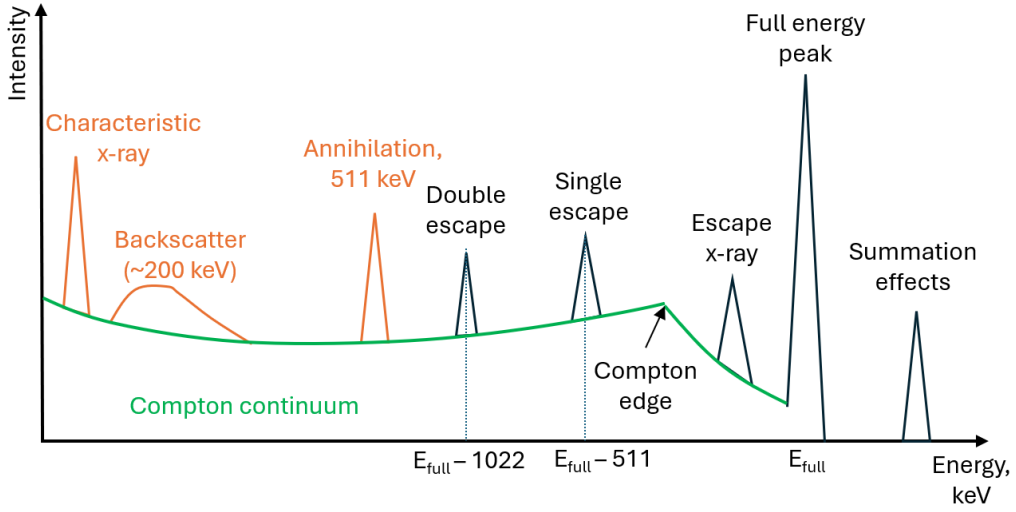


Figure 2.8: A schematic for the potential spectral features. Features originating externally to the detector are coloured orange. This figure is adapted from [13, 31]

Features resulting from interactions within the detector:

1. **Full energy photopeak** - occurs when the entire energy of the incident γ-ray is deposited in the detector, such as through the PE interaction.

2. **Compton continuum** - when higher energy γ rays interact within the crystal, they can Compton scatter out of the crystal and only deposit a fraction of their energy, which contributes to a continuum (CS interaction). This lies below the photopeak and covers the range from near zero up to the Compton edge.

3. **Compton edge** - the upper limit of the continuum (CS interaction) is marked by the Compton edge, corresponding to the maximum energy transfer in a single scatter (180° scatter angle).

4. **Single escape peak** - after PP, positron emission can result in one annihilation photon escaping the detector, resulting in a photopeak downshifted in energy by 511 keV.

5. **Double escape peak** - after PP, positron emission can result in both annihilation γ rays escaping, resulting in a photopeak downshifted in energy by 1022 keV.

6. **Summation effects** - some isotopes emit multiple γ rays in cascade during the decay process which can arrive at the detector almost simultaneously (in coincidence) creating a higher energy summed peak. There are two types of coincidence. An example of true coincidence is ^{60}Co which emits 1173 keV and 1332 keV γ rays to form a sum peak at 2500 keV. Random coincidence is where multiple γ rays arrive at the same time from different radionuclides.

Features resulting from interactions external to the detector:

1. **Annihilation peak (511 keV)** - full absorption of 511 keV annihilation γ rays from positron emission (e.g. from PP in the surrounding materials where the annihilation γ rays may enter the detector).

2. **Backscatter** - a γ ray may Compton scatter off surrounding materials (walls, shielding, detector housing) and re-enter the detector with reduced energy. This typically presents around 200 keV (from Equation 2.6) on the existing Compton continuum.

3. **Characteristic X-rays** - produced when inner shell electrons in atoms are ejected (e.g. by PE of γ rays), and vacancies are filled by electrons from higher shells. The energy difference between shells is emitted as an X-ray photon. These can be absorbed by the detector, producing small, low-energy peaks. Alternatively, they may escape the detector (rare) and appear as escape peaks at the original γ ray energy minus the X-ray energy. Examples include characteristic X-rays of detector materials (e.g. germanium at 9.9 keV) or shielding materials (e.g. lead at 75 keV or 85 keV).

4. **Natural background radiation** - background radiation arises from cosmic rays, naturally occurring radioactive materials (NORM), and residual radioactivity in laboratory materials and shielding. The NORM are commonly found in soil, building materials, and industrial by-products, includes isotopes such as potassium-40, uranium-238, and thorium-232. These environmental sources

of radiation introduce additional peaks in the spectrum, potentially interfering with the identification and quantification of target radionuclides and so must be accounted for - especially for low background applications or to distinguishing anthropogenic sources from natural ones. Methods for accounting for background radiation includes background subtraction, which involves taking a measurement without a sample present prior to measuring the sample, and shielding design.

2.2.8 Spectral analysis

Activity quantification

In nuclear physics, activity refers to the rate at which a radioactive substance undergoes nuclear decay, and is measured in becquerels (Bq), where 1 Bq corresponds to one disintegration per second. The activity, a, is calculated as

$$a = \frac{A_n}{\epsilon \times t \times Y}, \tag{2.9}$$

where A_n is net peak area (Equation 2.10), ϵ is the absolute efficiency, t is the live counting time (s), and Y is the γ emission yield. The parameter A_n is measured from the identified photopeaks in the spectrum, collected from a detector (HPGe detector in this work). For cases in which the peak lies on a background that cannot be subtracted by a background spectrum (as illustrated in Figure 2.9), the net peak area is derived as [32]

$$A_n = P - b_g, \tag{2.10}$$

where P is total counts (which includes the green and blue area in Figure 2.9) and b_g is

$$b_g = \frac{n_c}{2}(B_1 + B_2), \tag{2.11}$$

where B_1 is number of counts in the start channel, and B_2 is the number of counts at the end channel, and n_c is the number of channels in the peak region between B_1 and B_2 [32].

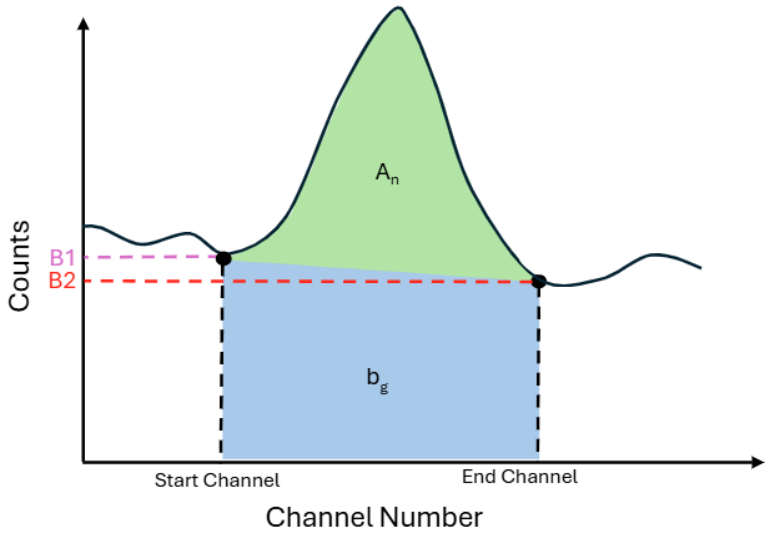


Figure 2.9: Visual representation of the terms in the activity and the signal-to-background ratio equation (Section 3.10). The green region (A_n) represents the net photopeak area, while the blue region (b_g) represents the background area estimated from the start and end channel counts (B_1 (pink dashed line) and B_2 (orange dashed line))

The absolute efficiency, ϵ, as discussed in Section 2.2.3, forms a key part of the activity calculation and is the ratio of counts in a full energy peak to the total γ-ray emission rate from the sample. The absolute efficiency, ϵ_{abs}, is expressed as

$$\epsilon_{\text{abs}} = \frac{N_{\text{detected}}}{N_{\text{emitted}}}, \tag{2.12}$$

where N_{detected} is the number of detected γ-rays in the full-energy peak, and N_{emitted} is the total number of γ rays emitted by the sample. Absolute efficiency is discussed in greater detail in Chapter 4.

Activity uncertainty quantification

The dominant contributions to the uncertainty in the activity calculation (Equation 2.9) arise from counting statistics, detector efficiency, measurement time, and nuclear data [33], all of which are included in the propagated uncertainty. Other generic sources of uncertainty in γ-ray spectrometry, such as thermal drift and dead time effects, are acknowledged but not treated explicitly, as they are either negligible for the measurements performed here or common to all approaches considered in this work. In this work, the symbol σ_x denotes the absolute uncertainty in the quantity x. The combined relative uncertainty is given by [33]

$$u_{\mathrm{rel}}(a) = \sqrt{\left(\frac{\sigma_{A_n}}{A_n}\right)^2 + \left(\frac{\sigma_\epsilon}{\epsilon}\right)^2 + \left(\frac{\sigma_t}{t}\right)^2 + \left(\frac{\sigma_Y}{Y}\right)^2}, \tag{2.13}$$

where σ_{A_n}/A_n is the relative uncertainty in net peak area (Equation 2.14), σ_ϵ/ϵ is the relative uncertainty in efficiency (Equation 4.4), σ_t/t is the relative uncertainty in live time (deemed negligible due to modern clock accuracy), and σ_Y/Y is the relative uncertainty in γ emission probability (nuclear data). The full uncertainty expression (Equation 2.14) formally includes a correlation coefficient [34] between the net peak area A_n and the live time t, this contribution is typically negligible in γ-ray spectrometry because the live time is measured with high precision and the statistical uncertainty in A_n dominates. For this reason, the correlation term is omitted here.

The relative uncertainty in the net peak area is defined as [33]

$$\left(\frac{\sigma_{A_n}}{A_n}\right)^2 = \left(\frac{\sigma_P}{A_n}\right)^2 + \left(\frac{n}{2}\cdot\frac{\sigma_{B_1}}{A_n}\right)^2 + \left(\frac{n}{2}\cdot\frac{\sigma_{B_2}}{A_n}\right)^2, \tag{2.14}$$

where σ_P/A_n is the relative contribution from peak counts (counting statistics), and σ_{B_1}/A_n and σ_{B_2}/A_n are the relative contributions from background counts. The underlying absolute uncertainties are estimated assuming Poisson statistics, with $\sigma_P \approx \sqrt{P}$, $\sigma_{B_1} \approx \sqrt{B_1}$, and $\sigma_{B_2} \approx \sqrt{B_2}$.

2.2.9 Summary

In summary, γ-ray spectrometry constitutes a complex set of γ-ray interactions within a detector, and it is important in nuclear applications to be able to accurately identify and measure radionuclides with a HPGe detector. One impact that hinders the ability to accurately quantify the low-energy γ rays is the CS interactions and the resulting Compton continuum. Reduction of this feature (called Compton suppression) has been investigated in literature, but with no perfect solution. A full literature review for existing methods is discussed in Chapter 3, Section 3.1. A second factor addressed in this thesis is the importance of the absolute efficiency calculation. Once radionuclides have been identified through γ-ray spectrometry the absolute efficiency is required as part of the activity calculation. Full details of the importance of this is included in Chapter 4, where a literature review of current methods has been undertaken as part of the production of a novel method.

Two HPGe detectors were used throughout the research presented in this thesis, with each being selected for its suitability to the specific experimental objectives. While this section has outlined general principles of HPGe detection, detailed de-

scriptions of these specific detectors, including their configurations and operational parameters, are provided in Chapters 3 and 4.

2.3 Machine learning theory

2.3.1 Introduction to machine learning

Machine learning is a sub-field of AI that involves developing algorithms and models that can learn general patterns and relationships from data and applying these relationships to new data to make predictions or decisions. The references [35] and [36] provide detailed background and methodology on how to use machine learning techniques and enables the reader to learn how to use machine learning methodically. This section provides background knowledge on the methodology for utilising machine learning models, which led to the development of two algorithms; the Machine Learning-based Compton Suppression Algorithm (MLCSA) and the Machine Learning-based Efficiency Calculator (MaLBEC).

Machine learning training type

Machine learning algorithms can be categorised into two types based on the training methods: supervised and unsupervised learning. In supervised learning, the model is trained using labelled data, where the correct outputs are already known and provided. This approach is effective when accurate mappings between input features and target variables are required, as in this thesis. In unsupervised learning, the model is trained using unlabelled data, where the objective is to find patterns or relationships in the data without any prior knowledge of what those patterns might be. The machine learning models in both technical chapters (Chapters 3 and 4) of this thesis are trained via supervised learning because the datasets consist of labelled measurements with known outputs, enabling the models to learn these mappings effectively. Therefore, the remainder of this section (and thesis) will only cover supervised learning methods.

Machine learning task type

Supervised machine learning models can be used for different types of tasks, including classification and prediction/regression. Classification refers to assigning inputs to discrete categories or classes. Prediction or regression involves estimating a continuous

numerical value based on input features [35]. The task in Chapter 3 is a classification task and Chapter 4 is a regression task.

Machine learning categories

Most machine learning models can be grouped into two categories: traditional and deep learning. Traditional learning methods require feature extraction, are often based on statistical algorithms, they are usually faster to train, and work well on medium sized datasets [35]. Deep learning methods make use of artificial neural networks (ANNs) and get their name from the common practice of training such models with many hidden layers. Deep learning methods typically work well on unstructured data (e.g. images, text, etc.), have feature extraction built-in, and require a large dataset and more computing power.

In addition to these model categories, ensemble learning methods can be applied to multiple weaker performing models, to form a single stronger performing model. Ensembles work via three strategies: bagging (Bootstrap Aggregating), Boosting, and Stacking. Bagging reduces variance by training multiple models independently on random subsets of the data and combines their predictions. For regression it averages outputs and for classification it takes a majority vote. Boosting reduces bias by training models sequentially, where each new model focuses on the mistakes of the previous ones. The predictions of all models are then combined, usually with weighted contributions. Stacking combines multiple different models (which can be any combination of models from the three categories) by training a combined model to learn how best to combine their predictions. The base models are usually trained in parallel. The combined model takes the outputs of the base models as input and produces the final prediction, effectively "learning how to learn" from the ensemble [35]. A combination of traditional and deep learning models, with and without ensemble methods, were considered in the selection process when building the MLCSA and the MaLBEC algorithms, full details of the models evaluated are discussed in Chapters 3 and 4 respectively.

Software implementation

All machine learning models in this thesis were implemented in Python. The *scikit* learn library [37] was used for classical machine learning algorithms and data preparation utilities. For the Convolutional neural network (CNN) framework used in Chapter 3, Keras was used for the implementation of the CNN. Data handling and processing were performed using NumPy and Pandas.

2.3.2 Model selection process

Selecting an appropriate model requires considering multiple factors from the previous section: the training type (e.g. supervised), the task the model is intended to perform (e.g. regression or classification), the category (e.g. traditional), the nature, structure, and amount of data available. This section discusses the process of selecting a machine learning model as used in this work, which includes evaluating many different models to identify the best performing one [35]. The selection process can be split into the steps listed below and are shown in Figure 2.10, with the outcomes of all steps are summarised in Chapters 3 and 4:

1. **Define the problem** - identify the training type, task type, and suitable categories.

2. **Data collection and exploration** - experimental or simulation data, understand the characteristics of the data, visualise and find correlations, identify what is useful, etc.

3. **Data preparation** - prepare the data for machine learning algorithms by creating data pipelines to take raw data to processed data. This could include data cleaning (remove missing values), feature engineering (creating useful features), data augmentation (create more varied data), feature selection (remove none-useful data), and feature scaling (e.g. normalising, standardising, etc.).

4. **Model selection and evaluation** - run the processed data on the appropriate machine learning models (identified in the first step) and evaluate the outcomes using held-out test data to select the best model.

5. **Fine tune the model** - optimise the hyperparameters through model tuning.

6. **Validation** - compare to an alternative method on an experimental sample as part of the validation process

7. **Interpret the results** - understand whether the results make physical sense. This has been included for this thesis work in the summary/discussion/limitation sections of Chapters 3 and 4.

8. **Prepare for large scale deployment** - build the final, tuned machine learning model into a full algorithm for wider use. This has been included for this thesis work in the future work sections of Chapters 3 and 4.

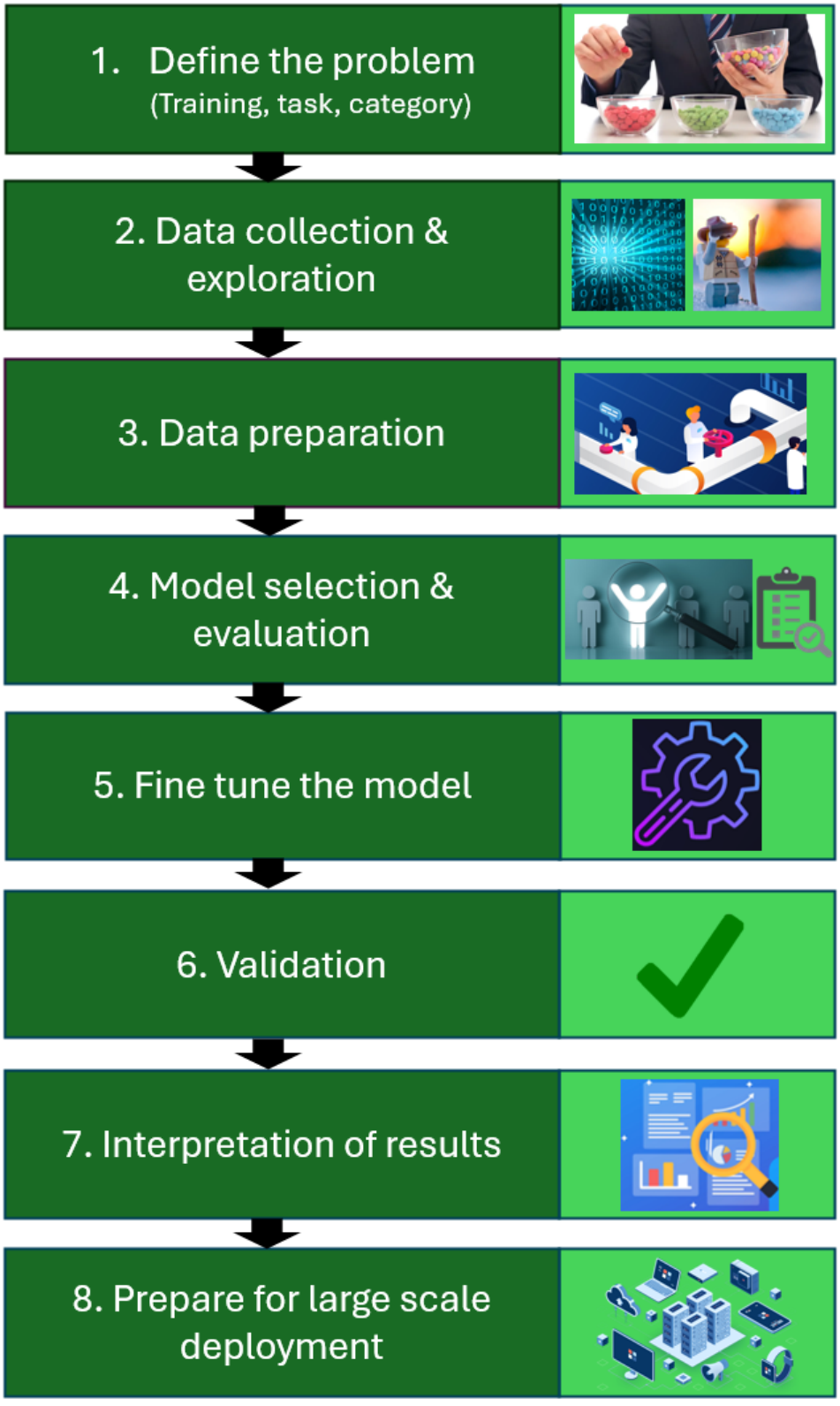


Figure 2.10: A diagram to summarise the steps machine in the learning selection process

Steps three, four, and five require further details which have been provided in the next few sections.

Data preparation (step 3)

Effective data preparation is a critical step for the successful application of machine learning-based algorithms to experimental physics data, to ensure that input data is representative and optimally structured for machine learning applications. Raw datasets, whether derived from simulations or measurements, often contain noise, missing values, or irrelevant features that can obscure underlying physical patterns and hence need to be removed. The data processing pipelines developed in this thesis were specific to each algorithm developed, the MLCSA in Chapter 3 and the MaLBEC in Chapter 4, therefore specific details are provided in those chapters, but some general techniques are listed here as they are relevant to both approaches:

- **Data splitting** is where the dataset are split into separate training and test subsets to enable reliable evaluation of model performance. The training set is used to fit the machine learning model, allowing it to learn patterns and relationships within the data. The test set, which remains unseen during training, is used to assess how well the model generalises to new, independent data. Typically the ratio of training to test data is 80:20, depending on the size and variability of the dataset. Care is required to ensure that the split preserved the distribution of key features and labels, which is discussed in the relevant chapters.
- **Data cleaning** is a technique that typically involves the removal of incomplete data and/or the smoothing of noisy signals.
- **Data augmentation** refers to the process of artificially expanding the training dataset, by applying controlled transformations to existing data. These transformations, such as shifting or adding noise, simulate realistic variations in experimental conditions while preserving the underlying physical characteristics. This improves model generalisation and robustness, especially in cases where collecting large volumes of labelled data is impractical.
- **Feature engineering** involves transforming raw data into more informative features that better represent the underlying physical processes. This may include extracting statistical summaries, ratios, or other domain-specific quantities that help the model learn more effectively from the available data.

- **Feature scaling techniques**, such as standardisation and normalisation, are applied when appropriate, for example for algorithms sensitive to feature magnitude [35].
- **Data labelling**, which is used in supervised training types, is where the data are labelled using known experimental outcomes or simulation metadata, ensuring that each input is paired with a reliable target value or class. This step is particularly important as mislabelled data could significantly degrade model performance.

Model selection and evaluation (step 4)

As part of the model selection process, a range of models were investigated using suitable techniques as identified in step one of the selection process. A brief description of the models evaluated in Chapter 3 and Chapter 4 is provided below, highlighting their key characteristics and differences. The models are grouped into traditional machine learning and deep learning approaches.
The following traditional models were evaluated:

1. Decision tree (DT) is an interpretable model that recursively partitions the feature space using threshold-based splits, serving as a simple baseline for classification in Chapter 3.

2. Gaussian naïve Bayes (GaussianNB) is a probabilistic classifier that assumes feature independence and models each feature with a Gaussian distribution, offering fast training and low complexity, evaluated for classification in Chapter 3.

3. Logistic regression (LR) is a linear probabilistic model that estimates class membership probability via a sigmoid function, providing an interpretable baseline for classification in Chapter 3.

4. Linear support vector (LinearSV) is a linear model that finds a maximum margin hyperplane, suited to linearly separable data and evaluated for classification in Chapter 3.

5. Support vector machine (SVM) extends the linear SVM with kernel functions to handle non-linear decision boundaries, evaluated for classification in Chapter 3, while its regression variant, support vector regression (SVR), is evaluated in Chapter 4.

6. K-nearest neighbours (KNN) is a non-parametric instance-based method that assigns a label or value based on the closest training points, applied to classification in Chapter 3 and regression in Chapter 4.

7. Random forest (RF) is an ensemble method that constructs multiple decision trees and aggregates their outputs, offering greater robustness against overfitting compared to a single tree, applied to classification in Chapter 3 and regression in Chapter 4.

8. Extreme gradient boosting (XGB) is a gradient boosting ensemble that builds trees sequentially, with each tree correcting the residual errors of the previous, applied to classification in Chapter 3 and regression in Chapter 4.

The following deep learning models were evaluated:

1. Single-layer perceptron (SP) is the simplest neural network architecture, consisting of a single layer of weights and capable of learning only linearly separable functions, evaluated for classification in Chapter 3.

2. Multilayer perceptron (MLP) consists of multiple fully connected layers capable of learning non-linear mappings, evaluated for classification in Chapter 3 and selected as the final model for regression in Chapter 4.

3. Convolutional neural network (CNN) employs convolutional filters to automatically extract hierarchical features, evaluated for classification in Chapter 3 and selected as the final model.

4. Stochastic deep graph convolutional (SDGC) incorporates graph-based convolution with stochastic regularisation, enabling it to capture relational structure in data, evaluated for classification in Chapter 3.

Machine learning models often have many parameters (called hyperparameters) that can be tuned to enhance the model performance and provide better results, so that the model is generalisable to new data and does not overfit or underfit. In machine learning, the concepts of overfitting and underfitting describe how well a model captures the underlying patterns in the data.

Overfitting occurs when a model learns not only the true signal but also the noise or random fluctuations in the training data, essentially memorising the training data, resulting in excellent performance on the training set but poor generalisation to unseen data (e.g. the test set) because it memorised noise instead of general patterns

[35]. To reduce overfitting and improve the generalisation performance of the model, the following techniques can be used (these are not in any specific order):

- Increase the amount of training data - providing more opportunities to learn general patterns and relationships in the data, making it less likely to memorise the training data.
- Simplify the model - e.g. by using simpler architectures, reducing the number of layers in neural networks, or by reducing the number of features used in the model.
- Regularisation - a technique used to simplify and constrain the complexity of the model and prevent overfitting. These are discussed in detail in the relevant chapters.
- Cross validation (CV) - a technique that splits the training dataset into multiple subsets (folds) and trains the model on some folds while validating on others, helping to assess generalisation and reduce overfitting.
- Data augmentation - artificially expanding the training dataset, as described above.
- Ensemble methods - the use of ensemble methods combine multiple models to make predictions, which can often lead to better performance and help prevent overfitting.
- Hyperparameter tuning - the process of adjusting the model's configuration parameters (e.g. number of layers) to reduce model complexity and prevent overfitting.
- Dropout - a regularisation technique commonly used in neural networks, where a random proportion of neurons are temporarily deactivated during training. This prevents the network from relying too heavily on specific pathways and encourages the learning of more robust, generalisable features.

Conversely, underfitting arises when a model is too simplistic or constrained to learn the relevant structure in the data, leading to poor performance on both training and test sets because it fails to capture the underlying patterns. To prevent underfitting, the following techniques can be used:

- Increase the complexity of the model - e.g. by adding more layers to a neural network, or increasing the number of features used in the model. This can help the model to capture more complex patterns and relationships in the data.

- Collect more data - if the model is underfitting, it may be because there is not enough data to learn the patterns and relationships in the data.
- Ensemble methods - bagging, boosting, and stacking can be used to improve the performance of underfitting models.
- Hyperparameter tuning - the process of adjusting the model's configuration parameters to improve performance and reduce underfitting.

There are many performance evaluation methods in machine learning to ensure the model generalises well, and these vary depending on the task [35]. For example, the evaluation metrics that are relevant to classification tasks may differ to those used for regression tasks. Therefore, only evaluation metrics that are relevant to both task types are included here, and more specific detail relevant to the MLCSA (classification) and the MaLBEC (regression) approaches is provided in Chapters 3 and 4 respectively.

One metric relevant to both classification and regression tasks is the CV score. The CV score is a widely used technique for assessing the generalisation performance of machine learning models, and relies on an underlying metric (e.g. accuracy). The core idea is to split the training dataset into multiple subsets, called folds, and iteratively train and evaluate the model across these folds. In k-fold CV (Figure 2.11), the data is divided into k equal folds; the model is trained on k–1 folds and evaluated on the remaining held-out fold, repeating this process k times so that each fold serves as the held-out test set once [35]. The CV score is typically reported as the mean performance across all folds, providing a robust estimate of the predictive capability of the model. This approach reduces the bias associated with a single train-test split and supports fair comparison between models. To further assess model behaviour, it is common to compare the mean CV score on the training folds versus the test folds. A significantly higher score on the training data relative to the test data may indicate that the model is overfitting. The CV method can be applied to both classification and regression tasks, with the choice of scoring metric tailored to the specific problem, these are explained in Chapters 3 and 4.

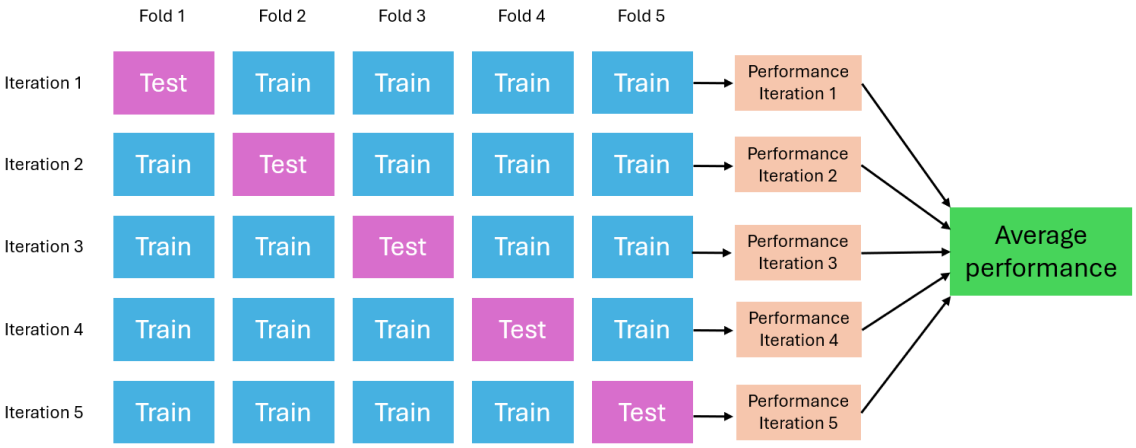


Figure 2.11: A diagram showing the CV score process, showing folds 1 to 5 (k-1 to k-5). Adapted from [38]

Another metric used in this work to assess both the performance of machine learning models and the agreement between measured and reference values, was the percentage difference calculation. This metric is defined as

$$\% \text{ Diff} = \left(\frac{a-b}{b}\right) \times 100, \tag{2.15}$$

where b refers to the reference value and a refers to the value being compared. The sign of this metric was retained to indicate whether the result was higher (positive) or lower (negative) than the reference, providing insight into systematic trends such as over-prediction or under-prediction.

Hyperparameter tuning (step 5)

Hyperparameter tuning/optimisation is the process of selecting the optimal hyperparameters for a given model (e.g. number of hidden layers in an ANN). Different combinations of hyperparameters are tested in training and their performance evaluated using CV and other evaluation metrics as previously discussed, where relevant. There are several methods for hyperparameter tuning, including grid search, random search, and Bayesian optimisation. Grid search involves specifying a set of values for each hyperparameter and trying all possible combinations of these values, which works best for a small parameter space. Random search involves randomly selecting hyperparameter values from a predefined range, which works best for a larger parameter space and it can help identify regions to specify for a grid search. Details of the implementation of grid search and random search processes are given in Chapters 3 and 4, where they have been used to develop the MLCSA and the MaLBEC.

2.3.3 Artificial neural networks

The final machine learning models identified from the selection process were both from the deep learning category, specifically ANNs; a CNN [39] for the MLCSA in Chapter 3 and a MLP for the MaLBEC in Chapter 4. Whilst the MLP consists of fully connected layers as described in this section, the CNN extends this architecture by replacing some hidden layers with convolutional layers, which apply learned filters to extract spatial or local features from the input data. The theory relating to ANNs that is relevant to both chapters is given here, with specific detail (e.g. CNN and MLP architecture) in the respective chapters.

Artificial neural networks are computational models inspired by the structure and function of biological neurons. They consist of layers of interconnected nodes/neurons (fundamental building blocks/computational units) that process inputs to produce outputs, and they are capable of learning complex non-linear relationships from data [39]. Artificial neural networks are widely used in the two task types discussed earlier (regression and classification).

The architecture of an ANN is organised into three layer types, as illustrated in Figure 2.12: an input layer, one or more hidden layers, and an output layer. Each neuron in a given layer receives signals from every neuron in the preceding layer. These signals are scaled by weights, which are scalar parameters assigned to each connection between neurons. The weight associated with a connection determines the strength and direction of influence that one neuron exerts on the next; a large positive weight amplifies a signal, a large negative weight suppresses it, and a weight near zero renders the connection negligible. Each neuron computes a weighted sum of its inputs and adds a bias term before passing the result through a non-linear activation function. The bias is a learnable parameter associated with each neuron (distinct from the connection weights) that shifts the activation function, allowing the neuron to control the threshold at which it activates independently of its inputs. Without a bias, the neuron's output would be entirely determined by its inputs, reducing the model's flexibility. Like weights, biases are initialised prior to training and updated iteratively through backpropagation (discussed later). It is the combination of weights and biases across the network that encodes its learned knowledge of the network; prior to training, weights are initialised randomly, and it is through the training process described below that they are iteratively refined to produce accurate predictions [40].

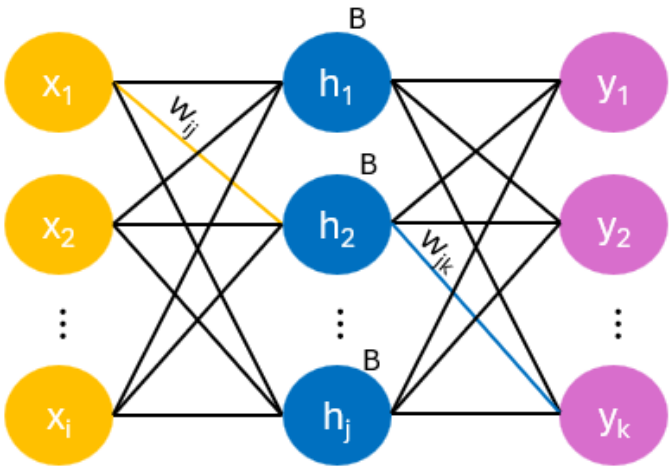


Figure 2.12: Schematic of a feedforward ANN with input neurons x_i (yellow), hidden neurons h_j (blue) each with an associated bias B, and output neurons y_k (pink). Weights w_{ij} and w_{jk} are shown as representative examples of the scalar parameters that scale the signal transmitted along each connection between the input-to-hidden and hidden-to-output layers, respectively

When training ANNs there are a number of critical hyperparameters that must be considered as they influence final performance. The first is the number of hidden layers - additional hidden layers enable the network to capture more complex features, but too many (deep architecture) risk overfitting, especially where there are limited data. A second architectural hyperparameter is the width of each hidden layer (the number of neurons per layer). Wider layers increase the model's capacity to learn complex features, but excessively wide networks can also overfit and may introduce unnecessary computational cost. The third is the mini-batch size, which refers to the number of samples that are processed prior to the model updating its weights, was tuned to identify the trade off between smaller mini-batches, which have the potential for better generalisation but slow down the convergence, and larger mini-batches which enable faster computation but poorer generalisation. The number of epochs is the fourth parameter, where one epoch is defined as one full pass through the training dataset. Too few epochs results in underfitting and too many results in overfitting. There are several additional hyperparameters that required tuning but beyond the ones discussed here, the hyperparameters are specific to each model and will be discussed in more detail in Chapters 3 and 4.

The algorithm used to train most ANNs is the backpropagation training algorithm [35], and is applicable to the MLP and CNN used in this thesis. Backpropagation is a gradient based optimisation algorithm which trains ANNs by iteratively optimising

the networks parameters to minimise the prediction error. There are four main stages to the process, listed below:

1. **Forward propagation** - the forward pass computes the output of the network.

2. **Error evaluation/loss calculation** - the output is compared with the known target using a loss function (specific to the task), which quantifies the difference between the predicted and actual values.

3. **Backwards pass/gradient computation** - where the algorithm computes the gradient of the loss; this process calculates how much each weight in the neural network contributes to the overall error. This is achieved by using the chain rule of calculus, which allows the algorithm to systematically trace the influence of each parameter through the layers of the network. Since neural networks are composed of multiple layers of interconnected neurons, each with its own weights and activation functions, the output of one layer becomes the input to the next. The chain rule enables the decomposition of the total error into partial derivatives, layer by layer, starting from the output and working backward toward the input. For each weight, the gradient tells us how a small change in that weight would affect the loss function. These gradients are essential for optimisation as they guide the direction and magnitude of the weight updates during training.

4. **Weight update** - the weights of the networks are the internal parameters that control the prediction from the model. These are updated using an optimiser, which applies the calculated gradients to adjust the weights in a direction that reduces the loss [35]. Optimisers control how the model learns by determining the step size and direction of each weight update (where the gradient acts like a slope on a loss surface, and each step moves the weights downhill toward lower error), helping the network converge toward a solution that minimises prediction error.

Details on each of these steps for the CNN and MLP in this thesis (e.g. loss functions used) are given in Chapters 3 and 4, respectively. This cycle of forward propagation, error evaluation, backward propagation, and weight adjustment is repeated across many iterations (epochs), enabling the network to learn complex patterns in the data. The weights are continually being updated after each mini-batch, which are then carried forward to the next epoch, so the model is incrementally improving [35]. As an example of the backpropagation process, consider a full dataset

to have 1000 samples, with each mini-batch containing 200 samples, and a total of 5 epochs:

1. Epoch 1
    (a) Randomly shuffle the data to avoid learning order based patterns.
    (b) Split into 5 mini-batches, each 200 samples each.
    (c) For each of the 5 mini-batches:
        i. Forward pass
        ii. Loss calculation
        iii. Backward pass for the gradient calculation
        iv. Weight updates (starting from the weights learned in the previous Epoch)
2. Repeat steps (a)-(c) for each of the 5 epochs.

2.3.4 Summary

This chapter has outlined the theoretical foundations underpinning the experimental work, covering nuclear fusion environments, γ-ray spectrometry principles, and relevant machine learning methodologies. The integration of these domains provides the basis for the diagnostic techniques and analytical approaches employed in this study. Here, γ-ray spectrometry is presented as a complementary tool for characterising fusion related materials, while machine learning offers a framework for enhancing data interpretation. These concepts, including some current applications of machine learning in fusion diagnostics, form the theoretical context for the work described in the following chapters.

Chapter 3

Machine Learning-based Compton Suppression Algorithm (MLCSA)

3.1 Introduction

3.1.1 Background

The materials commonly used in the construction of fusion machines mean it is often the case that detecting low-intensity, low-energy γ rays in the presence of high-intensity, high-energy γ rays is required. This poses a problem, as when γ rays interact within a detector they can Compton scatter [13, 26] out of the crystal and only deposit a fraction of their energy, as discussed in Section 2.2.6. Whilst this can occur at any energy, the probability of CS relative to PE increases significantly with γ-ray energy [13], meaning high-energy γ rays are more likely to contribute to the continuum than low-energy ones. This contributes to a continuum of detected energies, which extends to the lower energy part of the spectrum [14]. For example, the presence of ^{57}Co in activated steel from the ^{58}Ni(n,np)^{57}Co reaction [41] is indicated by a low-energy 122 keV γ ray, which may be masked by the Compton scattering contribution of high-energy ^{60}Co γ rays of 1173 keV and 1332 keV (from the ^{59}Co(n,γ)^{60}Co reaction). The detectability of a photopeak is determined by its absolute count and by its magnitude relative to the statistical uncertainty of the background beneath it. If the continuum beneath a 122 keV ^{57}Co photopeak contains, for example, several thousand counts per channel, a photopeak of only 300 counts may fall within the statistical noise of the background, rendering it indistinguishable from a fluctuation in the continuum rather than a true photopeak.

Figure 3.1 shows an experimental energy spectrum measured by a HPGe detector alongside example interaction scenarios and their resulting pulse shapes. Figure 3.1A

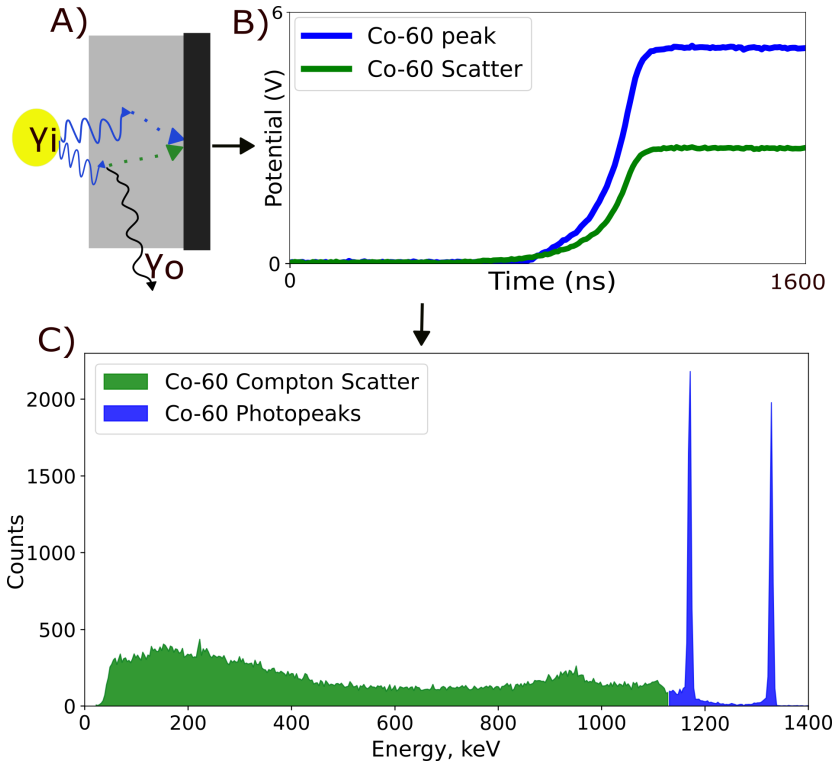


Figure 3.1: A) Shows example γ-ray interactions in a HPGe crystal from an incident γ ray, γ_{i}. In one scenario, γ_{i} (blue) is fully photoelectrically absorbed and produces a full-energy photopeak (blue/dark). In the second scenario γ_{i} Compton scatters out (shown as γ_{o}, black) and contributes to the Compton continuum (green/light), due to a partial energy deposition from the recoil electron. B) Example pulses from each interaction type. C) A typical energy spectrum showing Compton scattered and photoelectric events

illustrates two γ-ray interaction scenarios within the HPGe crystal: in the first, γ_{i} is fully photoelectrically absorbed deep within the crystal, depositing its full energy; in the second, γ_{i} Compton scatters near the front of the crystal and exits as γ_{o}, depositing only a fraction of its energy via the recoil electron. The position of the interaction within the detector influences the resulting pulse shape, as shown in Figure 3.1B. The PE interaction, occurring deeper in the crystal, produces a pulse with a faster rise time and higher amplitude. In contrast, the CS interaction, occurring closer to the front of the detector, produces a pulse with a slower initial rise time and lower amplitude, due to the reduced energy deposition and the position of the interaction within the crystal. This difference in pulse shape is the basis for pulse shape discrimination (PSD), a digital Compton suppression technique discussed later in this chapter. Figure 3.1C shows the resulting energy spectrum, where the Compton continuum extends across the lower energy region whilst the ^{60}Co photopeaks at 1173 keV and 1332 keV are clearly resolved.

This elevated background from the continuum increases the low-energy MDAs, which negatively impacts characterisation measurements. The MDA is the lowest activity of a radionuclide that can be reliably identified above the background continuum at a given confidence level, and is directly proportional to the square root of the background count in the energy region of interest [26]. The MDA is defined fully in Section 3.3.2. For waste disposal, this elevated background could result in waste being consigned with an inflated activity estimate for radionuclides that emit low-energy γ rays, potentially increasing the disposal cost if this leads to a higher consignment category (Table 2.2) or complexity of consignment (e.g. low-level waste rather than out of scope). The elevated background could also lead to radionuclides not being identified, for the case of unexpected radionuclides, such as those resulting from the activation of unknown impurities in material compositions. For large scale decommissioning projects like the JDR project at the UKAEA, inflated activities could mean a significant increase to project costs. For research work, higher MDAs mean that key nuclides for specific reactions may be difficult to quantify, limiting their usefulness and compromising research outputs. Therefore, some form of Compton suppression is imperative to reduce the impact of Compton scattered γ rays, to ensure the most accurate and reliable results. Existing solutions include both physical and digital solutions, each having advantages and disadvantages. Some examples are discussed below.

3.1.2 Existing physical solutions

An example of a physical system, meaning something that can be purchased and installed on a γ-ray spectrometry detector, is a Compton veto ring (CVe) [42], where a ring of γ-ray spectrometry detectors surround a single HPGe crystal. An example of such a system is shown in Figure 3.2, which shows the Mirion BEGe [42] detector with a CVe system, using NaI detectors, located in the RADLab, at the UKAEA site in Oxfordshire. If a γ ray is detected in the HPGe crystal and NaI crystal within a small time window (typically μs), the signals are rejected (vetoed) from the spectrum as a Compton scatter event, therefore not counting towards the final spectrum. The timing window must be experimentally optimised such that only appropriate signals are removed. Optimisation of the RADLab CVe was undertaken prior to this work and therefore the work presented in this thesis assumes the RADLab CVe is fully optimised.

To provide a comparison between a typical physical solution (the CVe) and the work developed in this chapter, americium-241 (^{241}Am, 59-keV photopeak) and ^{60}Co

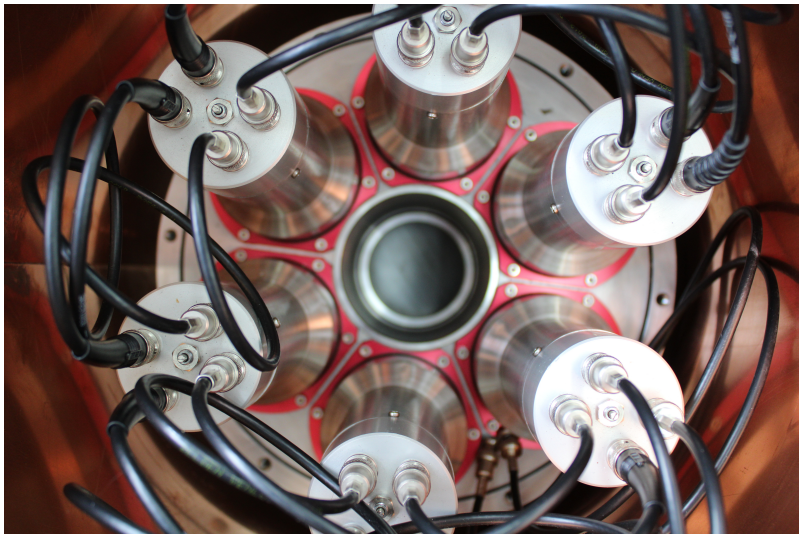


Figure 3.2: A top-down view of the Mirion BEGe coaxial detector used in this work (black central disk) [42], with the physical Compton suppression system surrounding the main HPGe crystal

(1173 keV and 1332 keV photopeaks) reference sources were measured using the RADLab BEGe detector with the NaI CVe. The two sources were measured at the same time (representing a mixed γ source), 10 cm from the end cap for 300 s. The resulting spectrum is shown in Figure 3.3, which shows the spectrum with and without the CVe system in use. This shows that the Compton background on an ^{241}Am and ^{60}Co spectrum was reduced on average (between 200 keV and 400 keV) by $\sim$ 88%, but this also reduced the ^{60}Co photopeaks by $\sim$ 77% and reduced the ^{241}Am photopeak by 18% (based on Equation 2.15 in Section 3.3.2, peak height reduction as a percentage for energy regions of interest, discussed in Section 3.3.2). This false reduction in the photopeaks has been seen in previous commissioning work for the RADLab CVe [6]. These results are included in Table 3.5, and will be used in benchmarking work in later sections. This false reduction likely arises due to coincidence summing; radionuclides such as ^{60}Co emit γ rays in cascade, where two γ rays are emitted in rapid succession following a single decay. When both γ rays interact within the main detector and a surrounding detector within the resolving time of the CVe system, the second interaction can be incorrectly identified as a Compton scatter event and vetoed, resulting in a false suppression of true photopeak counts [6]. This directly affects the activity (Section 2.2.8) results for these photopeaks, underestimating the activity. The physical Compton suppression system is bulky, expensive, and is not easily re-configurable for multiple detector types. These factors greatly limit its usefulness in a radiometric laboratory, which

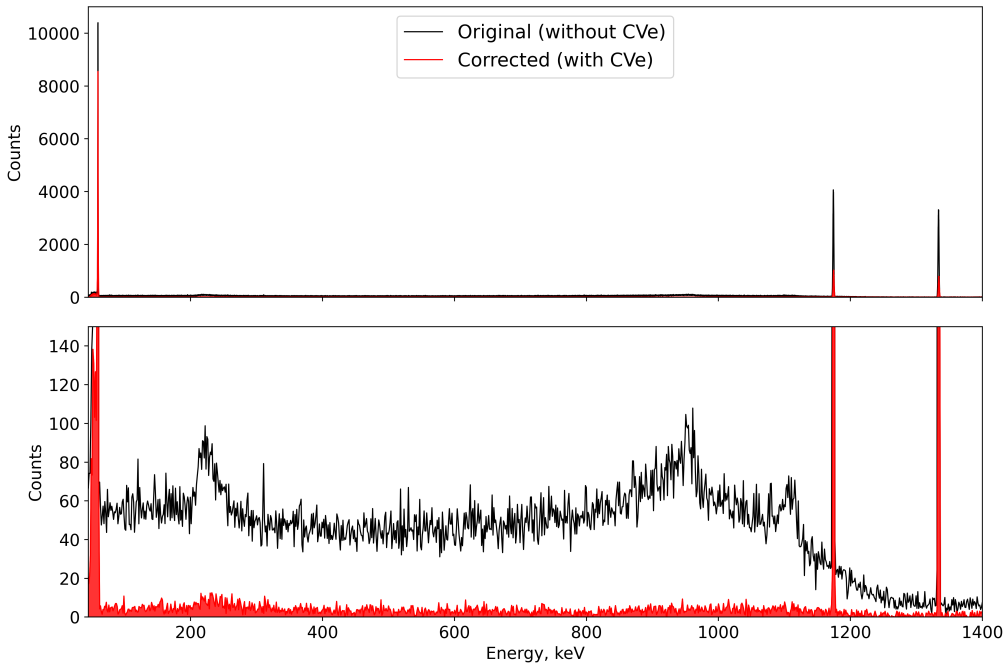


Figure 3.3: A γ-ray spectrum of ^{241}Am and ^{60}Co sources on the BEGe HPGe detector, where the spectrum in black is the original spectrum without the CVe system in use, and the red is the spectrum with the CVe system in use. The top shows the full spectrum and the bottom is zoomed along the vertical axis

typically has several γ-ray spectrometry detectors, and could limit any potential broader applications of the equipment which might value portability.

3.1.3 Existing digital solutions

An alternative technique to reducing the Compton continuum is to use digital methods. These offer greater flexibility in laboratories where there are multiple γ-ray spectrometry detectors, and for field measurements, as they do not require expensive and bulky physical hardware.

The first example of a digital solution is the Digital Compton Suppression Algorithm (DCSA) [28], which uses the method of PSD. The PSD method uses the theory of charge collection in a detector (see Section 2.2.4) and determines the origin of a pulse using features of the pulse shape, such as pulse height and rise time, as discussed in Section 3.1.1. The rise time of the preamplified pulse is the total time for all charge carriers to be collected from an interaction, counted as the region from 10% to 90% of the maximum pulse height. The region of 10% to 90% is used to reduce noise effects at the extreme ends and due to difficulties determining the 0% region. Low-energy γ rays are more likely to undergo PE (see Section 2.2.6) close to

the surface of the detector. In geometries where the charge-collection electrode is located at the base (as with the BEGe, which is discussed in detail in [28]), these surface interactions can result in a long charge collection time in the HPGe as the charge carriers (which in the p-type BEGe are the holes) drift further through the crystal to the collection anode. In contrast, high-energy γ rays more readily pass through the detector, and are more likely deposit their energy further inside the bulk of the crystal. Where high-energy γ rays Compton scatter within the bulk of the crystal with only partial energy deposition, the resulting pulses will have shorter rise time of the generated electrical pulse but a smaller pulse height. Here, the pulse height might appear as the result of a low-energy deposition, but the fast rise time can link it to a high-energy γ rays. Figure 3.4 used data following post processing, which shows the rise times of an example source dataset (comprising five individual sources), including the highest energy photopeak from each source and all scatter rise times, where it can be seen that only the lowest energy (59 keV ^{241}Am) γ ray has rise times above 400 ns. As ^{241}Am 59 keV γ rays typically interact at or very near the surface of the detector due to their limited penetration depth, they would be expected to produce broadly similar rise times. However, a wide spread of rise times is observed, which is attributed to noise in the low-amplitude 59 keV signals. The determination of 10% and 90% thresholds becomes unreliable for such noisy signals as previously discussed, artificially broadening the observed distribution.

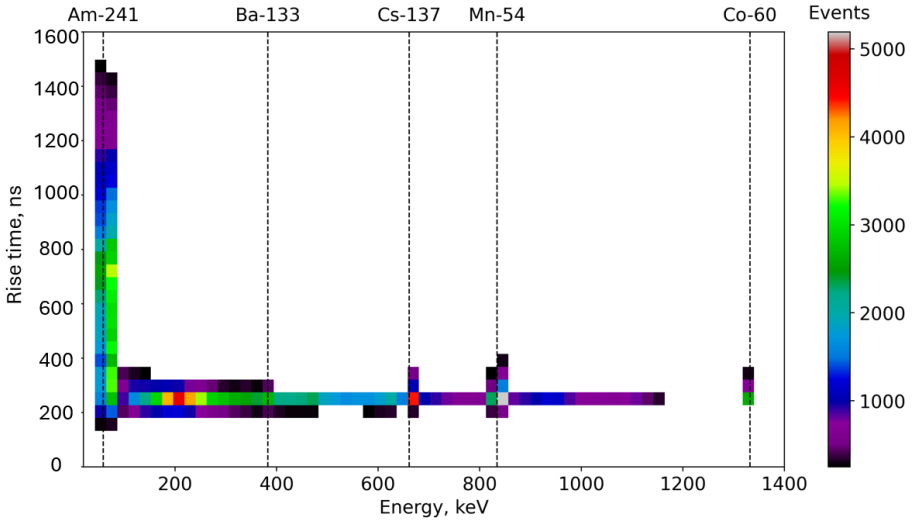


Figure 3.4: A plot showing the rise time for each of five radionuclides, including the rise time of the scatter pulses. The energies for the radionuclides include the highest energy photopeaks only, which are: 59 keV ^{241}Am, 383 keV ^{133}Ba, 661 keV ^{137}Cs, 834 keV ^{54}Mn, 1332 keV ^{60}Co, and all associated scatters. This plot shows that low-energy (e.g. ^{241}Am) γs have the longest rise time compared to higher-energies (such as ^{60}Co)

The application of the DCSA [28] utilises a γ ray pulse rise time cutoff to preserve low-energy pulses corresponding to PE, while removing pulses contributing to the Compton continuum. An example of the type of results this algorithm may produce is shown in Figure 3.5. It should be noted that this was created using a simple rise time cutoff which was not extensively modelled, where the value was chosen to demonstrate how the cutoff works, since the actual spectrum from [28] was not available. Figure 3.5 shows that the method is effective in preserving the data for low-energy photopeaks, but fully removes pulses from high-energy γ rays, resulting in a significant loss of data on any of the higher energies. While this may have some applications where only the low-energy γ rays are of importance (e.g. in some research applications), the method requires extensive detector modelling [28] to determine the optimal rise time cutoff which is non-trivial, is not practical for preserving a broad range of energies, or for deploying to multiple types of detectors.

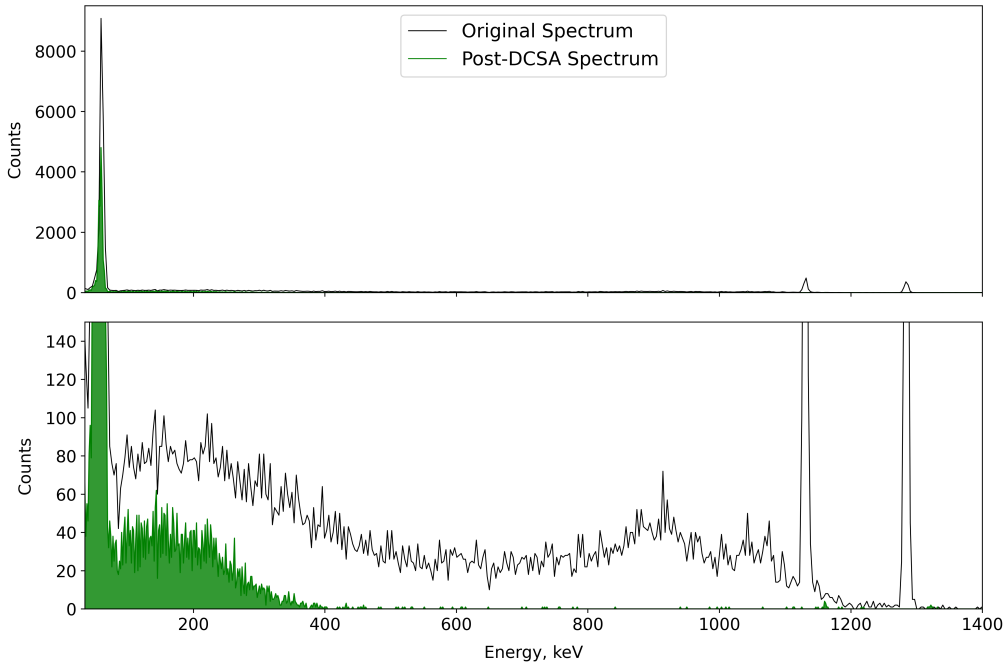


Figure 3.5: Kevin Tree [28] example results on a ^{241}Am and ^{60}Co spectrum, with before (black) and after (green) the application of the DCSA. This is not the actual data from [28], but a representation using a similar technique for illustration purposes

A second digital technique is the use of machine learning-based algorithms for PSD. Examples of this include γ particle tracking [43] and α-γ pulse shape analysis [44]. However, distinguishing a pulse generated by a Compton-scattered γ ray from a pulse due to a photoelectric interaction as a form of γ-γ PSD for Compton suppression in HPGe detectors has not been undertaken in the literature and is therefore the focus of this chapter.

3.1.4 Research aims

It has been identified through Chapter 2 and through exploring existing solutions in Section 3.1, that there is a requirement for an enhanced digital solution to the Compton continuum background problem. To this end, the work in this chapter explores the application of machine learning models for γ-γ PSD, comparing the effectiveness of many different models and includes the work on experimental data preparation. This work was in preparation for the creation of a novel machine learning-based algorithm, the MLCSA, which is compared to the existing solutions, to support the evaluation and validation process.
The research questions this chapter aims to address are:

1. Can the background from the Compton continuum typically found in the low-energy part of a spectrum from γ-ray spectrometry be reduced digitally, using state-of-the-art machine learning techniques?

2. Can the machine learning technique be implemented while preserving higher energy photopeaks and not rely on any detector modelling?

3. Can the techniques produced in this work be used in real-world applications?

3.2 Methods

Through step one of the model selection process in Section 2.3.2 and Figure 2.10, this work was identified as a classification task, as the machine learning models review each pulse and classify them as either a pulse from a PE event (to keep) or a pulse from a Compton scatter event (to discard). The training method for these machine learning models was supervised, as the training data could be pre-labelled as either photopeak or scatter, based on the pre-processed pulse heights. The final model used in the MLCSA comprises a CNN, which performs Compton suppression on pre-amplified pulses to reduce the Compton continuum, while preserving higher-energy photopeaks and not relying on any detector modelling or any additional (often expensive) hardware. The MLCSA is part of a signal processing workflow that comprises four components, as shown in Figure 3.6: detector (energy deposition and pre-amplification), digitiser (signal/pulse processing), MLCSA, and Compton-suppressed spectrum (results). The methodology, refinement, and benchmarking/validation for each are described in the following sections.

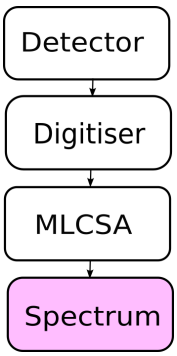


Figure 3.6: Digital Compton suppression flow diagram showing the four components, including the MLCSA [45]

3.2.1 Data collection

Step two of the the model selection process in Section 2.3.2 and Figure 2.10 is the data collection and exploration step, which is discussed in the sections below.

Detector

The high-resolution, HPGe detector used to gather data was the Mirion BEGe detector [42], shown in Figure 3.2. The technical information and specifications for this detector are in Appendix A.1.

Experimental data

Two datasets were collected. The first, called the 'individual source' dataset, comprised pulses from five single nuclide calibration sources (^{241}Am, ^{60}Co, barium-133 (^{133}Ba), ^{137}Cs, and ^{54}Mn) individually measured and counted for 300 s. All sources except for ^{54}Mn were positioned 10 cm from the end cap to reduce coincidence summing effects, where two γ rays enter the detector almost simultaneously [13]. The ^{54}Mn source was placed at 0 cm due to its lower activity. The individual source dataset was later split into two subsets, a training dataset and testing dataset - details of this are discussed later.

The second dataset, called the 'combined source' dataset, consisted of pulses from two calibration sources that were measured on the detector simultaneously (59 keV from ^{241}Am and 1173 keV and 1332 keV from ^{60}Co). The combined source dataset simulates a realistic sample, one which might be seen in a standard laboratory measurement of a mixed γ-ray sample and is used as a validation dataset. It reflects conditions typical of radioactive waste, where radionuclides emitting higher-energy γ rays are present alongside those emitting lower-energy γ rays. The Am-241 and

Co-60 radionuclides were deliberately chosen as representative extremes in energy to serve as a proof of concept for the model's ability to distinguish between mixed sources.

3.2.2 Digital pulse collection and processing

The pre-amplified pulses, referred to as 'raw' pulses, were collected and processed by a Red Pitaya digitiser (the STEMlab 125-14 [46], Figure 3.7). Examples of raw pulses are shown in step A of Figure 3.8. A simple voltage threshold trigger on the rising edge of the preamplifier signal was used, determined experimentally so that the full energy range of the calibration sources was visible on the spectrum; the corresponding settings are provided in Appendix A.2. This type of leading edge triggering can introduce amplitude dependent time walk in HPGe systems, though this does not affect the extracted pulse height.

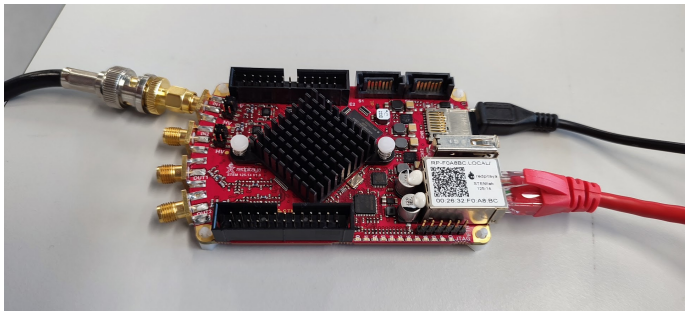


Figure 3.7: Image of the Red Pitaya digitiser, connected with power, Ethernet port to the laptop, and the detector cable

The raw pulses could not be used directly to train the machine learning models as the algorithms could use features such as pulse height (which is directly proportional to the energy) to characterise pulses in a way that would lead to overfitting. Therefore, measured raw pulses were modified so that the machine learning models could be trained purely on the shape of the pulses. A pulse processing pipeline (as per step three of the model selection process in Section 2.3.2 and Figure 2.10) with steps A to G, shown in Figure 3.8, was created and included filtering, labelling, transformations, and regularisation techniques to reduce overfitting. Figure 3.8 shows example ^{54}Mn pulses at each stage of the pipeline, and is referred to and discussed throughout the following sections. The figure shows that the difference between fast and slow rise times in pulses can be differentiated clearly. Some parts of the processing pipeline were not applied to the combined source dataset (greyed out steps in Figure 3.8) and this is discussed in the applicable sections below.

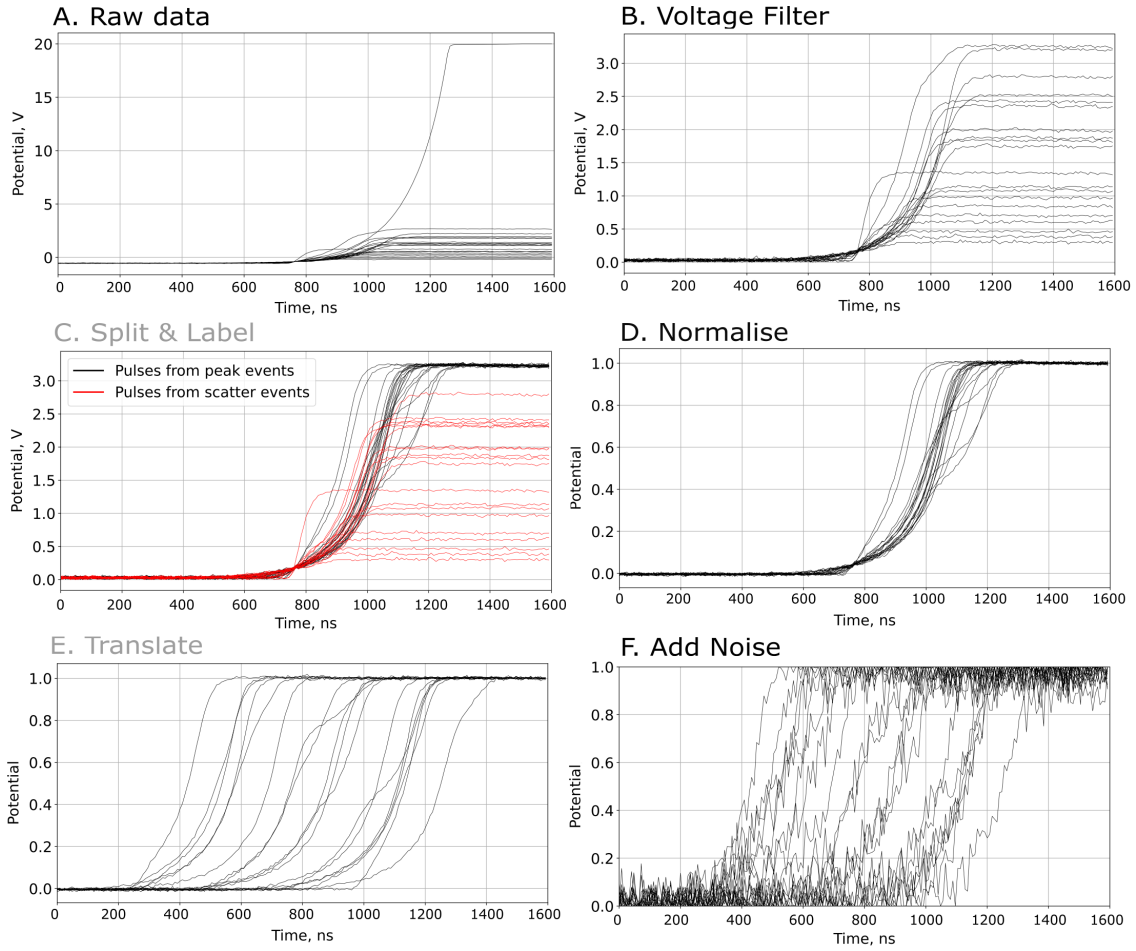


Figure 3.8: The pulse processing pipeline [45], is described in detail in the main text. The greyed out steps are not used when applying this process to the combined dataset. The plots show example pulses at each stage from the ^{54}Mn source, corresponding to steps A to F. Step G, rebalancing data step, is not included as a plot

Pipeline step A: Raw data and B: Voltage filter

The pipeline starts by importing raw data from the red Pitaya (step A), and then removing pulses that exceed the maximum voltage of the digitiser and those below a pre-determined noise level (step B).

Pipeline step C: Split & label

The individual source dataset's pulses were then split into *scatter* and *photopeak* categories and labelled accordingly (based on the pulse height/energy). For the photopeak categories, only the highest energy photopeak pulses from each nuclide were labelled as photopeaks, removing the lower-energy photopeaks to ensure no Compton scatter pulses from the higher-energies were incorrectly labelled as photopeaks. This was achieved by defining an energy window around the highest-energy photopeak for

each nuclide; pulses falling within this window were labelled as photopeak events, while all pulses below the lower bound of this window were labelled as scatter events. Pulses in the energy ranges corresponding to lower-energy photopeaks were excluded entirely from the dataset. The scatter category for higher-energy sources such as ^{60}Co (1332 keV) spans a broad continuum of energies down to the noise threshold, whereas the scatter distribution for ^{133}Ba (383 keV) is comparatively narrow due to the limited Compton scatter range at lower energies. The γ energies used from the dataset following this step were:

- 59 keV ^{241}Am photopeak,
- 383 keV ^{133}Ba photopeak,
- 661 keV ^{137}Cs photopeak,
- 834 keV ^{54}Mn photopeak,
- 1332 keV ^{60}Co photopeak,
- and all associated scatters.

The resulting spectra from this step, using the pulse heights to plot the energy, for the ^{60}Co source (as an example) is shown in Figure 3.9. This step was not applied to the combined source dataset because that information cannot be known for a mixed γ-ray source.

Pipeline step D: Normalise

Pulses were then normalised between 0 and 1 to remove pulse height information, as these could be memorised by the machine learning models which would lead to overfitting and poor generalisation.

Pipeline step E: Translate

The pulses for the individual source dataset were then randomly translated along the time axis to prevent the machine learning model memorising the time of the pulses. On average, for example, ^{241}Am pulses tend to exhibit a later rise onset within the recorded trace relative to ^{60}Co pulses - that is, the pulse rise begins later relative to t=0 of the trace. This occurs because 59 keV γ rays from ^{241}Am interact preferentially near the surface of the detector, resulting in a longer initial charge carrier drift distance to the collection point and therefore a delayed rise

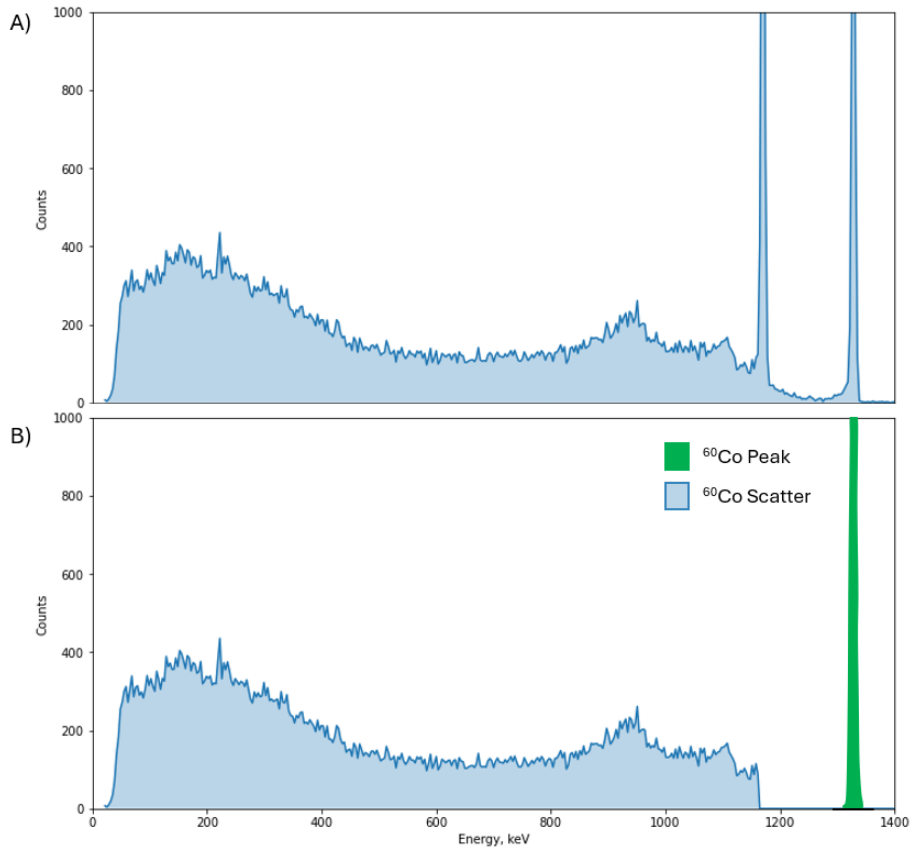


Figure 3.9: A) An example spectrum of the ^{60}Co source prior to splitting and labelling (step C). B) The ^{60}Co spectrum following step C of the pipeline, showing the scatter part of the spectrum in blue and the photopeak in green, noting that the lower energy ^{60}Co photopeak has been removed

onset. This difference in trigger times relates to the variation in how the signal forms (charge collection) and pulse rise times caused by the interaction depth and energy of the incident γ rays. Random translation is an example of data augmentation and increases the generalisability of the model. The minimum and maximum time (x-axis) was chosen to be -240 ns and +560 ns, respectively, to ensure the preservation of the bulk of the pulses (ensuring the full rise was kept) when they are translated to the extreme values. The start position of each pulse was changed by a random translation amount. This step was not applied to the combined source dataset as it is unnecessary due to the way the machine learning models are trained with data augmentation, aiming to cover many possibilities of 'natural' translation.

Pipeline step F: Add noise

Gaussian noise was then added so that every pulse had the same standardised signal-to-noise ratio, irrespective of its energy. The results of this step can be seen

in Figure 3.8. This was added as it was noted that lower energy pulses had a higher relative noise level, whereas the pulses from higher energies were significantly smoother. The amplitude of baseline electronic noise is independent of pulse height, the peak-to-peak noise has a fixed value regardless of the signal amplitude. However, following normalisation in step D, pulses from low-energy depositions have a larger noise component relative to the signal height, resulting in an energy dependent signal-to-noise ratio. The noise is primarily electronic in origin, with a secondary smaller contribution from statistical variation in charge carrier production, characterised by the Fano factor [27]. As this noise, after normalisation, is related to the pulse height, the noise level still indirectly contained the height information even though the pulses were normalised in amplitude. This enables the machine learning models to memorise and overfit by energy. Hence it was determined that the noise level should be standardised to increase generalisation to all energies. This relative noise is a problem since the suite of radionuclides (and therefore γ-ray energies) available to characterise the photopeak pulses is limited to a few discrete energies, rather than a continuous range of energies. This step was not included in the initial pipeline that was used for the training of the 12 machine learning models in Section 3.2.3, and was added to the pipeline during the hyperparameter tuning stage of the final model to improve physical realism - this is discussed later in Section 3.3.1.

To implement the addition of Gaussian noise, the standard deviation of the first 240 ns of the 50 noisiest ^{241}Am pulses (the 59 keV γ line) was evaluated in post processing (to obtain σ_{target}) as this captures the maximum expected variability in noise that may be encountered in all pulses. The ^{241}Am radionuclide was chosen as the lowest energy and therefore has the poorest signal-to-noise ratio after normalisation, and so σ_{target} was calculated as 0.06. The time window, t, of 240 ns (from 0 ns to 240 ns) was chosen as this is the part before the rise begins, to ensure that just the baseline fluctuations were considered.

$$\sigma = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(x_i - \mu)^2}, \tag{3.1}$$

where x_i are the pulse height data points, t is the number of pulse height samples (which is the time window; 240 ns), and the mean (μ) is

$$\mu = \frac{1}{t}\sum_{i=1}^{t} x_i. \tag{3.2}$$

For any pulse whose current standard deviation (σ_{current}) in the same time window of 240 ns was below σ_{target}, additional Gaussian noise was generated by

$$\sigma_{\text{noise}} = \sqrt{\sigma^2_{\text{target}} - \sigma^2_{\text{current}}}. \tag{3.3}$$

The final pulse was then obtained by adding random Gaussian noise with a standard deviation σ_{noise} to all bins of the full original pulse, i.e.

$$\text{final pulse} = \text{original pulse} + \sigma_{\text{noise}}. \tag{3.4}$$

Here, σ_{noise} is not a constant offset but the standard deviation of the Gaussian distribution from which random noise samples are drawn. These samples are added independently to each bin of the pulse, ensuring the variability in the first 240 ns matches the target while noise is applied across the full pulse. Pulses already exceeding σ_{target} were left unchanged. This ensures that all pulses have a consistent minimum noise level, reducing the likelihood that the machine learning models will overfit based on noise.

Pipeline step G: Rebalance data

Finally, the number of events in the photopeak and scatter categories were standardised (referred to as the rebalance step in Figure 3.8) for the individual source dataset. This approach is adopted because in a typical γ-ray measurement, the number of scattered pulses generally exceeds the number of photopeak pulses, which could lead to the machine learning models having more information about the scatter pulses (class imbalance). Class imbalance can cause models to favour the majority class, learning scatter features in detail while under representing the photopeak class, which can shift the decision boundary to the minority class and lead to misclassification of photopeak pulses. Therefore, the same number of photopeak pulses and scatter pulses were provided by down-sampling the number of scatter pulses from the highest energy source (^{60}Co). Table 3.1 summarises the number of photopeak and scatter pulses retained from each source following this rebalancing step. The ^{241}Am scatter pulses were excluded from the dataset, as the limited Compton scatter range at 59 keV results in very few scatter events, and their inclusion risked confusion with the lower-energy ^{241}Am photopeaks. It should be noted that this rebalancing approach may introduce a bias, since the scatter pulses retained for each source span a different range of energies depending on the photopeak energy of that source, for example, the scatter continuum for ^{60}Co extends over a much broader energy range than that for ^{133}Ba (383 keV), meaning the rebalanced dataset does not represent a uniform distribution of scatter energies across sources. This step was not applied to the combined source dataset as this information is not possible to know because the data

originate from a mixed source hence the number of each type of are present is an unknown.

Table 3.1: A table showing the number of photopeak and scatter pulses retained per source following pipeline step G

Source	Photopeak pulses	Scatter pulses	Scatter pulses retained
^{241}Am (59 keV)	73,239	0	0
^{133}Ba (383 keV)	895	7,121	7,121
^{137}Cs (661 keV)	5,285	22,410	22,410
^{54}Mn (834 keV)	10,582	28,240	28,240
^{60}Co (1332 keV)	4,050	64,292	36,280
Total	94,051	122,063	94,051

Pulse processing pipeline summary

After processing, the individual source dataset consisted of 188,102 pulses from five radionuclides. Following processing through the pipeline, the individual source dataset was divided (using 'train test split' from *sklearn*), as is common in machine learning, into a training set (80%, 150,482 pulses) and a testing set (20%, 37,620 pulses). The split was taken from the photopeak data and scatter data separately, so that the final 80/20 split contained an equal number of scatter and photopeak pulses. The final pulses following this process are what is provided to the machine learning models, in .CSV format.

3.2.3 Machine learning model selection

Once the individual source data were collected, processed through the pipeline, and split into training and testing data, they were ready to use with various machine learning algorithms to identify the most suitable model for use within the MLCSA (as per step four of the model selection process in Section 2.3.2 and Figure 2.10). As part of this process, 12 machine learning algorithms were chosen, based on their applicability for classification tasks. The 12 models chosen were split into traditional methods (DT, GaussianNB, LR, LinearSV, SVM, KNN, RF, and XGB) and deep learning methods (SP, MLP, CNN, and SDGC). More details are provided on each of these in step four in Section 2.3.2.

While Recurrent Neural Networks [35] could have been considered given that γ pulses have an inherent sequential nature, they were not evaluated here as the pulse data was treated as a complete snapshot in time rather than a temporal sequence. This approach captures the full pulse shape as a single feature representation,

making convolutional-based feature extraction more appropriate than sequential modelling. The CNN, which was selected as the final model, is well suited to extracting discriminative features from such fixed length pulse waveforms.

Machine learning model evaluation

The 12 machine learning models were trained without hyperparameter tuning at this stage for simplicity and speed (default parameters were used), and then evaluated based on the following metrics [47]: accuracy, precision, recall, F1-score, and CV score. In all metrics, the photopeak category is the positive value and the scatter is the negative.

Accuracy is a measure of how often the machine learning model is correct, calculated as the ratio of correct classifications to the total number of classifications, defined as

$$\text{Accuracy} = \frac{TP + TN}{N_p}, \tag{3.5}$$

where TP is true positives, TN is true negatives, and N_p is the total number of predictions [35].

Precision measures the proportion of true positive predictions among all positive predictions defined as

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{3.6}$$

where FP is false positives [35].

Recall (or sensitivity), also called true positive rate (TPR), measures the proportion of true positive predictions among all actual positive cases, defined as

$$\text{Recall (TPR)} = \frac{TP}{TP + FN}, \tag{3.7}$$

where FN is false negatives [35].

The F1-score provides a balanced measure that combines precision and recall into a single metric [35].

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3.8}$$

The final metric is the CV score, using 10 folds, which is detailed in Section 2.3.2. The CV score is computed using accuracy as the underlying metric, representing the mean classification accuracy across all 10 folds. The CV score is the most useful metric as it provides an all round indication of performance.

All of these metrics were applied to the training and testing data for each model to test for generalisability and to check for overfitting, as discussed in Section 2.3.2. The purpose of these tests is to identify the best performing model out of the initial 12 to develop further and validate. The results for the metrics for all 12 models are shown in Table 3.2.

Table 3.2: Performance metrics for various machine learning models. The models with the highest CV values have been highlighted. Acc = Accuracy, Prec = Precision, Rec = Recall, F1 = F1-score

Model	Training				Testing				CV
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	
RF	100.0	100.0	100.0	100.0	93.4	91.2	96.4	93.7	93.1
SGDC	88.5	87.4	89.8	88.6	88.7	87.2	91.5	89.3	86.7
DT	100.0	100.0	100.0	100.0	89.4	90.4	88.9	89.6	89.3
KNN	82.1	90.1	72.1	80.1	79.2	88.3	68.6	77.2	79.5
LinearSV	88.2	87.9	88.7	88.3	88.3	87.5	90.1	88.8	88.0
MLP	92.1	88.7	96.5	92.5	92.1	88.9	96.8	92.7	91.3
SVM	86.8	88.5	84.4	86.4	86.6	87.9	85.7	86.8	86.7
GaussianNB	50.6	90.8	1.2	2.3	49.0	90.9	0.8	1.5	50.7
LR	88.5	88.2	88.9	88.5	88.7	87.8	90.6	89.1	88.3
SP	84.9	91.4	77.1	83.6	85.1	90.6	79.2	84.5	81.5
XGB	100.0	100.0	100.0	100.0	93.4	91.1	96.5	93.7	93.3
CNN	92.9	88.9	97.9	93.2	92.7	88.7	98.5	93.3	92.6

The machine learning models with the highest CV score (highlighted pink in Table 3.2) were: XGB (93.3%), RF (93.1%), and CNN (92.6%). The XGB and RF are conceptually similar, as decision tree-based ensemble methods, but they differ in how they build trees. The RF uses bagging (parallel trees), while XGB uses boosting (sequential trees, where each tree corrects errors of the previous ones [35]). Both were included in the initial evaluation to benchmark performance and ensure robustness, and the XGB model offered only a marginal advantage over the RF. The RF and XGB models had an accuracy, precision, recall, and F1-score of 100% on the training data (from Table 3.2), but when compared to the test data there was a difference of around 7%. This pattern of the training data having a perfect score, and a lower score for the test data indicates that the model is likely overfitting and simply memorising the data, meaning when new, unseen data is presented, the models are not able to generalise as well. Therefore, despite appearing to perform well, those models have not been deemed suitable for use in this work. With hyperparameter tuning the

model may perform better and be less likely to overfit, but as other models worked well those were pursued instead.

The CNN came third in CV performance in Table 3.2 and was shown to perform well (pre-tuning), with minimal overfitting and a great generalisability to new, unseen data. Therefore, the CNN was chosen as the final model for further development.

3.2.4 CNN theory

Convolutional neural networks are designed for handling structured grid-like data. They are typically used in fields such as object recognition, image classification, and image segmentation, as they excel at feature extraction and pattern recognition. There are different types of CNNs, such as 2D CNNs which are commonly used for image recognition and 1D which are used for input data that is a time series - this work focuses on 1D CNNs as they suit the pulse data. The foundations of convolutional neural networks were established by [48], who introduced the LeNet architecture for handwritten digit recognition, demonstrating that learnable convolutional filters could automatically extract spatial features from raw input data. A significant milestone in CNN development was the introduction of AlexNet by [49], which demonstrated the capability of deep CNNs on large scale image classification tasks. The AlexNet architecture, illustrated in Figure 3.10, comprises multiple convolutional and pooling layers followed by fully connected layers, establishing the deep hierarchical structure that underpins modern CNN design. Since these foundational works, CNNs have been widely adopted across numerous domains beyond image recognition, including time-series classification and signal processing tasks.

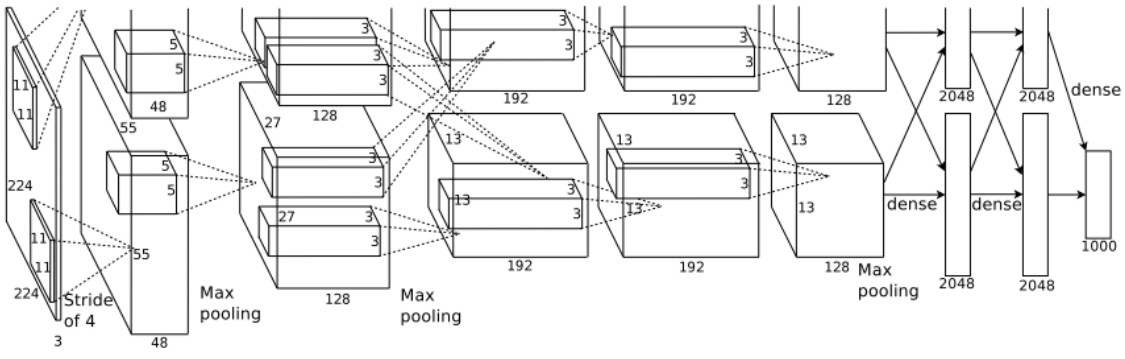


Figure 3.10: The AlexNet architecture [49], illustrating the deep hierarchical structure of convolutional and pooling layers followed by fully connected layers

The CNNs consist of multiple hidden layers, where each layer performs specific operations to extract and learn useful features from the input data. The backpropa-

gation method, as discussed in Section 2.3.3, is used during training within CNNs. The layers include:

- **Convolutional layers** - these use learnable filters to scan and identify local features, resulting in a set of feature maps that highlight different aspects of the input. Each convolutional layer is typically followed by a rectified linear unit (ReLU) activation function layer, which enables the network to capture complex relationships beyond simple linear mappings.
- **Pooling layers** - these occur after each convolutional layer, and these reduce spatial dimensions (e.g. simplify by summarising regions of feature maps). Pooling layers also reduce complexity while retaining the most important information, such as extracting the key features, enhancing computational efficiency and maintaining feature robustness [35].
- **Flatten layer** - this is where the output of the last convolutional and pooling layers is flattened into a 1-dimensional vector.
- **Fully connected layers** - these are fed the vector from the flatten layers, which then act as a traditional neural network. The fully connected layers learn to map the extracted features to the binary classification task. The last fully connected layer typically uses a sigmoid activation function to produce a probability value between 0 and 1, representing the likelihood of belonging to one of the classes. A threshold then needs to be selected (through hyperparameter tuning), and applied to determine the threshold for when a pulse should be classified as a photopeak or a scatter. The threshold defines the trade-off between precision and recall.

Uncertainty estimation was not addressed in this work; instead, the CNN's performance was assessed using conventional metrics such as accuracy and loss, as discussed.

CNN hyperparameter tuning

Prior to the CNN being ready for use, further hyperparameter tuning and validating was required, as per step five of the machine learning process in Section 2.3.2 and Figure 2.10. Details on what hyperparameter tuning is and the methods to do this are discussed in Section 2.3.2. Random search was chosen due to the large hyperparameter space to find the best values for each [35]. All other parameters use the defaults, such as kernel size for the convolutional layers, where the default kernel

size is 3×1 as provided by the *keras* library. The hyperparameters tuned for the CNN and their respective search ranges, were:

1. **Network architecture** – variations in the number of convolutional layers, filter sizes, and fully connected layers were tested to balance feature extraction and model complexity.
   Search range:

   - Convolutional layers: [1, 2, 3]
   - Filters per layer: [16, 32, 64, 128]
   - Fully connected layer neurons: [32, 64, 128]

   These ranges were selected to cover typical values used.

2. **Batch/mini-batch size** – controls the number of samples processed before the model updates its weights. Larger batches can improve computational efficiency but may require more memory and risk poorer generalisation, while smaller batches can lead to more stable updates but slower training.
   Search range: [32, 64, 128, 256]. These were chosen to balance memory constraints and training stability.

3. **Number of epochs** – determines how many times the entire training dataset is passed through the network. Too few epochs can lead to underfitting, while too many can cause overfitting.
   Search range: [10, 20, 30, 40, 50]. These were selected to allow sufficient training without excessive computation.

4. **Threshold** – the decision boundary for classifying outputs in binary classification. A threshold too low or too high can skew predictions toward one class.
   Search range: [0.3, 0.4, 0.5, 0.6, 0.7]. These values were tested to optimise classification performance.

5. **Number of filters** – controls the depth of feature maps in convolutional layers. More filters allow learning richer representations but increase complexity and risk overfitting.
   Search range: [16, 32, 64, 128]. These were selected to balance feature extraction and computational cost.

Each set of hyperparameters were evaluated by checking the effect on the CNN performance, using common methods described earlier such as accuracy, precision,

recall, and CV score (accuracy) [35, 36]. These evaluation methods all compare the known labels of the pulses to the model predicted labels and are recorded as a percentage. Two more evaluation metrics were used at this stage, the confusion matrix and the receiver operating characteristic (ROC) curve. A confusion matrix is a grid that compares the model's predicted labels against the true labels on the test data. The results for the confusion matrix for the CNN are in Section 3.3.1.

A ROC curve is plotted by varying the classification threshold used to convert predicted probabilities into class labels. For each threshold, the TPR (recall), and false positive rate (FPR, the proportion of negatives incorrectly classified as positives) are calculated. These values are then plotted with FPR on the x-axis and TPR on the y-axis, and connecting the points across all thresholds produces the ROC curve. Classifiers that give curves closer to the top-left corner indicate better performance, while a random classifier is expected to give points lying along the diagonal (where FPR = TPR). An extension of this metric is the area under the ROC curve (AUC-ROC), which quantifies the overall performance as a single quantitative measure. The AUC quantifies a classifier's ability to distinguish between classes, with values closer to 1 indicating stronger separation of positives and negatives: a perfect classifier achieves AUC = 1, whereas a random classifier yields AUC = 0.5. The results for the ROC curve for the CNN are in Section 3.3.1.

The tuning and evaluation process was repeated until there were no further improvements to the evaluation results (results didn't increase in value), and at that point the CNN was concluded to have reached close to peak performance. The final CNN architecture used in this work is a basic one-dimensional (1D) CNN [50], which is suitable for processing 1D sequences, such as time series data [35]. The model is designed for binary classification tasks, with one output of a probability between 0 and 1 (due to the sigmoid activation function), and a threshold is set to determine what the classification should be. The CNN model was trained in a CPU computing environment and used the Adam optimiser (the default in *keras*) and a binary cross-entropy loss function (default). The architecture was implemented through standard *sklearn* and *keras* Python libraries. Figure 3.11 shows an illustration of what the architecture layers look like, and this is described in the list below, with the other final hyperparameters:

- Final CNN architecture layers:
  - Input layer (size (200, 1))
  - Convolutional layer (32 filters, ReLU activation)
  - Max pool layer (pool size 2)
  - Second convolutional layer (64 filters, ReLu activation)
  - Second max pool layer (pool size 2)
  - Flatten layer
  - Fully connected layer (64 neurons, ReLU activation)
  - Output layer (1 neuron, sigmoid activation).
- Batch size of 128
- 30 epochs
- Threshold value of 0.5.

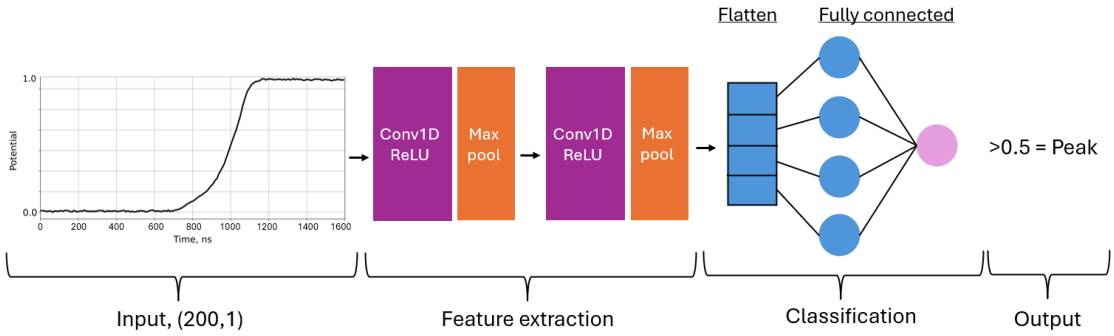


Figure 3.11: A simplified diagram of the CNN architecture used in the MLCSA. The diagram shows the progression from the input layer through convolutional and pooling layers, followed by flattening and fully connected layers, and ending with a single sigmoid output neuron that outputs a probability for binary classification, using a threshold of 0.5

CNN validation process

For validation of the trained CNN, as per step six, validation, from Section 2.3.2 and Figure 2.10, the trained CNN was incorporated into the digital machine learning Compton suppression flow process (Figure 3.6) as the MLCSA step. The processed pulses for the combined source dataset were passed to the MLCSA, where the output comprised pulses with predicted labels of photopeaks (1) or scatter (0). The pulses classified as scatter pulses are removed as part of the spectrum step in Figure 3.6, where the output is a spectrum from only pulses classed as photopeaks. The results from the combined source dataset are discussed in Section 3.3.2.

3.3 Results

3.3.1 Individual source, test dataset results

Confusion matrix

The confusion matrix for the final model, corresponding to the test data, is shown in Figure 3.12. The matrix shows that the model correctly predicts 74% of the scatters and 88% of the photopeaks, and misclassified 26% of scatters as photopeaks and 12% of photopeaks as scatters. While the model performs well, the photopeaks are more accurately classified than the scatters, which infers that the features used to identify the scatters may be less distinctive or may overlap with the photopeaks in the inputs. While this does highlight minor confusion between the two categories, this is expected as higher energy scatters are likely to be similar to high-energy photopeaks (as explained in Section 3.1.3). The misclassification percentages are only a small percentage, therefore, the results of the confusion matrix indicate that the model performs as intended, as preserving the photopeaks is a high priority to retain crucial spectral data, which the model achieves with a low misclassification (12%) of photopeak pulses.

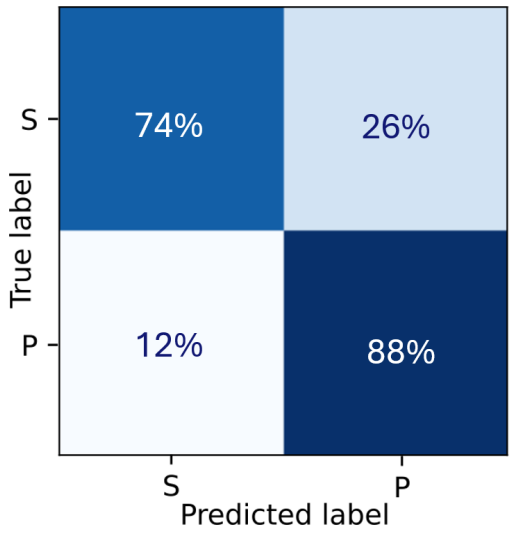


Figure 3.12: Confusion matrix for the pulses from the test dataset (37,620 pulses), where S is scatter and P is photopeak. The values represent percentages of the total number of pulses and the the rows represent the true label and the columns represent the predicted labels [45]. A perfect model would have non-zero values on the top left and bottom right only

The ROC

The ROC curve for the final model is shown in Figure 3.13. The ROC curve for the CNN in this work shows a good classification rate, with an AUC value of 0.92, suggesting the CNN is effective at distinguishing between photopeak and scatter events. The ROC curve complements the results from the confusion matrix, providing additional quantitative evidence of the CNN's strong classification performance.

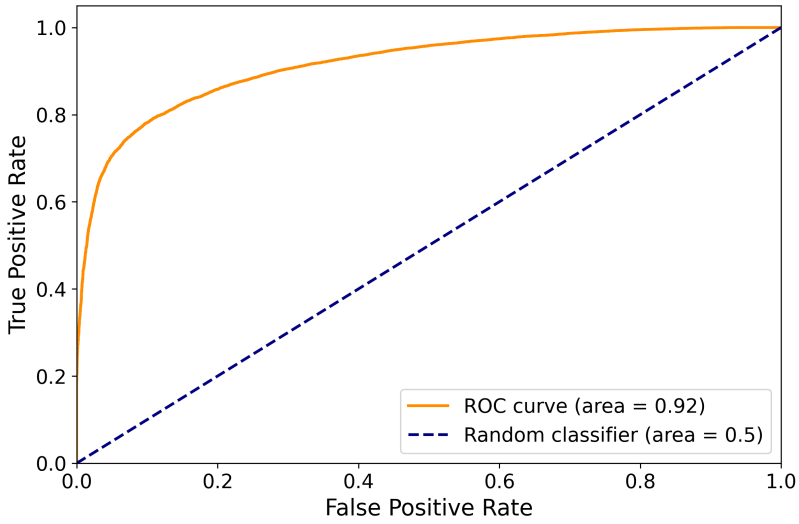


Figure 3.13: The ROC curve for the CNN, which shows the TPR and FPR for the test data. A random classifier is expected to give points lying along the diagonal (FPR = TPR)

Accuracy, precision, recall, F1-score

The evaluation results are shown in Table 3.3, which shows that the CNN performed strongly across the training and test data, with all evaluation metrics greater than 70%. The high CV score on the training data (this is not applied to the test data, as standard) indicated that the model generalises well to new data, as it performed well across the 10 different folds, and so is not likely to over-fit on new data. The model correctly identifies the vast majority of true photopeak events (high recall) while keeping the proportion of incorrectly classified scatter events low (precision). The results were similar for the training and test data, with the greatest difference being 3.5%, further illustrating good generalisation. This balance demonstrates that the CNN model effectively prioritises the retention of photopeaks while still filtering out the scatter pulses.

Table 3.3: Evaluation results of the trained CNN (trained with data from 5 nuclides) on the 80% training and 20% test dataset. All values are given as percentages [45]

Data	Accuracy	Precision	Recall	CV score
Training	82.2	75.4	95.6	84.5
Test	78.7	72.5	92.4	N/A

The results in Table 3.3 show a decrease in the CNN performance relative to Table 3.2, following hyperparameter tuning. This decrease can be attributed to the additional data processing step (step F in Figure 3.8), introduced during hyperparameter tuning to account for the observed correlation between noise level and pulse height. Full details on the requirement for this step is discussed in Section 3.2.2. This step was not retrospectively applied to the other models as the primary aim of this chapter was to develop and optimise a single algorithm for scatter reduction, rather than to perform a fully standardised comparison across all models. Although it is likely that all models would have experienced a similar reduction in performance under the same conditions, this does not affect the validity of the results because the initial benchmarking served only to identify the most promising model, which was then refined under the most realistic conditions.

3.3.2 Combined source dataset results

The CNN was used in the MLCSA with the combined source dataset for validation, comprising ^{241}Am (59 keV) and ^{60}Co (1173 keV, and 1332 keV) sources simultaneously positioned 10 cm away from the detector. The data were collected and processed through the digital Compton suppression flow from Figure 3.6, which included the second pulse processing pipeline without the greyed out processes from Figure 3.8. The result of the MLCSA is shown in Figure 3.14, showing the before and after spectra.

There were three metrics used to quantify and evaluate the Compton suppression performance of the MLCSA on the combined source spectrum, which are described below. The three metrics were also applied to the physical CVe system from Section 3.1.2, to provide a robust comparison to an existing system, and to further validate the MLCSA. The spectra used were collected using the same experimental conditions (count time, sources, and detector) and the spectra from Figure 3.3 were the final before and after spectra used in the evaluation. For comparison to the DCSA from Section 3.1.3 [28], only the MDA metric was used as this was the only similar metric reported in [28]. The reproduction spectra from Figure 3.5 demonstrates

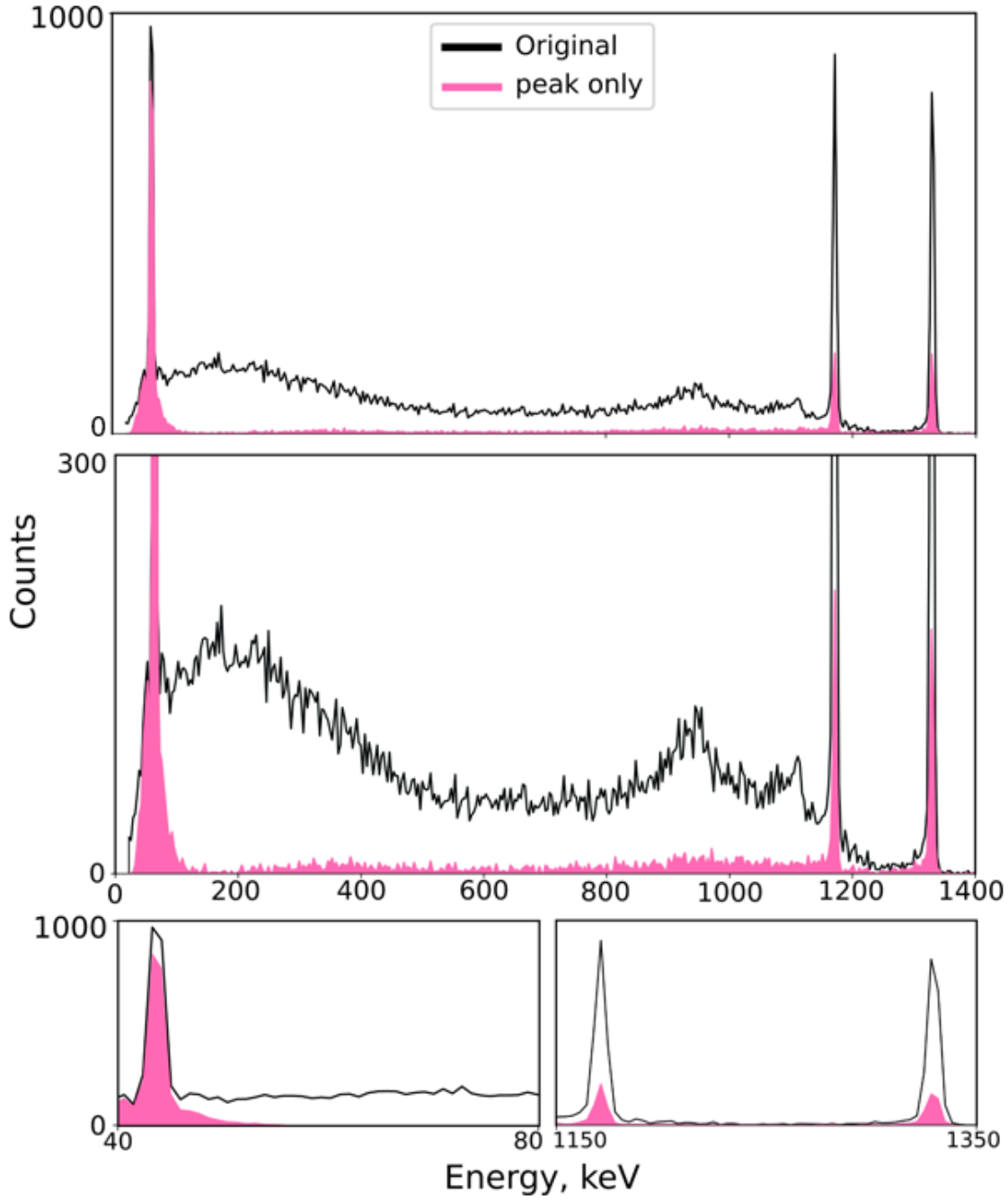


Figure 3.14: The combined source ^{241}Am and ^{60}Co spectrum, pre-MLCSA (black/line) and post-MLCSA (pink/filled). The top shows the full spectrum, middle is zoomed along the vertical axis, and bottom is zoomed along the horizontal axis (left shows around the ^{241}Am photopeak and right around the ^{60}Co photopeaks) [45]

how this digital method works but this did not involve the extensive modelling and testing work from [28] to find the best rise time cut off, and so this would not provide a fair comparison. For all three metrics to be applied to the rise time cut off method, the spectra would need to be attained, which was not possible at the time.

Spectra evaluation metric 1 - MDA

The first metric was the MDA improvement following the application of the MLCSA. This metric was only applied to the 59 keV ^{241}Am photopeak, as it is the main peak in the spectrum where high-energy scatter contributions need to be removed in this work, due to its low energy which sits in the continuum region. The improvement was calculated as percent difference, using Equation 2.15. The MDA represents the lowest level of radioactivity that a γ-ray spectrometry system can reliably distinguish from background radiation, for a specific set up. This is derived from the Currie method [51, 52], which defines a decision threshold (the count level above which the presence of activity is inferred) and the detection limit (the true net signal required to achieve a high probability of detection). The MDA accounts for the background, counting time, detector efficiency, and the γ emission probability of the radionuclide of interest. This relationship highlights that the MDA decreases (improves) with longer count times, lower backgrounds, and improved detector efficiency.

The MDA is expressed as [51]

$$\text{MDA (Bq)} = \left(\frac{2.71 + 4.65\sqrt{B}}{b_r \times \epsilon \times t} \right), \tag{3.9}$$

where B is the background counts sum, b_r is the yield/γ emission probability, ϵ is the full-energy peak efficiency of the detector for the nuclide of interest, and t is the live count time in seconds. To calculate B, the counts are summed from a region of 10 keV either side of the photopeak centroid (10 keV was chosen for the combined source spectrum, for the ^{241}Am photopeak; for other photopeaks this would differ depending on photopeak width).

The MDA equation is intended as an indicator of sensitivity for benchmarking between methods, rather than a precise metrological quantity; therefore, uncertainty propagation was not performed. The equation assumes ideal conditions, including stable Poisson-distributed background, accurate detector efficiency, and negligible systematic effects such as self-absorption or dead time. As a result, the reported MDA values should be interpreted as approximate lower bounds rather than rigorously uncertainty qualified limits. The MDA values before and after the MLCSA, the

percent difference improvement observed in the post-MLCSA and post-CVe, and the percent improvement for the DCSA are given in Table 3.4.

Spectra evaluation metric 2 - SBR

The second metric was the signal-to-background ratio (SBR) improvement following the application of the MLCSA. This metric was also only applied to the ^{241}Am 59 keV photopeak, as the main low-energy peak in the spectrum where high-energy scatter contributions need to be removed. The signal in this work is defined as the net area (a_n) of the photopeak above the background level, and the background (b_g) is defined as the underlying area beneath the peak; these are defined in Section 2.2.8 and illustrated in Figure 2.9. The SBR was calculated on the spectrum before and after the application of the MLCSA, and the percent difference was calculated using Equation 2.15, to determine the improvement. The SBR equation is

$$\text{SBR} = \left(\frac{a_n}{b_g}\right). \tag{3.10}$$

The SBR values before and after the MLCSA, the percent difference improvement observed in the post-MLCSA and post-CVe are given in Table 3.4.

Table 3.4: Evaluation of the photopeak SBR and MDA improvements for ^{241}Am. Values before and after the MLCSA [45] and CVe are provided, along with the percentage difference following the application of each method. The DCSA % Diff was calculated from the values in [28]

Method	Metric	Before	After	% Diff
MLCSA	MDA	156 Bq	76 Bq	51
	SBR	1.23	1.83	49
CVe	MDA	183 Bq	160 Bq	12
	SBR	20.77	24.78	19
DCSA [28]	MDA	-	-	37
	SBR	-	-	-

Spectra evaluation metric 3 - percent reduction in counts

The third metric represents the percentage difference (reduction) in counts before and after the MLCSA was applied to the live spectrum, using Equation 2.15. This was applied to several areas of the spectrum, including the three photopeaks of interest from the sources and two energies within the Compton continuum area

(200 keV, the back scatter region, and 400 keV to represent the lower and higher energy parts of the continuum). A higher percentage difference on the photopeaks represents misclassification and is therefore undesirable to prevent important data loss, indicating better performance of the MLCSA. The percent difference on the energies representing the Compton continuum should be as high as possible, as this indicates better Compton suppression performance through correct classification and removal of scatter pulses. The percent difference in the counts of the aforementioned energies of interest are given in Table 3.5.

Table 3.5: Evaluation of the photopeak counts reduction, and Compton counts reduction, as a percentage difference between pre- and post- MLCSA [45] and CVe. Values highlighted red indicated reductions where the optimal value would be 0% and values in green indicate where the optimal value would be 100%

Region	Energy, keV	MLCSA count reduction, %	CVe count reduction, %
^{241}Am	59	15	18
Low-energy scatter	200	**99**	81
High-energy scatter	400	**95**	94
^{60}Co	1173	79	77
^{60}Co	1332	77	76

3.4 Summary

In this chapter, a novel machine learning-based algorithm was created as a proof-of-concept tool, the Machine Learning-based Compton Suppression Algorithm (MLCSA), comprising a 1D CNN. The MLCSA was developed to reduce the effects of the Compton continuum on low-energy γ rays, and to improve the MDA and SBR of lower energy photopeaks, which positively impacts the γ-ray spectrometry process from Section 2.2.2 and Figure 2.3. The MLCSA was compared to an existing hardware Compton suppression solution, the CVe, and an existing digital solution, the DCSA, and was shown to perform similarly or better. The resulting spectrum from the MLCSA on the combined source data (^{241}Am and ^{60}Co) show that the Compton continuum from ^{60}Co scatter can be reduced by $\sim 97\%$ using the MLCSA, compared to $\sim 88\%$ for the CVe. Despite the MLCSAs high performance in reducing the continuum, the higher energy photopeaks were reduced in counts by $\sim 78\%$, however, this is similar to the CVe at $\sim 77\%$, but better than the DCSA which removed all high energy photopeaks. The SBR of the 59 keV ^{241}Am peak was improved by 49%, and the MDA improved by 51%, whereas for the CVe these were 19% and 37% respectively, and the MDA improvement of the DCSA was 37%. The MLCSA has demonstrated improved SBR and MDA performance with a digital advantage that it could be readily adapted to other detectors, eliminating the requirement for additional hardware. A higher SBR indicates that the photopeak stands out more clearly against the residual background, simplifying peak identification and reducing ambiguity in complex spectra. Similarly, reducing the MDA means that smaller amounts of activity can be confidently detected. Together, these improvements in SBR and MDA create conditions that would allow low-energy γ-rays and low-activity nuclides to be more readily detected. This could have benefits to fusion waste and research applications, for example by supporting accurate characterisation for large fusion-related projects such as the JDR project at the UKAEA, or by enabling reliable quantification of key radionuclides in research settings where higher MDAs would otherwise limit scientific output. Therefore, this novel method is valuable to fusion, and future developments could improve the algorithm into a highly useful laboratory tool.

The MLCSA development has addressed all three research questions from Section 3.1.4. The research questions (black) and how they have been addressed (pink) is summarised below:

1. Can the background from the Compton continuum typically found in the low-energy part of a spectrum from γ-ray spectrometry be reduced digitally, using state-of-the-art machine learning techniques? The MLCSA, a machine learning-based algorithm, was applied to the ^{241}Am 59 keV photopeak, where the SBR and MDA improved by approximately 50%. Since both metrics are strongly influenced by background counts, their enhancement demonstrates that the MLCSA successfully suppressed the low-energy Compton contribution.

2. Can the machine learning technique be implemented while preserving higher energy photopeaks and not rely on any detector modelling? The MLCSA does not involve any complex or detailed detector modelling, it simply requires data to be collected from each new detector. The higher energy photopeaks are significantly reduced, but at a similar level to the physical CVe system.

3. Can the techniques produced in this work be used in real-world applications? The MLCSA can, in theory, be used on any spectrum where there is significant background from Compton scattering, especially where low-energy γ rays may be hidden by the Compton continuum. This has application in waste classification measurements, such as in the UKAEA's JDR project, where a reduction in the Compton continuum is particularly important in waste classification to avoid mismeasurements. The MLCSA also has applications in research measurements, such as activation foil measurements [6].

3.4.1 Discussion

One discussion point is that it is not entirely clear what the CNN within the MLCSA uses to differentiate the pulses, due to the nature of the black box algorithm. It is likely that it is a combination of pulse features such as rise time, but there could be others. If rise time is the primary discriminating feature, this effectively partitions the detector volume into depth layers, since the average penetration depth of a γ ray is a function of its energy. This depth dependent behaviour means the CNN may be implicitly performing a depth of interaction classification. Also, if rise time is all the CNN uses, then any γ rays entering the detector from the back and deposit their energy immediately would have a short rise time, and so this may appear as a scatter γ ray. This is unlikely as the samples are positioned in front of the detector,

and back shielding can be used where required, and so has not been deemed to be of concern. This behaviour could also be advantageous, as sensitivity to rise time may reduce the influence of low-energy environmental γ rays (such as the 46.5 keV γ ray from lead-210), which often produce short rise time pulses, thereby minimising their contribution to background and improving spectral quality. The black box nature of the CNN means that such interpretations remain speculative without further investigation. The application of explainable artificial intelligence (XAI) techniques could provide insight into which features of the pulse the CNN uses to make its classifications, and is therefore identified as a direction for future work.

A second discussion point, which demonstrates a further strength of the MLCSA, is its robustness to signal noise compared to traditional digital pulse shape analysis methods such as the DCSA [28]. The DCSA relies on precise determination of timing markers, which are sensitive to noise on the pulse signal and can lead to erroneous rise time estimates. In contrast, the MLCSA trains on the general shape of the pulse, making it more tolerant of noise as the classification is based on the overall signal shape rather than precise timing points. Similarly, the MLCSA is not sensitive to coincidence γ cascades, such as those observed in ^{60}Co, which can distort the pulse shape. Since the CNN learns from the full pulse waveform rather than relying on specific timing characteristics, such coincidence events are less likely to cause misclassification. These properties represent practical advantages of the machine learning approach over traditional methods.

A third discussion point is that the MLCSA has only been shown to work on spectra that contains ^{241}Am and ^{60}Co. While this does have useful applications, most spectra are more complicated and contain a more complex network of radionuclides, which suggests a need for further development to broaden the applications that it can be used on. One way of achieving this would be to include significantly more radionuclides to the single nuclide dataset, but achieving this would require significantly more reference sources spanning a wide energy range to gather training and testing data, which is beyond the scope of this work due to costs and availability surrounding radioactive sources. Also, expanding the source set in this way would introduce greater complexity due to overlapping spectral features and more challenging classification scenarios. For the purposes of this proof-of-concept study, two sources representing the extremes of the energy range were selected to demonstrate feasibility, with the inclusion of intermediate energies left as a direction for future work. Although simulated γ-ray spectra could have been used to expand the dataset, this work focuses on measured data to ensure that the machine learning model is trained on realistic detector responses under experimental conditions. It is also

important to note that the MLCSA is able to generalise to new data, as evidenced by its treatment of the unseen 1173 keV photopeak, so its presence in the post-MLCSA spectrum suggests the CNN model is able to generalise to pulses from unseen energies.

A fourth discussion point is that the MLCSA does not run in real time, as currently the experimental data are gathered and later processed to perform the Compton reduction. Whilst this does not hinder the process, this would in theory be possible to run live as part of the data collection step, which would broaden the applications of the MLCSA. This would take more processing power than is currently available in the RADLab, such as a better digitiser with a high bandwidth stream to a graphical processing unit (GPU). Consideration would also be needed into how useful this would be for specific applications, therefore, this has been identified as a future development opportunity.

A fifth discussion point is that the MLCSA currently works for one specific detector, but with some future developments this could be expanded to be detector agnostic. The development work of the MLCSA did not involve complex modelling of a specific detector, but instead relied on measuring a set of reference sources with a detector and using those as training data. Therefore, this could be applied to other detectors as the CNN model within the MLCSA would only need re-training with pulses from that detector - this would require a different machine learning model for each detector. However, it may even be the case that one model could suffice for all of the detectors it was trained on, with enough validation work. This would require significant amount of training data, development, and access to enough detectors with the correct connections to be able to fully develop, and therefore is out of the scope of this thesis.

A sixth discussion point for future development could explore the integration of the CVe system with the MLCSA to create a hybrid solution that combines physical suppression with machine learning-based classification. The CVe system could not only improve suppression by further reducing the Compton continuum but could also serve as a source of ground truth for training and validating the MLCSA CNN model, providing high confidence labels for distinguishing true photopeaks from scatter events. This synergy could significantly enhance performance in complex spectra where residual Compton scattering remains. However, implementing this would require extensive experimental work, including substantial access to the CVe system at the UKAEA and the acquisition of paired spectra (with and without the CVe system) across multiple radionuclides and energy ranges. These resource and time constraints place this development beyond the scope of this thesis, but it represents a valuable direction for future research.

A seventh discussion point is that the lack of uncertainty estimation represents a limitation of the current work, as uncertainty estimation would provide additional confidence information alongside classifications, which could be valuable in a nuclear diagnostic context where misclassification carries practical consequences. Incorporating uncertainty estimation is therefore identified as a direction for future work.

Finally, further machine learning developments could include feature engineering, dataset balancing, or exploring alternative model architectures to further reduce the retention of scatter pulses or increase the retention of photopeak pulses.

Chapter 4

Machine Learning-based Efficiency Calculations (MaLBEC)

4.1 Introduction

In γ-ray spectrometry, analysis of the spectra measured from radioactive samples enables the identification and quantification of the radionuclides present, thus enabling further interpretation. As discussed in Chapter 2, the results of the analysis are vital for characterisation of radioactive waste and for research work of fusion samples. For waste measurements, an inaccurate efficiency value could lead to misclassification. For research measurements, accurate efficiency calculations and the corresponding calculation of activity results, are a key part in reaction rate calculations which can then used to determine fusion fuel consumption and fusion power. Quantification is done through the activity calculation, Equation 2.9 as discussed in Chapter 2, where the absolute efficiency, ϵ, forms a key part of the activity calculation.

The efficiency is not a fixed property of the detector but rather is influenced by experimental factors, and as such varies with energy. The efficiency curve of a detector shows a characteristic energy dependence: at very low energies the efficiency decreases due to absorption in the detector window and dead layers. A maximum efficiency is observed in the low to intermediate energy region, where γ rays are sufficiently penetrating but still have a high probability of full absorption. At higher energies, the efficiency decreases, reflecting the reduced probability of complete energy deposition in the crystal due to the dominant γ-ray interaction mechanisms. In this high-energy region, CS is the most likely process, as γ rays scatter within the detector and deposit only part of their energy before escaping. As the γ-ray energy increases above 1022 keV, PP becomes possible, but this process often leads to partial

energy loss if one or both of the annihilation photons escape the detector. Meanwhile, the PE effect, which is the only interaction that guarantees full-energy absorption in a single event, becomes less probable as energy increases. The combined effect of these processes leads to the gradual fall off of the efficiency curve at high energies [26]. Some of the primary experimental factors that influence efficiency are listed below [33]:

- **The crystal size** - a larger detector crystal has a greater volume of active material, which increases the probability that a scattered γ ray will undergo further interactions before escaping. This effect reduces the rate at which the efficiency falls with energy, leading to a flatter slope in the high-energy region. Smaller detectors have a lower chance of capturing all the scattered or secondary photons, which results in a steeper decline in efficiency [26].

- **The crystal shape** - as an example, coaxial HPGe detectors with thicker end caps or dead layers may show stronger losses at low energies (as these are particularly susceptible to absorption in these regions), while well-type detectors, which nearly entirely surround the source, provide much higher efficiency across a broad energy range.

- **The detector attenuating layers** - this includes the material, size, and any inactive/dead layers or entrance windows. These factors determine the likelihood that a γ ray entering the detector will deposit its full energy and contribute to the full-energy peak [26, 27].

- **The sample properties** - this includes activity, emission probability, and self-absorption within the source matrix (e.g. sample geometry and density). These characteristics influence the efficiency curve by altering the probability that γ rays escape the sample and reach the detector. For example, high-density or large-volume samples increase self-absorption, leading to stronger losses at low energies and a steeper decline in efficiency in that region. Conversely, thin or low-density samples minimise absorption effects, resulting in a flatter efficiency response across the energy range.

- **The geometry of the measurement** - a key factor is the source-to-detector distance. When a source is moved closer to the detector, this increases the solid angle subtended by the detector, raising the overall efficiency. This primarily shifts the curve upward or downward without fundamentally altering the slope. While extended or non-point-like sources further decrease the probability of

γ-ray detection [53]. The solid angle, Ω, represents the fraction of the total emission sphere that a detector subtends at the source (see Figure 4.1). For a circular detector of radius a (cm or m), positioned at a distance d (cm or m) from a point source along its axis, the solid angle is given by [53]

$$\Omega = 2\pi \left(1 - \frac{d}{\sqrt{d^2 + a^2}}\right). \tag{4.1}$$

This expression accounts for the geometry of the detector and is valid even when the source is close to the detector.

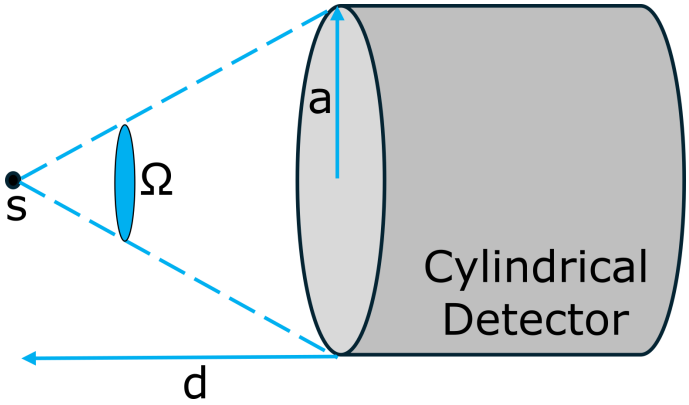


Figure 4.1: A diagram for calculating the solid angle (Ω) subtended by a cylindrical detector for an isotropic point source (s), adapted from [53]

- **Electronic effects** - it should be noted that overall system efficiency is also influenced by electronic effects such as dead time, sampling frequency, and signal processing, which can reduce the fraction of events successfully recorded. These effects are beyond the scope of this work as they are related to the detector hardware and acquisition electronics, and so they apply equally to all methods considered in this work.

This list is not exhaustive, but covers the primary factors that affect the efficiency curve. The complexity of these factors highlights the importance of calculating the efficiency calibration accurately in γ-ray spectrometry, since the relationship between the number of detected counts and the actual activity of a source is not straightforward. The efficiency calibration is therefore essential as it provides the quantitative link between the observed count rate in the detector and the true emission rate of γ rays from the source. Without calibration, quantitative analysis of γ spectra would be impossible, since raw counts do not account for the combined influence of these factors.

A key point to consider with existing efficiency calculation methods (which are discussed in the next section) is that it is typically time consuming to generate models of the samples in order to calculate the efficiency values. On the small scale this has minimal impact, but when there are many samples or many detectors (as each detector will have their own efficiency curve) this can be a time-intensive process. In waste, especially in large scale projects such as the decommissioning of nuclear machines (e.g. JDR at the UKAEA), significant volumes of the machine are required to be measured via γ-ray spectrometry. In JET, the volume of waste is expected to be around 1.03×10^7 kg (10300 tonnes). Some of this will be standardised into waste package items, but this won't be possible for all items. New efficiency models will be required for each new item shape, which could be extremely time consuming. Prior to measuring items of waste, a large sampling regime is required to understand the extent of the activation on the machine - these too will require the γ-ray spectrometry measurement of hundreds or potentially thousands of these items. Efficiency calculation models will be required to be created for all items, and so a solution where this can be done quickly, and where possible, automatically for minimal user input as possible will be important. Similarly, in research measurements, and measurements to continually determine fusion power and fuel consumption, many items are required to be measured daily (e.g. through "rabbit" systems [54, 55]), where again, the ability to quickly attain the efficiency and therefore activity calculations is important.

4.1.1 Existing solutions

There are several existing methods for calculating the efficiency values, including experimental, Monte Carlo based, or a hybrid approach (semi-empirical).

The most direct, standard method is experimental calibration using standard γ-ray reference sources of known activity and emission probabilities. By recording the count rate in the full energy photopeak for each γ line and normalising to the source activity and emission probability, the absolute efficiency at that energy can be calculated. Repeating this process for multiple energies allows construction of an empirical (measured) efficiency curve. While highly reliable when appropriate reference sources are available, this method has practical restraints. Only a small number of reference sources are easily manufacturable and they can be costly, therefore the range of energies from these sources will be limited. It is also preferable that the reference source geometry closely matches the geometry and size of the samples being measured, to provide a representative efficiency curve, this means measurements

need to be repeated to generate a new efficiency curve for different geometries and it is often not feasible. Samples from decommissioning, waste, or research work can be any geometry. This method works best for standardised samples and is limited to the availability of reference sources. An example commercial software that utilises this approach is Ortect's GammaVision (GV) software [56].

Modern approaches increasingly rely on Monte Carlo simulations, which are a class of computational algorithms that use repeated random sampling to approximate numerical results. Codes such as Monte Carlo N-Particle (MCNP) [57], GEometry ANd Tracking (GEANT4) [58], or OpenMC [59], are used to model the transport and interaction of γ rays within the detector and surrounding materials. These simulations can account for detailed detector geometry, including crystal size, dead layers, and shielding, as well as source distribution, sample container geometry, and attenuation in air or other materials. Monte Carlo methods can be used to generate efficiency curves for energies and geometries where experimental calibrations are impractical. However, their accuracy depends on careful modelling and validation against experimental data, typically using one or more reference sources to ensure that the simulated detector response is realistic and valid. Also, the accessibility and barrier to entry are difficult for Monte Carlo methods, which can require expensive licences, extensive training and expertise, and may need a significant amount of time to create and run models.

The hybrid/semi-empirical approaches combine physical models of γ-ray interactions with experimentally determined parameters to calculate efficiency. These models are often expressed as mathematical functions that describe efficiency as a function of γ-ray energy. For example, one common approach is to fit experimental efficiency data points with polynomial or logarithmic functions, allowing interpolation between calibration energies. Commercial software often uses semi-empirical methods to predict detector efficiency for a wide range of geometries without requiring full experimental calibration, by combining experimental measurements at a few key energies with simulation or analytical interpolation to produce a complete efficiency curve. For example, Mirion's Genie 2000 software integrates In Situ Object Counting System (ISOCS) [60], which can generate efficiency curves for arbitrary sample geometries using analytical modelling based on a factory generated detector characterisation file derived from detailed Monte Carlo simulations. This eliminates the need for extensive physical calibration sources. Similarly, Ortec's Isotopic [61] software employs a combination of empirical calibration and Monte Carlo simulation to estimate detector response for a wide range of sample sizes and distributions, enhancing accuracy for complex or irregular geometries. These commercial tools

illustrate how hybrid methods can use both experimental data and computational modelling to provide flexible, reliable efficiency determinations. However, these commercial systems often require expensive software (and licenses), extensive training and expertise, can be time consuming, and often require restrictive software that is difficult to use in tandem with other laboratory automation tools. The use of these commercial tools also require the user to be reliant on the company for updates and developments to the software.

There were, at the time of the development, no alternative solutions that utilise state-of-the-art machine learning as an alternative solution to the efficiency calculation process. The choice of method depends on the availability of reference sources, detector complexity, sample geometry complexity, the availability of software/licenses, the time demands to create models, and the expertise/training of those performing the calculations. Overall, the restrictions and difficulties with all of these methods, along with the gap in the literature identified here, have identified an opportunity for a novel method that provides an all round solution to calculating the absolute efficiencies, with an emphasis on improving the speed of attaining the efficiency values for the end user.

4.1.2 Research aims

The work in this chapter explores the application of novel machine learning-based algorithms for efficiency calculations, comparing the effectiveness of different models and includes work on data preparation. This work was in preparation for the creation of a novel Machine Learning-based Efficiency Calculator (MaLBEC) [62]. The final algorithm is compared to an existing method (MCNP), to support the evaluation and validation processes.
The research questions addressed in this chapter are:

1. Can the absolute efficiency calculation process be improved and simplified (e.g. automated, sped-up, etc.), via state-of-the-art machine learning methods?

2. How does the novel machine learning method compare to a traditional method, through comparing the efficiency values, the resulting activity calculations, the computational speed, and the usability (e.g. simplicity of model creation)?

3. In what ways can any techniques produced in this work be used for real world applications, and how can their effectiveness be demonstrated through experimental validation?

4.2 Methods

Through step one of the the model selection process in Section 2.3.2 and Figure 2.10, this work was identified as a regression task, as the machine learning models would predict a discrete value for each new input. The training method for these machine learning models was supervised, as the training data could be pre-labelled with an efficiency value, based on a simulated sample.

The MaLBEC predicts efficiency values for 11 γ-ray energies for radioactive items, based on just four inputs from the geometry of the sample and measurement. The MaLBEC comprises a MLP as a supervised, regression machine learning model. In the broader context of computational modelling, the MaLBEC can be characterised as a machine learning-based reduced order surrogate model. Surrogate models, first formalised by [63], are lightweight approximations of computationally expensive simulations, trained to reproduce their outputs as a function of key input parameters at a fraction of the computational cost. Rather than running a full simulation for each new input, a surrogate is trained on a set of simulation outputs and then used to predict new outputs in real time. Such approaches have been adopted in the nuclear industry to replace costly Monte Carlo simulations, for example in radiation transport modelling [64], where the computational burden of repeated high-fidelity simulations makes direct simulation impractical. The MaLBEC follows this paradigm, replacing costly Monte Carlo efficiency simulations with a trained MLP that predicts efficiency values directly from geometric inputs. The methodology, selection and refinement of the machine learning model, and benchmarking/validation are described in the following sections.

4.2.1 Data collection

Step two of the the model selection process in Section 2.3.2 and Figure 2.10 is the data collection and exploration step. For this work, a combination of experimental and simulated data were required, which are discussed in the following sections.

Detector

The high-resolution, HPGe detector system, owned by the RADLab at the UKAEA, used in this work was an Ortec Trans-SPEC-DX-100T, shown in Figure 4.2. This detector was used for the experimental data collection, and a model of it was used to generate simulated data, for consistency and direct comparisons. The detector has a built-in digital MCA. This design eliminates the need for liquid nitrogen, resulting

in a portable system that is particularly well-suited for applications such as waste characterisation. The technical information and specifications for this detector are in Appendix A.3.

Figure 4.2: A Photograph of the Trans-SPEC detector in the RADLab at the UKAEA, with the external power unit attached

Experimental data - reference calibration source

The mixed radionuclide source used in this work contained γ lines from cadmium-109 (^{109}Cd, 88 keV), ^{57}Co (122 keV), ^{137}Cs (661 keV), and ^{60}Co (1173 keV and 1332 keV). The source activity was around 40 kBq and was measured for 3,600 s at a distance of 0 cm (placed directly by the end cap), resulting in a dead time of around 7%. All activity results related to this source are decay corrected to the reference date of 01/11/2017, using the exponential decay law [27], to account for radioactive decay occurring between the reference date and the measurement time. The decay correction equation, for the activity at a reference time A_0, is given by

$$A_0 = a\, e^{\lambda t}, \tag{4.2}$$

where a is the activity (Bq) at the measured time (Equation 2.9) and t is the time since the reference date (s). The decay constant, λ is

$$\lambda = \frac{\ln(2)}{T_{1/2}}, \tag{4.3}$$

where $T_{1/2}$ is the nuclide half-life (s).

Experimental data - radioactive sample

The experimental sample used in the validation of the final machine learning model in this work was a stainless steel item approximated as a cylinder (radius 2 cm, height 13.5 cm, measured density of 6.89 g/cm^3), that was positioned within the JET torus hall and was activated by fusion neutrons during DD and DT campaigns over the JET operational period (40 years). This sample formed part of the JDR sampling and characterisation programme, and is used in this work as a representative radioactive waste sample.

The sample activity was around 3 kBq, it was positioned at a distance of 30 cm from the Trans-SPEC detector and data were acquired for 800 s, which resulted in a dead time of <1%. These values were selected to reflect typical measurement conditions for active samples in the RADLab, ensuring consistency with standard laboratory practice. A photograph of the sample in relation to the detector, and the corresponding spectrum are shown in Figure 4.3. The γ rays of interest identified in Figure 4.3 are from ^{54}Mn and ^{60}Co. These are likely due to the $^{59}\text{Co}(n,\gamma)^{60}\text{Co}$ and $^{54}\text{Fe}(n,p)^{54}\text{Mn}$ reactions in the steel. The other photopeaks are due to NORM (as discussed in Section 2.2.7).

Simulated γ-ray spectrometry equipment - MCNP model

An MCNP model for the Trans-SPEC HPGe detector was created and Figure 4.4 shows the MCNP plotter image of the detector. As part of the validation process, the model was used to calculate efficiency values for 11 γ-ray energies, which were: 59 keV, 88 keV, 122 keV, 150 keV, 200 keV, 300 keV, 400 keV, 500 keV, 661 keV, 1173 keV, and 1332 keV. The energies were selected to span a broad energy range, to incorporate those available from reference sources at the RADLab (specifically: 59 keV, 88 keV, 122 keV, 661 keV, 1173 keV, and 1332 keV). The remaining energies were included to ensure that efficiency values were sampled across the central portion of the spectrum. For energies between these values, interpolation (such as linear or cubic) can be used. Each γ-ray energy required a separate MCNP file, as the reference source was modelled as a point source emitting monoenergetic γ-rays isotropically (uniformly in all directions) to match the default of the commercial software, Ortec's GV, used in comparisons.

The MCNP $F8$ tally was used to calculate detector efficiencies as it simulates the pulse height distribution, which closely represents the response of a real detector, and accounts for partial energy deposition, escape peaks, and relevant detector physics. The $F8$ tally scores the energy deposited per event in a specified cell, corresponding

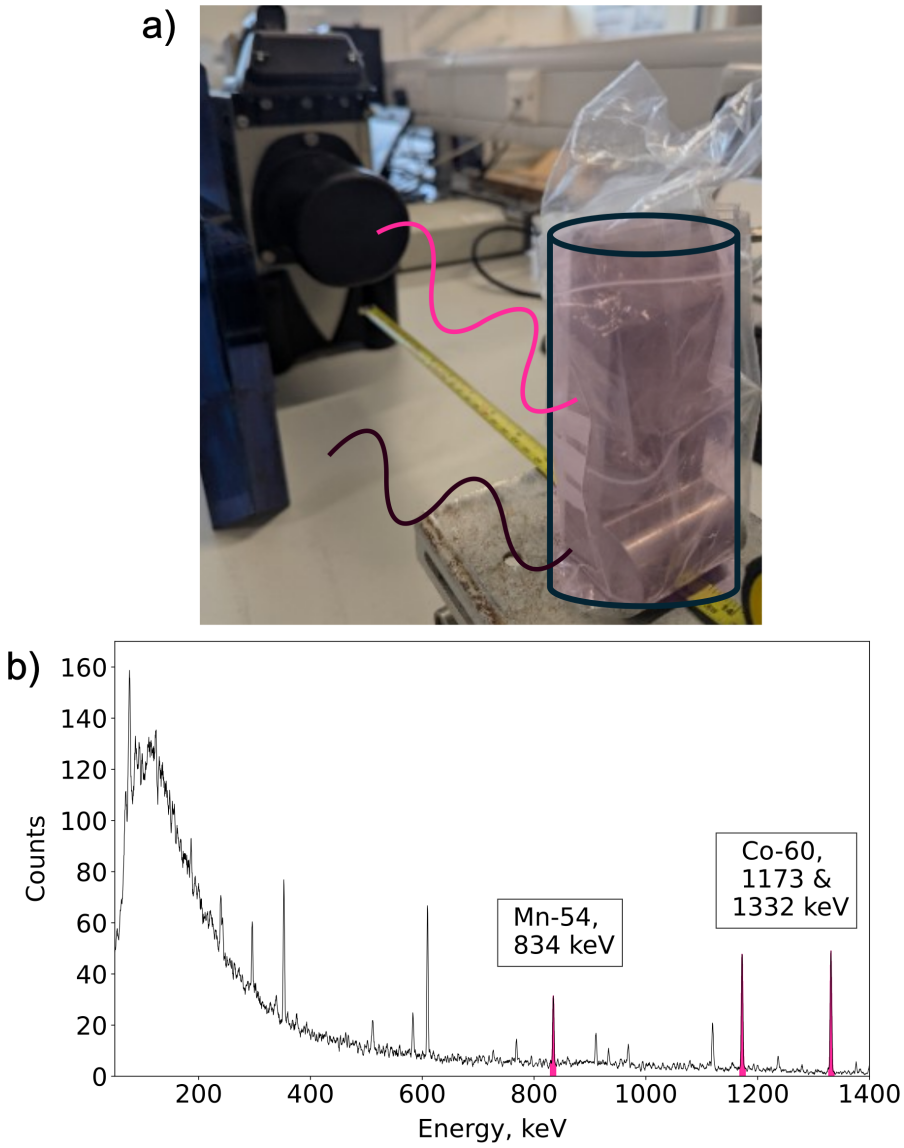


Figure 4.3: a) The experimental measurement set up, including the Trans-SPEC-DX-100T HPGe detector and a steel sample. A transparent cylinder is overlaid to illustrate the geometric representation used in the computational model. The pink (light) wavy line demonstrates a γ ray that reaches the detector and the black (dark) wavy line is one that is not incident on the detector and so will not be seen. b) The spectrum from the set-up shown in a), with the key photopeaks, from ^{54}Mn and ^{60}Co, filled in pink [62]

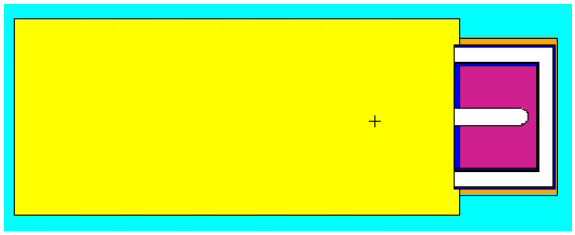


Figure 4.4: The MCNP model of the Trans-SPEC HPGe detector from the MCNP plotter, total length of 41.5 cm. Light blue is air, yellow is the aluminium detector housing, orange is PVC, dark blue is aluminium, white is a vacuum, green is steel (not visible at this scale), and purple is germanium crystal

to the detector crystal in this work, and produces a histogram of deposited energy. To determine a photopeak efficiency, MCNP integrates the counts within a defined energy window (photopeak energy) and normalises by the number of simulated source particles [57]. The output from MCNP is three values for the $F8$ tally bin: the bin identifier (e.g. energy), the mean tally result, and the relative error R. The relative error corresponds to one standard deviation (1σ) expressed as a fraction of the mean tally value, $R = \sigma/\text{mean tally result}$ [57]. It provides a measure of the statistical uncertainty associated with the tally result, based on the variance of scores accumulated over all particle histories. According to reference [57], results with relative errors greater than 50% are considered unusable, those between 20% and 50% are highly uncertain, and those below 10% are generally reliable. For detector tallies, a stricter criterion applies where the relative error should be less than 5% for high confidence. In this MCNP model validation work, all reported tallies met these criteria, with all relative errors below 1%.

The statistical uncertainties from MCNP were combined with measurement uncertainties to determine a full picture of the uncertainties in calculating the efficiencies. The relative efficiency uncertainty, σ_ϵ/ϵ, can be expressed as a combination of contributions and is derived as

$$\left(\frac{\sigma_\epsilon}{\epsilon}\right)^2 = \left(\frac{\sigma_{\epsilon_{\text{stat}}}}{\epsilon}\right)^2 + \left(\frac{\sigma_{\epsilon_{\text{geom}}}}{\epsilon}\right)^2 + \left(\frac{\sigma_{\epsilon_{\text{pos}}}}{\epsilon}\right)^2 + \left(\frac{\sigma_{\epsilon_{\text{dens}}}}{\epsilon}\right)^2, \tag{4.4}$$

where $\sigma_{\epsilon_{\text{stat}}}/\epsilon$ represents the relative statistical uncertainty (e.g. MCNP tally or model error), $\sigma_{\epsilon_{\text{geom}}}/\epsilon$ is the relative uncertainty due to geometry tolerances (0.5-2% for a ± 1 mm dimension uncertainty [33]), $\sigma_{\epsilon_{\text{pos}}}/\epsilon$ is the relative uncertainty due to positioning uncertainty (1-3% for a ± 1 mm positioning uncertainty [33]), and $\sigma_{\epsilon_{\text{dens}}}/\epsilon$ is the relative uncertainty due to density/composition variations (less than 0.5% for a $\pm 1\%$ density uncertainty [33]). The upper values for each of these were used in this work, for conservatism. When combined with the statistical relative errors for

the MCNP model, all relatives errors were below 4%. All uncertainties reported in this work are expressed as absolute values and correspond to one standard deviation (1σ).

As part of the MCNP model validation, the efficiencies from MCNP were compared with efficiency values obtained using commercial software GV, which uses the experimental approach (as described in Section 4.1.1). The resulting efficiency values and uncertainties are shown in Table 4.1 and the curves are shown in Figure 4.5, where dead time has been account for to account for the differences between simulation and experimental data. The uncertainties from GV (σ_{GV}) are provided by the software and includes all relevant contributions (e.g. counting statistics and geometry). The MCNP uncertainties (σ_{MCNP}) include statistical uncertainty from the simulation and other relevant contributions. The efficiency results for both methods were compared using Equation 2.15, with GV as the reference value.

Table 4.1: A comparison of GV and MCNP efficiency values and associated uncertainties for 11 γ-ray energies, with percentage differences

Energy, keV	GV ϵ	σ_{GV}	MCNP ϵ	σ_{MCNP}	ϵ % Diff
59	5.75×10^{-2}	1.88×10^{-3}	8.17×10^{-2}	3.00×10^{-3}	42.1
88	1.36×10^{-1}	7.66×10^{-3}	1.68×10^{-1}	6.15×10^{-3}	24.2
122	1.85×10^{-1}	5.40×10^{-3}	1.92×10^{-1}	7.01×10^{-3}	3.8
150	1.86×10^{-1}	4.66×10^{-3}	1.85×10^{-1}	6.74×10^{-3}	-1.0
200	1.62×10^{-1}	4.05×10^{-3}	1.57×10^{-1}	5.74×10^{-3}	-3.0
300	1.14×10^{-1}	2.86×10^{-3}	1.12×10^{-1}	4.11×10^{-3}	-1.7
400	8.57×10^{-2}	2.14×10^{-3}	8.67×10^{-2}	3.18×10^{-3}	1.1
500	6.85×10^{-2}	1.71×10^{-3}	7.13×10^{-2}	2.62×10^{-3}	4.2
661	5.23×10^{-2}	1.08×10^{-3}	5.62×10^{-2}	2.07×10^{-3}	7.5
1173	2.85×10^{-2}	6.40×10^{-4}	3.54×10^{-2}	1.32×10^{-3}	24.5
1332	2.57×10^{-2}	5.78×10^{-4}	3.20×10^{-2}	1.19×10^{-3}	24.6

Comparison of GV and MCNP efficiencies shows strong energy-dependent behaviour. In the mid-energy range (122-661 keV), agreement is excellent with differences typically within a few percent, indicating that the detector geometry and intrinsic response are well represented in the model. However, at low energies ($\leq$88 keV) MCNP overestimates efficiency by 24-42% and at high energies ($\geq$1173 keV) overestimates by approximately 25%. These discrepancies are consistent with physical and modelling limitations. At low energies, under representation of attenuating layers such as the end cap and dead layer can lead to inflated simulated efficiencies; the dead layer was experimentally estimated as 0.07 cm. At high energies, the absence of true coincidence summing in the MCNP source definition for cascade emitters (e.g.

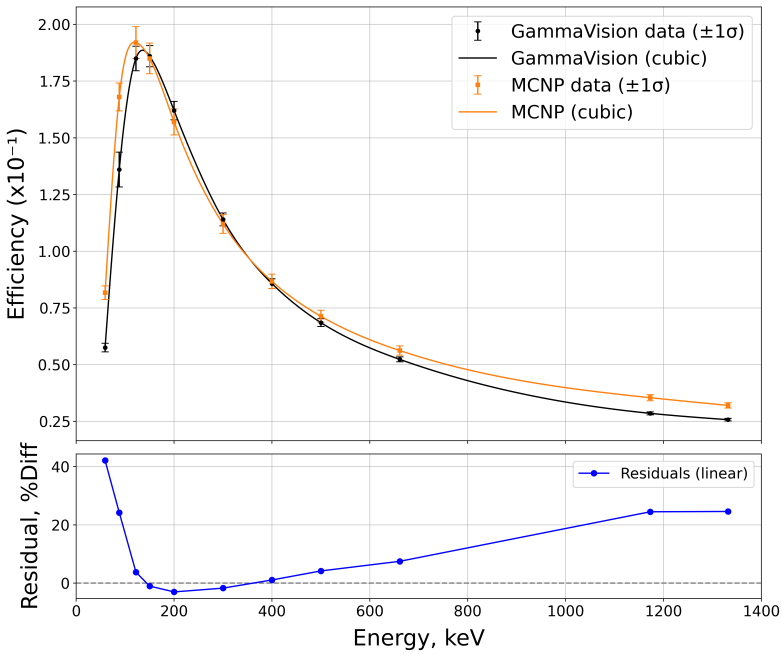


Figure 4.5: The efficiency curves for the Trans-SPEC model, comparing GV (black) and MCNP (orange), based on the reference source. The error bars represent $\pm 1\sigma$ uncertainties and the smooth curves are cubic spline fits. The bottom panel shows residuals as percent difference

^{60}Co) likely explains some of the observed over prediction, as GV does not apply summing corrections by default (this was off in the settings) and therefore reflects the actual detector response including summing losses. The MCNP method represents an idealised case without summing effects, while GV measurements incorporate these losses, leading to systematic differences for cascade lines. The MCNP model and GV assume a point source, whereas the actual calibration source was a small disk; this approximation can slightly bias simulated efficiencies, particularly at low energies and short source-to-detector distances. However, this was not considered significant enough to warrant correction. Uncertainties for both GV and MCNP are low (typically <5%), indicating that the observed differences are systematic (e.g. modelling assumptions) rather than statistical noise. Further refinement of the measurement set up (i.e. increase the source-to-detector distance) and MCNP model is recommended, but for the purposes of this work, it is deemed suitable for use.

Further validation of the Trans-SPEC MCNP model was performed by applying the MCNP derived efficiency values to calculate the decay-corrected activity of the mixed radionuclide reference source, as described in Section 4.2.1. The efficiencies from MCNP were loaded into GV for use in the automatic analysis sequence, meaning all activities and uncertainties were provided by GV. The activities, uncertainties, and

percent differences (with the known activity as the reference term in Equation 2.15) are presented in Table 4.2.

Table 4.2: Comparison of known and measured activities for different nuclides from the mixed radionuclide reference source. All activities and uncertainties are provided from GV

Nuclide	Energy, keV	Known activity, Bq	MCNP activity, Bq	σ_{MCNP}, Bq	Activity % Diff
^{109}Cd	88	1.53×10^4	1.70×10^4	9.59×10^2	10.95
^{57}Co	122	5.41×10^2	5.28×10^2	1.54×10^1	-2.44
^{137}Cs	661	2.89×10^3	2.68×10^3	5.56×10^1	-7.15
^{60}Co	1173	3.20×10^3	2.57×10^3	5.79×10^1	-19.58
^{60}Co	1333	3.20×10^3	2.57×10^3	5.79×10^1	-19.58

The results in Table 4.2 show underestimation for all nuclides except ^{109}Cd, which is overestimated, with differences all below 20%. Agreement for single-gamma emitters (not in coincidence) such as ^{57}Co is good, while cascade emitters like ^{60}Co exhibit the largest bias due to summing effects. This trend is consistent with the efficiency analysis; MCNP does not account for true coincidence summing, leading to inflated simulated efficiencies and consequently lower calculated activities. Uncertainties for MCNP activities are small (typically $< 3\%$), confirming that these discrepancies are systematic rather than statistical. Despite these limitations, the MCNP model provides sufficiently accurate activity estimates for the purposes of this work. Importantly, the level of accuracy demonstrated in the validation work is consistent with industry requirements and international standards for γ-ray spectrometry in waste characterisation and research measurements, where larger differences are generally considered acceptable unless the activity boundaries are near a consignment boundary (e.g. at the top of the OOS boundary) or where the consequences of a wrong decision are less critical [65, 66].

All MCNP calculations used in this chapter were performed using the UKAEA Intel Xeon E5-2640 high-performance computing cluster with 16 CPU cores, two sockets per physical node, and 125 GB of RAM.

Simulated data - randomly generated data

A broad dataset of simulated cylindrical samples was required to mimic the likely geometries and measurements encountered in the RADLab. Simulated data were employed rather than experimental measurements, as generating efficiency curves for over a thousand geometries, as required for training machine learning models,

would have been time consuming and resource intensive, and in practice it is difficult to obtain such a large and varied experimental dataset. Simulations, by contrast, allowed rapid exploration of the design space and ensured sufficient diversity across geometries. A total of 1300 geometries were selected, which was deemed sufficient to cover the typical range of sample sizes seen in the RADLab, providing enough variation to capture relevant trends while remaining computationally feasible. This balance ensured that the dataset was large enough to train robust machine learning models without incurring excessive runtime costs.

Each sample geometry required a separate MCNP model which had to be run at every γ-ray energy - given that every sample was run for 11 energies, approximately 14,300 MCNP files were needed in total. This required significant run time on the computer clusters. Data augmentation was considered as an option, as is common in machine learning problems where a large training dataset is required; however, for this work the four parameters that make up the new geometry were required to be used as labels to link them to the efficiency curve in machine learning training and testing, and so random efficiency curves could not be generated as this would lose the vital information linking the patterns to sample dimensions. The material and shape were fixed for this proof of concept work, as stainless steel and cylindrical, respectively (although future work could expand on this, as discussed in Section 4.4.1). Stainless steel was chosen as this is a material relevant to fusion waste and research measurements. Samples in the RADLab are often cylindrically shaped, due to the storage containers used for measurements and the sampling methods (e.g. drill cores), therefore this geometry was the focus of this work.

To generate the 1300 samples, four parameters were edited, namely: sample density, height, radius, and source-to-detector-distance. In order to achieve the MCNP modelling of these samples, a python script was written to randomly generate different cylindrical geometries in the form of an MCNP input file, within specified minimum and maximum limits over the four parameters. Uniform random sampling was used to generate the parameter combinations, whereby each parameter value was drawn independently from a uniform distribution within its specified bounds. The selection of a sampling strategy for surrogate model training is an established area of research in design of experiments [63]. More structured approaches, such as Latin Hypercube Sampling [67], have been shown to provide more uniform coverage of the input space with fewer samples than simple random sampling. While an alternative method would offer improved space filling properties, the uniform random sampling approach used here was considered sufficient given the relatively large number of samples generated, which provides reasonable coverage of the parameter

space. The use of other design of experiments strategies to improve sampling efficiency is identified as a direction for future work. An automation script was used due to the large volume of MCNP input files that were required. The upper and lower limits on the parameters were also chosen to mimic likely samples in the RADLab, these were:

1. Sample density, range: 2.5 g/cm^3 - 7.86 g/cm^3,
2. Sample height, range: 0.4 cm - 40 cm,
3. Sample radius, range: 0.5 cm - 15 cm,
4. Distance to the detector, range: 1 cm - 50 cm.

The sample height, radius, and distance to the detector all represent typical values seen within the RADLab, however the density limits require further consideration. When the volume and the mass of samples are measured, the calculated density is typically much smaller than the density of the uniform material, due to several factors. For solid items, one factor is due to small internal voids/air gaps within, which will reduce the density. A second factor arises from inaccuracies in measuring the samples; it is often difficult to determine the exact dimensions without overestimating, which in turn lowers the calculated apparent density. For packaged items (e.g. a steel container filled with many steel bolts) the packing ratio also needs to be accounted for. For stainless steel, the density is typically 7.86 g/cm^3 (hence this is the upper limit in the auto generator), and the lower limit accounts for a worst case poorly packed item. The equation for density (ρ, g/cm^3) and volume of a cylinder are given below,

$$\rho = \frac{m}{V} \tag{4.5}$$

where m is the mass in grams and V is the volume in cm^3. Volume is

$$V = \pi r^2 h, \tag{4.6}$$

where r is the radius (cm) and h is the height in (cm).

Once the 1300 samples (relating to a total of 14,300 MCNP input files) were auto generated and run on MCNP via the UKAEA clusters, the result was a .CSV file with the 4 input parameters and the resulting 11 efficiency values (one for each γ-ray energy) as the outputs. Figure 4.6 shows the efficiency curves of 12 geometries, corresponding to 1% of the full dataset so that the curves can be seen clearly without overlap. This figure also shows three of the samples in MCNP, to demonstrate the variation in sample dimensions and positions.

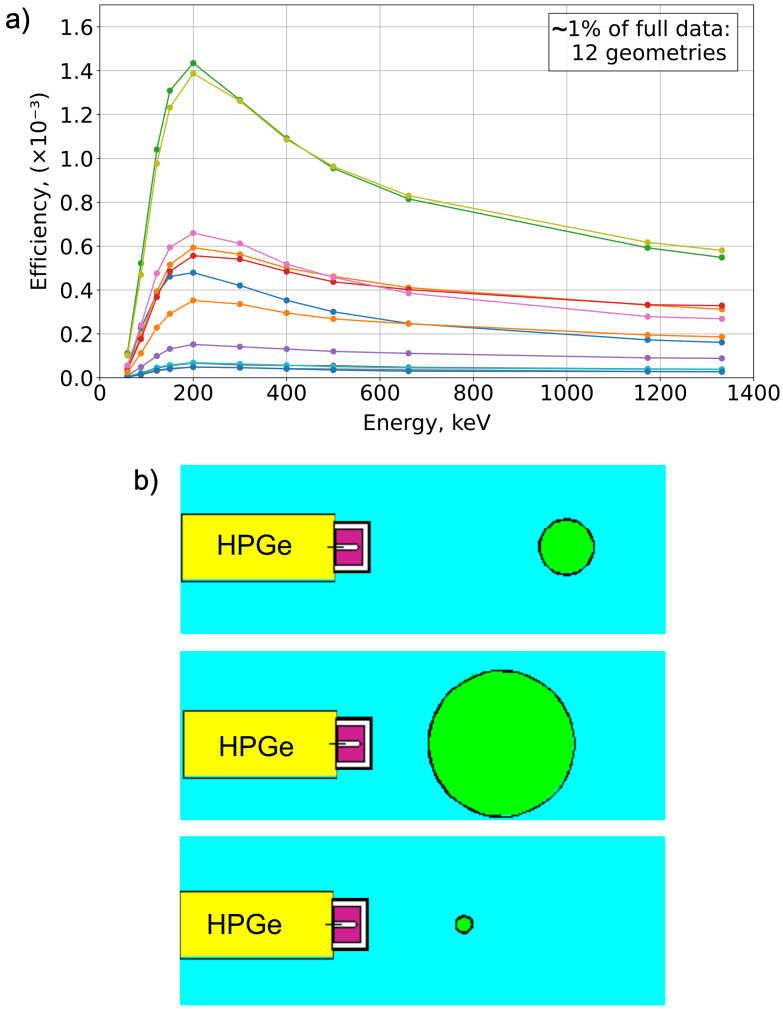


Figure 4.6: a) The efficiency curves of ∼ 1% of the full 1300 simulated cylindrical geometries. Each colour represents an efficiency curve of a different random geometry. b) Top down view from MCNP of three of the geometries simulated, showing example variations in the simulations, such as varying radius and detector-to-sample distance. The detector is on the left (yellow, with a total length of 41.5 cm) and the cylindrical samples are on the right (green) [62]

Data processing Pipeline

The data processing pipeline, as per step three of the machine learning model selection process from Section 2.3.2 and Figure 2.10, for this work is relatively simple. The pipeline was three steps:

1. Generate the training data in MCNP to produce a dataset with 1300 geometries.

2. Clean the data to remove any geometries with NaN data or 0 values for efficiency (e.g. from errors with MCNP simulations resulting from, as an example, geometry errors) leaving 1258 geometries.

3. Split the data into training (80%, 1006 geometries) and testing (20%, 252 geometries) datasets.

4. Scale the target variable (efficiency) by a constant factor of 10^5 to improve numerical stability during training. The original efficiencies were on the order of 10^{-5}, which would make the loss and gradients extremely small. Tiny gradients lead to very small weight updates, causing slow convergence or even stagnation during training. By scaling the targets, the gradients remain in a reasonable range, allowing the optimiser to make effective updates and train more efficiently. This also makes the scaled values easier to interpret during model development, but all predictions are inverse-scaled prior to evaluation, thus all root mean squared error (RMSE) values reported in this work are calculated on the original (unscaled) efficiency values.

The training data include the four input parameters and the labelled data are the efficiency values at the required energies for each geometry. For the test data, only the four input parameters were provided. For the experimental data, the four input parameters will be provided but these will be determined through experimental measurements, rather than through MCNP simulations, as discussed earlier.

4.2.2 Machine learning models

This task was a supervised, regression task, and therefore only machine learning models that were suitable for this type of work were selected. For step four of the model selection process in Section 2.3.2 and Figure 2.10, five machine learning models were evaluated. At this stage of the model selection process, all five models were evaluated using their default hyperparameters, as the purpose of this step was to compare model architectures rather than optimise individual models. Default hyperparameters provide a reasonable baseline for comparison, and hyperparameter tuning was reserved for the final selected model as per step five of the model selection process (Section 2.3.2). The models selected are detailed in step four in Section 2.3.2, and comprised four traditional regressors (RF, XGB, SVR, and KNN) and one deep learning regressor (MLP).

The five machine learning models were evaluated using the chi-squared (χ^2) statistic on the training dataset, which measures the discrepancy between observed and expected values. The χ^2 statistic was used as it is a single, unit-less measure, which provides a straightforward basis for comparing the models, regardless of the scale. In this work, this relates to the difference in the actual and predicted efficiency values for each energy, per geometry. The χ^2 statistic is defined as

$$\chi^2 = \sum_{i=1}^{n_g} \frac{(O_i - E_i)^2}{E_i}, \tag{4.7}$$

where O_i is the observed efficiency for geometry i, E_i is the expected efficiency for geometry i, and n_g is the number of geometries being compared. A larger χ^2 indicates a greater difference between observed and expected efficiencies, and therefore a poorer performance. The χ^2 value for each model are shown in Figure 4.7, where the MLP was identified as the best performing due to having the lowest total χ^2 value. This superior performance can be attributed to the inherent ability of neural networks to capture complex non-linear relationships between input features and output efficiencies. The efficiency response is influenced by multiple interacting parameters, including sample height, radius, density, and detector distance, which exhibit correlated non-linear dependencies on γ-ray attenuation and geometric effects. Unlike tree-based methods (RF, XGB) or distance based approaches (KNN), a MLP uses multiple hidden layers and non-linear activation functions to approximate these interactions effectively.

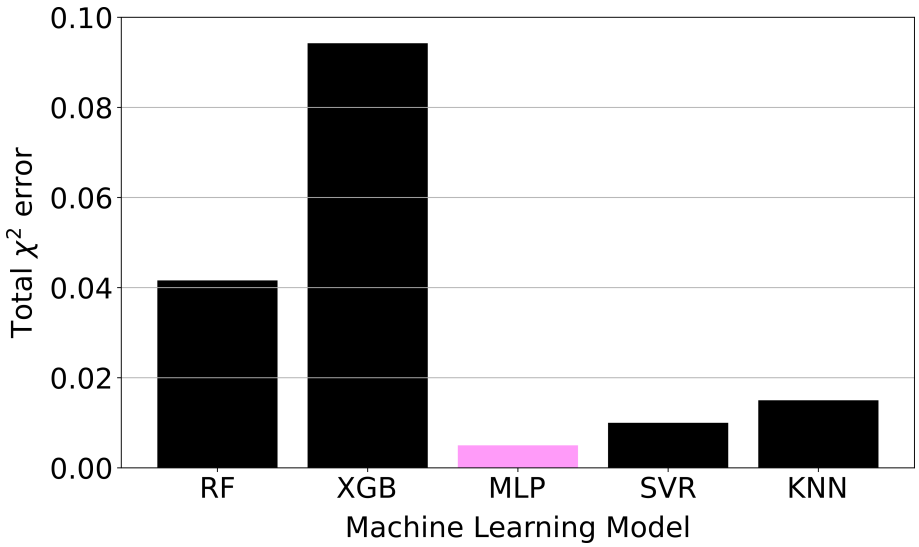


Figure 4.7: A bar chart showing the total χ^2 values for the five machine learning models evaluated

4.2.3 MLP theory

The final machine leaning model identified as most suitable for further development in the MaLBEC was the MLP, as it produced the lowest χ^2 value. An MLP is a type of feedforward ANN (discussed in Section 2.3.3). The MLP consists of one or more hidden layers of interconnected nodes, called neurons, that are fully connected

to the input layer and the output layer. Each neuron in the hidden layer applies a non-linear activation function to the weighted sum of its inputs, allowing the network to learn complex, non-linear patterns in the data [37]. During training, the model is optimised using backpropagation (see Section 2.3.3). The MLP has several parameters that require tuning/optimising, to ensure the best performing model, which are discussed in the next section.

Hyperparameter tuning process

As per step five of the machine learning model selection process from Section 2.3.2 and Figure 2.10, the hyperparameter tuning of the MLP was performed via a grid search [35] across five key parameters, with the objective of improving the evaluation metric of the RMSE. The RMSE was selected because for a unit-less target such as efficiency, it preserves the magnitude of the underlying metric, making the size of prediction errors directly interpretable. The RMSE is a measure of the average size of prediction errors, giving higher weight to larger errors because of the squaring operation, and is defined as [68]

$$\text{RMSE} = \sqrt{\frac{1}{n_d}\sum_{i=1}^{n_d}(y_p - y_a)^2}, \tag{4.8}$$

where n_d is the number of data points, y_p is the predicted efficiency for the ith geometry, and y_a is the actual (true) efficiency for the ith geometry. Lower RMSE values indicate that the predictions are closer to the actual values. The RMSE was calculated on the training and testing datasets, and it was also calculated on the mean CV score obtained from 10-fold CV (Figure 2.11). The CV score is computed using negative mean squared error (MSE) as the underlying scoring metric, which is the negated MSE returned by the CV procedure; the square root of the absolute value is then taken to yield the reported RMSE, ensuring consistency with the primary evaluation metric.

To select optimal hyperparameters a grid search was used (see Section 2.3.2) with a training loss of negative MSE and evaluation metric of RMSE. The combination that yields the lowest RMSE is selected as the optimal set of hyperparameters. Although computationally expensive, grid search ensures that the entire parameter space is explored, providing a robust approach to hyperparameter optimisation. Early stopping was enabled to prevent overfitting and over-optimisation, as well as to reduce computational cost.

The five hyperparameters specified in this work, and their respective search ranges, were:

1. **The loss regularisation parameter (alpha)** – a penalty term added to the loss function (MSE) based on the sum of the squares of the weights. These weights control how strongly a neuron's output influences the next neuron. Penalising large weights reduces over-reliance on individual connections, helping prevent overfitting and improve generalisation while maintaining necessary variation for learning. If alpha is too high, weights shrink excessively, limiting model capacity and causing underfitting; if too low, weights can grow large, increasing overfitting risk [35].
   Grid range: [0.0001, 0.001, 0.01, 0.1, 0.12, 0.2]. These initial values were based on typical ranges for neural networks, with additional values introduced through experimental tuning to refine regularisation.

2. **Number of iterations with no change** – stopping criterion for early stopping. If the MSE does not improve for a specified number of consecutive iterations (epochs), training is stopped.
   Grid range: [5, 10, 20, 40]. The values were chosen experimentally to balance early stopping responsiveness with allowing enough iterations for convergence.

3. **Network architecture** – number and size of hidden layers, tuned to provide sufficient capacity to capture complex relationships in the data while avoiding unnecessary complexity.
   Grid range: [(50,), (75,), (100,), (75, 40), (100, 50), (100, 100), (150, 75)]. The architectures were based on common practice for moderately complex regression tasks, with additional configurations introduced through iterative experimentation.

4. **Learning rate** – controls the step size of weight updates during training, balancing convergence speed with stability.
   Grid range: [0.0001, 0.001, 0.005, 0.01, 0.02]. The values were selected from typical ranges for MLPs and validated through experimental trials to avoid divergence (too high) or slow learning (too low).

5. **Maximum iterations (`max_iter`)** – limits the number of training epochs to ensure convergence without excessive computation.
   Grid range: [1000, 3000, 5000, 8000, 10000]. The values were chosen to allow sufficient training for complex architectures, with adjustments made based on observed convergence behaviour.

By comparing the RMSE values across the three datasets (training, testing, and mean CV score), the best values for each hyperparameter were identified, and the models' accuracy and generalisation capability was evaluated. These are discussed in the next section.

Final MLP hyperparameters and architecture

The MLP architecture and hyperparameters, which were implemented through standard *sklearn* and *keras* Python libraries and illustrated in Figure 4.8, were chosen via hyperparameter tuning through the grid search optimisation technique. The final version of the MLP is the machine learning model in the MaLBEC. The optimised architecture and parameters were:

- An input layer with 4 neurons
- Network architecture, two hidden layers with sizes (75, 40)
- An output layer with 11 neurons
- Activation function ReLU (default)
- Alpha of 0.12
- Initial learning rate of 0.01
- Number of iterations with no change set to 20
- Maximum iterations of 10,000 (with early stopping enabled),
- Solver of Adam (default in *keras*).

4.3 Results

4.3.1 Final MLP training, testing, and evaluation results

The RMSE was calculated for the training data, the test data, and for the mean CV score data, for step six of the machine learning model selection process from Section 2.3.2 and Figure 2.10. The results are shown in Table 4.3. To compare differences between the mean CV score and the training and testing data, the percentage difference, Equation 2.15, was used with the reference values as the training and test data. Table 4.3 shows that overall the RMSE values are very small;

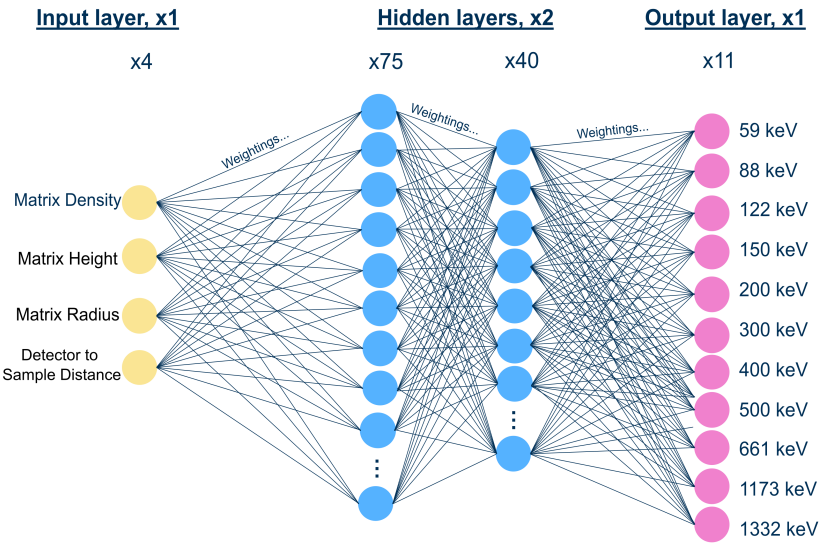


Figure 4.8: The architecture of the MLP used in the MaLBEC, showing the input layer with four neurons, two fully connected hidden layers with 75 and 40 neurons respectively, and an output layer with 11 neurons [62]

the efficiency values are approximately 1×10^{-4} , so an RMSE of less than 1×10^{-5} is desirable as this is within 10% of the true value and indicates strong predictive accuracy in absolute terms. This low training RMSE suggests that the model fit the training data well, but still performs very well on unseen data, as shown by the low test RMSE. The training and testing RMSE are within the same order of magnitude, which shows that the model is not overfitting and is able to generalise to new data. The increase in RMSE between training and mean CV is expected, as the training RMSE is often optimistic as the model sees those samples during training and can adjust parameters to minimise the error on the dataset. In contrast, the CV RMSE reflects out-of-sample performance, as each fold contains a held-out subset of the data on which the model must predict previously unseen points. This increase suggests that the model is not memorising the training data, but instead learned patterns and performs consistently across the 10 different folds. The CV RMSE is approximately 30% lower than the test RMSE (using the test value as the baseline), indicating a modest increase in error on the testing set. While this difference may reflect some sensitivity to data distribution, the test RMSE remains within the same order of magnitude as training and CV and well below the target efficiency scale. Therefore, the model demonstrates acceptable generalisation and accuracy for practical use in this application.

Table 4.3: The RMSE results for the training, testing and mean CV datasets, with a percentage difference comparison

Training RMSE	Mean CV RMSE	Testing RMSE	% Diff CV to Training	% Diff CV to Testing
1.17×10^{-5}	1.61×10^{-5}	2.30×10^{-5}	38	30

The model converged (no further improvements to performance) at 159 iterations, despite a maximum of 10,000 iterations being available, indicating efficient learning and rapid stabilisation. The selected 'strong' alpha (0.12) is considerably higher than typical defaults (e.g. 0.0001), indicating substantial loss regularisation was applied to shrink weights and reduce model complexity. This is an important factor in mitigating overfitting given the moderately large network architecture.

A final indication that the model performs well is the comparison of the actual and predicted efficiency, on the training and test data. This comparison is shown in Figure 4.9, in which both plots show a strong correlation around $y = x$, consistent with the low RMSE values reported in Table 4.3. The few outliers on the test data correspond to geometries not fully represented in the training set due to limitations in the random generation of sample geometries (discussed in Section 4.4.1). This highlights the importance of representative sampling but does not significantly impact overall model performance. Further investigation into the outliers, for example by analysing whether they correspond to particular parameter combinations such as extreme density values or large source-to-detector distances, would provide further insight into the specific limitations of the model and is identified as a direction for future work. Despite these isolated cases, overall alignment indicates that the model generalises effectively to unseen data and is suitable for practical use in this application.

Final MLP uncertainty estimation

Machine learning models, such as the MLP within the MaLBEC, do not provide intrinsic uncertainty estimates by default. To approximate the uncertainty of the MaLBEC based efficiencies, the test set RMSE was used as an estimate of the model's absolute uncertainty, corresponding to one standard deviation ($\pm 1\sigma$). This RMSE value was then applied in Equation 4.4 to calculate the relative uncertainty for each efficiency prediction. This approach assumes that the test set adequately represents the operational range of sample geometries, meaning it is assumed that the MLP has captured the signal (the true, structured relationship between input

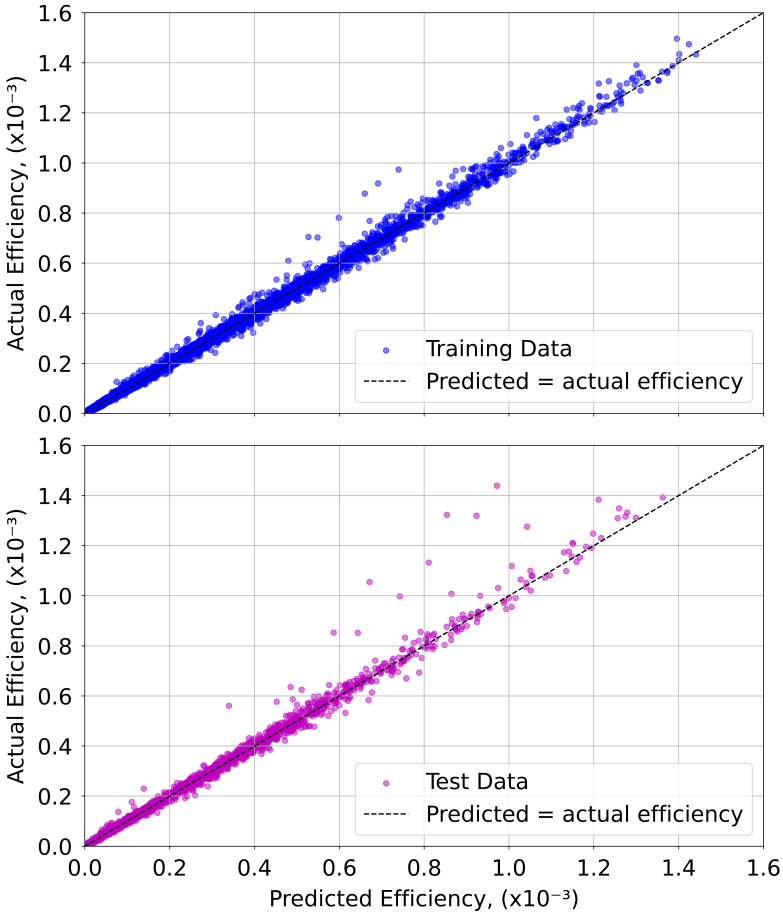


Figure 4.9: A plot showing the actual vs predicted efficiency values for the training data (top/blue) and test data (bottom/pink), for all energies and samples) [62]

features and efficiency) very well. Under this assumption, the only remaining source of uncertainty is random noise in the data. Random noise cannot be predicted by any model because it is non-correlated and purely stochastic. Therefore, the RMSE calculated on the test set mostly reflects this random noise, and in this ideal case the RMSE is approximately equal to the true uncertainty due to noise. However, in practice, the model may not perfectly learn the signal, especially if the training data is limited or the model architecture is imperfect. This means the RMSE includes both noise error and residual signal uncertainty. As a result, the RMSE is larger than just the noise component, making it a conservative estimate that likely overstates the uncertainty compared to the ideal case. While this conservatism is generally desirable for validation purposes, it should be noted that this method provides only an approximate uncertainty for the MaLBEC based efficiencies and should be interpreted with caution, as MLP regressors do not inherently provide uncertainty estimates.

Once the MLP was optimised and shown to perform well on the training and testing datasets, the next steps were to fully evaluate the final MLP. Four metrics were used to compare the MaLBEC to a traditional method, in this case MCNP. The metrics used and the results are discussed in detail in the following sections. Where the percentage difference equation is used (Equation 2.15), the reference term b is the value from MCNP.

4.3.2 Metric 1: Comparison of the efficiency values

For the first metric, a direct comparison of the efficiency values output from the two methods was compared (the MaLBEC and MCNP) for the experimental steel sample. The geometry of the experimental sample as modelled in the MCNP plotter can be seen in Figure 4.10. For the MaLBEC, the sample dimensions were input into the python script and all 11 efficiency values were produced as a .CSV and as a plotted efficiency curve.

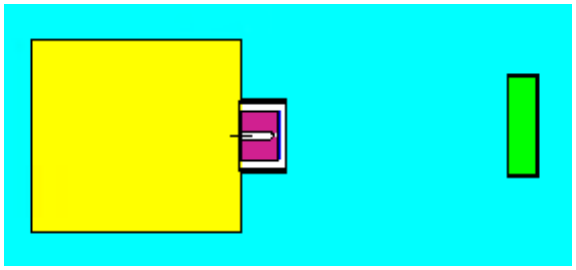


Figure 4.10: A side-on view from the MCNP plotter, containing the Trans-SPEC HPGe detector (total length of 41.5 cm) from Figure 4.4 and the experimental sample modelled as a cylinder standing on its base, in green

Metric 1 results

The efficiency curves for both methods, along with the residuals, are plotted in Figure 4.11 for the steel sample. This shows the efficiency values and uncertainties for each of the 12 energies - 11 efficiencies directly calculated from the relevant method, plus an additional efficiency at 834 keV which was obtained using cubic spline interpolation between the 661 keV and 1173 keV γ rays. Cubic interpolation was selected to better reflect the non-linear efficiency trend. As the 834 keV efficiency value was not included for direct calculation in the training data for the MaLBEC, it demonstrates that intermediate efficiencies can be accurately estimated without training the model on every energy point. For the MCNP results, the efficiency at 834 keV was directly calculated in MCNP. The percentage difference between the

two methods was, on average, approximately 6%, and all values are within 1 σ. The closeness of the curves in the efficiency plot, and the rapid decrease in percentage difference in the residual plot further illustrates how well the MaLBEC performs compared to MCNP at the energies above 100 keV, where the efficiencies all have percentage difference of around 10% or significantly less. The curve highlights that the lower energy efficiencies are less closely matched, with a percentage difference of approximately 20% (but are still within 1 σ, compared to the higher energies which are much closer together). Although the MaLBEC does not perform explicit physics based modelling of air attenuation, it is trained on MCNP simulation data which inherently accounts for air attenuation through detailed particle transport modelling. The MaLBEC therefore implicitly learns these attenuation effects from the training data. The positive bias observed at low energies may reflect the limitations of the training data coverage at extreme parameter combinations, or the greater sensitivity of low-energy efficiency predictions to small geometric variations that are harder to fully represent in the training set. At intermediate energies (300-500 keV) both models predict similar detector response, resulting in near-zero differences. The impact of this is discussed in Section 4.4.1 and 4.4.1.

The absolute uncertainties reported for MCNP and the MaLBEC correspond to relative uncertainties of approximately 4-8% for MCNP and 6-12% for the MaLBEC across the energy range. Overall, the MaLBEC predictions are consistent with MCNP within the stated uncertainty bounds, and the observed energy dependence reflects expected physical behaviour. It is important to note that the MaLBEC was trained on MCNP outputs and therefore any systematic bias or modelling limitation present in MCNP will propagate into the MaLBEC predictions. This dependency reinforces that the observed differences largely reflect the underlying physics and the reference model assumptions rather than errors introduced by the machine learning approach.

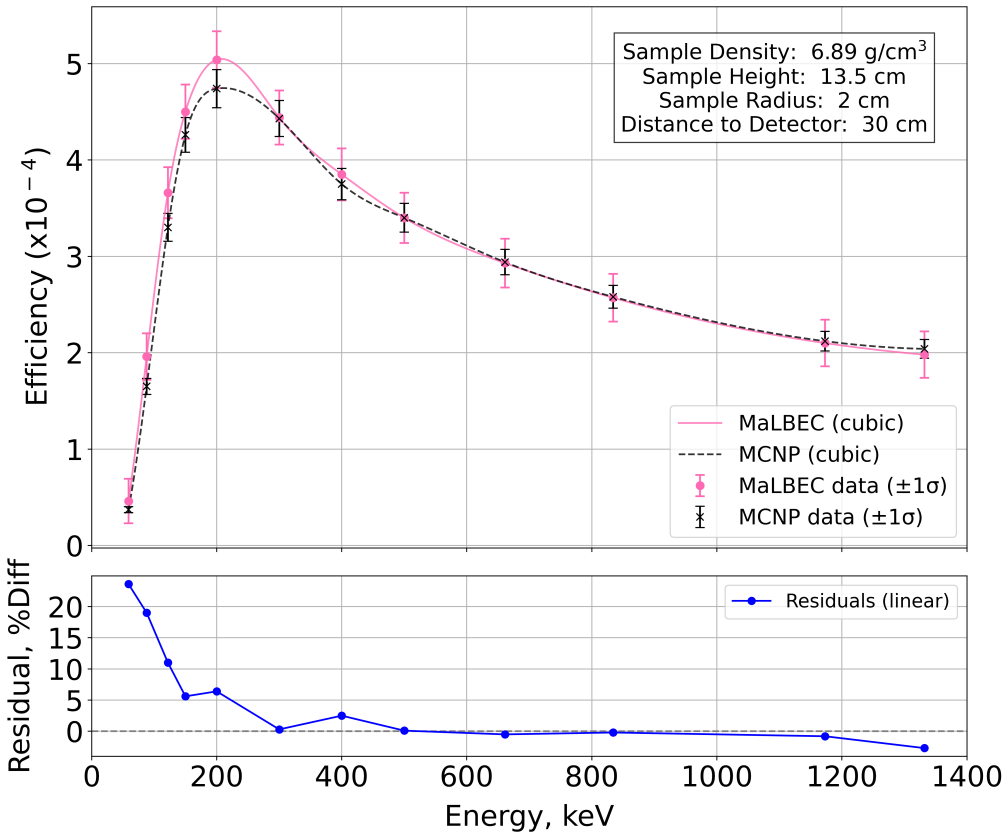


Figure 4.11: Metric 1 results (top): A comparison of the efficiency curves produced by MCNP (black) and the MaLBEC (pink), for the steel sample γ-ray spectrometry measurement. The error bars represent $\pm 1\sigma$ uncertainties and the smooth curves are cubic spline fits. The bottom panel shows residuals as percent difference [62]

4.3.3 Metric 2: Comparison of the activities

The second metric is a percentage difference comparison of the activities of the relevant radionuclides from the steel sample as described in Section 4.2.1. The calculated efficiencies from MCNP and the MaLBEC were used within the activity Equation 2.9, whilst this is similar to metric one, it is intended to demonstrate how an end user would apply the MaLBEC in practice. The absolute uncertainties provided were calculated using Equation 2.13 and correspond to one standard deviation (1σ). The spectrum, from Figure 4.3, contained photopeaks expected from activated steel, comprising NORM radionuclides and the activation products ^{54}Mn and ^{60}Co (as discussed in Section 2.2.7).

Metric 2 results

The results are shown in Table 4.4. The percentage difference is less than 3% for the three energies that were investigated, and all activities are in agreement within their

uncertainties. This level of agreement demonstrates excellent consistency between the MaLBEC and MCNP at these energies and supports the validation of the MaLBEC as a reliable tool for activity determination in fusion samples. These results are expected because the percent difference reflects the efficiency variation, reporting this metric illustrates its impact on the calculated activity and provides valuable context for fusion sample measurements.

Table 4.4: Metric 2: Percentage difference for the steel sample using the efficiency from MCNP and the MaLBEC, including uncertainties

Energy, keV	Nuclide	MCNP activity, Bq	σ_{MCNP}, Bq	MaLBEC activity, Bq	σ_{MaLBEC}, Bq	Activity % Diff
834	^{54}Mn	9.99×10^2	1.09×10^2	1.00×10^3	1.38×10^2	0.2
1173	^{60}Co	2.14×10^3	2.09×10^2	2.15×10^3	3.09×10^2	0.8
1332	^{60}Co	2.25×10^3	2.17×10^2	2.31×10^3	3.41×10^2	2.7

4.3.4 Metric 3: Computational speed comparison

The third metric is a comparison of the computational speed to calculate the efficiency values for the steel sample, from the point of submission. This excludes the physical model build time (the human effort) for either method as this is subjective, however, an approximate estimate of the human effort has been included to enable broad qualitative comparisons. For MCNP, the computational time was taken for each of the 11 γ-ray energies per geometry and summed (as these were 11 separate MCNP files). The run time for MCNP is provided in the output files, in minutes, and is dependent on the computational power of the machine used (the machine for this work is described in Section 4.2.1). The run time for MCNP also depends on the number of source particles used, this was set at number of particle histories (nps) 5×10^6 following the validation. For the MaLBEC, this was calculated as the time taken from running the code to the results being produced for all 11 efficiencies.

Metric 3 results

The results are listed in Table 4.5. For the MaLBEC, the speed to get the efficiency for all 11 energies, on a standard computer, was near instant and difficult to measure, so for the comparison it was given the total value of one second (divided by 11 to provide a value for each energy in Table 4.5). The MCNP results were generated within around 40 minutes, using the UKAEA high-performance computing cluster.

The MaLBEC represents a reduction in computation time of over 99.96% (using Equation 2.15). Although the nps in the MCNP input files influences runtime, the reduction would need to be substantial to challenge the MaLBEC's advantage. Even a total runtime of six minutes (as an example) for MCNP would still represent a reduction of over 99%, and lowering nps would increase MCNP uncertainties. These results demonstrate a significant improvement in computational time for the MaLBEC (not accounting for training time). To account for the human application time, an estimate of 15 to 30 minutes is required for an MCNP expert user to create a model. In contrast, the estimated time for an end user to generate an equivalent model using the MaLBEC is approximately 5 minutes.

Table 4.5: Metric 3: Comparison of runtime measurements for MCNP and the MaLBEC for the 11 different energies for the steel sample

Energy, keV	MCNP Runtime, s	MaLBEC runtime, s
59	40.2	0.09
88	65.4	0.09
122	102.6	0.09
150	136.8	0.09
200	184.8	0.09
300	244.2	0.09
400	274.2	0.09
500	288.0	0.09
661	302.4	0.09
1173	334.2	0.09
1332	333.6	0.09
Total	2306.4	1

4.3.5 Metric 4: Method usability/simplicity comparison

The final metric was designed to compare the usability and simplicity of each method. This was done by comparing the number of code or file edits required to produce efficiency values for all 11 γ-ray energies of a new cylindrical sample (the steel sample). For MCNP, this metric depends on how the MCNP file is created and so in general is difficult to measure, but for this work it has been standardised by utilising a template MCNP file for cylindrically shaped samples to enable a robust comparison.

Metric 4 results

For the MCNP files, 23 edits were required to set up the sample dimensions and position, and a further three edits for each of the 11 γ-ray energies. Therefore, to obtain a full efficiency curve for a new sample, 56 edits were required. In contrast, only four edits were required by the end user for the MaLBEC, once initial training and validation had been performed. This represents a significant difference and improvement of 93% for the MaLBEC, drastically improving usability.

4.4 Summary

This chapter introduced a novel algorithm, the Machine Learning-Based Efficiency Calculator (MaLBEC), which utilises state-of-the-art machine learning to calculate the absolute detection efficiencies at 11 fusion relevant γ-ray energies for a cylindrical steel sample. The MaLBEC positively impacts the γ-ray spectrometry process from Section 2.2.2 and Figure 2.3, and was shown to excel in performance when compared against MCNP using four metrics: efficiencies were within an average of 6% of MCNP and are all within 1 σ, activity calculations within 3%, computation time reduced by over 99.96%, and with only four user inputs (sample density, height, radius, and detector-to-sample distance) required compared to 56 code edits for MCNP. When both computational and human application time are considered, the MaLBEC provides efficiency values in approximately an order of magnitude faster than the MCNP workflow, highlighting a significant overall improvement for the end user. Combining the performance of all four metrics, the results show that the MaLBEC offers an excellent balance of accuracy, extreme computational efficiency, and greatly improved usability. These results confirm the MaLBEC as a robust alternative to traditional methods, with its MLP regressor showing good generalisation to unseen data and experimental validation without overfitting. This lowers the barrier for end users of γ-ray spectrometry to perform efficiency calculations independently of costly, complex, or inaccessible traditional methods, that are instead performed at an earlier training stage by experts.

By simplifying and accelerating efficiency calculations, the MaLBEC has the potential to significantly improve fusion diagnostics and enable rapid analysis of samples relevant to fusion calculations. As discussed in Section 4.1, a new method to quickly, simply, accurately, and more automatically calculate efficiency values has important applications in nuclear fusion research and waste/decommissioning measurements. The information gathered from the example JDR waste sample in this chapter provides vital insights for fusion research. For example, the γ-ray data can be used to infer the number of incident neutrons (neutron flux) during irradiation [69], providing insights into fusion power.

The research questions that were addressed in this chapter are listed below, and how they were addressed is in pink:

1. Can the absolute efficiency calculation process be improved and simplified (e.g. automated, sped-up, etc.), via state-of-the-art machine learning methods? The MaLBEC is a machine learning-based algorithm that produces a full efficiency curve for a new cylindrical sample within a one second runtime, from just four code changes. The results are comparable to a traditional method, and represent improvements in speed and simplicity of over 90%, making the efficiency calculation process accessible to non-experts.

2. How does the novel machine learning method compare to a traditional method, through comparing the efficiency values, the resulting activity calculations, the computational speed, and the usability (e.g. simplicity of model creation)? The MaLBEC provides efficiency and therefore activity results that are, on average, within approximately 6% of the traditional method MCNP, and it gets these results with over 90% improved speed and over 90% improvement in simplicity for the user.

3. In what ways can any techniques produced in this work be used for real world applications, and how can their effectiveness be demonstrated through experimental validation? The MaLBEC was used to calculate the activity results of a real sample that was in the JET fusion machine at the UKAEA for three DT fusion campaigns. The improvements to the efficiency calculation process could have significant impacts for analysis of future samples like this; during the JDR program at the UKAEA there will be thousands of samples and items like this that will require bespoke models to be created to attain the efficiency calculations. A method as fast and flexible (with automation potential) as the MaLBEC could dramatically simplify and speed up this process. This method also has the potential to significantly improve the modelling for research samples too.

4.4.1 Discussion

The first discussion point is that metric to evaluate the results (accuracy of the efficiency results) shows larger discrepancies (10-20% Diff) at lower energies, likely reflecting limitations in the modelling of detector response, which is particularly complex in the low-energy region. However, all values remain within 1 σ, suggesting that the apparent disagreement is largely an artefact of the percent-difference metric

itself. Percent difference becomes unstable at very small absolute efficiencies, where even tiny absolute deviations produce large relative percentages.

A second discussion point is based around the training data used in the training of the MLP within the MaLBEC. The data were simulated using MCNP, as thousands of training data were required and this would be extremely difficult to gather experimentally. However, this does create the issue that the MaLBEC will only be as accurate as MCNP (the MCNP model used to create the data was shown to be around 20% out from expected experimental results). For the proof of concept for this novel method, this has been deemed acceptable, as the intention of this work was to create a novel efficiency calculation method and use MCNP as a benchmark. Work on improving MCNP models (or using a combination of different methods) is treated separately to this thesis work, but is an important consideration for future developments.

Similar to this, a third discussion point for future development for the MaLBEC is reducing the dependency on MCNP for end users by making the algorithm more detector agnostic. This could involve introducing additional input parameters, such as detector crystal size and properties like the dead layer, so that the MaLBEC can adapt to different detector configurations without requiring new MCNP simulations for each case. While a validated Monte Carlo model (e.g. MCNP) would still be needed for the initial training phase, which should be carried out by an expert, subsequent applications for end users could be performed without further MCNP runs. This improvement would not only enhance usability for fusion applications but also make the MaLBEC more broadly applicable and useful in laboratory environments where multiple different detector systems are used. Implementing this feature was beyond the scope of this work, as a significant number of detectors would need to be modelled and validated in MCNP, but represents a promising direction for future development.

A fourth discussion point is that the lack of uncertainty estimation represents a limitation of the current work, as uncertainty estimation would provide additional confidence information, which would be valuable in a nuclear diagnostic context. Incorporating uncertainty estimation is therefore identified as a direction for future work.

A final discussion point is regarding the limited output of the MaLBEC. This is set at 11 energies to cover a broad range of relevant energies, meaning any other energies outside of these are required to be interpolated from the results, potentially reducing the accuracy and limiting the usability of the results. This was done to reduce the computation effort in generating training data (each new energy is required to be

ran for each of the around 1300 training geometries), however further work could limit the impact of this. One option for future work to improve this is to gather training data across a greater number of discrete γ-ray energies, rather than at the 11 considered here. Currently, the sampling approach does not prioritise regions where efficiency predictions are more sensitive to changes in geometry or energy. For example, high-energy γ rays measured at long distances may require less training data than low-energy γ rays measured close to the detector, where small variations in geometry can have a large impact on efficiency. Implementing a more targeted sampling strategy could improve model accuracy in the more sensitive regions. Whilst future work could increase the usability of the algorithm, the interpolation method is a commonly used method in the nuclear industry and so it was deemed suitable for this work.

Chapter 5

Conclusion

The overall question this thesis addresses is *'can machine learning be applied to enhance γ-ray spectrometry an7alysis, with a focus on nuclear fusion applications?'*. The short answer to this question is yes. This work has demonstrated that machine learning can be successfully integrated into γ-ray spectrometry workflows, offering improvements in accuracy, efficiency, and accessibility. Part of the success of integrating machine learning into the γ-ray spectrometry process is through the use of a robust machine learning selection process, as presented in Section 2.3.2, applying rigorous techniques and validation metrics to all stages. This thesis has demonstrated that integrating machine learning into other areas of γ-ray spectrometry is possible through the development of two novel algorithms, validated using fusion related sources or samples (samples activated within the JET fusion machine). To arrive at this conclusion, two areas of γ-ray spectrometry were identified as areas where enhancement would be beneficial, each having unique challenges to overcome throughout the process of developing the algorithms. These areas were reducing the Compton scattering effect on spectra and improving the absolute efficiency calculation process. This chapter summarises the journey taken in addressing the challenges and reflects on the implications of the work. Although the core question of this thesis has been answered, there is still promising developments to be undertaken, which will be discussed as future directions.

For reducing the Compton scattering effect on spectra, a novel machine learning-based algorithm was developed in Chapter 3: the Machine Learning-based Compton Suppression Algorithm (MLCSA). This algorithm uses a convolutional neural network (CNN), a form of artificial neural network (ANN), to categorise γ-ray pulses from Compton scatter events versus those corresponding to full photopeaks, thus enabling the removal of Compton scatter events. The algorithm was validated using fusion relevant γ-ray sources to demonstrate a proof-of-concept. The MLCSA achieved a

Compton continuum count reduction of around 97%, outperforming a traditional hardware suppression system by around 10% under comparable conditions. Although the Compton continuum was significantly reduced and low-energy photopeaks were mostly preserved, counts in the higher energy photopeaks decreased by roughly 70%, which is comparable to other suppression methods. When the MLCSA was applied to evaluate improvements in the minimum detectable activity (MDA) for a low-energy radionuclide, it achieved a 51% improvement, compared to 37% for an alternative digital suppression method and 12% for a traditional hardware suppression system. These results demonstrate that machine learning not only matches but can exceed the performance of established techniques while offering greater flexibility. However, the algorithm currently only works with three γ-ray energies, a low-energy γ ray and two higher energy γ rays, whereas fusion relevant samples typically have multiple γ-ray energies present. Therefore, future work should focus on expanding the γ-ray energies that the algorithm can process and evaluating its robustness under mixed-radionuclide environments.

For improving the absolute efficiency calculation process, a novel machine learning-based algorithm was developed in Chapter 4: the Machine Learning-based Efficiency Calculator (MaLBEC). This algorithm uses a multilayer perceptron (MLP), a form of artificial neural network (ANN), to predict efficiency values for γ-ray energies for a specific geometry. The MaLBEC was experimentally validated using a sample that had been located within the JET fusion machine at the UKAEA for 40 years and thus had been activated with fusion neutrons. The validation work showed that the predicted efficiency values were within an average of approximately 6% of those obtained using the Monte Carlo N-Particle (MCNP) method across the energy range of 59–1400 keV. Computational time was reduced by roughly 99% compared to MCNP simulations, and used only 4 lines of code compared to 56 in MCNP, highlighting the potential for simple and rapid analysis in operational environments. While the algorithm performed similarly to other methods, it is somewhat limited in its application as only a set number of energies are provided as an output. Expansion of the energy range would be valuable future work to enable broad laboratory-scale applications.

Prior work in this area predominantly uses standard algorithms or physics-based methods, these are often slow, complex, or difficult to access (licenses, export control, training requirements, etc.). In contrast, the two algorithms developed here are in the process of being made open access, Python-based, and thus more freely available to the scientific community. As fusion progresses toward power producing machines such as ITER and DEMO, γ-ray spectral analysis will be required on large scales

to quickly evaluate the presence of 14.1 MeV fusion neutrons in an active reactor environment. Equally, as current or past fusion machines (such as JET) enter decommissioning, there will be large volumes of radioactive waste requiring efficient processing. Therefore, the development of these algorithms has the potential to transform current and future fusion operations by enabling faster, more accessible, and more scalable analysis.

Beyond fusion, the implications of this work extend to other areas of the nuclear industry, including fission radioactive waste characterisation and medical physics, where γ-ray spectrometry is a cornerstone technique. Machine learning offers a pathway to modernise these workflows (including all aspects of Section 2.2.2), reducing reliance on resource intensive simulations and hardware solutions, thus paving the way for γ-ray spectrometry to be brought into the modern era. However, the adoption of machine learning in nuclear contexts is not without challenges. One critical consideration is the "black box" nature of many machine learning models. Regulatory frameworks demand interpretability and traceability, particularly when algorithms are used to inform safety critical decisions. Future work should therefore incorporate explainable artificial intelligence (XAI) techniques, such as uncertainty quantification, to satisfy these requirements and build confidence among stakeholders. Additionally, validation against international standards and benchmarking across diverse datasets will be essential for widespread acceptance.

In conclusion, this thesis provides a foundation for the integration of machine learning into γ-ray spectrometry processes, demonstrating clear benefits in accuracy, efficiency, and accessibility. By addressing two critical challenges, Compton suppression and efficiency calculation, this work contributes novel tools that can accelerate progress in nuclear fusion diagnostics and beyond. As fusion technology advances and the nuclear industry faces increasing demands for rapid, reliable analysis, machine learning offers a pathway to transformative change. The algorithms developed here represent an initial step toward that future, where data driven methods complement traditional physics based approaches to meet the evolving needs of science.

Appendix A

Experimental equipment specifications

A.1 The detector used in the MLCSA

The information and specifications for HPGe, BEGe, detector from the MLCSA Chapter 3 are:

- Detector model - BE3825
- Serial number - b13135
- Crystal cooling - liquid nitrogen (77 K)
- Crystal diameter - 71 mm
- Crystal length - 26 mm
- End cap to crystal - 5 mm
- Absorbing layer - carbon epoxy, 0.6 mm
- Operating bias - positive 5000 V
- Relative efficiency - $>$ 26%
- Measured energy resolution (FWHM) - 0.334 keV at 5.9 keV, 0.583 keV at 122 keV and 1.688 keV at 1332 keV
- Crystal type - p-type
- Crystal geometry - BEGe
- Preamplifier model - 2002C

A.2 Digitiser used in the MLCSA

The trigger settings for the Red Pitaya digitiser (from the MLCSA Chapter 3) were:

- Voltage range: (0, 20) V,
- Voltage trigger: (-0.35, -0.4) V,
- Trigger edge: positive,
- Length of pulse collected: 200 ns.
- Raw pulse polarity: positive

A.3 The detector used in the MaLBEC

The information and specifications for the HPGe, Trans-SPEC, detector from the MalBEC Chapter 4 are:

- Serial number - 8142
- Crystal cooling - integrated Stirling-cycle mechanical cooler
- Crystal diameter - 66.5 mm
- Crystal length - 67.7 mm
- End cap-to-crystal distance - 10 mm
- Absorbing layer - aluminium 1.5 mm
- Absorbing layer - inactive germanium μm
- Operating bias - positive 4500 V
- Relative efficiency - $> 50\%$
- Measured energy resolution (FWHM) - 1.45 keV at 60 keV, 1.57 keV at 122 keV and 2.37 keV at 1332 keV
- Crystal type - p-type
- Crystal geometry - coaxial
- Weight - 11.1 kg

References

[1] I. T. Chapman and A. W. Morris, "UKAEA capabilities to address the challenges on the path to delivering fusion power," in *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 377, Royal Society Publishing, 2019.

[2] F. G. Rimini, "40 years of JET operations: a unique contribution to fusion science," *Plasma Physics and Controlled Fusion*, vol. 67, no. 3, 2025.

[3] S. Geng, "An Overview of the ITER Project," in *Journal of Physics: Conference Series*, vol. 2386, Institute of Physics, 2022.

[4] G. Federici, W. Biel, M. R. Gilbert, R. Kemp, N. Taylor, and R. Wenninger, "European DEMO design strategy and consequences for materials," *Nuclear Fusion*, vol. 57, no. 9, 2017.

[5] UK Government, "The environmental permitting regulations 2016, chapter 23," 2016.

[6] C. L. Grove, C. R. Shand, L. W. Packer, N. Fonnesu, E. Łaszyńska, S. Loreti, J. W. Mietelski, M. Pillon, M. I. Savva, I. E. Stamatelatos, A. Turner, T. Vasilopoulou, R. Villari, and A. Zohar, "Initial Gamma Spectroscopy of ITER Material Irradiated in the JET D-T Neutron Environment," 2024.

[7] M. Zehtabvar, K. Taghandiki, N. Madani, D. Sardari, and B. Bashiri, "A Review on the Application of Machine Learning in Gamma Spectroscopy: Challenges and Opportunities," *Spectroscopy Journal*, vol. 2, no. 3, pp. 123–144, 2024.

[8] IAEA, "AI for accelerating nuclear applications," 2022.

[9] UK Government, "National AI Strategy 1 2 National AI Strategy Presented to Parliament by the Secretary of State for Digital, Culture, Media and Sport," 2021.

[10] F. F. Chen, *Introduction to Plasma Physics and Controlled Fusion*. Springer, 2016.

[11] J. Sheffield, "Additional heating in JET," 1976.

[12] D. Moiraf, A. Morace, and Y. Arikawa, "The use of Plasma Mirror for Relativistic Electron Generation Relevant to Fast Ignition in Inertial Confinement Fusion," 2020.

[13] K. S. Krane, *Introductory Nuclear Physics*. Wiley, 1988.

[14] J. J. Goodell, C. M. Egnatuk, S. W. Padgett, B. B. Bandong, K. E. Roberts, and A. C. Mignerey, "Determining the Activation Network of Stainless Steel in Different Neutron Energy Regimes and Decay Scenarios Using Foil Activation Experiments and FISPACT-II Calculations," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors, and Associated Equipment*, 2018.

[15] DMAMC, "Data Mining Analysis and Modeling Cell Compendium of Material Composition Data for Radiation Transport Modeling," tech. rep., Pacific Northwest National Laboratory, 2021.

[16] K. Lennon, A. Turner, C. Shand, R. Smith, A. Turner, T. Berry, S. Bradnam, and H. Chohan, "Simulated Activation and Characterisation of Fusion Wastes to Support JET Decommissioning and Repurposing-24402," 2024.

[17] A. Litnovsky, I. Duran, J. Coenen, Y. Gasparyan, M. Gilbert, E. Hollmann, C. Linsmeier, S. Nogami, C. Skinner, and S. Zinkle, "Fusion—Reactor Materials," in *Encyclopedia of Nuclear Energy*, pp. 594–619, Elsevier, 2021.

[18] S. Chae, J. Y. Lee, and Y. S. Kim, "Neutron diagnostics using nickel foil activation analysis in the KSTAR," *Nuclear Engineering and Technology*, vol. 53, no. 9, pp. 3012–3017, 2021.

[19] D. Chiesa, M. Nastasi, C. Cazzaniga, M. Rebai, L. Arcidiacono, E. Previtali, G. Gorini, and C. D. Frost, "Measurement of the neutron flux at spallation sources using multi-foil activation," *Nuclear Instruments and Methods in Physics Research*, 2018.

[20] M. D. Coventry and A. M. Krites, "Measurement of D-7Li Neutron Production in Neutron Generators Using the Threshold Activation Foil Technique," in *Physics Procedia*, vol. 90, pp. 85–91, Elsevier B.V., 2017.

[21] L. W. Packer, P. Batistoni, N. Bekris, S. C. Bradnam, M. Fabbri, Z. Ghani, M. R. Gilbert, R. Kierepko, E. Łaszyńska, D. Leichtle, I. Lengar, S. Loreti, J. W. Mietelski, C. R. Nobs, M. Pillon, M. I. Savva, I. E. Stamatelatos, T. Vasilopoulou, A. Wójcik-Gargula, and A. Zohar, "Technological exploitation of the JET neutron environment: Progress in ITER materials irradiation and nuclear analysis," *Nuclear Fusion*, vol. 61, no. 11, 2021.

[22] T. Stainer, M. R. Gilbert, L. W. Packer, S. Lilley, V. Gopakumar, and C. Wilson, "14 MeV neutron irradiation experiments - Gamma spectroscopy analysis and validation automation," in *International Conference on Physics of Reactors: Transition to a Scalable Nuclear Future, PHYSOR 2020*, vol. 2020-March, pp. 1786–1795, EDP Sciences - Web of Conferences, 2020.

[23] O. Wong, R. Smith, C. R. Nobs, and A. M. Bruce, "Optimising Foil Selection for Neutron Activation Systems," *Journal of Fusion Energy*, vol. 41, no. 1, 2022.

[24] C. W. Barnes, H.-S. Bosch, H. W. Hendel, A. G. A. Huibers, D. L. Jassby, R. W. Motley, E. B. Nieschmidt, T. Saito, J. D. Strachan, M. Bitter, R. V. Budny, K. W. Hill, D. K. Mansfield, D. C. Mccune, R. Nazikian, H. K. Park, A. T. Ramsey, S. D. Scott, G. Taylor, and M. C. Zarnstorff, "Triton burnup measurements and calculations on TFTR," tech. rep., 1988.

[25] C. W. Barnes, H.-S. Bosch, H. W. Hendel, J. Kallne, P. Batistoni, G. Gorini, G. B. Huxtable, M. Pillon, S. Podda, and M. Rapisarda, "Triton burnup measurements in JET using a neutron activation technique," tech. rep., 1988.

[26] R. Gilmore, G, *Practical Gamma-Ray Spectrometry*. Wiley, 2008.

[27] G. F. Knoll, *Radiation detection and measurement*. Wiley, 2000.

[28] K. A. Tree, *Enhanced Nuclear Waste Assay*. PhD thesis, University of Liverpool, 2019.

[29] W. Shockley, "Currents to conductors induced by a moving point charge," *Journal of applied physics*, vol. 9, pp. 635–636, 1938.

[30] Mirion Technologies, "Germanium Detectors," 2025.

[31] A. N. Turner, *Gamma spectrometry for forward and reverse problems in radioisotope identification*. PhD thesis, University of Birmingham, 2020.

[32] Mirion Technologies, "Nuclear measurement fundamental principle spectrum analysis," 2023.

[33] M. C. Lépy, A. Pearce, and O. Sima, "Corrigendum: Uncertainties in gamma-ray spectrometry," *Metrologia*, vol. 54, no. 6, pp. 883–883, 2017.

[34] JCGM, "Evaluation of measurement data -Guide to the expression of uncertainty in measurement," tech. rep., 2008.

[35] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras & Tensor Flow*. OReilly, 2 ed., 2019.

[36] M. Jordan, J. Kleinberg, and B. Schölkopf, *Pattern Recognition and Machine Learning*. Springer New York, NY, 2006.

[37] SciKit-Learn, "Scikit-learn Supervised Neural Networks," 2025.

[38] J. Ebner, "Cross-Validation in Machine Learning: Enhancing Model Performance with Confidence," 2023.

[39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[40] M. Kamuda, J. Stinnett, and C. J. Sullivan, "Automated Isotope Identification Algorithm Using Artificial Neural Networks," *IEEE Transactions on Nuclear Science*, vol. 64, no. 7, pp. 1858–1864, 2017.

[41] A. Žohar, I. Lengar, P. Batistoni, S. Conroy, A. Čufar, R. Kierepko, B. Kos, S. Loreti, J. W. Mietelski, C. R. Nobs, L. W. Packer, M. Pillon, V. Radulović, M. I. Savva, L. Snoj, I. E. Stamatelatos, Štancar, T. Vasilopoulou, and A. Wójcik-Gargula, "Long Term Neutron Activation in JET DD Operation," *EPJ Web of Conferences*, vol. 253, p. 03005, 2021.

[42] Mirion Technologies, "BEGe™ Broad Energy Germanium Detectors," 2021.

[43] F. Holloway, L. Harkness-Brennan, and V. Kurlin, *The Development of Novel Pulse Shape Analysis Algorithms for AGATA*. PhD thesis, University of Liverpool, 2022.

[44] M. Yoshino, T. Iida, K. Mizukoshi, T. Miyazaki, K. Kamada, K. J. Kim, and A. Yoshikawa, "Comparative pulse shape discrimination study for Ca(Br, I) scintillators using machine learning and conventional methods," *Nuclear Institute and Methods in Physics Research*, 2021.

[45] K. Lennon, C. Shand, and R. Smith, "Machine Learning Based Compton Suppression for Nuclear Fusion Plasma Diagnostics," *Journal of Fusion Energy*, vol. 43, no. 1, 2024.

[46] Red Pitaya, "Red Pitaya STEMlab 125-14," 2023.

[47] D. Powers, "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation," tech. rep., School of Informatics and Engineering, 2007.

[48] Y. Lecun, L. Eon Bottou, Y. Bengio, and P. Haaner, "Gradient-Based Learning Applied to Document Recognition," tech. rep., IEEE, 1998.

[49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," tech. rep., 2012.

[50] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, 2021.

[51] Currie. L. A, "Limits for Qualitative Detection and Quantitative Determination. Application to radiochemistry.," *Analytical Chemistry*, vol. 40.3, pp. 586–593, 1968.

[52] U.S. Nuclear Regulatory Commission, "Minimum Detectable Activity (MDA) calculations.," tech. rep., 2007.

[53] H. R. Baghani and B. Aminafshar, "An inter-comparison between calculated solid angle for different geometrical correlations of source-detector through analytical and Monte Carlo approaches," *Indian Journal of Physics*, vol. 99, no. 6, pp. 2207–2215, 2025.

[54] P. Batistoni, S. Popovichev, Z. Ghani, A. Cufar, L. Giacomelli, P. Hawkins, K. Keogh, S. Jednorog, E. Laszynska, S. Loreti, A. Peacock, M. Pillon, R. Price, A. Reed, D. Rigamonti, J. Stephens, J. Bielecki, S. Conroy, J. Dankowski, and V. Krasilnikov, "14 MeV calibration of JET neutron detectors - Phase 2: In-vessel calibration," *Nuclear Fusion*, vol. 58, no. 10, 2018.

[55] R. Villari, X. Litaudon, J. Mailloux, M. Dentan, N. Fonnesu, Z. Ghani, L. W. Packer, F. Rimini, R. Vila, R. Afanasenko, J. Alguacil, P. Batistoni, P. Beaumont, S. C. Bradnam, P. Carman, J. P. Catalan, M. Cecchetto, A. Colangeli, D. Croft, J. Cufe, M. De Pietri, M. Fabbri, J. Figueiredo, D. Flammini, C. L. Grove, A. Hjalmarsson, V. Ioannou-Sougleridis, L. Jones, R. Kierepko, M. Klosowski, A. Kolsek, B. Kos, E. Laszynska, E. Lerche, Y. L. Tonqueze, D. Leichtle, E. Leon-Gutierrez, I. Lengar, M. Loughlin, R. Lobel, S. Loreti, G. Mariano, S. Mianowski, J. Milnes, S. Moindjie, F. Moro, J. Mietelski, R. Naish, S. Noce, J. Peric, J. B. Pontier, S. Reynolds, V. Radulović, M. I. Savva, P. Sauvan, C. R. Shand, L. Snoj, I. E. Stamatelatos, Z. Štancar, T. Stokes, N. Terranova,

A. N. Turner, T. Vasilopoulou, and A. Wójcik-Gargula, "Overview of deuterium-tritium nuclear operations at JET," *Fusion Engineering and Design*, vol. 217, 2025.

[56] ORTEC (Ametek), "GammaVision ® 8.10 Gamma Spectroscopy Software "Compatible, Efficient, and Defendable Results for Gamma Spectroscopy Applications."," 2023.

[57] Los Alamos National Laboratory, "MCNP Code Version 6.3.1 Theory and User Manual," tech. rep., Los Alamos National Laboratory, 2024.

[58] CERN, "Introduction to Geant4," tech. rep., 2025.

[59] P. K. Romano, N. E. Horelik, B. R. Herman, A. G. Nelson, B. Forget, and K. Smith, "OpenMC: A State-of-the-Art Monte Carlo Code for Research and Development," p. 06016, EDP Sciences, 2014.

[60] Mirion Technologies, "Model S573 ISOCS™ Calibration Software Description," tech. rep., 2017.

[61] ORTEC (Ametek), "ISOTOPIC A Practical, Flexible, and Defendable Solution for Gamma Spectrometry Waste Assay Measurements," tech. rep., 2022.

[62] K. Lennon, C. Shand, G. Wilson, and R. Smith, "Machine Learning-Based Efficiency Calculator (MaLBEC) for Nuclear Fusion Diagnostics," *Journal of Fusion Energy*, vol. 45, no. 1, p. 6, 2026.

[63] J. Sacks, W. Welch, T. Mitchell, and H. Wynn, "Design and analysis of computer experiments," *Statistical science*, vol. 4, pp. 409–435, 1989.

[64] R. Chowdhury, J. Zamora, G. Bollen, and T. Ginter, "Surrogate modeling of Monte Carlo radiation transport with convolutional neural networks for shielding optimization," *Beam interactions with materials and atoms*, vol. 570, 2025.

[65] Nuclear Decommissioning Authority, "Solid Radioactive Waste Good Practice Characterisation Guide," tech. rep., Jacobs, 2022.

[66] International Organization for Standardization, "Guidance for gamma spectrometry measurement of radioactive waste. ISO 19017:201.," tech. rep., ISO, Geneva, 2015.

[67] M. D. Mckay, R. J. Beckman, and W. J. Conover, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," Tech. Rep. 1, American Society for Quality, 2000.

[68] J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 3 ed., 2021.

[69] R. Worrall, B. Colling, M. R. Gilbert, E. Litherland-Smith, C. R. Nobs, L. W. Packer, C. Wilson, and A. Zohar, "The development, testing and comparison of unfolding methods in SPECTRA-UF for neutron spectrometry," *Fusion Engineering and Design*, vol. 161, 2020.