

Real-time radionuclides detection using artificial intelligence

Filipe André Carvalho Mendes

Thesis to obtain the Master of Science Degree in

Engineering Physics

Supervisors: Doctor Bruno Miguel Soares Gonçalves
Doctor Alberto Manuel Martinho Vale

Examination Committee

Chairperson: Prof. Horácio João Matos Fernandes
Supervisor: Doctor Bruno Miguel Soares Gonçalves
Member of the Committee: Doctor Raul Fernandes Luís

December 2020

Acknowledgments

First of all, I would like to thank my parents, my brother and sister for their friendship, encouragement and caring over all these years. Without their unconditional support both financial but more importantly emotionally this project would not be possible. I would also like to thank my family, aunts, uncles, cousins but specially my grandparents and godparents for their caring and encouragement.

I would also like to acknowledge my dissertation supervisors Doctor Alberto Vale and Doctor Bruno Gonçalves for their insight, support and sharing of knowledge that has made this Thesis possible. I would like to thank Doctor Pedro Vaz who enabled me to have access to GADRAS software and Master Jorge Borbinha for his insights on Gamma Spectroscopy.

A word of appreciation for Instituto Superior Técnico and Instituto de Plasmas e Fusão Nuclear for providing the materials and infrastructures that made this whole project possible.

Special thanks to my colleague Miguel Barros for all the brainstorming about machine learning and his companionship throughout this whole process.

Last but not least, to all my friends and colleagues that helped me grow as a person and were always there for me during the good and bad times in the last 5 years. To all of you "meftlineblingers" and "nano-engenheiros" my sincere thank you for all the support.

Resumo

A espectroscopia de raios gama é normalmente o método utilizado para identificar regiões radioativas. A espectroscopia de raios gama clássica envolve inúmeras fases, longos tempos de análise e, na maioria das vezes, um perito na matéria até chegar a uma identificação. Dado este longo processo, desenvolver e melhorar os sistemas de identificação existentes tem sido um desafio para organizações de defesa e segurança tais como Departamentos de Segurança Interna, Equipas de Resposta, Alfândegas e Controlo de Fronteiras.

A abordagem proposta neste trabalho propõe a utilização de técnicas de Aprendizagem Automática para implementar um sistema de identificação de fácil utilização, sendo possível que qualquer pessoa, mesmo sem experiência no campo, consiga obter resultados. Esta proposta utiliza redes neurais artificiais de forma a produzir uma classificação para um determinado espectro, adquirido através de um sensor CZT. O sistema é treinado com dados simulados e testado posteriormente com espectros reais. É explorada a identificação de um único ou múltiplos isótopos para cada amostra, realçando os benefícios de uma implementação deste género e também possíveis melhoramentos.

É também sugerido um exemplo para uma possível aplicação utilizando um Raspberry Pi. A rede neuronal desenvolvida podia ser também implementada por qualquer outro sistema semelhante, tal como um telemóvel ligado a um sensor. A portabilidade e independência de um sistema como este permite a sua utilização no terreno por pessoas ou veículos não tripulados como drones.

Palavras Chave

Espectroscopia de Raios Gama; Identificação de Isótopos radioactivos; Aprendizagem automática; Redes Neurais Artificiais.

Abstract

Gamma-ray spectroscopy is the usual method to identify detected radioactive hot-spots. Classical Gamma spectroscopy involves many phases, longer analysis and usually an expert to reach an identification result. Thus, developing and upgrading the identification systems available has been a challenge for security and defence organisations such as Departments of Homeland Security, Emergency Response Teams, Customs and Border Control.

This work proposes an approach using machine learning techniques that is intended to be implemented as an easy to use identification system, meaning that it can be used by anyone without experience in the field. The proposed solution makes use of artificial neural networks to produce a classification to a given spectrum obtained with a Cadmium Zinc Telluride (CZT) sensor. The system is trained using simulated data and is then tested with real acquisition spectra. Single and multiple isotope identification on each sample is explored, highlighting the benefits of an implementation of this kind as well as possible improvements.

Additionally, an example of a portable application is suggested using a Raspberry Pi. It is noteworthy that the artificial neural networks developed could be implemented in other devices such as a mobile phone with a connection to a detector. This kind of standalone and portable system could be used on site by humans or even by unmanned vehicles such as drones.

Keywords

Gamma-ray Spectroscopy; Radioactive Isotope Identification; Machine Learning; Artificial Neural Networks.

Contents

1	Introduction	1
1.1	Motivation	2
1.2	Radioactivity	3
1.2.1	Types of Decay	3
1.2.2	Identification of Radionuclides	4
1.2.3	Interaction of Gamma Radiation with matter	5
1.2.4	Gamma ray detection systems	6
1.3	Objectives	7
1.4	Organization of the Document	8
2	Problem Statement	9
2.1	Gamma-ray Spectroscopy	10
2.1.1	Gamma-ray Spectrum Components	10
2.2	State of the Art Methodology	14
2.2.1	Peak search identification methods	15
2.2.2	Template Matching identification methods	17
2.2.3	Artificial Neural Networks	17
3	Proposed Solution	19
3.1	Artificial Intelligence	20
3.1.1	Machine Learning	20
3.1.2	Supervised Learning	21
3.1.3	Artificial Neural Networks	21
3.1.4	Activation Functions	23
3.1.5	Loss Functions	25
4	Solution Implementation	27
4.1	Training Data Generation	28
4.1.1	Simulator Description	28
4.1.2	Simulated Samples	30

4.1.3	Selected radionuclides and simulation parameters	33
4.2	Artificial Neural Network Structure Development	34
4.3	Artificial Neural Network Training	36
4.3.1	Training Features	36
4.4	Proof-of-Concept Hardware	37
4.4.1	Acquisition Module	37
4.4.2	Analysis Module	38
4.4.3	Support Module	39
5	Results	40
5.1	Performance Analysis Criteria	41
5.1.1	Evaluation Data	41
5.1.2	Metrics	42
5.2	Single Radionuclide Classification	44
5.2.1	Number of Hidden Layers	44
5.2.2	Regularisation Methods	47
5.2.3	Batch size variation	49
5.2.4	Total Epochs	50
5.2.5	Background Subtraction effect	51
5.2.6	Acquisition Time effect	52
5.2.7	False Positives Analysis	53
5.2.8	Final ANN Results	55
5.3	Multi Radionuclide Classification	59
5.3.1	ANN Results and Discussion	59
6	Conclusion	63
6.1	Final Remarks	64
6.1.1	Contributions	64
6.2	Future Work	65

List of Figures

2.1	Example of Cesium-137 spectrum [1].	11
2.2	Cesium-137 decay scheme.	11
2.3	Multi FEP Compton continuum. FEP and Compton continuum from E_1 is overlapping the Compton Continuum from E_2	12
2.4	Possible scenarios for pair production inside the detector [2].	13
2.5	Escape peaks for hypothetical FEP of 1600 KeV [3].	14
2.6	Derivative peak search [4].	15
3.1	Example of simple ANN with 5 input neurons, 1 hidden layer with 3 neurons and 1 output neuron.	22
3.2	Sigmoid function. [5]	23
3.3	ReLU function. [6]	25
4.1	GADRAS Detector Tab.	29
4.2	GADRAS Inject Tab.	30
4.3	Base model for the artificial neural network structure.	35
4.4	Detecting hardware used: Acquisition Module (CZT) - Yellow; Analysis Module (RasPi) - Red; Support Module (Power-Bank) - Green.	37
4.5	μ Spec500 detector. [7]	37
4.6	Screenshot from RaspberryPi during acquisition process using CZT sensor.	38
5.1	Image from samples used for SS2 data-set.	42
5.2	Background subtraction example using $SS1_{137Cs}$ sample ([0,450] Channels).	42
5.3	Precision and Recall visual demonstration [8].	43
5.4	Accuracy and Loss evolution for $SoMx1$ network.	45
5.5	Comparison between Real and Simulated spectrum from ^{152}Eu ([0,250] Channels). The number of counts is normalised to the maximum value.	46

5.6 Accuracy and Loss evolution for <i>SoMx2</i> for regularisation of $L1 : 0.1$ and $L2 : 0.1$. Final loss value of 3.36.	48
5.7 F1-score values for each batch size. Regularisation of $L1 : 0.001$ and $L2 : 0.001$	49
5.8 F1-score values for different total epochs for <i>SoMx1</i> . Regularisation of $L1 : 0.001$ and $L2 : 0.001$	50
5.9 Classification of each sample along acquisition time.	53
5.10 Comparison between Simulated spectra from ^{238}U and ^{228}Ac ([0,250] Channels). The number of counts is normalised to the maximum value.	54
5.11 <i>SS1_60Co</i> spectra with classification peaks.	55
5.12 <i>SS1_137Cs#1</i> spectra with classification peaks.	56
5.13 <i>SS1_137Cs#2</i> spectra with classification peaks.	56
5.14 <i>SS1_22Na</i> spectra with classification peaks.	57
5.15 <i>SS2_60Co#1</i> spectra with classification peaks.	58
5.16 <i>SS2_60Co#2</i> spectra with classification peaks.	58
5.17 <i>SS2_241Am</i> spectra with classification peaks.	59
5.18 Precision, Recall and F1-score values for each ANN configuration using single isotope evaluation data-set.	61
5.19 Comparison between Simulated Spectra from ^{204}Tl and ^{137}Cs . ([0,450] Channels). The number of counts is normalised to the maximum value in the later plot.	62

List of Tables

4.1	Values from fitting of the equations.	32
5.1	Data-sets used for evaluating the performance of the ANNs. * Expected radionuclides.	41
5.2	Starting point for the ANN structures under analysis.	44
5.3	Learning parameters for hidden layer variation.	44
5.4	Final number of epochs and loss value for each of the ANNs.	45
5.5	Results of the classification using each ANN for the evaluation samples regarding hidden layer number variation. A classification threshold of 15% was used to filter lower results.	46
5.6	Final number of epochs and loss value for each of the ANNs for regularisation of $L1 : 0.001$ and $L2 : 0.001$	47
5.7	Results for each ANN for regularisation of $L1 : 0.001$ and $L2 : 0.001$. A classification threshold of 15% was used to filter lower results.	48
5.8	Final number of epochs and loss value for each batch size. Regularisation of $L1 : 0.001$ and $L2 : 0.001$	49
5.9	Final number of epochs and loss value for each stopping criteria for <i>SoMx1</i> using a regularisation of $L1 : 0.001$ and $L2 : 0.001$	50
5.10	<i>SoMx1</i> classification results with and without background subtraction. Regularisation of $L1 : 0.001$ and $L2 : 0.001$	51
5.11	False positives results for <i>SoMx1</i> using a threshold of 50%	54
5.12	Learning parameters for multi isotope identification.	60
5.13	Final number of epochs and loss value for each of the ANNs tested.	60

Acronyms

CZT	Cadmium Zinc Telluride
FRIENDS	F leet of dR ones for radlological insp E ction commu N ication an D re S cue
ANN	Artificial Neural Network
IPFN	Instituto de Plasmas e Fusão Nuclear
IST	Instituto Superior Técnico
AI	Artificial Intelligence
FEP	Full-Energy Photopeak
MCA	Multi-channel Analyzer
FWHM	Full Width at Half Maximum
TP	True Positives
FP	False Positives
TN	True Negatives
FN	False Negatives

1

Introduction

Contents

1.1 Motivation	2
1.2 Radioactivity	3
1.3 Objectives	7
1.4 Organization of the Document	8

1.1 Motivation

Every one around the world is constantly exposed to radiation arising from their phones, microwaves, routers, communication antennas, high exposure to the sun. More often than not, this kind of exposure is not harmful to living beings, however not every type of radiation is equal. Radiation with higher energy is usually referred to as ionising radiation. These energetic beams represent a real threat to our well being since they are energetic enough to modify the DNA structure in human cells, causing damage to tissues and organs. For that matter, it is of high importance to detect and identify possible radioactive sources.

There has been a growing concern on this matter of radiation control especially due to the raise of awareness among people, fuelled by disasters like Chernobyl and Fukushima. Nuclear activated components can be found in a variety of different scenarios such as reactor outages, laboratory test facilities, storage areas for contaminated fusion and fission remains and even particular equipment used, for instance, in health care. Criminal and unauthorised acts related to nuclear and other radioactive material such as illegal dropout of activated substances in remote locations and terrorism are also to be taken into consideration. Naturally-Occurring Radioactive Materials (NORM) [9] is a term used to specify all naturally occurring radioactive materials which occur naturally or where human activities have increased the potential for exposure compared with an unaltered situation. These materials potentially include all radioactive elements found in the environment such as uranium, thorium and potassium. Any of their decay products also represent a serious threat. Moreover, with the current rates of using nuclear reactions, namely for electricity production, the possibility of nuclear accidents cannot be neglected, thus control and monitoring must be enhanced and improved. Not only detecting but also identifying the detected radioactive sources is important. Identifying an unknown source is usually relevant since this information might hint the cause for such radiation detection and also define how to mitigate its risks.

The project FRIENDS (**F**leet of **dR**ones for **rad**iological **inspE**ction **commuN**ication **anD** **reScue** [10]) headed by Instituto de Plasmas e Fusão Nuclear (IPFN) [11] and Instituto Superior Técnico (IST) aims to fulfil some of these previously mentioned necessities in a modernised way. FRIENDS preposition is to develop three main tasks in parallel: drone navigation, data collection and radiological analysis. The work presented in this thesis contributes to the development of the radiological analysis task of the project. Within the scope of this data collection and radiological analysis a first iteration of the project arised: Maria [12]. This is a project developed with the goal of providing the ability to easily create a radiological heatmap of our surroundings and to create a database to store this information, all of this, using portable devices such as a smartphone and a low-cost commercial GMC.

It was during a summer internship at IPFN that I was presented to MARIA. Here I was proposed to integrate a Geiger-Muller counter with a GPS module and a mobile analysis platform in order to obtain some geographical information on the activity of the areas scanned. Since smartphone implementation

was already on the go, I was instead proposed to implement this in a different platform: a Raspberry Pi. After concluding this integration, it was decided that additionally to the radioactive activity it would be useful to know the nature of the radiation of the hot spots. Making use of the extra versatility of the Raspberry Pi, I transitioned to the implementation of a Cadmium zinc telluride (CZT) which would provide enough information to identify the radionuclides after some analysis. Upon the conclusion of this implementation it was realized that there could be some easily available methods to identify the data collected and the idea of this thesis began to appear.

Artificial Intelligence (AI) applications have been increasing in the scientific and even commercial areas in the past few years. Following this huge dissemination of AI methods such as Artificial Neural Network (ANN) and Deep Learning, it was figured out that there could be some room to explore these recent developments in this area and establish a connection between AI and gamma-ray spectroscopy, contributing to the improvement of current Gamma-ray analysis applications.

1.2 Radioactivity

Atoms found in nature can be either stable or unstable. If the forces among the particles that form the nucleus of an atom are balanced it is considered stable. On the other hand, if these forces are unbalanced, it can cause the atom to become radioactive. This instability of an atom's nucleus may result from several causes such as excess of either neutrons or protons, or even excessive energy. In an attempt to reach stability the atom can eject nucleons (protons or neutrons), as well as other particles, or release energy in other forms [13].

This process of attempting to reach a more stable form for the atom is denominated radioactive decay. Depending on the type of process itself, this decay can be sorted out in different categories: Alpha, Beta and Gamma.

1.2.1 Types of Decay

1.2.1.A Alpha Decay

Alpha particles, also called alpha rays or alpha radiation, consist of two protons and two neutrons bound together into a particle identical to a helium-4 nucleus [14]. Consequently an Alpha decay consists in the transformation of the atom nucleus (parent) into a new nucleus (daughter) by the emission of an alpha particle. Thus, the daughter nuclei has a mass number 4 units less and an atomic number 2 units less than the parent atom. This type of event can be described by the following equation.



where X , Y and α represent the parent nuclei, daughter nuclei and alpha particle respectively.

1.2.1.B Beta Decay

Similarly, a beta decay consists in the emission of a beta particle (electron or positron) from the parent nucleus. For example, beta decay of a neutron transforms it into a proton by the emission of an electron accompanied by an anti-neutrino or, conversely, a proton is converted into a neutron by the emission of a positron with a neutrino in a so called positron emission [15]. Both beta particle and neutrino/anti-neutrino are created during the decay process, leading to a more stable ratio of protons to neutrons in the atom.



where $\nu, \bar{\nu}$ are neutrino and anti-neutrino, respectively; ${}^0_{-1} e$ is an electron; ${}^0_1 e$ is a positron.

1.2.1.C Gamma Decay

Finally, in a gamma decay the nucleus simply transitions from a higher energy state to a lower energy state releasing energy by means of the emission of electromagnetic radiation (gamma ray). Gamma rays are simply photons that have extremely high energies which are highly ionising. Following a α or β decay, the daughter nucleus may be left in one of these high energetic states, progressing to a more stable state via a supplementary gamma decay. Since a gamma-ray carries no charge nor does it have an associated mass, there is no change in the element as a result its emission.



where X^* , X and γ represent the excited nucleus, the stable nucleus and the gamma ray respectively.

1.2.2 Identification of Radionuclides

In an attempt to identify a specific radionuclide we must analyse its emissions deriving from the different types of decay previously mentioned. A classification based only on the type of radiation emitted by the source is clearly not possible, since the number of existing distinct radionuclides is far greater than the number of different types of emissions and also these emissions are usually a mixture of different ones. For example, a beta emission is often accompanied by a gamma emission. The energy levels of all existing radionuclides have different values and, consequently, their decay schemes have individual properties that makes it possible to use them as a "fingerprint" to identify them by the measurement of the energy of the emitted radiation. When compared with other types of radiation such as alpha and beta

particles, gamma rays have a much higher penetration power, enabling it to escape from target chambers and pass surrounding structures/barriers in order to reach the detectors. Additionally they are relatively easy to detect with good efficiency and resolution. These characteristics combined with the fact that almost every radioactive decay is accompanied by gamma emission, makes this kind of spectroscopy universal and one of the most widely used methods for the identification of radionuclides [16].

1.2.3 Interaction of Gamma Radiation with matter

The most straightforward way to analyse gamma ray radiation is to understand how it interacts with matter, in this case the detectors, and find ways in which we can quantify its energy. Differently from the other types of emission (beta and alpha), gamma radiation does not carry any charge meaning that its interaction with matter diverge from such charged particles. The majority of gamma ray interactions with matter can be described by 3 main processes: Photoelectric effect, Compton scattering and Pair production [16].

1.2.3.A Photoelectric effect

In the photo electric effect the incident photon gives all of its energy to a bound electron in an atom, leading to the ejection of the specific electron with kinetic energy (E_e) equal to the difference between the incident photon energy (E_γ) and the binding energy (E_b) of such electron.

$$E_e = E_\gamma - E_b \quad (1.4)$$

This process allows the detector to accurately measure the energy that the incident photon transferred to the electron, corresponding to a well defined energy peak in the spectrum. These peaks are denominated Full-Energy Photopeaks (FEP). In an ideal world this would be the desired interaction since this way the spectrum would be composed only by well defined peaks, making them easily identifiable.

Other events may occur after the photoelectric effect. The ejected electron creates an empty slot which is soon filled by electron rearrangement. As a result, there is an emission of the excessive energy in the form of either an X-ray or an Auger electron. This energy varies with the material of the detector itself.

1.2.3.B Compton Scattering

Compton scattering occurs when an incident photon transfers part of its energy to a free or loosely bound electron via collision process. The amount of energy transferred is dependant on the angle between the direction of the incident photon and the direction of the scattered photon. The scattered photon leaves

the detector and the detected energy is the kinetic energy of the electron. The energies of the scattered photon and electron are given by:

$$E_{\gamma'} = \frac{E_{\gamma}}{1 + E_0(1 - \cos \theta)}$$

$$E_e = E_{\gamma} - E_{\gamma'} = E_{\gamma} - \frac{E_{\gamma}}{1 + E_0(1 - \cos \theta)}$$
(1.5)

where E_{γ} is the incident gamma ray energy, $E_{\gamma'}$ is the scattered photon energy, E_e is the electron energy, E_0 is $\frac{E_{\gamma}}{m_e c^2}$ and θ represents the scattering angle for the scattered photon.

1.2.3.C Pair Production

Pair production consists in the creation of a positron-electron pair when the gamma ray is travelling through matter, usually in the vicinity of an atomic nucleus. To make this process possible, the incident gamma ray must have at least 1.022 MeV of energy which corresponds to the combined rest mass of those two particles. The positron is unstable causing it to lose its kinetic energy and find an available electron to annihilate. During this annihilation, two gamma photons with the energy of 511 KeV are emitted in opposite directions.

1.2.4 Gamma ray detection systems

The detectors normally used for this kind of measurement can be classified in two different categories: scintillator detectors or semiconductor detectors.

1.2.4.A Scintillator detectors

Scintillator detectors are transparent materials (crystals) that, when hit by gamma rays, produce light pulses with intensities that are proportional to the energy of the incident photons [2]. As described by the processes in (1.2.3), the interaction of gamma rays with the crystal produces electrons with high kinetic energy. These electrons move inside the detector and gradually lose their energy by transferring it to the nearby electrons, generating a number of lower energy photons. As a result, light is produced being its intensity (or number of photons) dependant on the kinetic energy of electrons and therefore on the energy of the incident gamma ray. A photocathode absorbs the light coming from the scintillator and emits electrons through photoelectric process into the photomultiplier tube, which converts these flashes into electrical pulses. These pulses are fed into a linear amplifier for further signal processing and then a Multi-channel Analyzer (MCA) which sorts the pulses into different channels with respect to the magnitude of the voltage.

These kind of detectors are relatively easy to use and have a high detection efficiency. Despite that, they lack resolution power, meaning that they can not clearly separate peaks that have similar energies.

1.2.4.B Semiconductor detectors

Semiconductor detectors are made from specially processed crystalline material. When entering the detector, the gamma ray transfers all or part of its energy to the electron, which gradually loses its kinetic energy due to interaction with surrounding electrons and produces a large number of electron-hole pairs [17]. The number of electron-hole pairs produced in this process is proportional to the amount of energy passed to the electron by the gamma photon. Negatively charged electrons and positively charged holes migrate to their respective electrodes and are collected. This results in a charge pulse which is then fed to the pre-amplifier to produce a voltage pulse, its height being proportional to the incident energy of the absorbed photon. Finally and similarly to the above detector mechanism, this pulse enters a MCA that sorts the pulses into channels obtaining a final discrete spectrum. The energy resolution of a semiconductor detector exceeds that of all other radiation detectors because the energy needed to produce a pair of charge carriers is very low, meaning that there is a clear distinction between similar energy peaks. This excellent energy resolution makes them suitable for more demanding analyses such as in laboratory facilities.

1.3 Objectives

The main goal of this thesis is to develop a machine learning solution capable of identifying the radionuclides present in a spectrum. The proposed solution uses an ANN that receives the spectrum data and proceeds to provide a classification for the source in real time. In order to achieve such results, some objectives have to be met.

First of all, it is important to acquire or generate enough relevant data to both train and evaluate the ANN. This is an extremely important step towards the main objective of the thesis. Since real acquisition data is scarce and not widely available, generating training data by means of a simulator is necessary. Correctly defining the simulation parameters is a major milestone.

After acquiring the simulated data, the next stage is to define the ANN that is going to be used for the classification problem. This process includes the learning phase of the ANN, making use of the simulated data, as well as fine tuning some of the network parameters.

The final objective consists in developing a performance analysis procedure that evaluates the classification capabilities of the final algorithm. Some evaluating metrics are chosen and a relevant data-set is created using real data acquisitions.

As an additional benefit, a portable and fully featured proof-of-concept is developed using a Raspberry Pi and a CZT detector. This setup highlights the flexibility of the developed solution, providing an idea for possible useful implementations.

1.4 Organization of the Document

In [Chapter 2](#), an introduction to Gamma-ray spectroscopy is provided along with the current state of the art methodology on the subject.

[Chapter 3](#) aims to briefly introduce the topic of artificial intelligence describing the basis of ANNs. Some important parameters of neural networks are also explained since they will be part of the proposed solution presented by this thesis.

In [Chapter 4](#), the whole process of implementation of the proposed solution is described. The chapter starts by describing the whole data generation procedure, followed by the ANN structure development and training mechanisms chosen. The hardware setup used for the acquisition/detection procedure is also detailed in this chapter.

[Chapter 5](#) contains all the results, providing some comments and explaining the ANN structure optimisation. The evaluation criteria is also described, as well as the evaluation data-set chosen for this matter.

[Chapter 6](#) provides a final conclusion, summarising all the accomplished work and proposing some future improvements.

2

Problem Statement

Contents

2.1 Gamma-ray Spectroscopy	10
2.2 State of the Art Methodology	14

2.1 Gamma-ray Spectroscopy

Gamma-ray spectroscopy is a non-destructive analytical technique that can be used to identify various properties of the radioactive isotopes present in a sample of a specific radioactive substance. The energy of incident gamma-rays produced by the sample are acquired and measured by a detector, being then compared to the known energy of gamma-rays produced by radioisotopes and determining the identity of the emitter. This technique has many applications, such as in material analysis, geological exploration or even computer tomography.

Gamma rays are found in the majority of radioactive sources and can be of various energies and intensities. This type of rays are produced by the decay of nuclei as they transition from a high energy state to a lower state and typically range from energies of a few keV up to ~ 8 MeV, corresponding to the typical energy levels in nuclei with reasonably long lifetimes [18]. In order to analyse the sources of radiation, the detected gamma emissions are measured and used to produce an energy spectrum. A detailed analysis of this spectrum is useful to determine the identity and quantity of gamma emitters present in a sample.

Apart from being non-destructive and providing good insight of the radioactive sources, gamma spectroscopy is interesting since, due to their penetrating nature, gamma rays are able to escape from their surrounding and reach the detectors making this a good choice for radioactive sources analysis.

A common way to identify a radionuclide is to search for peaks in the spectrum and correlate them with the energy transitions present in radioactive substances. Most of the useful information within a gamma-ray spectrum can be extracted manually by a skilled and experienced professional. However this is an extremely time consuming and inefficient process so other more automated analytical techniques can be used to perform such tasks of spectrum analysis. Conventional analytical peak location approaches include: using regions-of-interest, using channel differences, derivative peak searches, correlation methods and checking the acceptability of peaks [4]. Some of these methods can be quite computational demanding (derivative peak searches for example) and depend heavily on the tuning of a great number of parameters in order to be accurate.

2.1.1 Gamma-ray Spectrum Components

Gamma ray spectrum is basically an histogram that represents the number of occurrences detected by the spectrometer system for a certain energy. Each individual radionuclide has a different decay scheme and consequently a distinct spectrum that is used to identify them. Gamma rays entering the detector can undergo any of the possible interaction processes described in (1.2.3). It is noteworthy that gamma rays move at speed of light and the detector size is in the range of centimeters, thus independently of the type interaction processes that occur, the time-frame of the events make it impossible to distinguish

between separate processes. The system sends a single pulse in the detector output signal. With that being said, the measurement can only point out how much energy of the gamma photon was lost inside the detector. These processes that occur inside the detector are responsible for the several features that can be encountered in the spectrum. The ability to recognise these features in a certain spectrum is what enables the identification of the radionuclides in question.

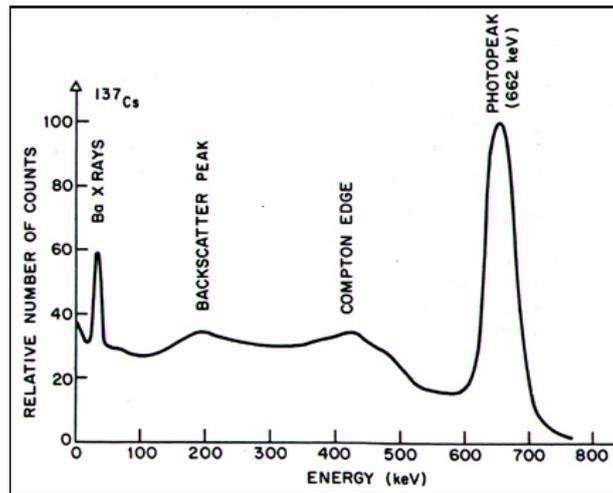


Figure 2.1: Example of Cesium-137 spectrum [1].

2.1.1.A Full Energy Photopeak

A FEP is visible in the spectrum when the incident gamma ray is completely absorbed by the detector. This peak corresponds to the energy of the emitted gamma photon from the radioactive source. It includes all cases when the entire energy is deposited in the sensor, independently of the number or type of interaction processes that the gamma ray went through [2]. In the case of Cesium-137 the FEP should be located at the 662 KeV mark (see Fig. 2.1), since it is the energy value of the gamma decay of this radionuclide (see Fig. 2.2). Thus, these are the peaks that we want to locate in the spectrum since they provide us with useful information on the gamma decay energies.

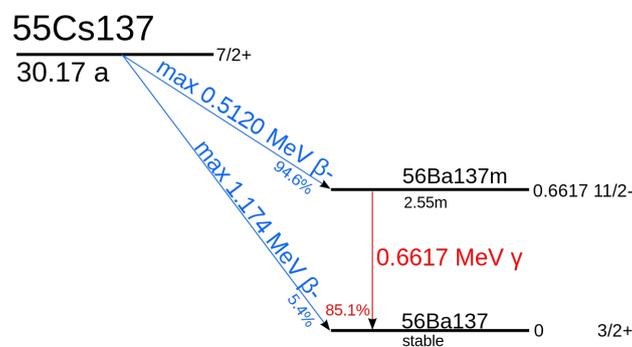


Figure 2.2: Cesium-137 decay scheme.

A particular case that contributes to this peak is the absorption of incident gamma photons by the photoelectric effect, meaning that its energy is fully absorbed by the electron inside the detector. Nonetheless, this could also be caused by a sequence of other interactions such as Compton scattering followed by full absorption via photoelectric effect. In the ideal case where the detector would be infinitely large, every single incoming gamma photon would deposit all its energy inside the detector, contributing to the FEP. This would mean that the spectrum would be composed only by FEPs and the radionuclide would be easily identifiable. In real world applications the FEP might be mixed with other peaks, making it difficult to recognise.

2.1.1.B Compton Edge and Compton Continuum

When Compton scattering occurs inside the detector, the scattered gamma photon usually leaves the detector. This implies that the energy absorbed by the detector corresponds only to the kinetic energy transferred to the electron. Consequently, the energy deposited in the detector varies with the scattering angle (θ) of the gamma ray. This energy is minimum when the angle equals to 0° meaning that the photon barely touches the electron and is maximum when the angle is equal to 180° , corresponding to the case when the photon collides frontally with the electron, being scattered back in the same direction.

The Compton Edge (see Fig. 2.1) designates the "peak" corresponding to this maximum energy ($\theta=180^\circ$) while Compton continuum is represented by the plateau of counts located on the left side of the Compton edge, corresponding to all of the other possible scattering angles ($\theta < 180^\circ$). Usually, every FEP has an associated Compton edge and Compton continuum which means that they can overlap each other creating a higher plateau (see Fig. 2.3).

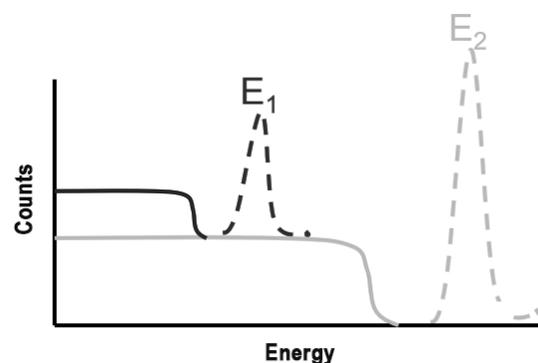


Figure 2.3: Multi FEP Compton continuum. FEP and Compton continuum from E_1 is overlapping the Compton Continuum from E_2 .

2.1.1.C Backscattering peak

A backscattering peak appears in the spectrum due to the interaction of gamma rays with materials outside the detector. When the gamma ray passes through or near the detector without any interaction, it can still reenter after being backscattered by the surrounding materials. The backscattered gamma photons are then absorbed inside the detector depositing an energy close to the energy difference between the FEP and the Compton edge (see Fig. 2.1). This occurs since the backscattered photons correspond to the ones that scatter at angles around 180 degrees, meaning they go back inside the detector. Consequently the backscattering peak appears at energies equal to backscattered photon energy, which overlaps with the Compton continuum.

2.1.1.D X-ray peak

Upon interacting with surrounding materials, the gamma photon can provoke ionisation, removing tightly bound inner shell electrons (the lowest energy states). The ejected electron creates an empty slot which is compensated by an electron from higher energy states. When this rearrangement occurs, the replacing electron emits a characteristic X-ray with energy corresponding to the difference between those two energy states.

2.1.1.E Annihilation peak

As previously mentioned in (1.2.3.C), when a gamma ray with energy greater than two electron masses (1022 KeV) passes near a nucleus a pair production phenomena may occur. If pair production happens outside of the detector only one of the annihilation photons may enter it and deposit its energy (both photons travel in opposite directions), leading to the appearance of a peak corresponding to the energy of 511 KeV (Annihilation peak).

2.1.1.F Escape peaks

Contrasting to the Annihilation peak, escape peaks emerge when pair production phenomena occurs inside the detector. When this happens only three possible outcomes are on sight:

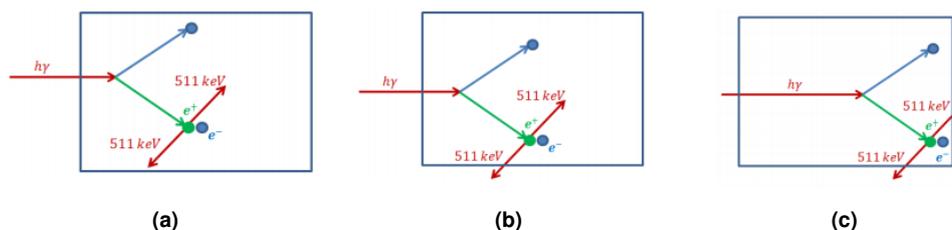


Figure 2.4: Possible scenarios for pair production inside the detector [2].

- (a) Both of 511 KeV gamma rays interact with the detector being absorbed and (together with the kinetic energy of the electron) contributing to the FEP. All of the energy is deposited in the detector.
- (b) If one of the annihilation photons escape, the total energy absorbed would decrease by 511 KeV meaning that there would be a peak corresponding to the energy of the FEP minus 511 keV. This peak is called single escape peak (see Fig. 2.5).
- (c) If both of the annihilation photons escape, the total energy absorbed would decrease by 1022 KeV meaning that there would be a peak corresponding to the energy of the FEP minus 1022 keV. This peak is called double escape peak (see Fig. 2.5).

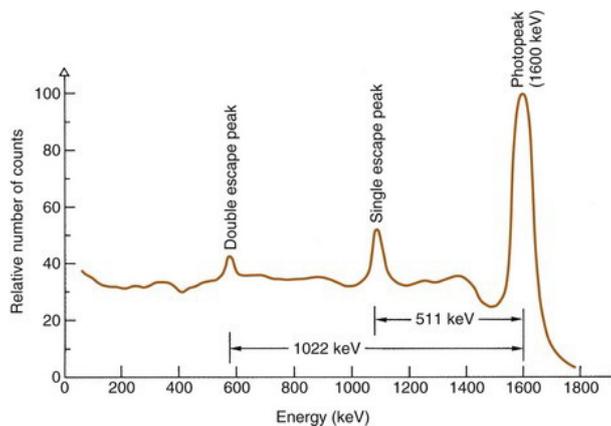


Figure 2.5: Escape peaks for hypothetical FEP of 1600 KeV [3].

2.1.1.G Sum peak

In the eventuality that more than one photon reaches the detector at approximately the same time, they can both interact with the detector. Such event gives rise to the designated sum peaks which correspond to the sum of all the energy deposited during those interactions. Sum peaks are usually present for highly active isotopes since there is a higher probability of emission of multiple gamma rays in a single decay.

2.2 State of the Art Methodology

After understanding the several features that may be present in a gamma spectrum it is important to understand the methodology used in order to apply that knowledge in a way that enables the user to identify the radionuclide in question. First of all, it is worth mentioning that identification procedures are not usually accomplished by a single "mechanism" but by a series of sequential methods. Approaches to radionuclide identification can be divided in two wide-ranging categories: peak search and template matching .

2.2.1 Peak search identification methods

Peak search identification methodology can be defined by a sequence of three procedures: peak search, feature extraction and classification.

2.2.1.A Peak Search

The method begins by trying to locate FEP peaks present in the spectrum and take note of their energies. Compton edges can be included in this search since they are associated with each individual FEP, acting as secondary confirmation.

The most straightforward and simple way to locate the peaks is to manually identify the peaks by "hand", however this is not an efficient nor agile process. The most commonly used algorithm to perform a peak search is called derivative search [4]. This procedure consists in applying the first and second derivatives to the entire spectrum and identify possible peaks by the signal variation of such derivatives. Taking a closer look at Fig. 2.6 it is possible to note that the first derivative sign changes as it crosses the peak centroid. On the other hand, the second derivative reaches a minimum at the centroid, being preceded and followed by a maximum. Please note that a gamma ray spectrum is not a continuous function but a histogram, meaning that its gradient must be approximated to the differences between the channels.

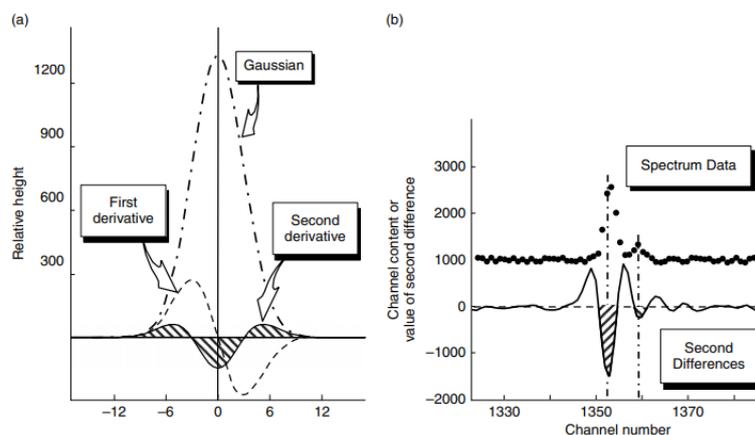


Figure 2.6: Derivative peak search [4].

In an ideal world where all spectra were a smooth and continuous curve, a derivative peak search could be immediately applied to the raw spectrum. Unfortunately, this is not the case since discrete Poisson fluctuations in the data usually lead to undesired noise that must be filtered/smoothed before performing such analysis. Several different smoothing techniques may be applied for gamma-ray spectroscopy [19]. From simple moving average to far more intricate and complex polynomial approaches

such as B-splines [20] or Savitzky-Golay filter [21]. Fourier transforms [22] have been utilised for this matter as well as some more recent procedures such as wavelet analysis [23] [24].

Independently of the way of implementation, all these previously mentioned methods require a fine tuning of the parameters in order to function optimally. Reaching a balance between filtering the existing noise and preserving relevant features in the spectrum is not always as straightforward as it might seem. This task gets even more challenging when dealing with low count data, where there is a smaller distinction between real features and background noise.

Another possible alternative to derivative peak search is to use "deconvolution" methods. Deconvolution is an algorithmic approach that tries to reverse the effects that a physical detection system has on the original theoretical source spectrum. A good analogy is to look at the obtained energy spectrum as being the convolution between the original source energy spectrum and a response function for a specific detector system. For example, this can be mathematically represented as the product of two matrices. Deconvolution algorithms aim to obtain the transformation that inverts the detector response function, which can then be applied to the observed spectrum, resulting in the original source spectrum. Having obtained a simplified spectrum, the localisation of the peaks should be far more simple. Some examples of deconvolution techniques include Maximum Likelihood Expectation Maximization (ML-EM), Maximum Entropy Method (MEM) and linear regularization [25]. The overall performance of these algorithms heavily depends on the complexity and accuracy of the detector response function provided by the user.

2.2.1.B Feature extraction

Despite not being strictly necessary, feature extraction may provide extra relevant information about the acquired spectrum. These features can provide useful input for the classification algorithm making it more robust. For example, measuring the counts under the peaks could help the classifier to understand which peak is more relevant and which peak should be discarded. This stage of the process depends on the type of classification algorithm that the user designed.

2.2.1.C Classification

The classification procedure consists in gathering all the data obtained in the previous steps and matching it against a known nuclide library in order to possibly identify the radionuclides present in the sample. Classifier algorithms that have been applied to radionuclide identification include [19]: Expert systems (essentially hard coded decision trees) [26], Naïve Bayes [27] [28] [29], Nearest neighbour [30] [31] and also Support vector machines [32] [33].

2.2.2 Template Matching identification methods

Template matching requires a separate type of algorithms that can search the space of possible radionuclides in order to find the correct mixture that contributes to a certain spectrum. These types of algorithms fall into two broad categories [19]: Heuristic and Systematic.

Heuristic algorithms involve a strategy based approach that compares the sampled spectrum with a huge number of possible spectra. One solution is to start from the full set of possible radionuclides and sequentially eliminate them on the way to a solution (strip down) [34] [35]. Another possible method is to begin with background only and start to add nuclides until a the final solution is similar to the given spectrum (build up). The removal or addition of candidate radionuclides must be succeeded by an optimisation step that evaluates the newly potential set of nuclides as a possible solution. This approach can be efficient at finding the optimal solution, provided that certain conditions are met, however the problem size must not be too large, otherwise the classification process could take too much time.

Differently from heuristic ones, systematic algorithms consider multiple possible solutions at each decision node, therefore decreasing the problem of path dependence in heuristic methods [19]. Unlike the heuristic approach, which can contain supplementary strategies, this is a more mathematically rigorous approach for solving the combinatorial optimisation.

In order to be able to apply this type of algorithms, an extremely well defined detector response information is required since relevant variations in the templates can harshly affect their performance. When using high energy resolution detectors, slight variations in the shape of the spectrum can be very noticeable, therefore this techniques are preferred for low to moderate energy resolution detection systems.

2.2.3 Artificial Neural Networks

The majority of the aforementioned algorithms require specific data pre-processing in order to be properly utilised. Some of them inclusively require special tuning of parameters, making those types of approaches unsuitable for being used by non expert users. This kind of problems can be overcome by making use of artificial intelligence algorithms such as ANNs. Although ANN implementation can be extremely consuming both time and resource wise due to training and the usual need for high performance computers, all of these difficulties are condensed in the development stage. After development, the algorithm requires a relatively low computational power to be used (this heavily depends on the purpose of the ANN) and more importantly, presents a far more user friendly experience since the user provides the input and receives an output without any additional parameters to tune. Additionally, the successful implementation of ANNs in other complex areas such as computer vision and image classification [36], further enhances the significance of exploring such approach.

ANN approaches have already been applied in several previous works using low to medium energy resolution detector systems (scintillator detectors) [37] [38] [39] [40]. Since the input for the ANN is the entire spectrum, there is no need for more complex pre-processing procedures, such as transformations or filters, making these applications easy to operate by the end user. To the best of my knowledge, no work was found on ANN development for radionuclide identification using high resolution detectors such as Cadmium zinc telluride (CZT) ones. Taking this into consideration, the goal of this thesis is to further explore the possibilities of using ANNs to analyse gamma-ray spectra from a CZT sensor and identify radioactive isotopes.

3

Proposed Solution

Contents

3.1 Artificial Intelligence	20
-----------------------------------	----

In this work an ANN approach is proposed as a possible solution to perform isotope identification on samples acquired with a CZT detector. Taking this into consideration it is important to understand some concepts of Artificial Intelligence and Machine learning before proceeding to the implementation of such algorithm.

3.1 Artificial Intelligence

Artificial intelligence (AI) is wide-ranging branch of computer science that emphasises the creation of intelligent machines that work and react similarly to humans. These human like processes include learning (acquiring information and making rules on how to use it), reasoning (use defined rules to reach conclusions) and self-correction (fine-tuning the existing rules to provide better results) [41]. Some relevant examples of common implementations of AI in our daily routines are: speech recognition, self-driving cars, search results, product recommendations, etc [42].

With the invention of powerful computational machines (the first computers) in the 20th century some scientists from a variety of fields began to discuss the possibility of creating an artificial brain and, with that, started the idea of AI [43]. However, they soon enough faced some critical setbacks, especially computational power limitation and lack of data, that led to the highs and lows of AI during its lifespan. On a more positive note, with the increasing in computational power and the loads of information available today, development of AI software is becoming more and more popular and accessible not only for big corporations but also for the small average researcher.

3.1.1 Machine Learning

Machine Learning is the scientific study of algorithms and statistical models that computer systems use to perform a specific task by solely relying on patterns and inference and without using any explicit instructions [44]. This is one of the most widely active subsets of AI [45]. Machine learning algorithms focus on defining a model based on sample data (training data), through a process known as "learning", in order to make predictions or decisions for a huge variety of input data, but without being explicitly programmed to perform such task.

There are mainly 3 types of learning processes: supervised learning, unsupervised learning and reinforcement learning [46]. Supervised learning describes a type of problem that involves using a model to learn a mapping between input examples (training data) and the available possible outputs. This type of learning is really useful to solve classification and regression problems. On the other side of the spectrum, unsupervised learning describes the class of problems that involve using a model to describe or extract relationships in the data, without previous identification or output examples. Finally we have reinforcement learning that consists in learning how an environment works from the feedback

that it provides to the learning algorithm actions. A good example is when a machine learning program is trying to learn how to play a game and tries to understand how it works by performing random actions and recording its feedback.

3.1.2 Supervised Learning

There are several supervised learning examples such as Neural Network, Decision Trees, Random Forest, Linear Regression, etc. The majority of these have proven to be really useful in pattern recognition and solving problems that are really difficult to define in a normal program. Supervised Learning problems can be divided in 2 major types: Classification problems and Regressions problems [47]. A Classification problem is defined by having a labelled output (discrete value) contrary to the regression one where the output is a continuous value. Identifying the colour of a shirt is Classification problem (the output is a colour), while defining a price for a house based on several characteristics is a regression problem (the output is a value that can vary widely).

In a normal computer program, the programmer defines every step that needs to be taken in all the calculation/processes in order to obtain a desired result. This means that every scenario that lays outside of the scope of the programmed instructions can not be evaluated properly and moreover, the complexity of the calculations performed throughout the running of the program depend heavily on the complexity of the input data. On the other hand, for supervised learning techniques the high demanding part of the process is laying on the training procedure of the model. After the training, the model can provide an output for any given input even for those cases that were never presented to the model before. Another huge advantage of this type of implementation is that the resources needed to reach a certain output are completely independent on the complexity of the input problem provided, often requiring little effort from the machine that is running it. Additionally the results of the training can be easily exported for other machines, making it easy to share, reproduce and retrain the models.

3.1.3 Artificial Neural Networks

Artificial Neural Networks (ANN) are one of the most widely used tools of Machine Learning today. As the “neural” part of their name suggests, the way they work is inspired in the human brain as they are intended to replicate the way that we (humans) learn. From the beginning of our lives, we are faced with distinct scenarios and are constantly adjusting our reaction to every each of them according to the feedback we receive from the environment. Even when presented with a completely new problem, we can usually figure it out based on previous events. Our brain learns things from experience, developing connections between our neurons and afterwards using these established connections to make decisions. This learning and decision making processes can be reproduced by ANNs.

A simple ANN is composed by an input layer with as many neurons as there are input signals, at least one hidden layer (this may vary for each network) and an output layer with as many neurons as output signals (see Fig. 3.1).

Each neuron-neuron interaction can be described by receiving the input signals, multiplying them by each respective weight and sending these values to an activation function that delivers a final output signal. These weights are the constants that we get from the training/learning procedures. There are numerous activation functions available that can be chosen according to the main goal of the ANN. These functions, also known as transfer functions, act as thresholds for the signal usually outputting values between 0 and 1 or -1 and 1.

After defining the structure of the ANN, the data-set needs to be prepared for training. When setting out to train a neural network, the data is usually divided into three sets: training set, validation set and test set [48]. The majority of the data is allocated to the training set (normally 70-80%) where it is used to define the value of the weights and take part in the learning procedure. The reminiscent data is divided between the validation set, being useful to fine-tune hyper-parameters of the training methods such as batch size and epoch number, and the test set which is used to evaluate the final results and the performance of the ANN. The test set is usually not included in the training in order to represent "new" and independent data, trying to examine the resulting ANN's capability to evaluate scenarios that were never presented to it before.

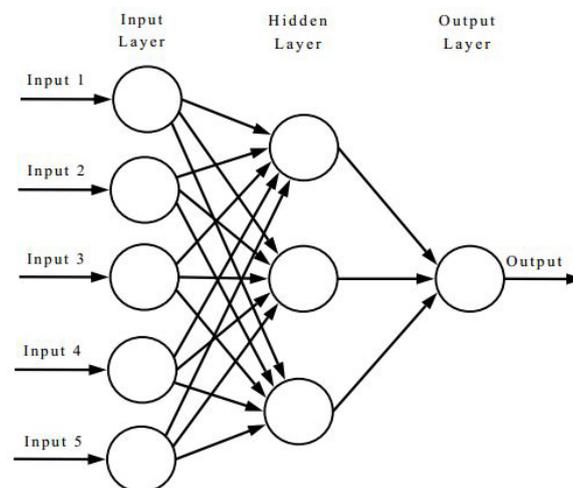


Figure 3.1: Example of simple ANN with 5 input neurons, 1 hidden layer with 3 neurons and 1 output neuron.

During the training process, the weights of the ANN are updated with the intention of minimising the error in relation to the expected output for each of the training data examples. This error can be evaluated and measured by what is called a loss function. Commonly used loss functions include Mean Squared Error (MSE), Mean Absolute Error (MAE) normally used for regression problems or even Hinge Loss or Cross Entropy that are usually applied to classification problems [49]. This process of minimisation of a

function with respect to a set of parameters (in this case the loss function in respect to the weights) is at the root of many computer science issues [50] and can be carried out by optimisation algorithms like Gradient-based learning. Popular algorithms include Stochastic Gradient Descent (SGD), Momentum based GD, Nesterov Accelerated Gradient Descent (NAG), RMSprop and Adaptive Moment Estimation (ADAM), every single one of them being some variant of the classical Gradient Descent Algorithm. ADAM, which is a combination of RMSprop and Momentum, is considered to be current state of the art [51]. This kind of algorithms can significantly impact the duration of the training, meaning that more optimised approaches can reduce drastically this period and even avoid possible hold backs such as finding local minima instead of global minima. The learning process can still be optimised by a fine choice of the learning rate and by the application of regularisation, something that can have a positive effect on the mitigation of over-fitting.

3.1.4 Activation Functions

An activation function, also known as transfer function, is basically an operation that is applied to values received from the input connections of the node and provides an output for the node itself. There are numerous different activation functions available to be used in ANN. The choice of the activation function is based on the purpose of the designed ANN. For this work, the transfer functions used were: Sigmoid, Softmax and ReLu.

3.1.4.A Sigmoid

The Sigmoid function, also known as Logistic function, is a S-like shaped function which can be used as the output for ANNs that aim to predict the probability of something.

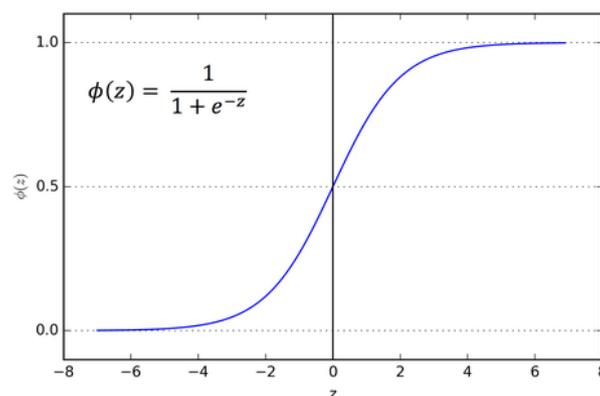


Figure 3.2: Sigmoid function. [5]

This is the case since the output of this function is a value between 0 and 1, something that relates

particularly well with probabilities (values between 0 and 1 as well). Sigmoid is differentiable, making it suitable for Gradient Descent Algorithms, however it can cause problems to the learning process due to lower slopes for values far from 0. This type of activation function is often used for binary classification algorithms.

3.1.4.B Softmax

The Softmax function is a generalization of the logistic function to multiple dimensions [52]. Similarly to the sigmoid, this activation function is often used in the output layer of an ANN to predict the probability distribution over the possible distinct classes. Softmax differs from Sigmoid in the fact that it normalises the probability function among the possible classes, meaning that the output values are related to each other. This is useful for multi-class classification where only one of the possible outcomes is correct and the predicted class is then the one with the higher value. The formula for the Softmax function is presented in the following equation.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (3.1)$$

where σ represents the softmax function output for a class i , \vec{z} the input vector, K the total number of classes and e^{z_j} the standard exponential function for j element of \vec{z} .

The Softmax function takes as input a vector \vec{z} of K real numbers (output from the neuron of each class), and normalises it into a probability distribution consisting of K probabilities proportional to the exponential of the input numbers. Before applying the Softmax function, the elements of \vec{z} can be any real number, however after applying Softmax, each component will be in the interval between 0 and 1, and the total sum of \vec{z} components will add up to 1, so that they can be interpreted as probabilities.

3.1.4.C ReLU

The rectified linear unit (ReLU) is one of the most used activation functions in the area of ANN [53]. this function can be represented as:

$$f(x) = \max(0, x) \quad (3.2)$$

where x is the input value of the function.

This equation means that the function returns the value 0 if it receives any negative input, but for any positive value x it returns the same value. When compared to Sigmoid, ReLU has the particular advantaged of solving the so called vanishing gradient problem. When the derivative of the activation function is very small, the learning algorithm might get "stuck" and have problems evolving in the direction of a better solution.

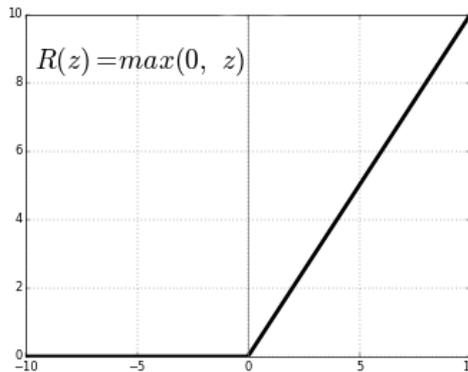


Figure 3.3: ReLU function. [6]

Since the derivative of ReLU is always 0 (for negative values) and 1 for positive values (see Fig. 3.3), and taking into consideration that for a reasonable sized batch the average derivative is rarely close to 0, the gradient descent is allowed to keep progressing [53].

3.1.5 Loss Functions

A Loss function is a method for evaluating the performance of a specific algorithm for a given data-set. This type of functions receive both the prediction and ground truth for each of the samples as input and produce a value that is inversely proportional to how well the specific algorithm performs. During the learning process, the optimisation algorithm tries to minimise the loss function since a lower loss value indicates a better model. The choice of a specific loss function depends directly on the type of problem that we are trying to solve, as well as the type of algorithm chosen to solve it. Generally, loss functions can be classified into two major categories: Classification losses and Regression losses. Since we are dealing with a classification problem, Cross-Entropy loss function was used.

3.1.5.A Cross-Entropy Loss

Cross-entropy loss, or log loss, is usually applied in models whose outputs can be described by a probability value between 0 and 1. Cross-entropy loss increases as the predicted probability diverges from the actual label.

The use of cross-entropy for classification is often given different specific names based on the type of application. For example, when facing a multi-class problem, where only one of the classes is correct, Categorical Cross entropy is used (see Eq. 3.3).

$$\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(\hat{y}_i) \quad (3.3)$$

where N is the total number of different observations, y_i is the target value and \hat{y}_i is the predicted probability of observation i .

On the other hand, when solving a binary classification problem we can use Binary Cross-Entropy (see Eq. 3.4). This can be applied even when we have more than two classes but they are independent of each other, each one representing an individual binary problem between being identified or not (1 and 0).

$$\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) = -\frac{1}{N} \sum_i^N [y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)] \quad (3.4)$$

where again N is the total number of different observations, y_i is the target value (0 or 1) and \hat{y}_i is the predicted probability of observation i .

4

Solution Implementation

Contents

4.1 Training Data Generation	28
4.2 Artificial Neural Network Structure Development	34
4.3 Artificial Neural Network Training	36
4.4 Proof-of-Concept Hardware	37

The desired implementation consists in a fully featured system that not only provides classification capabilities but also enables the user to acquire the spectrum of the radionuclides under analysis. Taking that into consideration, the solution implementation can then be divided in two parts: acquisition of the spectra and classification. The classification solution proposed in the previous chapter consists in training an ANN in order to receive the spectra information and provide an identification of the radionuclides present. On the other hand, the acquisition of the spectrum is made by a CZT detector that is connected to an analysis module (Raspberry Pi), where the classification procedure is then completed.

4.1 Training Data Generation

One of the challenges of this work is to find proper training data for the ANN. A really important factor on having a successful ANN is the quality of the training data that we feed to our system during the learning process. Ideally we would want to get a set of experimental data for every nuclide available, containing a wide variety of examples from distinct acquisition times, sources, detectors, backgrounds and even combinations of several nuclides. Creating the conditions to acquire the spectrum data required for a proper training of the ANN would be extremely difficult. The acquisition time for each spectrum ranges from several minutes up to several hours, creating some time constraints. Additionally, the restrictions and availability of some types of radionuclides prove such task unfeasible.

Since obtaining a data-set composed of real spectra was not possible, the selected approach was to resort to simulated data and use them as training samples. In order to obtain such material, a simulator was used: GADRAS-DRF from Nuclear Energy Agency [54].

4.1.1 Simulator Description

Gamma Detector Response and Analysis Software – Detector Response Function (GADRAS-DRF) is a software that contains a range of tools related to radiation detection. Its primary function is the simulation of gamma-ray and neutron detector signals to radiation sources. It also packs some limited analysis tools, although this section focus on the simulation capabilities, more specifically on the gamma-ray ones. According to bibliography [55], GADRAS-DRF uses a Spherical 1D Model [56] for the simulation, being the features in a gamma-ray detector spectrum such as photopeaks and the Compton continuum derived from first-principles calculations based on interaction cross sections. Validation studies can be found in such works as [57] and [58].

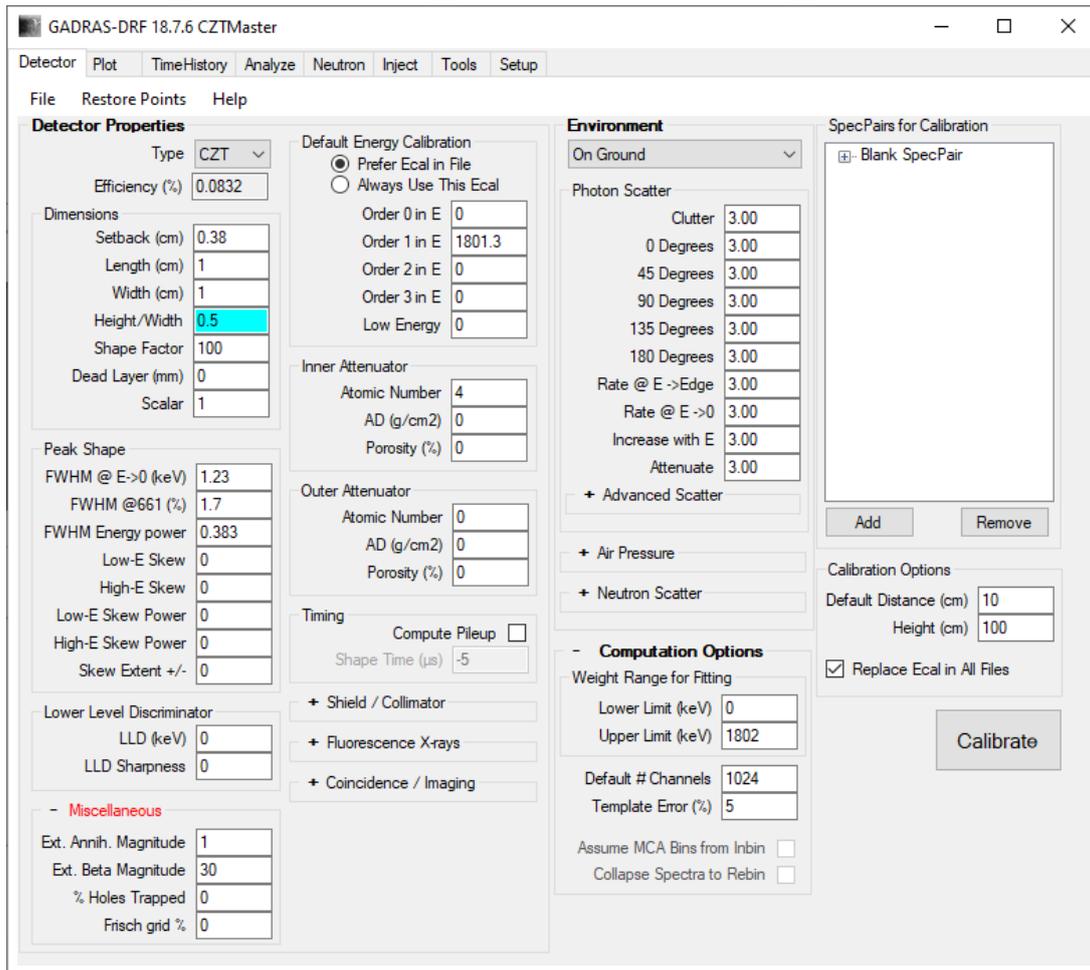


Figure 4.1: GADRAS Detector Tab.

Using the "Detector" tab of GADRAS-DRF software illustrated in Fig. 4.1 we are capable of defining a wide range of properties of the detector, such as: Dimensions, Peak Shape, Energy Calibration, Attenuation and some more specific characteristics that will not be used.

After defining the characteristics of the detector, simulated spectrum data can be obtained using the "Inject" tool present in the third tab counting from the end. Using this tool we are able to define the parameters of our synthetic acquisition and acquire the corresponding simulated spectra.

As we can see in Fig.4.2, this tab is divided in 4 main sections: General Settings, Neutrons, Gamma Detector and Background. The neutrons section was not used in the simulations. In the "General Settings" section we can define the main properties of the simulation: sources of radiation and corresponding activity rates, distance from the source to the detector, source height relative to the floor and acquisition time. There is also the choice of using Poisson statistics on the data, simulating spectra more similar to real measurements.

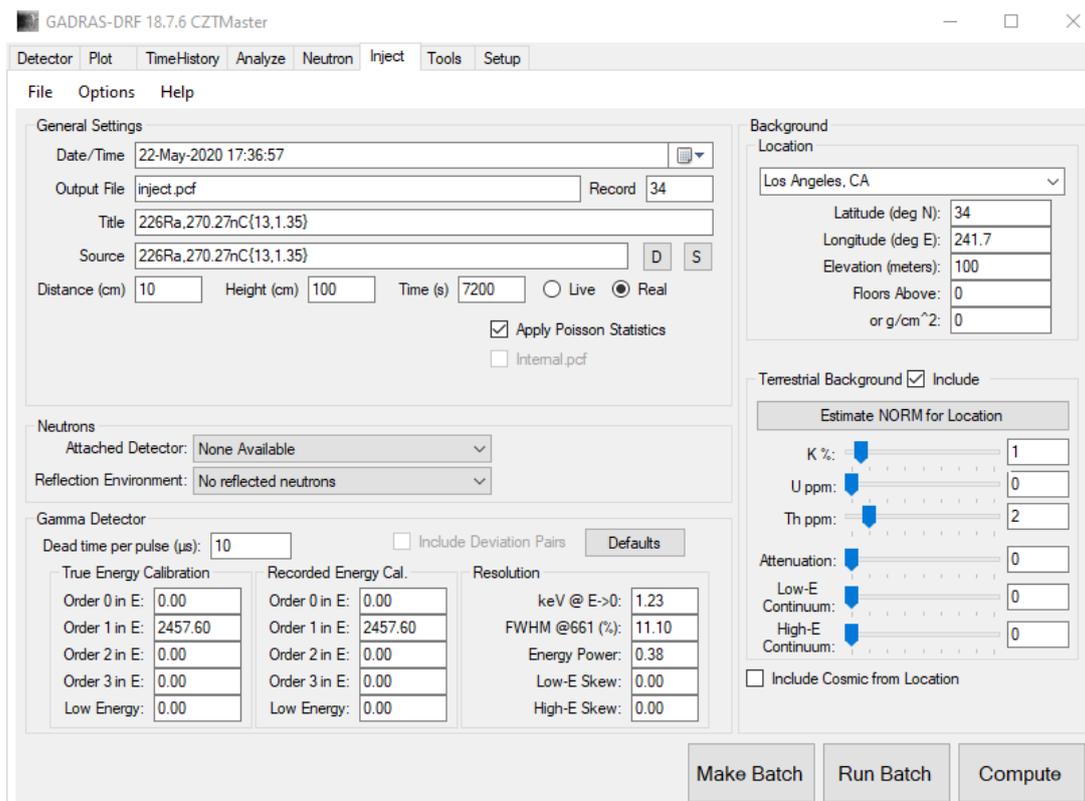


Figure 4.2: GADRAS Inject Tab.

Proceeding to the "Gamma Detector" module we encounter some of the same properties of the detector that could be found on the previous tab with the addition of a dead time per pulse input that can be used to define the dead time of the detector. If no changes are applied, the detector properties remain the same as the ones pre-defined on the "Detector" tab. Finally, the "Background" segment enables us to add background contribution to the simulated spectra. The options presented allow us to define the levels of background contribution from the isotopes: potassium (^{40}K), uranium (^{238}U , ^{235}U) and thorium (^{232}Th); as well as the attenuation of such background noise and even low-energy and high-energy continuum contributions from the terrestrial background.

An important feature provided by this software is the ability to perform batch inject calculations, meaning that a list of desired simulation parameters can be defined and then performed without having to individually set each run.

4.1.2 Simulated Samples

A solid and high quality training data-set should try to be as diverse as possible, containing samples that cover numerous different scenarios similar to the ones that we want to classify. Usually, having a more wide-ranging training data-set corresponds to better results of the algorithm.

When planning the different simulation scenarios it was realised that the parameters that focus should be on the parameters that lead to significant changes on the shape of the output spectrum and not on the total number of counts. This is the case since the idea is to normalise the spectrum to its maximum, thus disregarding the variation on the total counts for each channel and focusing on the relative counts. The maximum normalisation is chosen so that the ANN can more easily learn the spectrum pattern without having to deal with a varying maximum number of counts for different samples. Taking this into consideration, varying parameters such as acquisition time and activity of the source are not relevant, although it seemed a reasonable choice at first glance. Since these parameters will only affect the total number of counts in the spectra, after normalisation, the resulting spectrum samples would look almost identical, assuming an acquisition time large enough to fully develop the shape of the spectrum.

4.1.2.A Detector definition

In Fig.4.1 is shown all the different parameters that can be tuned regarding the simulated detector. In this work, the idea is to develop an ANN that would identify the data acquired with a specific CZT detector. Thus, the detector dimensions were given a fixed set of values that were obtained by looking at the data-sheets [59] of the detector in use (see 4.4.1).

Peak shape and energy calibration were defined using data provided by [60] where they performed an analysis of the CZT detector used in this work, obtaining a characterisation of its energy resolution and calibration.

In the peak shape area, only the first three inputs were used, setting all the remaining values to 0 (default). This area is designed to define the resolution of the simulated detector. Upon taking a closer look at the manual of the simulator [55] we can see that these values are described as follows:

- **FWHM @ E→0 (keV)**: zero-energy asymptote of the Full Width at Half Maximum (FWHM) in keV (can be negative);
- **FWHM @661 (%)**: percent FWHM of the photopeak at 661 keV (or for HPGe detectors, it is displayed as the width of the photopeak at 1332 in keV);
- **FWHM Energy power**: describes the rate at which the FWHM increases as a function of energy.

In order to obtain the values for each of the input boxes, the following equation was used to fit the peak data provided in [60]. This is the system of equations provided by GADRAS manual for FWHM definition:

$$FWHM(E) = \begin{cases} 6.61r \left(\frac{E}{661}\right)^p & E > 661 \\ \sqrt{\left[r_0 \left(\frac{661-E}{661}\right)\right]^2 + \left[6.61r \left(\frac{E}{661}\right)^p\right]^2} & E \leq 661, r_0 \geq 0 \\ 6.61r \left(\frac{\max(20,E)}{661}\right)^{p(\log(1-r_0))^{-1}} & E \leq 661, r_0 < 0 \end{cases} \quad (4.1)$$

	$E < 661$	$E > 661$
r	1.96	1.36
p	0.51	0.83

Table 4.1: Values from fitting of the equations.

where E is the photopeak energy in keV, r is the percent resolution at 661 keV, p is the FWHM energy power (unitless) and r_0 is the zero-energy asymptote in keV.

Fixing r_0 to 1.23 keV (value from [60]), r and p were obtained by fitting the first two equations from Eq. 4.1 since it is assumed that r_0 is positive.

The values obtained were slightly different both for r and p (Table 4.1). Having said that, a group of values was chosen for each of the parameters, using all of them in the simulation of the training data. The set of values include both of the values obtained for each fitting equation in order to cover the majority of relevant configurations.

Using multiple values provides the ANN with more possibilities of acquired spectrum since these parameters can fluctuate marginally even for the same model of detectors. Since every sensor can have a slightly different resolution, I proceeded to define minimum and maximum values for both r and p and chose some extra values in between those limits. For each combination of r and p , a different detector was simulated. It is worth mentioning that the step of each variation is important since on one hand a large step can exclude some relevant samples but on the other hand a tiny step can lead to samples too similar, possibly causing over fitting.

The values used for the three previously mentioned input boxes were:

- **FWHM @ $E \rightarrow 0$ (keV):** 1.23;
- **FWHM @661 (%):** [1.4, 1.7, 2];
- **FWHM Energy power:** [0.5, 0.65, 0.8]

Regarding the energy calibration, in [60] is shown a linear dependence between the energy and the channel number, which can be described by the equation 4.2.

$$E(\text{keV}) = 1.759 \cdot c - 0.0779 \quad (4.2)$$

where E and c correspond to the energy of a certain channel and the channel number, respectively.

Having noticed this linear relation, we can obtain the values for the first two input boxes (the only ones that are used) located in the energy calibration area. *Order 0 in E* corresponds to the bias of the energy calibration or, in other words, the offset energy for "channel 0". In this case, this value can be neglected since it is extremely low, so it was set to 0. *Order 1 in E* corresponds to the maximum value of the energy scale of the detector in use, meaning it should be set to a value close to 1801 keV

(1024 channel acquisition). The chosen possible values for the energy calibration were [1790.3, 1801.3, 1811.3].

After a close analysis of the GADRAS Manual, all remaining inputs were set to default apart from the channel number which was set to 1024 as previously mentioned.

4.1.3 Selected radionuclides and simulation parameters

Although simulating spectra for all existing radionuclides would be ideal, it would also be computationally unviable due to time and data management constraints. In order to reduce the amount of data generated it was defined that the ANN would only be trained to detect a specific set of isotopes, a set which can be expanded if needed in the future. The procedure for choosing the isotopes that would be included in the training set consisted in two phases. The first group of isotopes to be included were the ones which had available real world acquisitions. Having experimental data is important since this is the preferable type of data to be used for testing and evaluating the trained ANN. This first batch of isotopes includes: ^{22}Na , ^{60}Co , ^{137}Cs , ^{152}Eu , ^{241}Am . Since this first batch had a relatively small number of isotopes a second batch of isotopes was added in order to increase the complexity of the classification problem avoiding possible biased results due to learning for only a few radionuclides. This second group was chosen based on the list of common radionuclides presented on [61]. These were chosen taking into consideration that they would be more relevant to trace and identify. The final radionuclide list used for learning is:

- ^{22}Na , ^{60}Co , ^{137}Cs , ^{152}Eu , ^{241}Am , ^{226}Ra , ^{228}Ac , ^{235}U , ^{40}K , ^{133}Ba , ^{222}Rn , ^{57}Co , ^{54}Mn , ^{204}Tl , ^7Be
- 15 Isotopes

Having picked the radionuclides to simulate, the following step was to define the simulation parameters namely: acquisition time, distance from source to detector, activity of the source and background noise. In General Settings area (see Fig. 4.2), we have the possibility to define the source that is being chosen as well as its activity, using a specific code word. The number of counts is proportional both to the activity of the source as well as the acquisition time. A combination of an activity of 1 μCi and an acquisition time of 3600 real seconds was chosen, this choice being based on characteristics of previous experimental acquisitions. Distance to the source was set to 10 cm, height was set to 100 cm and Poisson statistics were applied. The neutrons section was not used and the gamma detector section configuration was inherited from the *Detector* tab with dead time per pulse set to default.

Regarding the background segment, there is the option of choosing a desired location (available only to USA) which automatically estimates the background activity. Taking into consideration that the ANN should be able to identify the radionuclides regardless of the location and background noise, a set of possible background values were picked for the simulations. In order to establish the domain of these

values, an heuristic scan was performed on the available locations defining minimum and maximum values for each input. Only "K %", "U ppm" and "Th ppm" values were modified, setting the remaining values to zero. The values used for the three previously mentioned inputs were:

- **K %**: [0, 2, 4];
- **U ppm**: [0, 2, 4];
- **Th ppm**: [0, 2, 4];

An independent spectrum was produced for each isotope for all the different detector configurations and simulation parameters previously described, scaling the number of single isotope spectrum samples to a total of **10 935 samples**. The same process was performed for multiple radionuclide simulations, where a combination of up to 3 isotopes were simulated in a single sample. In this case all radionuclides were simulated as having equal activity in the combination, since having different combinations of activity would lead to a not viable huge number of samples. A total of **76 545** and **331 695 samples** were simulated for dual and triple isotope combinations respectively. Each simulation of both single and multiple isotope data-sets, took approximately 3-4 hours (several iterations of this data-sets were generated along this work).

Several data management executables were created in order to handle this huge quantity of data files. The process of simulation of the data was divided into:

1. **Creation of simulation info files (.inj)** - These files containing the parameters of each individual simulation sample were moved to a specific folder inside GADRAS directory and then read by the program in order to produce the simulated spectrum.
2. **Decoding of the resulting files (.pcf)** - GADRAS outputs an individual file for each simulation sample and then decoded using a specific decoder program written in python. This program was developed by using [62] data-sheet.
3. **Creation of single file containing all data (.npy)** - In order to facilitate moving such a huge data set, all samples were put together in a Numpy array [63] and saved to a single file.
4. **Creation of file with classification information (.npy)** - A secondary Numpy array was created containing the identification array (array with a length of 15 values consisted of 0 and 1) for each simulated sample.

4.2 Artificial Neural Network Structure Development

When designing the structure of an ANN, two important definitions are required: the input and output formats. Deciding the way that we pretend to feed the data to the ANN has a direct impact on the number

of neurons of the input layer of the network. Similarly, the same is applied to the form of the output layer, which derives straightly from the type of information we pretend to extract from the input data.

Taking into account that the idea for this implementation is to receive the full acquired spectrum after a maximum normalisation, the size of the input layer must be equal to the number of channels of the detector itself. That being said, the input layer of the ANN needs to be composed of 1024 neurons, each one receiving the normalised number of counts for the respective channel, a number between 0 and 1.

Regarding the output of the ANN, the chosen approach was to have an output neuron for each of the possible identifiable radionuclides, totalling 15 neurons. Each neuron must return a number between 0 and 1, which is proportional to the probability of the respective radionuclide being present in the analysed sample, according to the ANN classification. Differently from the input layer, the output layer uses an activation function in order to provide the output. The activation function depends heavily on the type of classification desired for the ANN. For the case of single-isotope identification (multi-class classification) the activation of choice is usually Softmax (see 3.1.4.B) since we want to predict the most probable radionuclide in the sample. This means that the probability of each class must be related with the other classes as well.

On the other hand, in the case of trying to identify more than one radionuclide in the same sample (multi-label classification) the approach must be distinct. Each neuron output must be independently treated as a binary problem consisting in realising if the radionuclide is either present (value close to one) or not present (value close to zero) in the sample. For this case the activation function chosen was Sigmoid (see 3.1.4.A).

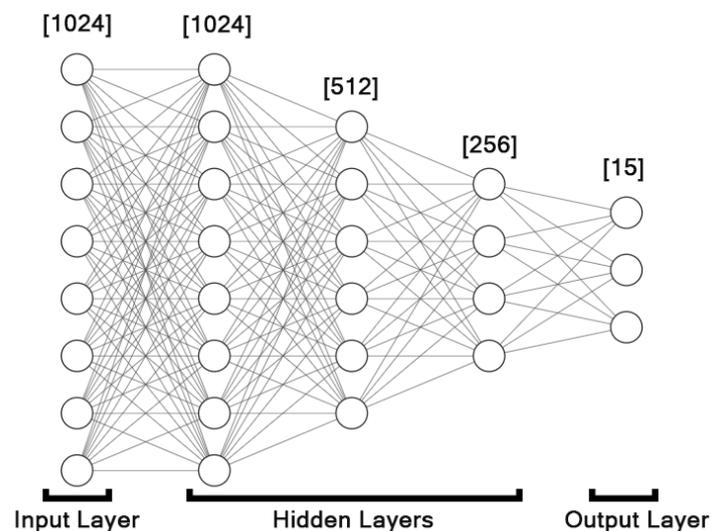


Figure 4.3: Base model for the artificial neural network structure.

Having defined the format of both input and output layers, the next step was to define the hidden

layers. In this work, three different hidden layers structures were tested for each of the desired classification options (single and multi radionuclide identification). Each independent approach used one, two or three hidden layers. The base model is shown in Fig. 4.3.

After defining the number of hidden layers of the ANN and their corresponding number of neurons, the following step is to choose the activation function to be used. For this approach, ReLu activation (see 3.1.4.C) was chosen for every neuron belonging to the hidden layers.

Every neuron from each layer makes a connection to all neurons in the next layer, meaning that the implemented ANN is fully connected.

Tensorflow [64] and Keras API [65] were used for the implementation of the mentioned ANNs.

4.3 Artificial Neural Network Training

As mentioned in 3.1.3, the training process consists in an iterative process of changing the weights of the neurons in order to decrease the loss function value, thus obtaining a more correct classification. With this in mind, one step is to choose a specific loss function and a learning algorithm that attempts to minimise it. The loss function can be usually related to the activation function that was chosen. For this particular implementation, binary cross-entropy was used for Sigmoid activation and categorical cross entropy for Softmax activation. The learning algorithm that was used for the training process was ADAM optimiser.

The use of regularisation methods is also important to try to mitigate the occurrence of over-fitting for the training data. Regularisation mechanisms that were analysed consisted in L1 and L2 regularisation. In addition, features such as batch size and number of epochs were studied in order to obtain the best possible results.

4.3.1 Training Features

When preparing to train the ANN, the best approach is to divide the whole data in at least two parts: Training data and Test data. The idea is to test the ANN using data that was not used in the training process in order to avoid a biased classification derived by the fact that the ANN might have already learned how to classify that specific example. In this case, the entire simulated data-set was used, since the results for the test data-set and training data-set were almost identical, being preferable to train with the whole data. The starting point for stopping mechanism implemented consisted in evaluating the training loss value and stop the learning process when this value did not reduce at least 10^{-4} over 10 consecutive iterations. Regarding the batch size, common values include 32, 64, 128, 256 [66]. The value of 128 was chosen as a first approach.

4.4 Proof-of-Concept Hardware

One important step towards the goal of this work is to develop a hardware solution that would enable the implementation and running of the Machine Learning solution proposed. There are mainly 3 important modules for the hardware implementation: Acquisition module, Analysis module and Support module (see Fig. 4.4). This modular architecture is designed to be as portable as possible in order to ease its integration in portable platforms such as cars, drones or other unmanned vehicles.

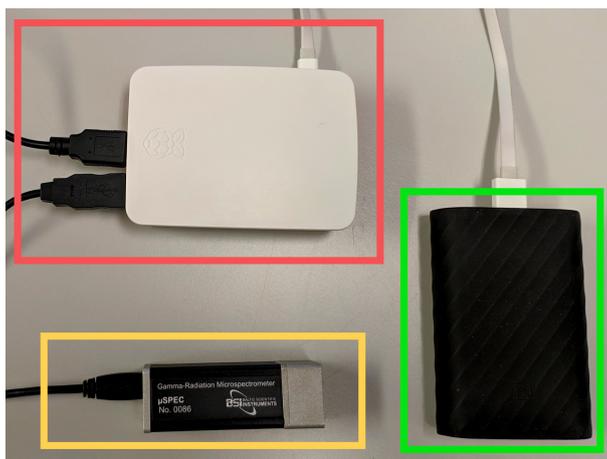


Figure 4.4: Detecting hardware used: Acquisition Module (CZT) - Yellow; Analysis Module (RasPi) - Red; Support Module (Power-Bank) - Green.

Note that this is one of many possible applications of the developed ANN. Other devices such as smartphones or laptops could be used as well as other type of interfacing languages.

4.4.1 Acquisition Module



Figure 4.5: μ Spec500 detector. [7]

When aiming to obtain the identity of a certain radioactive mixture/substance it is important to have the correct detecting systems. The quality of the acquired spectrum is usually directly related to the grade of

the detector used, meaning that a higher quality detector is always preferred. In this work, the detector used for the acquisition of the spectrum is the μ Spec500 [59] from *Ritec* [7].

This gamma-radiation spectrometer uses a quality CdZnTe semiconductor detector with a volume of 500 mm³ coupled to a small and low consuming digital multi channel analyser (MCA) *MicroMCA527* [67]. This micro-spectrometer allows measuring gamma-radiation spectra and storing it for processing in a PC through the USB port, making it extremely viable for portable, room temperature applications [7]. The choice of this detector can be attributed to the fact that it is a reliable component that was already field tested with real radiation sources [60].

4.4.2 Analysis Module

The identification of the acquired spectrum is performed in this module. This analysis could be performed by any device with enough processing power to run the developed ANN namely: computer, smartphone or micro-controller. For this particular implementation, it was used a compact and portable device that could be easily transported by a person or any unmanned vehicle and perform all the desired analysis. A RaspberryPi 3 (model B) [68] was used for this purpose, due to its huge versatility derived from the Linux type OS, excellent IO interface that enables it to connect to any additional gadget and its high portability.

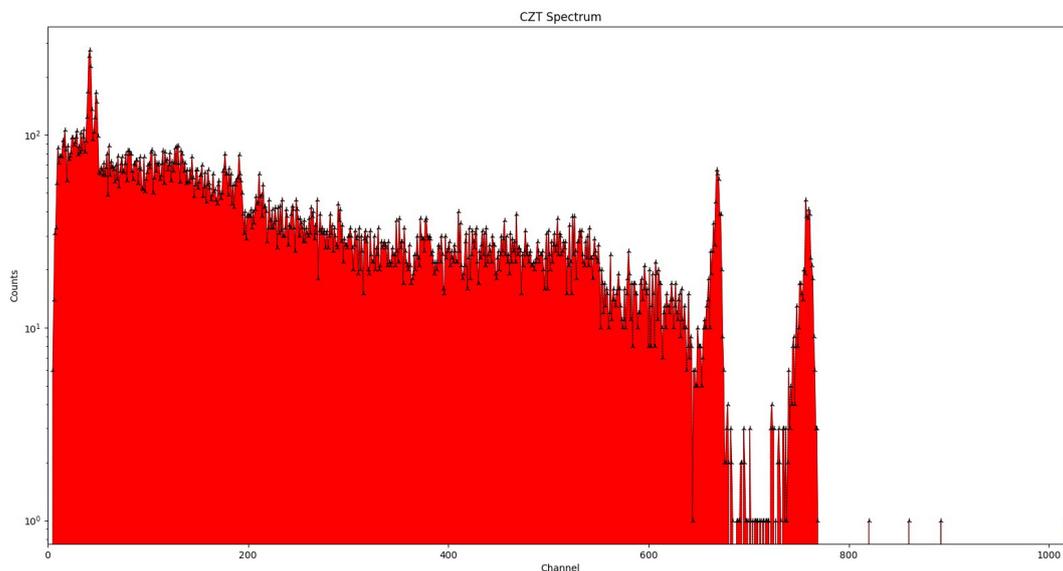


Figure 4.6: Screenshot from RaspberryPi during acquisition process using CZT sensor.

The entire software interface between the detector and the analysis module was coded using Python3 and based on MCA 527 Firmware Commands [69]. The current state of the code enables real-time readings from the CZT detector with the data being presented in a Counts vs Channel spectrum (see Fig.

4.6). This code has already been field-tested with real radioactive sources in the IPFN laboratory.

4.4.3 Support Module

This module includes all the peripheral components needed for the whole system to work such as a power-bank, interface peripherals (mouse + keyboard + screen) and a possible future storage box that englobes all the gadgets. Due to the lack of input/output characteristics of the RaspberryPi, in order to interact with it, we need a feedback output component such as a screen and some input components such as a mouse and keyboard. Carrying all these peripherals in a real world scenario would not provide a quality experience, so it was decided that swapping them by a smartphone would be a better option. Using a program such as VNC [70] enables us to connect the smartphone to the RaspberryPi and mirror the GUI output into our smartphone screen as well as use its touch capabilities to mimic mouse and keyboard interaction. The low consumption of the equipment used, means that the only power supply needed for this system to work is a simple power bank that can be resized according to the demand.

5

Results

Contents

5.1 Performance Analysis Criteria	41
5.2 Single Radionuclide Classification	44
5.3 Multi Radionuclide Classification	59

In order to obtain an optimal identification system, several steps of optimisation procedures must be taken. The results of all relevant phases are analysed in this chapter. Furthermore some improvement suggestions are also referred.

5.1 Performance Analysis Criteria

Validation of results is one hugely important step of developing any kind of artificial neural network. Acquiring an independent data-set from the training, for performance evaluation as well as defining a logical and understandable criteria is part of result demonstration.

5.1.1 Evaluation Data

The performance of the resulting ANNs was evaluated using a set of real spectra acquired in 3 different scenarios with the CZT detector previously introduced. The 3 groups of samples can be categorised as Single or Multi radionuclide acquisitions and include the radionuclide samples mentioned in Table 5.1.

Set Name	Radionuclides	Sample Name	Aquisition Time	Background sample?
Single_Set #1 (SS1)	^{60}Co	<i>SS1_60Co</i>	30 min	Yes
	^{137}Cs	<i>SS1_137Cs#1</i>	30 min	Yes
	^{137}Cs	<i>SS1_137Cs#2</i>	30 min	Yes
	^{152}Eu	<i>SS1_152Eu</i>	30 min	Yes
	^{22}Na	<i>SS1_22Na</i>	30 min	Yes
Single_Set #2 (SS2)	^{60}Co	<i>SS2_60Co#1</i>	90 min	Yes
	^{60}Co	<i>SS2_60Co#2</i>	90 min	Yes
	^{241}Am	<i>SS2_241Am</i>	90 min	Yes
Multi_Set #1 (MS1)	$^{226}\text{Ra}, ^{214}\text{Pb}, ^{214}\text{Bi}$ *	<i>MS1_#1</i>	30 min	Yes
	$^{226}\text{Ra}, ^{214}\text{Pb}, ^{214}\text{Bi}$ *	<i>MS1_#2</i>	30 min	Yes
	$^{226}\text{Ra}, ^{214}\text{Pb}, ^{214}\text{Bi}$ *	<i>MS1_#3</i>	30 min	Yes
	$^{226}\text{Ra}, ^{214}\text{Pb}, ^{214}\text{Bi}$ *	<i>MS1_#4</i>	30 min	No

Table 5.1: Data-sets used for evaluating the performance of the ANNs. * **Expected radionuclides.**

SS1 and *SS2* were acquired at the Lab using sources with well-known identities, meaning that the radionuclides attributed to each of the samples are correct. For *MS1* data-set the spectra were acquired in a real scenario of an old ore mine, so the attributed radionuclides to each sample correspond to the expected identities obtained with other analysis.



(a) Sources from SS2 data-set.



(b) Acquisition setup for sample SS2_60Co#2.

Figure 5.1: Image from samples used for SS2 data-set.

Before classification, each spectrum was normalised and subject to a background subtraction procedure consisting in taking the original spectrum and subtracting the entirety of the background noise acquisition (proportional to acquisition time of each sample).

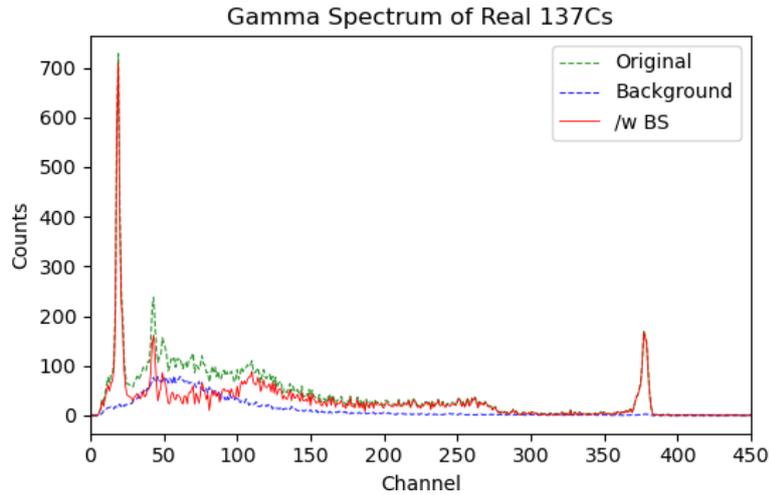


Figure 5.2: Background subtraction example using SS1_137Cs sample ([0,450] Channels).

An example of this procedure can be observed in Fig. 5.2, where both processed and unprocessed spectra are represented. This background subtraction phase was intended to remove some of the background noise and highlight real features of the spectra.

5.1.2 Metrics

When evaluating the performance of a resulting ANN trained for classification, the usual approach is to run the algorithm to analyse a specific set of "problems" and attribute an evaluation of correct (if the classification was correct) or incorrect (otherwise) to each of the processed examples. In the end, the *Accuracy* of a given algorithm is obtained based on the ratio of right guesses against the total number of "problems" that were analysed from that specific test set. Additionally to this accuracy metric, *Precision*

and *Recall* [8] (Fig. 5.3) are commonly used to analyse the performance of multi-label classification algorithms.

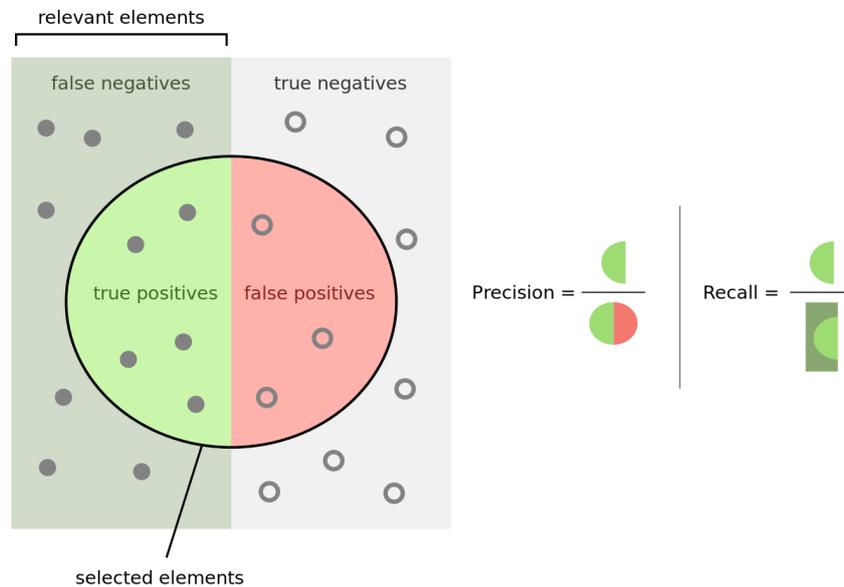


Figure 5.3: Precision and Recall visual demonstration [8].

These two metrics can be obtained by the ratio between True Positives (TP) and all identified positives and the ratio between TP and all existing positives respectively. High precision means that an algorithm returns more relevant results than irrelevant ones, while high recall means that an algorithm returns most of the relevant results. In this specific case, the output of the ANN is a list containing the reference for the identified radionuclides along with the certainty value associated to each identification. This output is then compared to the expected classification and these three metrics are evaluated. Precision and recall provide fine measurements to analyse the performance of the ANN, sometimes better than the Accuracy, which is obtained by the ratio between correct guesses, TP and True Negatives (TN), and total guesses.

Finally, another interesting metric to analyse is the F1-score which essentially combines precision and recall values in order to obtain a final score. This metric can be obtained by the following equation:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (5.1)$$

Due to the low number of evaluation samples, the metrics presented are always referring to micro-averaged values [71], meaning that they are computed considering all samples together. In other words, instead of analysing class by class cases, all the TP, False Positives (FP) and False Negatives (FN) are summed up and used to calculate the metrics previously mentioned. In this case in particular, a Positive refers to the isotope being detected in the sample and Negative otherwise.

5.2 Single Radionuclide Classification

In order to reach the best ANN, a sequence of steps was followed trying to optimise every feature of the network. When training the ANN to learn single radionuclide classification, only the single isotope simulated training data-set was used.

The starting points for the ANN structures analysed are described in Table 5.2.

ANN Codename	Input Layer	Hidden Layers		Output layer	
	neurons	neurons	activation	neurons	activation
SoMx3	1024	1024*512*256	ReLu	15	Softmax
SoMx2	1024	1024*512	ReLu	15	Softmax
SoMx1	1024	1024	ReLu	15	Softmax
Sig1	1024	1024*512*256	ReLu	15	Sigmoid
Sig2	1024	1024*512	ReLu	15	Sigmoid
Sig1	1024	1024	ReLu	15	Sigmoid

Table 5.2: Starting point for the ANN structures under analysis.

5.2.1 Number of Hidden Layers

For a first approach, 6 different ANN structures were tested. Every single training process used the same learning parameters described in Table 5.3. No regularisation methods were used in this section. Each learning process continued until the early stopping mechanism was activated. For every different configuration, this procedure was repeated at least several times in order to avoid possible fluctuations due to different weight initialisation.

Learning Parameters	
Batch_size	128
Early stopping	10^{-4} , patience: 10 epochs
Regularisation	None

Table 5.3: Learning parameters for hidden layer variation.

Looking at Table 5.4, we can see that without any type of regularisation the ANN rapidly converges to a solution after less than 100 epochs, reaching loss values in the order of 10^{-4} . These low values indicate that there could almost certainly be some sort of over-fitting for the training data. Another detail that hints at this over-fitting phenomena is the fact that the ANNs with more hidden layers are requiring less epochs in order to activate the stopping mechanism, something that is counter intuitive since these have a higher number of neurons, hence weights, that need to be tuned.

ANN Codename	Final Epochs	Final Loss
SoMx3	32	$7.9 \cdot 10^{-5}$
SoMx2	46	$1.4 \cdot 10^{-4}$
SoMx1	77	$2.7 \cdot 10^{-4}$
Sig3	34	$4.0 \cdot 10^{-5}$
Sig2	40	$7.3 \cdot 10^{-5}$
Sig1	79	$1.2 \cdot 10^{-4}$

Table 5.4: Final number of epochs and loss value for each of the ANNs.

The evaluation of the ANNs using the training-set resulted in a perfect score for all 3 metrics (Accuracy, Precision and Recall) for all of the different structures analysed. Thus the confusion matrix is an identity matrix.

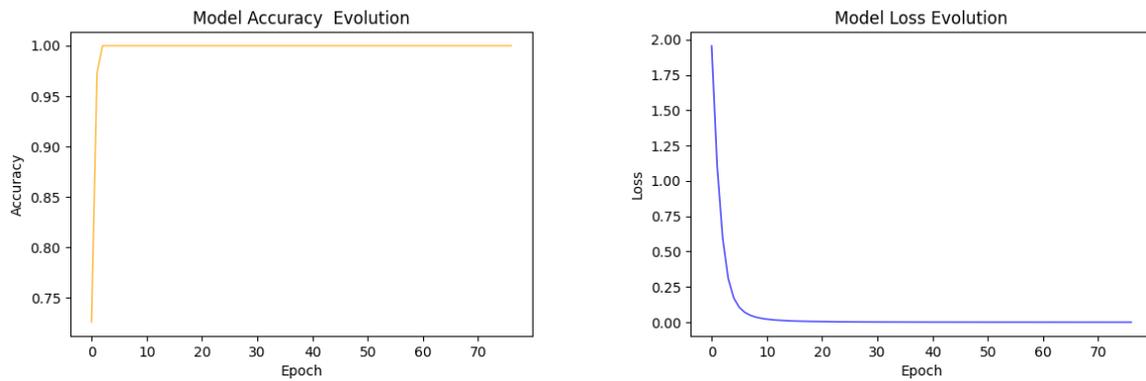


Figure 5.4: Accuracy and Loss evolution for *SoMx1* network.

This was somewhat expected, since the simulated data represented a favourable scenario for identification for each of the samples. Additionally, since the variation of the parameters for the simulation of the spectrum samples is small, the ANNs are capable of learning particularly well for the simulated samples. Although some background noise features were simulated, the resulting spectra were easily identifiable. Another possibility for this perfect score is over-fitting for the training data.

On the other hand, the results obtained for the samples from the evaluation data-set (see Table 5.5) were not so great, as could be expected. Table 5.5 represents the classification provided by each of the trained ANN configurations for each of the samples from the evaluation data-set. Moreover, the accuracy, precision, recall and F1-score values are also present.

Sample name	SoMx3	SoMx2	SoMx1	Sig3	Sig2	Sig1
SS1_60Co	$^{137}\text{Cs} - 88\%$	$^{137}\text{Cs} - 98\%$	$^{137}\text{Cs} - 98\%$	$^{137}\text{Cs} - 99\%$	None	None
SS1_137Cs#1	$^{137}\text{Cs} - 99\%$	$^{137}\text{Cs} - 100\%$	$^{137}\text{Cs} - 100\%$	$^{137}\text{Cs} - 99\%$	$^{137}\text{Cs} - 99\%$	$^{137}\text{Cs} - 100\%$
SS1_137Cs#2	$^{137}\text{Cs} - 99\%$	$^{137}\text{Cs} - 99\%$	$^{137}\text{Cs} - 99\%$	$^{137}\text{Cs} - 98\%$	$^{137}\text{Cs} - 99\%$	$^{137}\text{Cs} - 99\%$
SS1_152Eu	$^{152}\text{Eu} - 99\%$	$^{152}\text{Eu} - 99\%$	$^{152}\text{Eu} - 99\%$	$^{152}\text{Eu} - 99\%$	$^{152}\text{Eu} - 99\%$	$^{152}\text{Eu} - 99\%$
SS1_22Na	$^{22}\text{Na} - 99\%$	$^{22}\text{Na} - 99\%$	$^{22}\text{Na} - 99\%$	$^{22}\text{Na} - 73\%$	$^{22}\text{Na} - 98\%$	$^{22}\text{Na} - 97\%$
SS2_60Co#1	$^{40}\text{K} - 96\%$	$^{40}\text{K} - 92\%$	$^{40}\text{K} - 55\%$ $^{137}\text{Cs} - 36\%$	None	None	None
SS2_60Co#2	$^{40}\text{K} - 72\%$	$^{40}\text{K} - 40\%$ $^{137}\text{Cs} - 40\%$	$^{40}\text{K} - 25\%$ $^{137}\text{Cs} - 54\%$	None	None	None
SS2_241Am	$^{152}\text{Eu} - 72\%$	$^{133}\text{Ba} - 37\%$ $^{152}\text{Eu} - 24\%$	$^{133}\text{Ba} - 19\%$ $^{152}\text{Eu} - 38\%$	None	None	None
Accuracy	0.50	0.50	0.50	0.50	0.50	0.50
Precision	0.50	0.50	0.50	0.50	0.50	0.50
Recall	0.50	0.40	0.36	0.80	1.00	1.00
F1-score	0.5	0.44	0.42	0.62	0.67	0.67

Table 5.5: Results of the classification using each ANN for the evaluation samples regarding hidden layer number variation. A classification threshold of 15% was used to filter lower results.

The ANNs using Softmax activation in the output layer reveal a higher tendency to provide more FPs when the classification provided is wrong. This is the case of the sample *SS160_Co* which was consistently classified with high confidence as ^{137}Cs , despite representing a ^{60}Co spectrum.

By contrast, Sigmoid activation ANNs present more False Negatives having great difficulty to identify ^{60}Co in any of the provided samples, similarly to the Softmax ones. It is noteworthy the fact that all ANNs can correctly identify the ^{137}Cs , ^{152}Eu and ^{22}Na samples, despite the fact that these show some noise differences when compared with the simulated samples.

Fig. 5.5 reveal both real and simulated spectra for ^{152}Eu . It is possible to identify some differences regarding noise values, however the location and even resolution of the peaks is significantly similar.

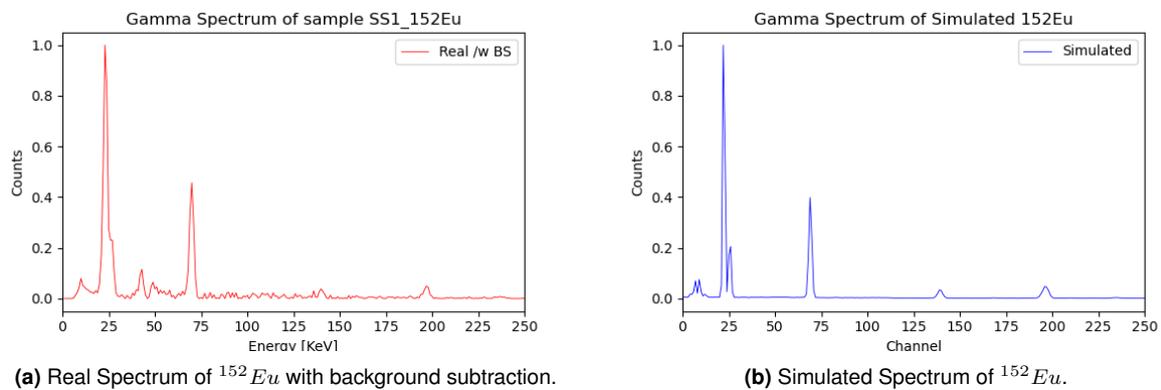


Figure 5.5: Comparison between Real and Simulated spectrum from ^{152}Eu ([0,250] Channels). The number of counts is normalised to the maximum value.

Assuming a possible over-fitting of the training-data, these results suggest that the simulation of the samples was properly executed. However they do not allow any concrete conclusions about the effect of the number of hidden layers in the classification process.

5.2.2 Regularisation Methods

Regularisation is a technique that works by constraining/regularising the learning parameters of the model, discouraging complexity and avoiding the risk of over-fitting. The two most common methods of regularisation are Lasso (L1) regularisation, and Ridge (L2) regularisation. They penalise the model by either its absolute weight (L1), or the square of its weight (L2) [72].

In an attempt to deal with the over-fitting phenomena present in the previous section, L1 and L2 regularisation measures were applied. When implementing regularisation methods it is important to start with lower values in order to avoid under-fitting. For this reason the first approach was to use regularisation values of 0.001 for both L1 and L2.

ANN Codename	Final Epochs	Final Loss
SoMx3	960	0.027
SoMx2	809	0.018
SoMx1	384	$3.0 \cdot 10^{-3}$
Sig3	1054	0.030
Sig2	924	0.020
Sig1	449	$4.0 \cdot 10^{-3}$

Table 5.6: Final number of epochs and loss value for each of the ANNs for regularisation of $L1 : 0.001$ and $L2 : 0.001$

Taking a closer look on the loss values shown in Table 5.6 we can notice that they are far greater when comparing to the previous results. This suggest that the over-fitting was somewhat reduced. Another factor that changed was the number of epochs necessary to reach the early stopping condition. We now observe that the number of epochs is proportional to the number of hidden layers (complexity) of the network.

Regarding the regularised ANNs, it is possible to observe that the some of them output highly satisfactory classification results, especially the networks with fewer hidden layers (Table 5.7). *SoMx1* network is clearly the best performer for the evaluation data-set, obtaining a F1-score value of 0,94. Another particularity is the fact that *SS1_152Eu* sample was correctly identified by all analysed ANNs.

The results also suggest that using either fewer hidden layers can lead to better classification outputs for this particular problem.

Sample name	SoMx3	SoMx2	SoMx1	Sig3	Sig2	Sig1
SS1_60Co	^{54}Mn – 44% ^{57}Co – 22% ^{137}Cs – 17% ^{152}Eu – 16%	^{54}Mn – 33% ^{235}U – 46%	^{60}Co – 91%	^{54}Mn – 94%	^{60}Co – 40% ^{152}Eu – 96% ^{235}U – 21%	None
SS1_137Cs#1	^{137}Cs – 99%	^{137}Cs – 99%	^{137}Cs – 100%	^7Be – 99% ^{40}K – 99% ^{60}Co – 99% ^{137}Cs – 100% ^{222}Rn – 99%	^{137}Cs – 100% ^{222}Rn – 60% ^{228}Ac – 99%	^{137}Cs – 99%
SS1_137Cs#2	^{137}Cs – 99%	^{54}Mn – 18% ^{137}Cs – 61%	^{137}Cs – 72%	^{137}Cs – 99%	^{152}Eu – 97%	^{137}Cs – 78%
SS1_152Eu	^{152}Eu – 99%	^{152}Eu – 99%	^{152}Eu – 99%	^{152}Eu – 99%	^{152}Eu – 99%	^{152}Eu – 99%
SS1_22Na	^{235}U – 89%	^{226}Ra – 46% ^{228}Ac – 45%	^{22}Na – 42%	^{54}Mn – 99%	^{137}Cs – 92% ^{228}Ac – 96%	None
SS2_60Co#1	^{54}Mn – 79%	^{54}Mn – 54% ^{60}Co – 22% ^{152}Eu – 21%	^{60}Co – 99%	^7Be – 80% ^{40}K – 78% ^{60}Co – 99% ^{137}Cs – 79% ^{222}Rn – 80%	^{60}Co – 99%	^{60}Co – 99%
SS2_60Co#2	^{40}K – 89%	^{54}Mn – 56% ^{60}Co – 36%	^{60}Co – 55%	^7Be – 91% ^{40}K – 90% ^{60}Co – 99% ^{137}Cs – 90% ^{222}Rn – 90%	^{60}Co – 94% ^{152}Eu – 21%	^{60}Co – 64%
SS2_241Am	^{222}Rn – 47% ^{226}Ra – 52%	^{222}Rn – 98%	^{22}Na – 15% ^{241}Am – 65%	None	^{222}Rn – 98%	None
Accuracy	0.38	0.25	0.88	0.25	0.25	0.63
Precision	0.50	0.63	1.00	0.63	0.63	0.63
Recall	0.33	0.36	0.84	0.26	0.36	1.00
F1-score	0.40	0.46	0.94	0.37	0.46	0.77

Table 5.7: Results for each ANN for regularisation of $L1 : 0.001$ and $L2 : 0.001$. A classification threshold of 15% was used to filter lower results.

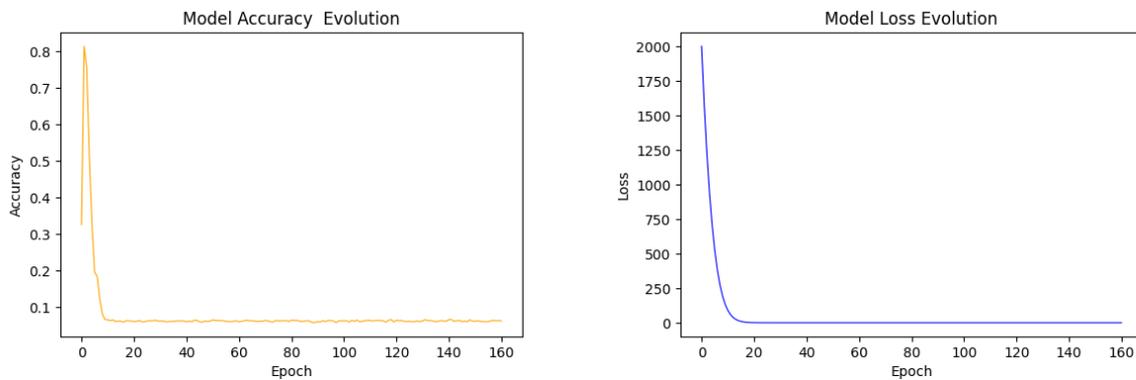


Figure 5.6: Accuracy and Loss evolution for SoMx2 for regularisation of $L1 : 0.1$ and $L2 : 0.1$. Final loss value of 3.36.

Other values of L1 and L2 were also tested, including using only one of the regularisation methods at a time. However none of them presented better results than the ones in Table 5.7. For ANNs with more hidden layers, higher values of regularisation led them not to converge to a proper solution at all (see Fig. 5.6).

5.2.3 Batch size variation

Due to the reduced size of the available evaluation data-set, it is difficult to identify the best and most optimised ANN parameters, however some optimisation attempts were taken.

	SoMx1			SoMx2			Sig1			Sig2		
Batch size	64	128	256	64	128	256	64	128	256	64	128	256
Final Epochs	269	384	571	614	809	1037	311	449	653	712	924	1204
Final Loss (10^{-2})	0.24	0.32	0.45	1.58	1.81	2.10	0.32	0.40	0.57	1.72	1.99	2.41

Table 5.8: Final number of epochs and loss value for each batch size. Regularisation of $L1 : 0.001$ and $L2 : 0.001$.

Regarding the variation of batch size, the first observation we can make by looking at Table 5.8 is that for higher batch sizes, the number of epochs required to reach the early stopping criteria was higher. However, the loss values were higher for every case.

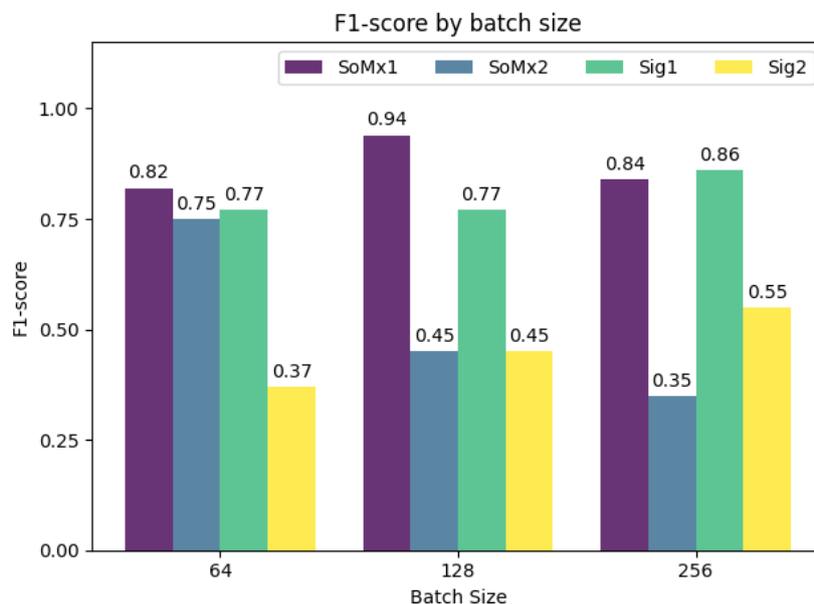


Figure 5.7: F1-score values for each batch size. Regularisation of $L1 : 0.001$ and $L2 : 0.001$.

In Fig. 5.7 is shown a bar plot representing the F1-score obtained for each of the ANNs with fewer hidden layers, using 3 different batch sizes. Notice that even though the batch size of 64 represents an

overall higher F1-score, the best results are still obtained for *SoMx1* for a batch size of 128. This is an additional confirmation that networks with fewer hidden layers might perform better than more complex ones. The F1-score does not take in consideration the variation in the confidence levels of each identified isotope, so if the F1-Score values were close to identical we could use Binary Cross-Entropy (Eq. 3.4) formula to compare those configurations.

5.2.4 Total Epochs

It is also pertinent to analyse if the correct stopping criteria was applied. In order to verify this, a number of different stopping mechanisms were tested for *SoMx1* configuration. Each stopping criteria value results in a particular number of total epochs. An extra configuration with 2000 epochs was added to the analysis to represent a larger number of iterations.

Loss variation	10^{-3}		10^{-4}				
Early Stopping (epochs)	10	5	10	20	30	40	-
Epochs	162	288	384	772	900	1038	2000
Final Loss (10^{-2})	0.98	0.46	0.30	0.31	0.27	0.23	0.13

Table 5.9: Final number of epochs and loss value for each stopping criteria for *SoMx1* using a regularisation of $L1 : 0.001$ and $L2 : 0.001$.

Results in Table 5.9 indicate that the change in loss values for early stopping criteria larger than 10 (for variation of 10^{-4}) is relatively small when compared to the extra epochs that the network used to learn. At first glance these data suggest that previously used criteria was qualitatively adequate.

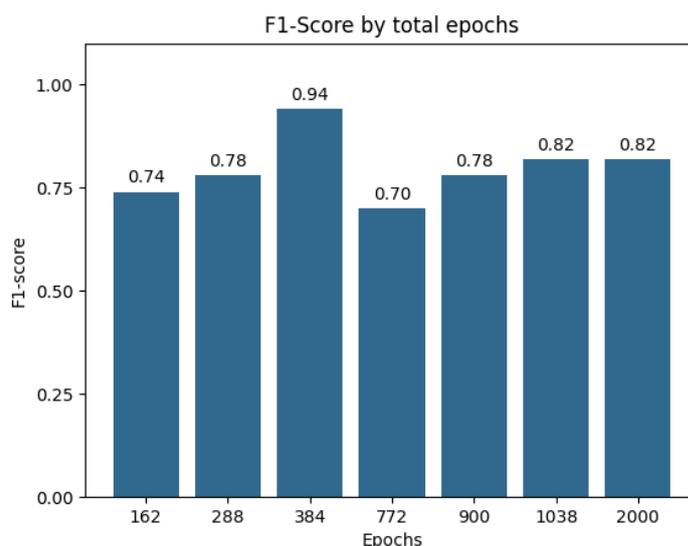


Figure 5.8: F1-score values for different total epochs for *SoMx1*. Regularisation of $L1 : 0.001$ and $L2 : 0.001$.

The bar plot in Fig. 5.8 confirms the pre-analysis made before by revealing the F1-score values for the different epoch number. The previously used early stopping criteria of 10 epochs for a variation of 10^{-4} presented the best results.

5.2.5 Background Subtraction effect

Background subtraction is a mechanism that intends to filter some of the noise present in the spectra of the samples. This process consists basically in acquiring a sample spectrum of the area where little to none radioactivity from the source under analysis is detected and subtracting it from the acquired spectra. Background noise sample acquisition is not always possible, making it relevant to analyse the performance of the ANN when classifying samples without background subtraction.

Background Subtraction	Yes	No
<i>SS1_60Co</i>	^{60}Co – 91%	^{22}Na – 53% ^{40}K – 41%
<i>SS1_137Cs#1</i>	^{137}Cs – 100%	^{137}Cs – 99%
<i>SS1_137Cs#2</i>	^{137}Cs – 72%	^{22}Na – 18% ^{204}Tl – 53%
<i>SS1_152Eu</i>	^{152}Eu – 99%	^{152}Eu – 99%
<i>SS1_22Na</i>	^{22}Na – 42%	^{22}Na – 96%
<i>SS2_60Co#1</i>	^{60}Co – 99%	^{60}Co – 91%
<i>SS2_60Co#2</i>	^{60}Co – 55%	^{60}Co – 56%
<i>SS2_241Am</i>	^{22}Na – 15% ^{241}Am – 65%	^{22}Na – 15% ^{241}Am – 65%
Precision	0.88	0.75
Recall	1.00	0.55
F1-score	0.94	0.63

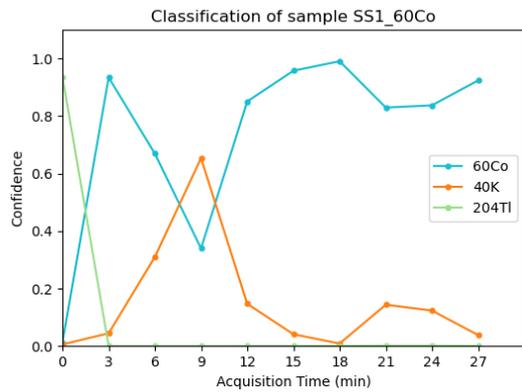
Table 5.10: SoMx1 classification results with and without background subtraction. Regularisation of $L1$: 0.001 and $L2$: 0.001

Table 5.10 reveals the classification output for each sample with and without background subtraction, making it possible to compare both results. Note that not having background subtraction affects especially the isotopes with lower activity counts or with acquisitions made farther from the source, where the ratio between source and background counts is lower. This might be the case of samples *SS1_60Co* and *SS1_137Cs#2* which show significantly better results when background subtraction is applied. Unexpectedly, the confidence value for *SS1_22Na* increased without background subtraction, despite it being the sample with lower activity.

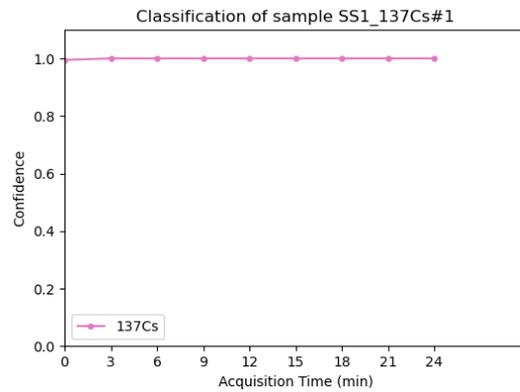
Based on these results, background subtraction should be performed whenever possible since it led to an overall better classification.

5.2.6 Acquisition Time effect

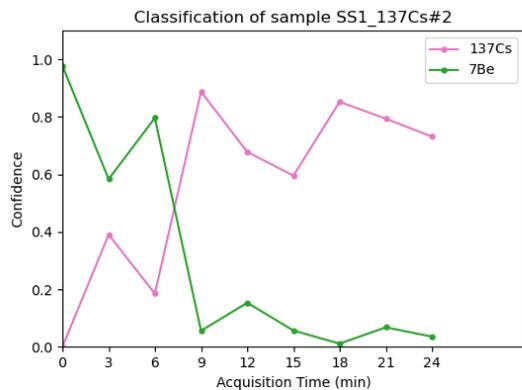
The Acquisition time for a given sample has an impact on the form of the spectrum, since some features might take longer to be defined. It is then relevant to examine how the acquisition time of each sample spectrum would affect the classification output provided by the ANN. For this matter, the spectrum of each sample was classified every 3 minutes, creating a graph that represents the ANN output in function of acquisition time.



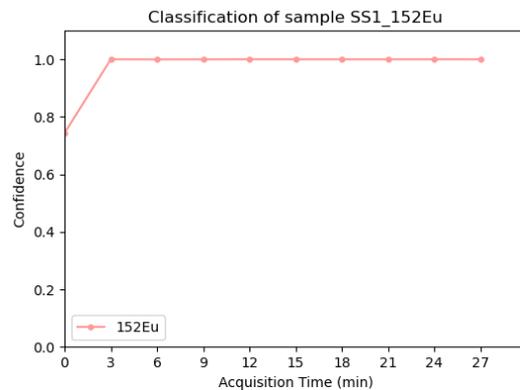
(a) Sample SS1_60Co.



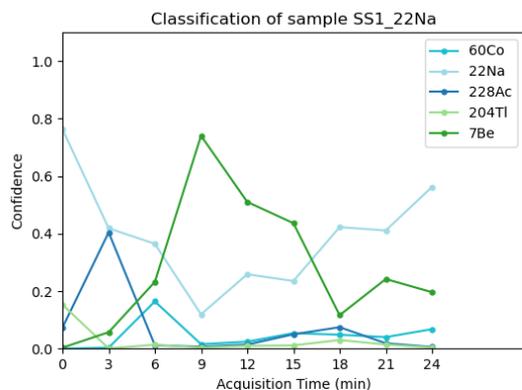
(b) Sample SS1_137Cs#1.



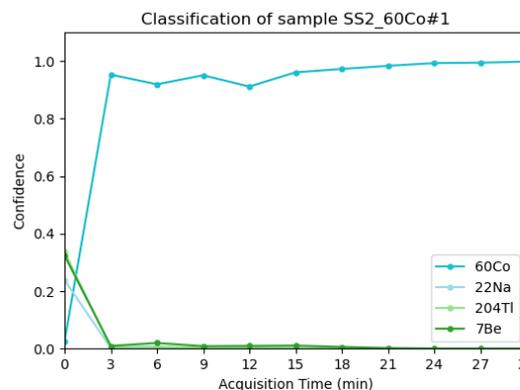
(c) Sample SS1_137Cs#2.



(d) Sample SS1_152Eu.



(e) Sample SS1_22Na.



(f) Sample SS2_60Co#1.

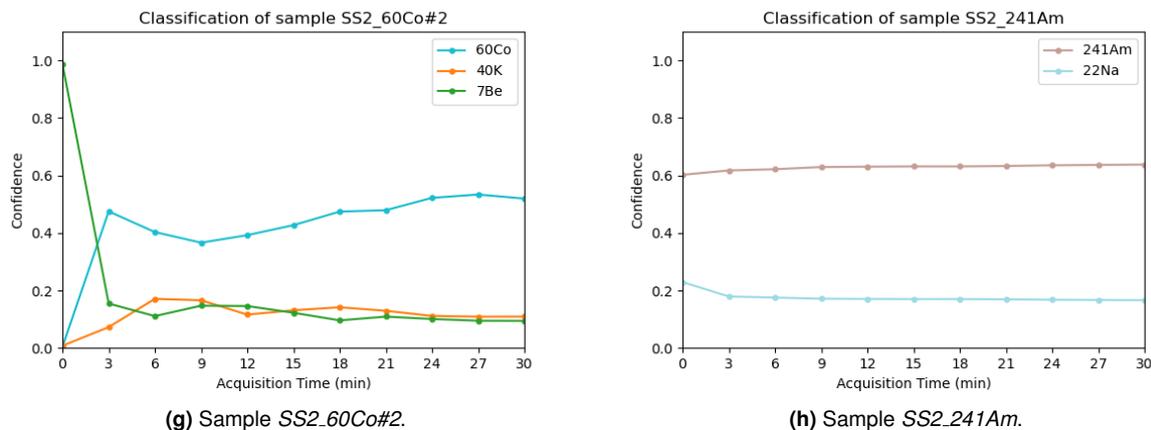


Figure 5.9: Classification of each sample along acquisition time.

Fig. 5.9 reveals that the majority of the samples take between 3 and 9 minutes of acquisition time to reach their final classification confidence levels. Sample *SS1_60Co*, *SS1_137#2* and *SS1_22Na* are the ones that exhibit noticeable fluctuations until reaching the final classification. This can possibly be explained by the fact that these spectra present higher background noise than the other samples, hindering the classification of the ANN. Note that background noise subtraction was applied in proportion to the acquisition time to every classified spectra.

Overall, this analysis suggest that the designed ANN is suitable for quick isotope identification with subjectively sufficient confidence levels.

5.2.7 False Positives Analysis

In order to test the possibility of obtaining False Positive outputs when using the resulting ANN, a data-set containing 4 new isotopes was created. These isotopes were: ^{65}Zn , ^{243}Am , ^{238}U and ^{109}Cd . The new data-set was generated equally to the training set, totalling 729 samples for each isotope. None of these samples was fed to the ANN during training, so the expected classification output would be to get no identification at all.

When using Softmax activation function, the sum of all the output values provided by the ANN is always equal to one regardless of the input. This means that, the ANN will always classify the sample as belonging to one of the available classes. Ideally there should be a class that represents "none of the above". Unfortunately, a class like this does not always make sense or is even achievable. In this particular case, creating a class like this would require the simulation of a large number of samples consisting of all the other possible radionuclides. If we were to simulate a data-set like this one, it would make more sense to label all these extra samples and define new classes for them. Since a class of this kind is not possible, the ideal way of solving this problem is to establish some thresholds for the

confidence levels output by the ANN.

The perfect output for "none of the above" cases would be to have the probability distributed over the total classes. In this case, each output neuron should give a value of 6.66%. A usual threshold for classification is the 50% value.

Isotopes	False Positives	Total
^{65}Zn	571 114	^{40}K ^{60}Co 685
^{109}Cd	415	^{241}Am 415
^{238}U	729	^{228}Ac 729
^{243}Am	27 216	^{22}Na ^{222}Rn 243

Table 5.11: False positives results for *SoMx1* using a threshold of 50%

Looking at Table 5.11 it is possible to notice that the results for each of the tested isotopes present a high number of false positives relatively to the total 729 samples. Again this can be elucidated by the fact that an ANN with Softmax activation function in the output layer will always try to provide a classification result.

These results can actually be explained by the form of each spectrum. For example, the ANN classifies all ^{238}U samples as being ^{228}Ac since they have two highly relevant peaks in the same energy range (Fig.5.10). All the remaining results could be explained similarly.

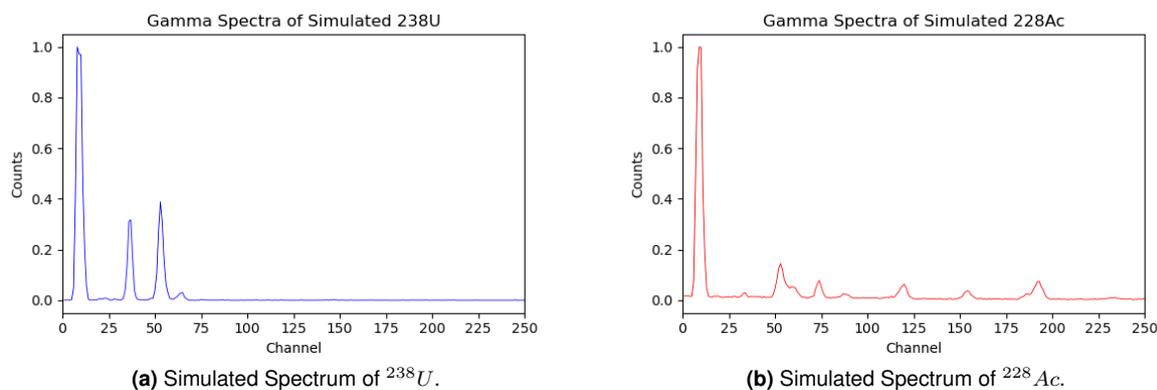


Figure 5.10: Comparison between Simulated spectra from ^{238}U and ^{228}Ac ([0,250] Channels). The number of counts is normalised to the maximum value.

In order to avoid this kind of problem, a possible solution would be to either train the ANN with more data or develop different ANN that can identify specific radionuclides. The ANN should always be trained with all the isotopes that we want to identify and should only be used when radioactivity has been previously detected for example by a Geiger-Muller counter, in order to avoid false positives when analysing background noise only.

5.2.8 Final ANN Results

The ANN that provided better results for single radionuclide identification was *SoMx1*, obtaining very interesting F1-score values for the utilised evaluating data-set. Additionally, it is interesting to analyse more closely the classification of each of the 8 samples that are part of the evaluating data-set. Each final output produces a graph containing the final acquisition spectra and highlighting the expected peaks from each of the identified isotopes.

5.2.8.A *SS1_60Co* sample

This sample in particular is difficult to classify even for an expert since the resulting spectrum contains a large amount of background noise (see Fig. 5.11). It is difficult to describe exactly how the ANN classifies a specific sample, since it simply receives an input and gives an output. However, in this case the correct classification of the isotope is possibly due to the network being able to identify the photo-peaks present in the 1173 keV and 1332 keV areas. If the photo-peaks were from a lower energetic area, they could be more difficult to distinguish from the noise.

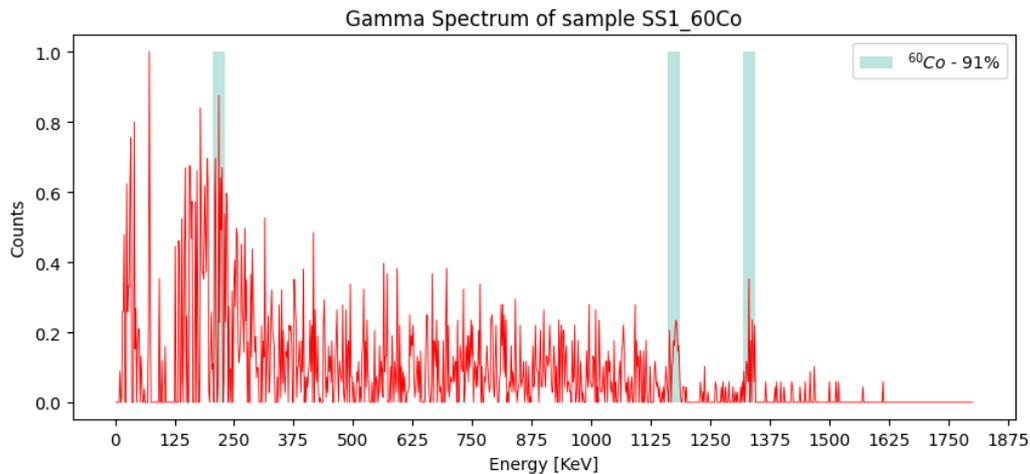


Figure 5.11: *SS1_60Co* spectra with classification peaks.

Signalling relevant peaks from the identified radionuclides in the spectrum provides relevant information that may help to visually confirm the classification obtained by the ANN.

5.2.8.B *SS1_137Cs#1* and *SS1_137Cs#2* samples

The spectra of these samples were acquired in similar conditions, so a discrepancy in the classification output is something that should be explained. The correct classification of sample *SS1_137Cs#1* is observed among almost all the analysed ANN structures. In fact, the only relevant difference in the

measurements of these two sources was the distance between the detector and the source.

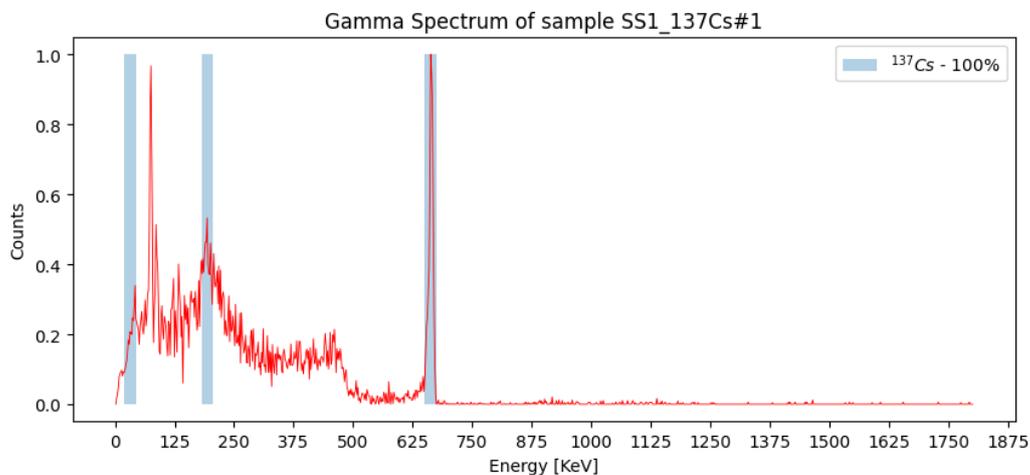


Figure 5.12: *SS1_137Cs#1* spectra with classification peaks.

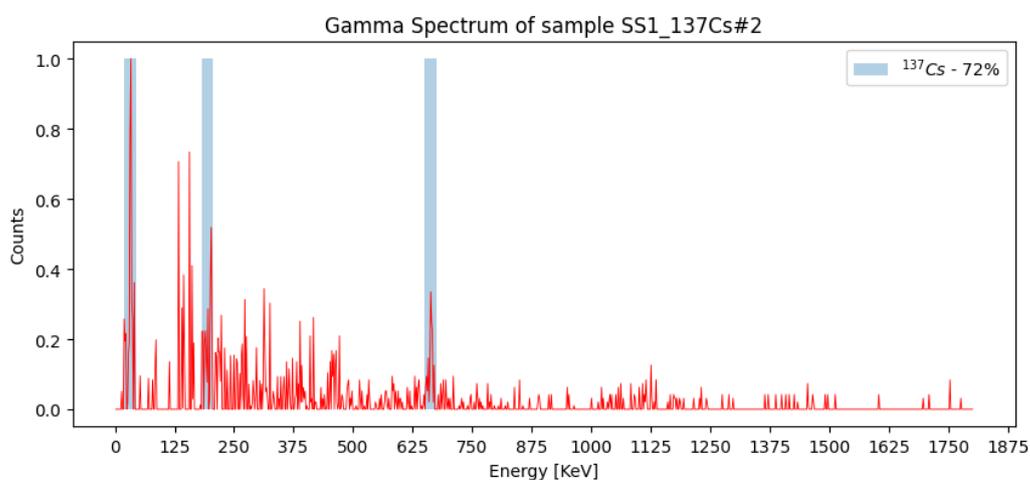


Figure 5.13: *SS1_137Cs#2* spectra with classification peaks.

Sample *SS1_137Cs#1* was acquired much closer to the source, revealing a much more consistent classification result. Being further from the source causes the ratio between actual source counts and background noise counts to be lower, making it more difficult for the ANN to correctly classify the spectrum. Although a good background subtraction procedure can attenuate this problem, we can clearly observe that sample *SS1_137Cs#2* (5.14) is much noisier and consequently has a lower confidence level output than sample *SS1_137Cs#1* (5.12).

5.2.8.C SS1_152Eu sample

Similarly to *SS1_137Cs#1*, the *SS1_152Eu* sample was correctly identified by almost every ANN configuration. This can be explained by the fact that the acquired spectrum is quite similar to the training ^{152}Eu spectrum (see Fig. 5.10a), having well defined photo peaks.

5.2.8.D SS1_22Na sample

The final ANN struggles to correctly identify the ^{22}Na radionuclide present in the sample. From Fig. 5.10a it is possible to observe that the identifying peaks are quite difficult to distinguish from the noise.

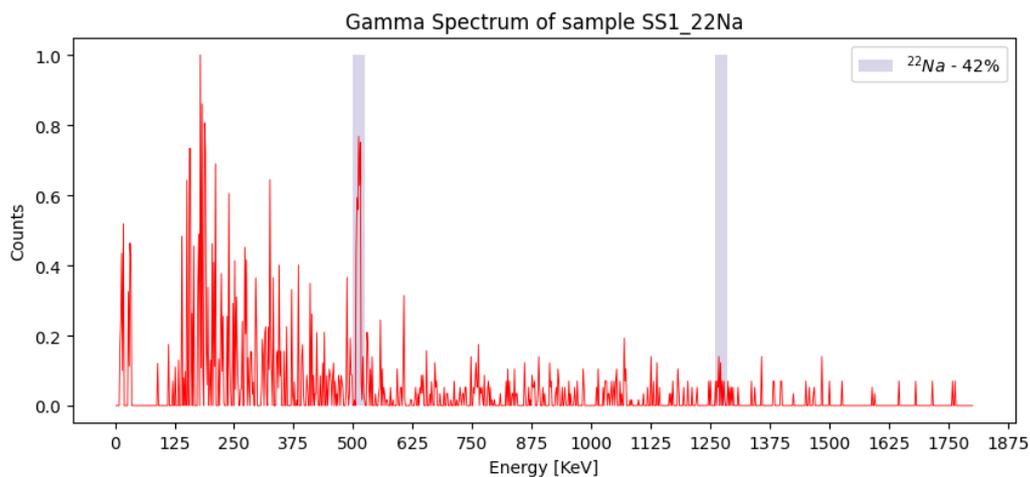


Figure 5.14: *SS1_22Na* spectra with classification peaks.

This difficulty is probably the reason for the lower confidence level of only 42%. Although the 1274 keV photo peak is extremely difficult to recognise, the annihilation peak of 511 keV is quite visible.

5.2.8.E SS2_60Co#1 and SS2_60Co#2 samples

Regarding the samples of *SS2*, all of the acquisitions were made equally, meaning that the disparity between the classification results for samples *SS2_60Co#1* and *SS2_60Co#2* must be analysed.

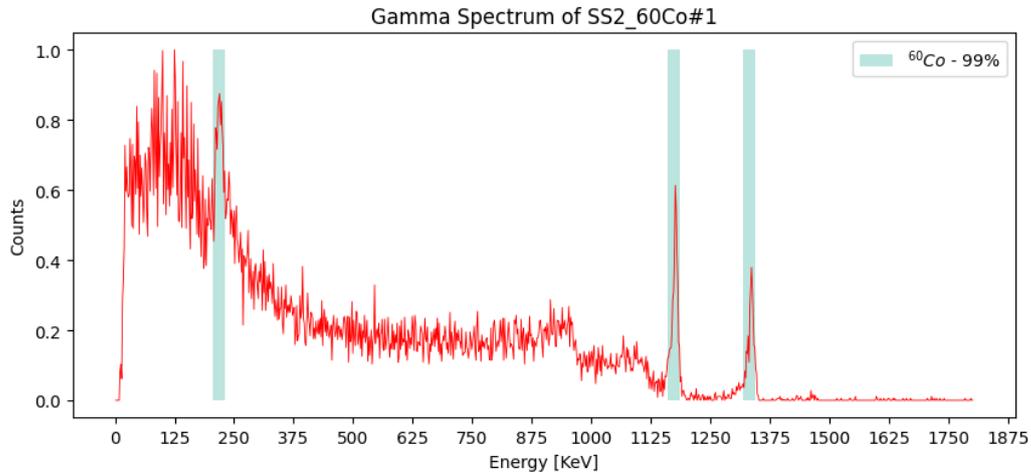


Figure 5.15: *SS2_60Co#1* spectra with classification peaks.

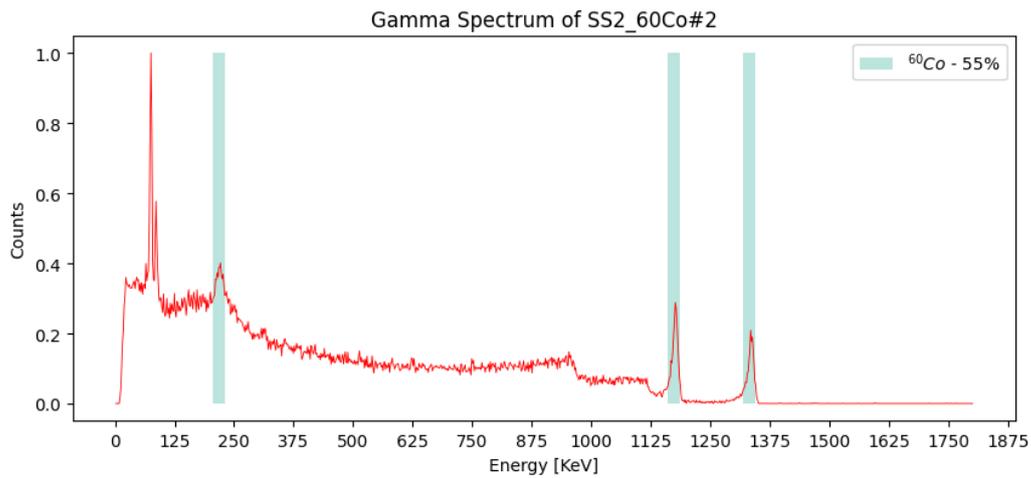


Figure 5.16: *SS2_60Co#2* spectra with classification peaks.

From Fig. 5.15 and Fig. 5.16 we can see that the spectrum from sample *SS2_60Co#2* contains a peak in the 70-80 keV region. This is the characteristic X-Ray peak from the lead casing that is covering the source. It is possible that this extra peak can cause some confusion to the ANN, leading to a worse classification confidence in the sample that is encased in lead.

5.2.8.F *SS2_241Am* sample

Finally, the classification result for sample *SS2_241Am* is correct but with a relatively low confidence.

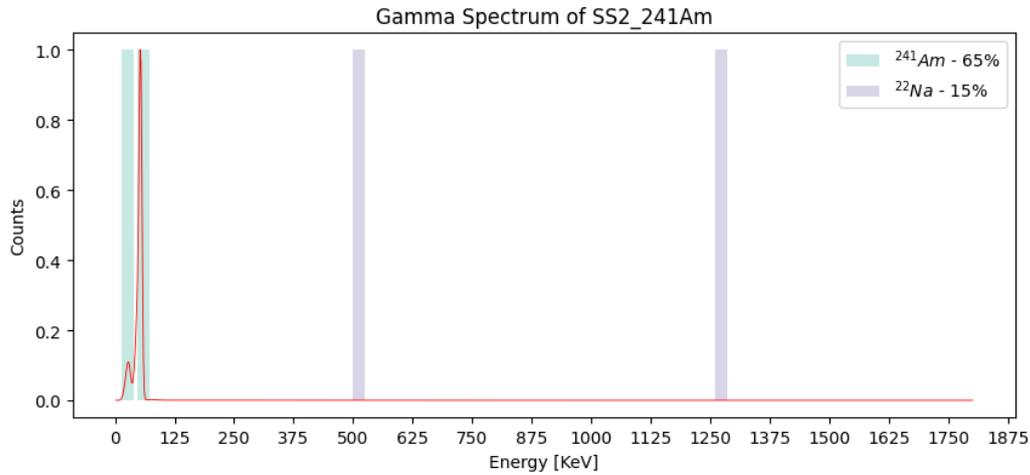


Figure 5.17: SS2_241Am spectra with classification peaks.

This spectrum is really well defined so a possible reason for this confidence value is related with the fact that the photo peaks of ^{241}Am are encountered in the lower energetic region of the spectra. Peaks in this area can be mistakenly considered as noise by the ANN. Consequently, the ANN might try to identify another isotope in the sample. This is likely the case of misidentification of ^{22}Na in this sample, since its photo peaks are far away from any relevant structure in the spectrum as seen in Fig. 5.17.

In this case, signalling relevant peaks enables us to quickly discard ^{22}Na from the possible isotopes by visually understanding that it is most likely not present.

5.3 Multi Radionuclide Classification

For multi radionuclide classification the same principles used in the previous section can be applied. One of the major difficulties in this case is dealing with such a huge number of data samples (more than 400 thousand). With the intention of reducing the implementation and optimisation process, it was assumed that the same ANN structure analysis made for single isotope could be applied to the multi radionuclide ANN.

5.3.1 ANN Results and Discussion

The parameters chosen for regularisation and early stopping criteria were the same as the final ANN for the previous section (Table 5.12).

Learning Parameters	
Early stopping	10^{-4} , patience: 10 epochs
Regularisation	$L1 : 0.001, L2 : 0.001$

Table 5.12: Learning parameters for multi isotope identification.

Based on the results of the previous section, the preferred ANN structure was using only one or two hidden layers. Softmax activation function can no longer be used for this classification since more than one isotope can be identified, meaning that the output values of the ANN do not sum up to 1.

ANN Codename	Final Epochs	Final Loss (10^{-2})
Sig1 - Bs: 32	66	0.14
Sig1 - Bs: 64	94	0.15
Sig1 - Bs: 128	117	0.19
Sig2 - Bs: 32	403	1.49
Sig2 - Bs: 64	424	1.60
Sig2 - Bs: 128	562	1.75

Table 5.13: Final number of epochs and loss value for each of the ANNs tested.

Both *Sig1* and *Sig2* were trained using the whole single and multi data-set for batch sizes of 32, 64 and 128. In Table. 5.13 is presented the total number of iterations as well as the final loss values for each training session.

Similarly to single isotope identification, evaluating the resulting ANN using the training data set resulted in a perfect score, obtaining an F1-score value equal to 1. However, this was not expected and hints that the ANN might be over-fitting.

The performance analysis on the evaluation data-set started with only the single isotope samples, since a multi isotope classification solution should perform adequately with single identification as well.

The results shown on Fig. 5.18 present a low F1-score value for every ANN structure tested. Note that the precision metric shows an overall high number, suggesting that the ANN is capable of identifying the present radionuclides. However, the recall values are low due to the huge number of FPs that the networks output.

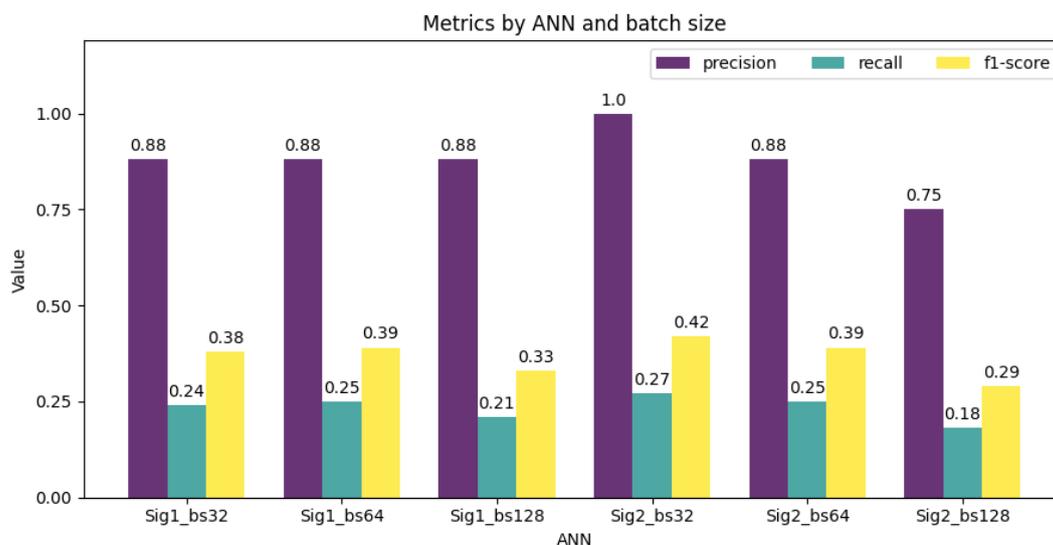
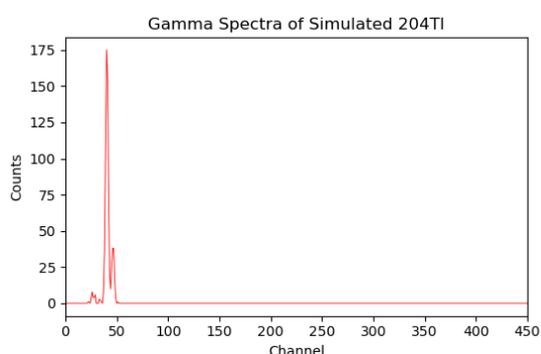


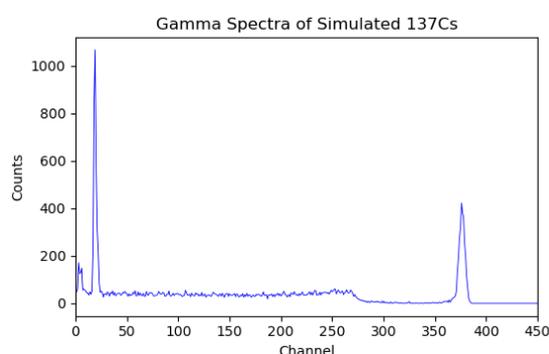
Figure 5.18: Precision, Recall and F1-score values for each ANN configuration using single isotope evaluation data-set.

After testing several other configurations and even different and newly simulated data-sets, the values remained closely identical, high precision but low recall. These results might be related to how the expected output is fed to the ANN during training.

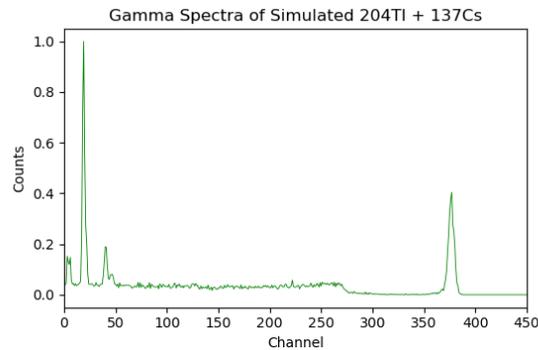
The expected output fed to the network is an array containing 0s and 1s. The 0 value indicates that the isotope is not present in the spectra and the 1 value indicates that the isotope is present. While in single identification this is an adequate approach, it was concluded that for multi isotope identification, representing the existing radionuclides with a simple 1 value might not be ideal. The problem arises when one of the present isotopes contributes with many more counts than the other, making the second isotope almost unnoticeable. When this happens, a spectrum with only the more active radionuclide might look nearly identical to a spectrum containing both radionuclides.



(a) Simulated Spectrum of ^{204}Tl .



(b) Simulated Spectrum of ^{137}Cs .



(c) Normalised simulated Spectrum of $^{204}\text{Tl} + ^{137}\text{Cs}$.

Figure 5.19: Comparison between Simulated Spectra from ^{204}Tl and ^{137}Cs . ([0,450] Channels). The number of counts is normalised to the maximum value in the later plot.

In Fig. 5.19 is shown the spectrum of ^{204}Tl on the left and ^{137}Cs on the right, both simulated by GADRAS before any normalisation is applied. We can see that for the same $1\mu\text{Ci}$, the total counts of each spectra is completely different. When merged into the same spectra and then normalised to the maximum number of counts (Fig. 5.19c), the features from ^{204}Tl are almost completely masked. With that being said, defining the expected output containing equal confidence values for both isotopes can possibly lead the ANN to output more false positives, since the ANN is supposed to learn something that is not as noticeable. This makes the network look for smaller variations in the spectra, possibly mistaking noise for a radionuclide and leading to incorrect classifications.

Additionally, for multi isotope identification, the training data labels should take into account the relative contribution of each of the present radionuclides in the sample. A similar approach can be found in [38], where the number of counts that each radionuclide contributes to the spectra is controlled and taken into consideration. Another possibility would be to normalise each individual spectra and then sum the normalised counts of each one, creating a new spectra with well defined features from each of the desired radionuclides.

Nonetheless, an implementation like the one presented in this section could still be possible if the total training samples included several distinct activity contribution ratios for each possible combination of isotopes. Despite this, such approach would not be advisable since the number of necessary training samples would significantly increase and the probability of obtaining a well performing ANN would be small.

6

Conclusion

Contents

6.1 Final Remarks	64
6.2 Future Work	65

6.1 Final Remarks

The work presented in this thesis provide a preliminary indication that ANN might be a promising solution for radionuclide identification. Although there is a need for further analysis and testing, especially due to the dimensions of the relatively small evaluation data-set available, the results provide a positive insight on this subject revealing some potential. Both the flexibility, as well as the accessibility of use by non expert users of the ANN makes this type of solution particularly interesting. All in all, radionuclide identification based on machine learning algorithms such as artificial neural networks should be further explored.

6.1.1 Contributions

The following contributions were provided:

- Creation of an ANN capable of identifying single isotope spectrum samples with relative high degree of confidence. This ANN was tested using real data acquisitions. The results show the ANN provides fine classification outputs even for small acquisitions times and could be possibly applied to real-time identification systems.
- It was shown that using Softmax activation function in the output layer resulted in a better classification as opposed to Sigmoid activation which presented worse results. The best performing network was the one containing a single hidden layer, since fewer hidden layers also resulted in better classification outputs.
- Despite providing relevant outputs for single isotope samples, the results for multi isotope classification were not so conclusive and additional work is required. It was possible to understand the importance of data-set labelling since this is the possible reason for such results. More samples need to be acquired since even if results were better, it would be difficult to correctly validate the performance of the ANN due to the lack of real data samples for multiple isotopes.
- A proof-of-concept was developed using a Raspberry Pi and CZT detector. This real world application presented high portability and flexibility, being perfect for real-time acquisitions either on foot or using unmanned vehicles.
- Development of detailed data-set generation explanation using GADRAS-DRF software. This description can be used to create different data-sets for other detectors, making this approach easily scalable.
- Paper ready for submission: "**Radioactive Hot-spot Localization and Identification using Deep Learning**" - Sensors, 2021. *Mendes, F; Barros, M; Vale, A.; Gonçalves, B.*

6.2 Future Work

Several possible enhancements and upgrades came up while making this thesis. Some future ideas include:

- Explore the labelling of the training samples for multi radionuclide identification applications. Training data labels should take into account the relative contribution of each of the present radionuclides in the sample. The same could be applied to generate more samples with various types of background contribution.
- Merge existing algorithms with this application. Using filtering techniques to filter some of the noise of the acquired spectrum could enhance the classification capabilities of the developed ANN. Nonetheless, it is worth mentioning that for a real time analysis, the pre-processing of the data should always be simple enough to be implemented in real time.
- Acquiring more real data would be ideal for upgrading the performance evaluation of the ANN and even contribute to possible optimisations. If big enough, some of the real data-set could be used to train the ANN.
- False positive occurrences should also be further analysed. Possibility of more complex threshold systems or even other ANN structures can be explored.

Bibliography

- [1] “¹³⁷Cs Spectrum image,” <http://www.people.vcu.edu>, Accessed: 2020-01-10.
- [2] *LAB-26 Analysis of γ spectrum*, LINKÖPINGS UNIVERSITET - IFM – The Department of Physics, Chemistry and Biology.
- [3] “Pulse-height spectrometry,” <https://radiologykey.com>, Accessed: 2020-12-08.
- [4] G. R. Gilmore, *Practical Gamma-ray Spectrometry*, 2008.
- [5] S. SHARMA, “Activation Functions in Neural Networks.”
- [6] K. Sarkar, “Relu : Not a differentiable function: Why used in gradient based optimization? and other generalizations of relu.” , Accessed: 2020-01-10.
- [7] “ μ SPEC from Ritec,” <http://www.ritec.lv>, Accessed: 2020-01-10.
- [8] “Precision and Recall,” https://en.wikipedia.org/wiki/Precision_and_recall, Accessed: 2020-01-10.
- [9] “Naturally-Occurring Radioactive Materials (NORM),” <https://www.world-nuclear.org>, Accessed: 2019-12-29.
- [10] “FRIENDS Project,” <https://www.ipfn.tecnico.ulisboa.pt>, Accessed: 2019-12-26.
- [11] “IPFN - Instituto de Plasmas e Fusão Nuclear,” <https://www.ipfn.tecnico.ulisboa.pt>, Accessed: 2019-12-26.
- [12] “MARIA - Mobile Application for Radiation Intensity Assessment,” <https://www.ipfn.tecnico.ulisboa.pt>, Accessed: 2019-12-26.
- [13] Arpansa, Australian Government, “Radioactivity,” <https://www.arpansa.gov.au>, Accessed: 2020-12-08.
- [14] “Alpha decay,” https://en.wikipedia.org/wiki/Alpha_decay, Accessed: 2020-12-08.
- [15] “Beta decay,” https://en.wikipedia.org/wiki/Beta_decay, Accessed: 2020-12-08.

- [16] *Identification of Unknown Radionuclides*, TECHNICAL UNIVERSITY DRESDEN - Institute of Power Engineering Training Reactor.
- [17] R Core Team, *Radioisotope Identification Devices (RIIDs)*, U.S. Homeland Security - Science and Technology, 2009.
- [18] Nuclear Power, "Gamma Spectroscopy," <https://www.nuclear-power.net>, Accessed: 2020-12-08.
- [19] M. Monterial, K. E. Nelson, S. E. Labov, and S. Sangiorgio, "Benchmarking algorithm for radio nuclide identification (barni) literature review," 2 2019.
- [20] M.-H. Zhu, L. Liu, Z. You, and A. Xu, "Least-squares fitting of gamma-ray spectra with b-spline basis functions," vol. 1, 06 2008, pp. 691 – 695.
- [21] M. Deriche and M. Raad, "Real-time implementation of a deconvolution filter for gamma ray spectroscopy," *Instruments and Experimental Techniques*, vol. 54, pp. 30–36, 01 2011.
- [22] A. Abdel-Hafiez, "Fourier transformation methods in the field of gamma spectrometry," *Pramana*, vol. 67, pp. 457–466, 09 2006.
- [23] C. J. Sullivan, S. E. Garner, and K. B. Butterfield, "Wavelet analysis of gamma-ray spectra," in *IEEE Symposium Conference Record Nuclear Science 2004.*, vol. 1, 2004, pp. 281–286 Vol. 1.
- [24] G. Yu, J. Gu, L. Hou, Z. Li, Y. Wang, and Y. Zhang, "Application of wavelet transform in -ray spectra analysis," *Science China Physics, Mechanics and Astronomy*, vol. 56, 09 2013.
- [25] L. Meng and D. Ramsden, "An inter-comparison of three spectral-deconvolution algorithms for gamma-ray spectroscopy," *Nuclear Science, IEEE Transactions on*, vol. 47, pp. 1329 – 1336, 09 2000.
- [26] H. T. T. Aarnio, P A and J. T. Routti, "Expert system for nuclide identification and interpretation of gamma-spectrum analysis," Jul 1992.
- [27] J. V. Candy, "Detection, classification and estimation of radioactive contraband from uncertain low-count measurements," 2010.
- [28] C. Sullivan and J. Stinnett, "Validation of a bayesian-based isotope identification algorithm," *Nuclear Instruments and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment*, vol. 784, 12 2014.
- [29] J. B. Stinnett, "Automated isotope identification algorithms for low-resolution gamma spectrometers," Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, Illinois, 2016.

- [30] D. Boardman, M. Reinhard, and A. Flynn, "Principal component analysis of gamma-ray spectra for radiation portal monitors," *IEEE Transactions on Nuclear Science*, vol. 59, no. 1, pp. 154–160, 2012.
- [31] R. C. Runkle, M. F. Tardiff, K. K. Anderson, D. K. Carlson, and L. E. Smith, "Analysis of spectroscopic radiation portal monitor data using principal components analysis," *IEEE Transactions on Nuclear Science*, vol. 53, no. 3, pp. 1418–1423, 2006.
- [32] M. Alamaniotis, S. Lee, and T. Jevremovic, "Intelligent analysis of low-count scintillation spectra using support vector regression and fuzzy logic," *Nuclear technology*, vol. 191, 07 2015.
- [33] H. Hata and K. e. a. Yokoyama, "Application of support vector machine to rapid classification of uranium waste drums using low-resolution -ray spectra," *Applied radiation and isotopes: including data, instrumentation and methods for use in agriculture, industry and medicine*, vol. 104, pp. 143–146, 06 2015.
- [34] e. a. T. B. Gosnell, J. M. Hall, "Gamma-ray identification of nuclear weapon materials," Lawrence Livermore National Laboratory, Tech. Rep. W-7405-ENG-48, 1997.
- [35] M. Hogan, S. Yamamoto, and D. Covell, "Multiple linear regression analysis of scintillation gamma-ray spectra: Automatic candidate selection," *Nuclear Instruments and Methods*, vol. 80, no. 1, pp. 61 – 68, 1970. [Online]. Available:
- [36] "YOLO: Real-Time Object Detection," <https://pjreddie.com/darknet/yolo/>, Accessed: 2020-01-10.
- [37] D. Liang and P. G. et al., "Rapid nuclide identification algorithm based on convolutional neural network," *Annals of Nuclear Energy*, vol. 133, pp. 483 – 490, 2019. [Online]. Available:
- [38] M. Kamuda and C. J. Sullivan, "An automated isotope identification and quantification algorithm for isotope mixtures in low-resolution gamma-ray spectra," *Radiation Physics and Chemistry*, vol. 155, pp. 281 – 286, 2019, iRRMA-10. [Online]. Available:
- [39] L. Chen and Y.-X. Wei, "Nuclide identification algorithm based on k-l transform and neural networks," *Nuclear Instruments Methods in Physics Research Section A-accelerators Spectrometers Detectors and Associated Equipment - NUCL INSTRUM METH PHYS RES A*, vol. 598, pp. 450–453, 01 2009.
- [40] M. Kamuda, J. Stinnett, and C. Sullivan, "Automated isotope identification algorithm using artificial neural networks," *IEEE Transactions on Nuclear Science*, vol. PP, pp. 1–1, 04 2017.
- [41] "Artificial Intelligence Focus," <https://searchenterpriseai.techtarget.com/>, Accessed: 2020-01-10.
- [42] "Artificial Intelligence - Applications," <https://www.edureka.co>, Accessed: 2020-01-10.

- [43] Wikipedia, "Artificial Intelligence," https://en.wikipedia.org/wiki/Artificial_intelligence, Accessed: 2020-01-10.
- [44] K. Demertzis, "Machine learning," 06 2019.
- [45] Wikipedia, "Machine Learning," https://en.wikipedia.org/wiki/Machine_learning, Accessed: 2020-01-10.
- [46] "14 Different Types of Learning in Machine Learning," <https://machinelearningmastery.com>, Accessed: 2020-01-10.
- [47] "ML — Types of Learning – Supervised Learning," <https://www.geeksforgeeks.org>, Accessed: 2020-01-10.
- [48] "What is an artificial neural network?" <https://www.digitaltrends.com>, Accessed: 2020-01-10.
- [49] "Common Loss functions in machine learning," <https://towardsdatascience.com>, Accessed: 2020-01-10.
- [50] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.
- [51] "A Deeper Look into Gradient Based Learning for Neural Networks," <https://towardsdatascience.com>, Accessed: 2020-01-10.
- [52] "Softmax function," https://en.wikipedia.org/wiki/Softmax_function, Accessed: 2020-01-10.
- [53] D. Becker, "Rectified Linear Units (ReLU) in Deep Learning," <https://www.kaggle.com/dansbecker>, Accessed: 2020-01-10.
- [54] "NEA - Nuclear Energy Agency," <https://www.oecd-nea.org/>, note = Accessed: 2020-06-06.
- [55] *GADRAS-DRF 18.6 User's Manual*. Albuquerque, New Mexico 87185 and Livermore, California 94550: Sandia National Laboratories, 2016.
- [56] G. Thoreson and L. Harding, "Inject calculations with gadras," Sandia National Laboratories, Tech. Rep., 2016.
- [57] S. S. e. a. Ron Jeffcoat, "A comparison of gadras-simulated and measured gamma-ray spectra," Savannah River National Laboratory.
- [58] D. M. e. a. Chuck Rhykerd, "Benchmarks for gadras performance validation," Sandia National Laboratories, Tech. Rep., 2009.

- [59] *Gamma-Radiation CdZnTe Microspectrometer with exchangeable detector modules μ SPEC*, ZRF RITEC SIA, Gustava Zemgala gatve 71A.
- [60] Jorge Borbinha, Yuri Romanets et al., "Performance analysis of geiger–müller and cadmium zinc telluride sensors envisaging airborne radiological monitoring in norm sites," *Sensors*, vol. 20, p. 1538, 03 2020.
- [61] Wikipedia, "Radionuclide," <https://en.wikipedia.org/wiki/Radionuclide>, Accessed: 2020-01-10.
- [62] G. G. Thoreson, *PCF File Format*, Sandia National Laboratories, 2017.
- [63] "Numpy," <https://numpy.org/>, Accessed: 2020-01-10.
- [64] "TensorFlow," <https://www.tensorflow.org/>, Accessed: 2020-01-10.
- [65] "Keras," <https://keras.io/>, Accessed: 2020-01-10.
- [66] J. Brownlee, "A Gentle Introduction to Mini-Batch Gradient Descent and How to Configure Batch Size," <https://machinelearningmastery.com>, Accessed: 2020-01-10.
- [67] "Multi Channel Analyzer MCA527 micro / micro+," <https://www.gbs-elektronik.de>, Accessed: 2020-01-10.
- [68] "Raspberry Pi website," <https://www.raspberrypi.org/>, Accessed: 2020-01-10.
- [69] *Description of the MCA527 Firmware Commands*, GBS Elektronik GmbH, 2020.
- [70] "Real VNC Software," <https://www.realvnc.com>, Accessed: 2020-01-10.
- [71] B. Shmueli, "Multi-Class Metrics Made Simple, Part II: the F1-score," <https://towardsdatascience.com>, Accessed: 2020-01-10.
- [72] P. Gupta, "Regularization in Machine Learning," <https://towardsdatascience.com>, Accessed: 2020-01-10.