Journal of Radioanalytical and Nuclear Chemistry (2023) 332:3421–3429
https://doi.org/10.1007/s10967-022-08568-8

Artificial neural networks for NAA: proof of concept on data analysed with k0-based software

N. Pessoa Barradas[1] · N. Farjallah[1] · A. Vieira[2] · M. Blaauw[3]

Received: 16 August 2022 / Accepted: 16 September 2022 / Published online: 1 October 2022

Abstract
Artificial intelligence methods such as artificial neural networks, Bayesian networks, genetic algorithms, and others, have shown great potential for application, not only as classification schemes, but also in numerical data analysis. In this work, we explore how, from a limited number of spectra (around 200), an ANN could be efficiently developed, using data augmentation techniques and optimized architecture, and used to analyse neutron activation analysis (NAA) data. The IAEA Collaborating Centre Research Institute Delft (RID), Netherlands, has collected NAA data sets consisting of one single spectrum per sample to determine one single element (selenium), with addition of a marker (caesium) for flux normalization, all irradiated and measured the exact same way and analysed with k_0-based software. The problem studied is one of the simplest that can be addressed with NAA; therefore the present work is intended merely as proof of concept that ANNs can perform well in NAA data analysis of simple problems. We present the results and discuss how to extend the present work to more demanding problems in NAA.

Keywords Neutron activation analysis · Artificial neural networks · Nuclear analytical techniques · Artificial Intelligence

Introduction

Neutron activation analysis (NAA) is the technique most commonly found in research reactors, with half of the 223 operating research reactors reporting activity in this application [1]. The level of activity varies widely among laboratories, with some analysing a limited number of samples, while others may analyse many hundreds or even thousands of samples per year. The International Atomic Energy Agency supports NAA laboratories through different activities, including the organization of annual proficiency test exercises [2], coordinated research projects (e.g. promoting automation of the NAA process, including process management, experiments and data processing [3], or the establishment of Large Sample NAA [4]) and quality assurance and quality control [5].

Data analysis is one of the crucial steps in the NAA process, and how it is done depends on the calibration method employed. This can be, for instance, absolute calibration, the relative (also named comparator) method, single comparator method, or the k_0 method. Single comparator methods such as the k_0 method allow for panoramic analysis without using multi-element standards [6]. The k_0-method in particular offers efficient methods for irradiation facility and detector characterization, as well as for necessary corrections incurred by sample geometry, counting geometry and sample matrix effects, such as true-coincidence summing corrections, neutron self-shielding and gamma self-absorption.

Nevertheless, the data analysis process can be complex and may require a great deal of expertise by the user. Therefore it would be desirable to have a method that would allow reliable, push-button, fast data analysis without requirement of specific knowledge by the analyst. Artificial neural networks (ANNs) are a flexible scheme capable, among other things, of approximating an arbitrary function via training with an appropriate data set. The programming of ANNs does not include any knowledge of the physics that applies to the system to be studied; in the case of NAA; this means that even the law of radioactive decay is not explicitly used. Instead, the ANNs learn by example:

✉ N. Pessoa Barradas
N.Pessoa-Barradas@iaea.org

[1] International Atomic Energy Agency, Vienna, Austria

[2] Hazy, London, UK

[3] Reactor Institute Delft, Delft University of Technology, Delft, Netherlands

in an initial training phase, they are trained with available examples, for instance sets of data (e.g. spectrum yields, irradiation, decay and counting times, calibration parameters, etc., named "ANN inputs") previously analysed with conventional analytical methods, for which the results (e.g. elemental amounts, named "ANN outputs") are known. In supervised training, a given ANN is optimized such that the inputs lead to calculated outputs as close as possible to the examples given. This often requires the availability of a database with thousands (or more) of datasets. The trained ANN is then ready to be applied to new datasets. Once trained, the analysis is nearly instantaneous.

We have in the past successfully applied ANNs to other analytical techniques, such as Rutherford backscattering [7], elastic backscattering [8], elastic recoil detection analysis [9], and even to automated determination of optimal experimental parameters for a given sample [10]. The problems studied ranged from the very simple to the very complex, including multilayered films with multiple elements [9]. ANNs have also been applied to analysis of X-ray [11] and gamma ray spectra [12]. ANNs have been applied to NAA by Roshani et al., using 29 spectra constructed with a Monte Carlo code, without testing to real experimental data [13]. Medhat [14] used a dataset consisting of eight samples, which is usually not sufficient to train an ANN sufficiently. Artificial neural networks have been successfully applied to prompt gamma analysis, either by using large extensive databases of experimental results [15] or by using a large number of simulated spectra – 4900 in the work by Bilton et al. [16], 1478 in the work by Hossny et al. [17]. They have also been applied to unfolding neutron spectra measured by means of a Bonner Sphere Spectrometer set [18], using up to 52 spectra.

In this work, we took advantage of an NAA data set collected and analysed at the IAEA Collaborating Centre Research Institute Delft (RID), Netherlands, to establish a methodology for development of an ANN applied to NAA, even when only a limited amount of data is available. The data set consists of 216 single measurements per sample to determine one single element (Se) from a short-lived isotope, with addition of a marker (Cs) for flux normalization. All samples were irradiated and measured the exact same way and analysed with k_0-based software. In this work, we show how an ANN could be efficiently developed using data augmentation techniques and optimized architecture. The problem studied is one of the simplest that can be addressed with NAA and is intended as a proof of concept that ANNs can perform well in NAA data analysis of simple problems. The potential of extending this work to more demanding problems, and possible methodologies for doing so, is discussed. The code and data used in this work are publicly available.

Experimental data

216 toenail samples were weighed into 10 mm tall, 9 mm diameter high-density polyethylene capsules. Caesium flux monitor samples were prepared by gravimetrically pipetting 50 µl of Cs solution containing 60 µg of caesium onto filter paper in 5 mm tall and 9 mm diameter polyethylene capsules. The toenail masses were in the 30–200 mg range.

Flux monitor and sample were stacked inside a polyethylene rabbit and irradiated together in the CAFIA facility at RID [19], at a neutron fluence rate of about 5×10^{16} m^{-2} s^{-1}, and measured together, rabbit included, at a 2 cm distance from a horizontal-looking 40% relative efficiency Ge detector. Each sample was irradiated for 17 s and measured after 3-s decay time, then the next cycle restarted after a 3-s wait and was repeated 7 times in a cyclic protocol, where the counts from all 7 measurements were accumulated in a single spectrum. This protocol yields ^{134m}Cs 127 keV peak areas of about 2×10^4 with relative uncertainties below 1%, and a limit of detection (LOD) for selenium of about 0.02 mg kg^{-1}.

The spectra were acquired with an 8192 channel Ortec DSPEC digital spectrometer with zero-dead time counting enabled, at dead times of about 15%.

The measured spectra were converted to lists of peak energies and areas using the in-house RID software [20], the neutron fluence rate determined using the ^{134m}Cs peak at 127 keV, and then the elemental mass fraction, again using the in-house RID software [21]. All detected peaks were interpreted, but only the result for selenium was reported and used for the purpose of the current study. The observed selenium concentrations were typically near 0.5 mg kg^{-1}, with absolute uncertainties of about 0.02 mg kg^{-1}. The system was specifically calibrated for these analyses, eliminating the need for an efficiency curve, the dependency on nuclear data et cetera. Timing errors were considered to be negligible, even in view of the 17.45 s halflife of ^{77m}Se. The counting statistics of the ^{77m}Se peak at 161.8 keV are the remaining, dominant source of uncertainty by far.

Artificial neural networks

ANN architecture

There is a wide range of different types of ANNs. Most modern ANNs architectures are based on transformers [22] that have enabled major breakthroughs, for instance in natural language processing [23] and computer vision [24]. Transformers have become the dominant network

architectures, replacing many previous models like Convolutional Neural Networks (CNN), Long Short Term Memory Networks (LSTM) and Multilayer Perceptrons (MLP) [25, 26]. The transformer architecture is based on two concepts: a) recurrent-free architecture which computes the representations for each individual token in parallel, and b) multi-head self-attention blocks which aggregate spatial information across tokens.

Despite their success, recently [27] Liu et al. showed that a modular structure of an MLP, consisting of a channel projection and spatial projections with static parameterization, is a competitive alternative to transformers without the need for self-attention and having a much lighter design.

For an analytical problem as the one at hand, we have used a similar architecture consisting of a modular multilayered deep neural network [28]. As inputs we used the channel yields and experimental parameters, which in this case were the Cs comparator mass, the sample mass, and the fill height of the sample during the irradiation. As outputs we consider three parameters of interest, namely the Se amount, its uncertainty and limit of detection.

Deep neural networks have a number of intermediate processing nodes, with typically hundreds of thousands, or even millions, of connections or parameters, that have to be learned. Normally, learning of these parameters is performed through a gradient descendent technique by presenting to the ANN thousands of training examples, in which both the inputs and outputs are given, and the connection parameters are adjusted through backpropagation until optimum agreement is found between the calculated outputs and the actual values. In this case we use the metric Root Mean Square Error (RMSE) as the loss function.

The architecture of the network is shown in Fig. 1. It consists of blocks of fully connected layers at the input layer that

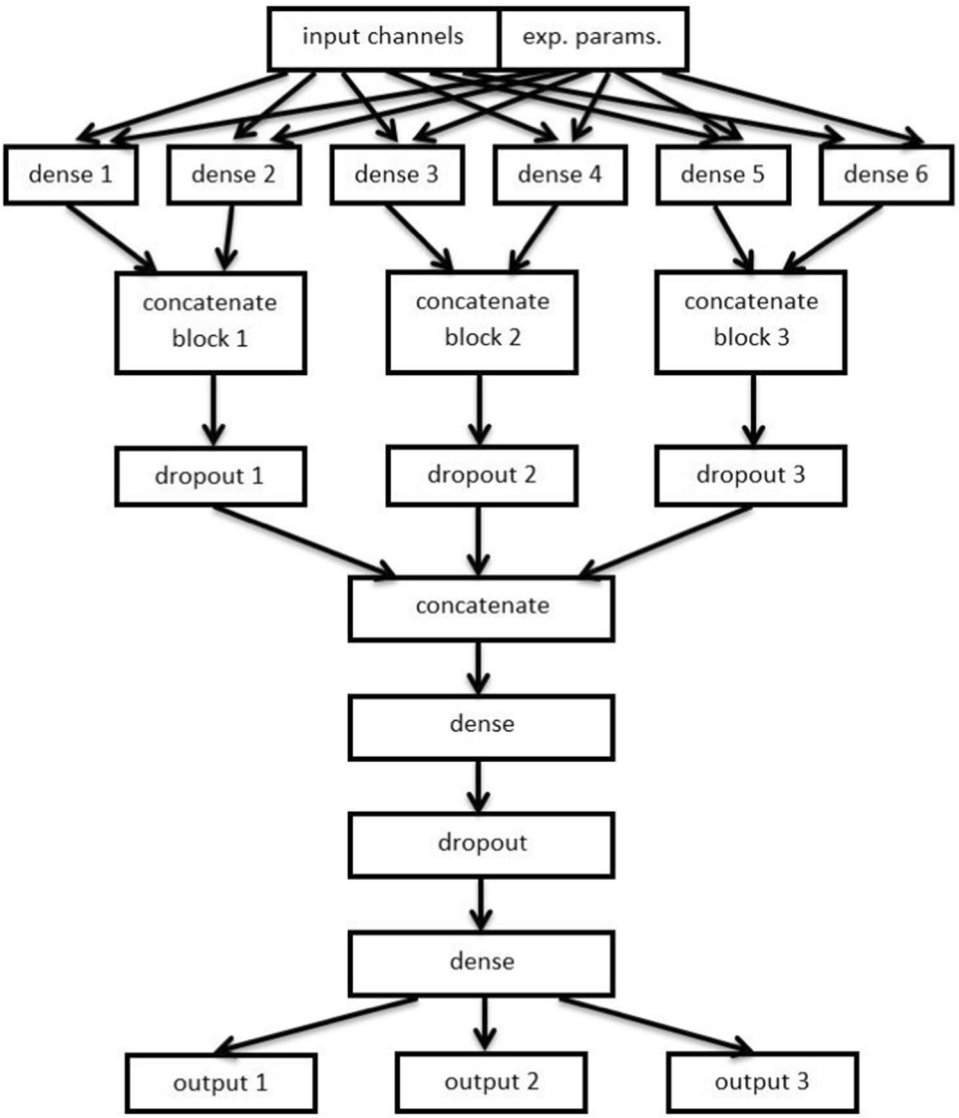

Fig. 1 ANN architecture used in this work. It consists of a set of 3 blocks that are concatenated and then fed to an intermediate layer

are then concatenated and followed by a number of intermediate layers. Since each spectrum has $Nc = 8192$ channels (or 414 after data compression as explained below), if we haven't used a modular approach, we would end up with a very large ANN that would overfit on the small amount of data available. In the figure, the first (top) layer is the spectrum and the relevant parameters (Cs comparator mass, the sample mass, and the fill height of the sample during the irradiation).

In order to facilitate the training, we used a modular approach to the network, where the input layer is sliced first into six blocks (layer 2 from the top) and then into $K = 3$ blocks (layer 3), each of size $S = Nc/K$, allowing for initial pre-processing of each block separately, which contain only a limited number of peaks. The comparator mass, the sample mass and the fill height were concatenated to each of the blocks. In the following layers, the blocks are then connected to each other. The "dense" layers enhance the deep processing capability of the model. The "dropout layers", in which connection nodes are randomly eliminated, are used to reduce overfitting (i.e. over specialization of the ANN to the cases used in the training) [29]. The three outputs are the Se amount, its uncertainty and LOD.

For the activation function we used the gelu [30], and the optimization method was the Adam [31]. We also applied Dropout of 0.1 to avoid overfitting. After the first stage, two intermediate layers with 30 and 10 neurons are added before the outputs. The loss function to minimize had weights (1.0, 0.1 and 0.2), corresponding to the error on the Se content, uncertainty and LOD. This was driven by the knowledge that the Se content can be determined from the data with much higher accuracy than the other two parameters. Giving slightly smaller or higher weights to the uncertainty and detection limit did not significantly change the results.

The model was implemented using Tensorflow Keras version 2.2 [32]. All values were normalised to values between 0 and 1. From the original 216 points, we use the first 173 points for training and the remaining 43 for testing. In order to deal with the large dynamic range of counts, we applied the following transformation:

$$y = y_0 + log\left(y - y_0 + 1\right), y > y_0 = 10000 \quad (1)$$

However, even with these optimizations, a small amount of training data can result in overfitting, where the ANN can reproduce the data used in the training, but performs poorly when confronted with new data (the so-called test data). The high dimensionality of the data as in the present case makes this risk even more likely. Moreover, the gradient descent algorithm may become unstable when only a small amount of data are available.

Therefore, we used a limited degree of knowledge about the technique to design a learning strategy to circumvent these difficulties, as detailed below. This consisted in data compression, data augmentation and removal of outliers.

Data compression

Data compression was achieved by summing a number of adjacent channels. This has the advantage of reducing the number of input parameters and therefore reducing the complexity of the ANN and making its training easier and more effective. It has the disadvantage that it can lead to worse discrimination of neighbouring gamma lines. We chose a compression of five, i.e. channels were added five by five. Given the FWHM energy resolution of the detector and the energy width of the original channels, the resulting compressed channels have an energy width very close to one FWHM, which should be sufficient in many cases. This leads to a reduction from 8192–1638 channels.

Furthermore, a simple peak recognition routine was implemented using the average spectrum, allowing easier identification of small peaks or peaks only present in some of the spectra. Only channels in the vicinity of the 72 peaks that could be identified were used by the ANN. This led to a further reduction to 414 channels.

Data augmentation

Data augmentation was achieved by generating 10,000 synthetic spectra by doing linear combinations of the real data. The spectra yields were combined linearly:

$$Y_t = a_1 Y_1 + a_2 Y_2, \quad (2)$$

where Y_t is the array of channel contents for the synthetic spectrum, and Y_1 and Y_2 are the channel contents corresponding to the same energy for the two selected spectra. Each spectrum is given a weight, a_1 and a_2 such that

$$a_1 + a_2 = 1. \quad (3)$$

The spectra were selected randomly, but with higher probability in parameter regions where the data were very sparse. As an example, very few data points had high Se content, so the density of data points for high Se values was increased in the synthetic data, which aids the ANN learning process.

The Se content, its uncertainty and LOD for each synthetic spectrum were determined from the corresponding values for the two selected spectra. Some degree of extrapolation was introduced, whereby these parameters could be 20% outside the range in the real data, which also aids the ANN learning process.

The distribution of Se in the original and in the synthetic data is shown in Fig. 2. The extension of the Se range on the extremes of the distribution, as well as the increased density

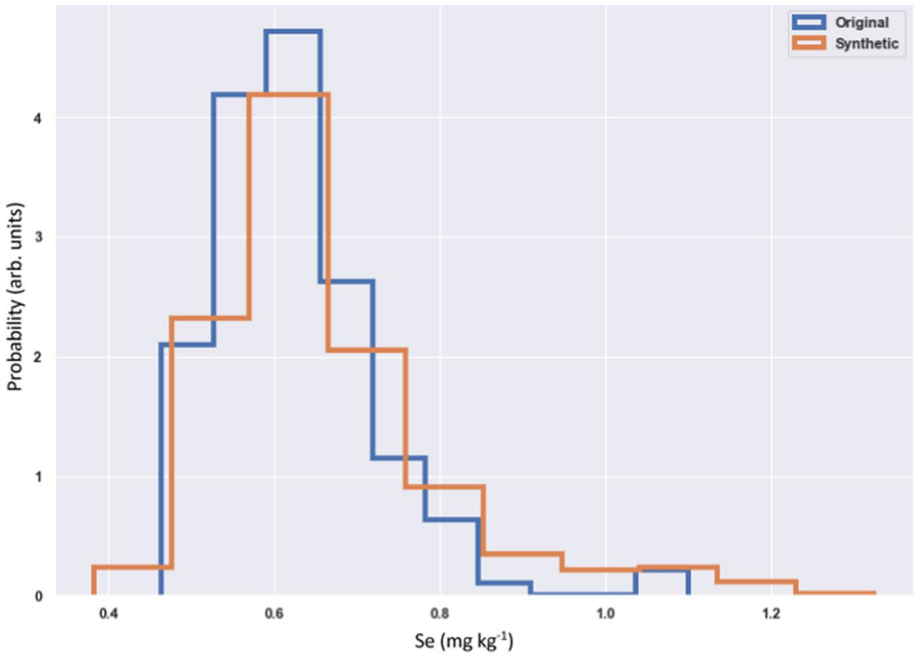

Fig. 2 Distribution of Se in the original and synthetic data

of data points in the intermediate region between 0.85 and 1.05 mg kg^{-1} are apparent.

The validity of the synthetic data was checked. A representative sample of the synthetic spectra (200 spectra) was analysed with the in-house RID software, which yielded values of the Se content, uncertainty and detection limit fully consistent with those provided by the data augmentation routine.

Removal of outliers

The original data may have outliers, which given the data augmentation procedure could be amplified in the training set. Therefore, we removed the outliers from the data used to generate the synthetic data. We did not remove them from the test set used to test the final ANN.

In first order approximation one should have:

$$I_{Se}/\left(f\, m_{Se} c\right) = \text{constant} \tag{4}$$

$$I_{Cs}/\left(f\, m_{Cs}\right) = \text{constant} \tag{5}$$

where I_{Se} and I_{Cs} are the Se and Cs peak integrals, respectively, m_{Se} and m_{Cs} are the sample and comparator masses, respectively, c is the Se concentration, and f is the flux, which was available from the RID analysis, but not used by the ANN given that the information it provides is redundant with the comparator mass.

Equations (4) and (5) were used, first to check the data and remove any outliers from the training (taken as having values given by Eqs. (4) and (5) outside three sigma of the entire data set) and second to ensure that all synthetic data generated respected Eqs. (4) and (5) within three standard deviations of the mean calculated for the data. This is also a further check to the validity of the synthetic data, at least for the comparator mass and for the Se content.

Results and discussion

The ANN was trained as indicated above, over a number of epochs. In each epoch, the entire training data set is presented to the ANN, one by one, and the ANN internal parameters are adjusted accordingly. At the end of an epoch, the error is calculated for both the training and test sets. The training and test set errors as a function of epoch are shown in Fig. 3. The ANN is considered trained when the test set error stops decreasing.

The results obtained for the Se content, its uncertainty and LOD are given in Figs. 4, 5, and 6. The figures present the normalized values (range 0 to 1) obtained by the ANN vs the real values. The first observation is that the ANN seems to reproduce the values obtained with the standard software (which we will call the "real values" from now on) very closely. Numerically, the average absolute deviation between the ANN prediction and the real values for the Se content, considering only the test set, is 0.008 mg kg^{-1} and its standard deviation is 0.006 mg kg^{-1}. This is within the expected uncertainty of the experimental Se values. It needs to be stressed that ANNs are interpolation devices, and the spread in the training data inevitably must also be found in the ANN results. Additional spread that comes from the

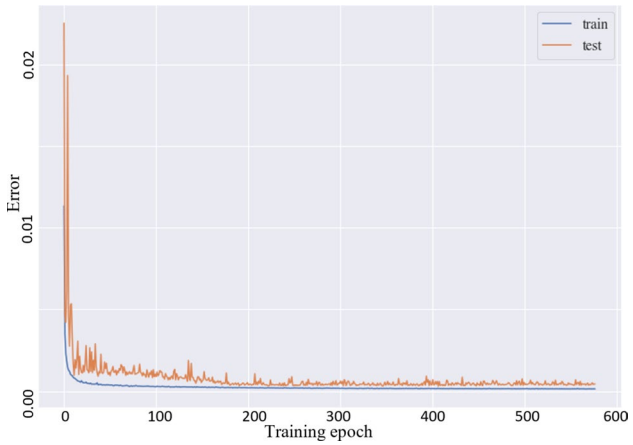

Fig. 3 ANN error for the train and test sets as function of training epoch

interpolation mechanism learned during the training will be small when the ANN architecture and the training process are effective, as is the case in this study.

The variation in the results obtained for the uncertainty and for the LOD (Fig. 5 and 6) is significantly higher than for the Se amount (Fig. 4). By observing Figs. 5 and 6, particularly the data for the LOD, some outliers in the ANN prediction can be seen, even for the training data. One should keep in mind that the figures show more than 10,000 points, so the outliers are in fact a small percentage of the total number of points (0.3% and 1.4% of data points for the uncertainty and the LOD, respectively).

Nevertheless, the higher variation in the uncertainty and LOD results, as compared to the variation in the Se content results, is clear. This meets expectations: the relative accuracy with which the Se amount is determined is much higher than the relative accuracy with which its uncertainty and LOD are. Numerically, and considering only the test set, the average absolute deviation between the ANN prediction and the real values for the uncertainty is 0.0013 mg kg^{-1}, and its standard deviation is 0.0018 mg kg^{-1}. The average value of the uncertainty in the original data is 0.0154 mg kg^{-1}. For the LOD, the corresponding values are 0.0027 mg kg^{-1}, and 0.0038 mg kg^{-1}, respectively. The average value of the LOD in the original data is 0.02791. This means that the standard deviation of the ANN results is slightly below 10% of the values to be predicted. All in all, this variation in the ANN prediction can be taken as representing closely the uncertainty with which the Se uncertainty and LOD can be determined from the data.

Summary and outlook

We have shown that ANNs can effectively analyse simple NAA data with an accuracy that is close to what is achievable with analytical methods. The problem studied is one of the simplest cases in NAA, with one single spectrum for each sample, and all samples measured under nearly the same experimental conditions. In particular, the decay and measurement times were equal.

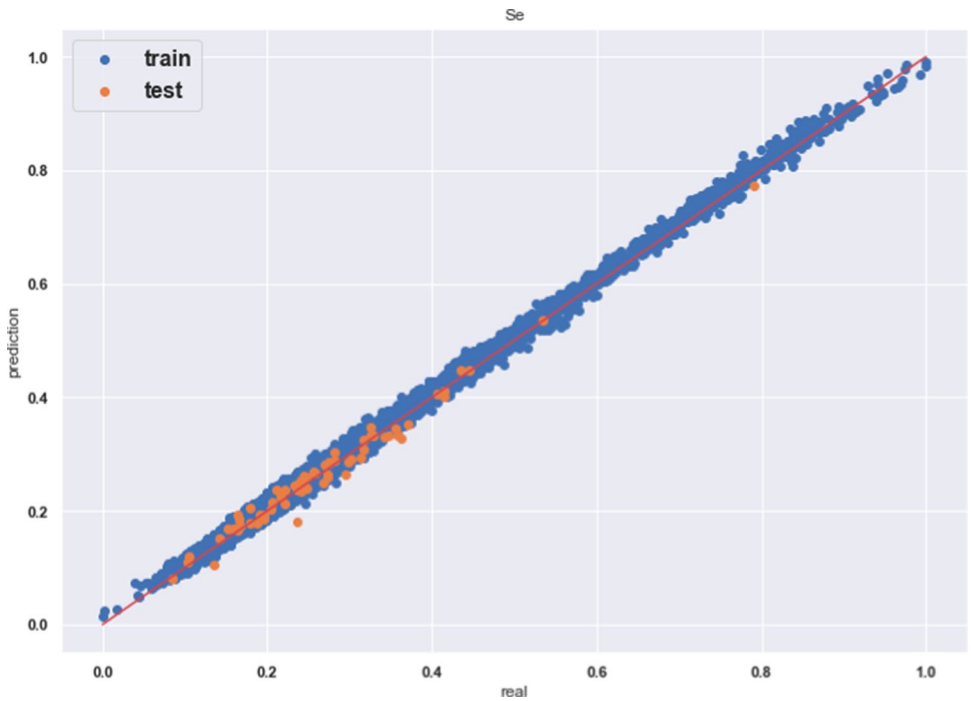

Fig. 4 Normalised ANN (prediction) results versus real values for the Se content. The blue dots are the results for the training set, i.e., the 10,000 synthetic data plus the 80% used for their construction. The orange dots are the results for the 43 real spectra that were not used to generate the synthetic spectra

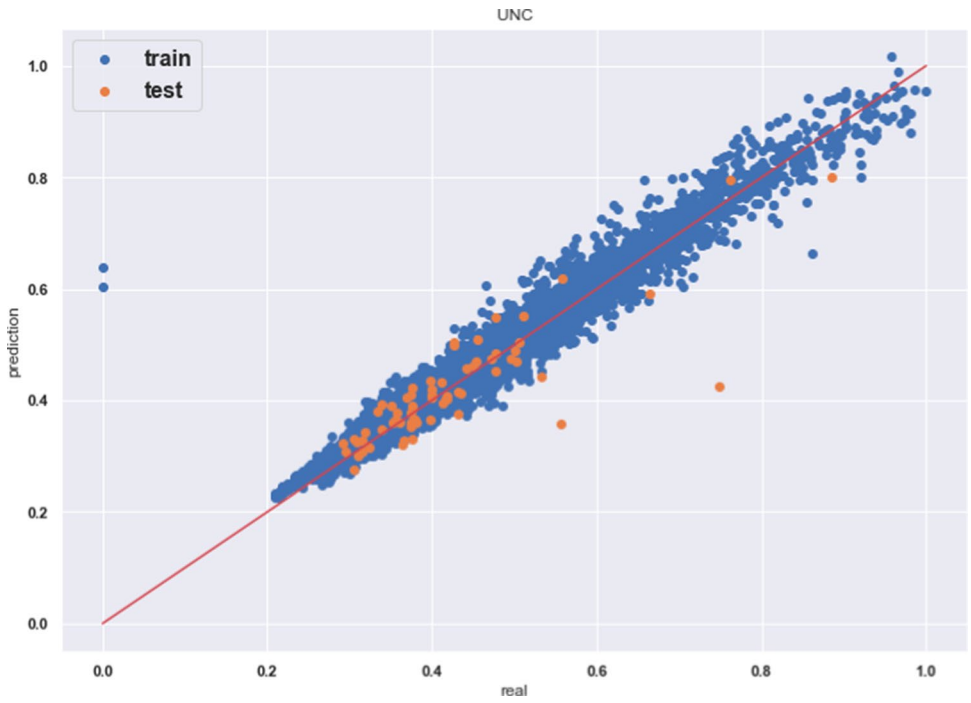

Fig. 5 Normalised ANN (prediction) results versus real values for the uncertainty on the Se content. The blue dots are the results for the training set, i.e., the 10,000 synthetic data plus the 80% used for their construction. The orange dots are the results for the 43 real spectra that were not used to generate the synthetic spectra

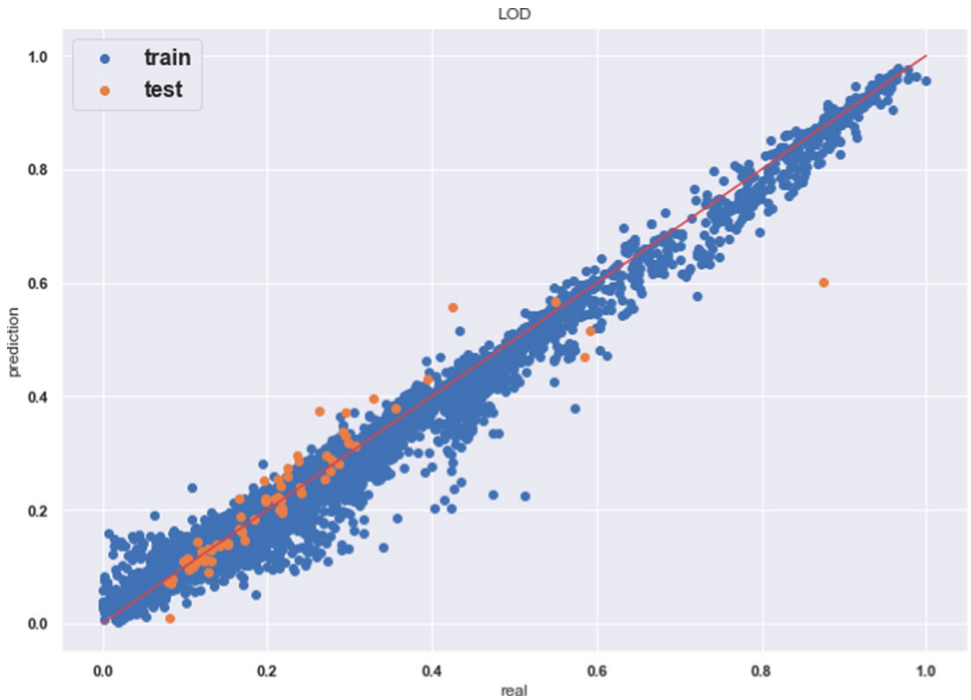

Fig. 6 Normalised ANN (prediction) results versus real values for the Se detection limit LOD. The blue dots are the results for the training set, i.e., the 10,000 synthetic data plus the 80% used for their construction. The orange dots are the results for the 43 real spectra that were not used to generate the synthetic spectra

The question arises of whether ANNs can also analyse common NAA data where the dimension of time plays a role, for instance in single spectra with varying decay times, typical for samples irradiated simultaneously and measured sequentially, or when more than one spectrum is collected after successively longer decay times.

In principle, if the information is present in the data, an ANN with a suitable architecture and suitably trained can retrieve it. The modular approach to the network architecture applied in the present case can be extended for the case of multiple spectra. For instance, each spectrum can be kept disconnected from the other ones in the initial ANN layer,

allowing for initial pre-processing of each spectrum separately. This technique would prevent a strong increase of the number of nodes in the network, which helps in the training process and has been successfully employed in analysis of multiple ion beam analysis spectra from the same sample, collected with different techniques and experimental conditions [9, 33]. In any case, the training data needs to be sufficiently representative, covering the range of parameters that is found in the data to be analysed, including irradiation, decay and measuring times. Appropriate data augmentation methods, similar to those employed in this work, could be used to generate synthetic training data.

One possibility of bypassing the need to use data augmentation techniques, and at the same time allowing one to train the ANN before any experimental data is collected, would be to use calculated spectra for the training. These would have to be realistic, i.e. including Poisson noise and realistic continuum. As many thousands of spectra are needed, Monte Carlo methods might not be appropriate. However, Monte Carlo methods could be used to generate, for a given detector configuration, a set of continua for single energies, from which the continuum for a given sample with given elemental amounts could be interpolated. The continuum could then be added to the continuum-free spectra calculated with fast analytical methods. Alternatively, the production of synthetic spectra using generative neural networks might be explored. Generative models, capable of producing new data from a learned model, have made tremendous progress in recent years, since the initial proposal of Generative Adversarial Networks (GANs) [34]. Initially applied to image processing, generative models have been also successfully applied in other domains such as particle physics [35].

The data, the augmentation code and its outputs, and the ANN code have been made publicly available [36].

Declaration

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. International Atomic Energy Agency (2022) Research reactor database. https://nucleus.iaea.org/rrdb/#/home
2. International Atomic Energy Agency (2018) Proficiency testing by interlaboratory comparison performed in 2010–2015 for neutron activation analysis and other analytical techniques. IAEA-TECDOC-1831, IAEA, Vienna
3. International Atomic Energy Agency (2018) Development of an integrated approach to routine automation of neutron activation analysis. IAEA-TECDOC-1839, IAEA, Vienna
4. International Atomic Energy Agency (2018) Advances in neutron activation analysis of large objects with emphasis on archaeological examples results of a coordinated research project. IAEA-TECDOC-1838, IAEA, Vienna
5. International Atomic Energy Agency (2022) Quality assurance and quality control in neutron activation analysis: a guide to practical approaches. IAEA technical reports series 487, IAEA, Vienna
6. Simonits A, Corte F, Hoste J (1975) Single-comparator methods in reactor neutron activation analysis. J Radioanal Chem 24:31–46. https://doi.org/10.1007/BF02514380
7. Pessoa Barradas N, Vieira A (2000) Artificial neural network algorithm for analysis of rutherford backscattering data. Phys Rev E 62:5818–5829. https://doi.org/10.1103/PhysRevE.62.5818
8. Vieira A, Pessoa Barradas N (2001) Composition of NiTaC films on Si using neural networks analysis of elastic backscattering data. Nucl Instrum Method Phys Res B 174:367–372. https://doi.org/10.1016/S0168-583X(00)00621-2
9. Nené NR, Vieira A, Pessoa Barradas N (2006) Artificial neural network analysis of RBS and ERDA spectra of multilayered multielemental samples. Nucl Instrum Method Phys Res B 246:471–478. https://doi.org/10.1016/j.nimb.2006.01.016
10. Pessoa Barradas N, Vieira A, Patrício R (2002) Artificial neural networks for automation of rutherford backscattering spectroscopy experiments and data analysis. Phys Rev E 65:066703. https://doi.org/10.1103/PhysRevE.65.066703
11. Li F, Gu Z, Ge L et al (2019) Application of artificial neural networks to X-ray fluorescence spectrum analysis. X-Ray Spectrom 48:138–150. https://doi.org/10.1002/xrs.2996
12. Yoshida E, Shizuma K, Endo S, Oka T (2002) Application of neural networks for the analysis of gamma-ray spectra measured with a Ge spectrometer. Nucl Instrum Method Phys Res A 484:557–563. https://doi.org/10.1016/S0168-9002(01)01962-3
13. Roshani GH, Eftekhari-Zadeh E, Shama F, Salehizadeh A (2017) Combined application of neutron activation analysis using IECF device and neural network for prediction of cement elements. Radiat Detect Technol Method 1:23. https://doi.org/10.1007/s41605-017-0025-z
14. Medhat ME (2015) Artificial neural network: a tool for rapid quantitative elemental analysis using neutron activation analysis. Int J Adv Res Electr Electron Instrum Eng 4:5497–5501
15. Lee D (2019) Application of artificial neural network to prompt gamma neutron activation analysis for chemical warfare agents identification. https://doi.org/10.2172/1565918
16. Bilton KJ, Joshi THY, Bandstra MS et al (2021) Neural network approaches for mobile spectroscopic gamma-ray source detection. JNE 2:190–206. https://doi.org/10.3390/jne2020018
17. Hossny K, Hossny AH, Magdi S et al (2020) Detecting shielded explosives by coupling prompt gamma neutron activation analysis and deep neural networks. Sci Rep 10:13467. https://doi.org/10.1038/s41598-020-70537-6
18. Braga CC, Dias MS (2002) Application of neural networks for unfolding neutron spectra measured by means of bonner spheres. Nucl Instrum Method Phys Res, Sect A 476:252–255. https://doi.org/10.1016/S0168-9002(01)01464-4
19. Bode P, Korthoven PJM, de Bruin M (1987) Microprocessor-controlled facility for INAA using short half-life nuclides. J Radioanal Nucl Chem 113:371–378. https://doi.org/10.1007/BF02050509
20. Blaauw M (1999) The reference peak areas of the 1995 IAEA test spectra for gamma-ray spectrum analysis programs are absolute and traceable. Nucl Instrum Method Phys Res A 432:74–76. https://doi.org/10.1016/S0168-9002(99)00257-0
21. Blaauw M (1994) The holistic analysis of gamma-ray spectra in instrumental neutron activation analysis. Nucl Instrum Meth A 353:269–271
22. Vaswani A, Shazeer N, Parmar N et al (2017) Attention is all you need. In: Guyon I, Luxburg UV, Bengio S et al (eds) Advances

in neural information processing systems. Curran Associates Inc, New York
23. Yang Z, Dai Z, Yang Y et al (2019) XLNet: generalized autoregressive pretraining for language understanding. In: Wallach H, Larochelle H, Beygelzimer A et al (eds) Advances in neural information processing systems. Curran Associates Inc, New York
24. Touvron H, Cord M, Douze M et al. (2021) Training data-efficient image transformers & distillation through attention. In: Proceedings of the 38th international conference on machine learning. PMLR, pp 10347–10357
25. Chollet F (2021) Deep learning with python, 2nd edn. Manning Publications, Shelter Island
26. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. The MIT Press, Cambridge, Massachusetts
27. Liu H, Dai Z, So DR, Le QV (2021) Pay attention to MLPs. Adv Neural Inf Process Syst 34:9204–9215. https://doi.org/10.48550/ARXIV.2105.08050
28. Tang C, Zhao Y, Wang G et al. (2021) Sparse MLP for image recognition: is self-attention really necessary. https://doi.org/10.48550/ARXIV.2109.05422
29. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958
30. Hendrycks D, Gimpel K (2016) Gaussian error linear units (GELUs). https://doi.org/10.48550/ARXIV.1606.08415
31. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. https://doi.org/10.48550/ARXIV.1412.6980
32. Abadi M, Agarwal A, Barham P et al. (2015) Tensorflow: large-scale machine learning on heterogeneous systems
33. Pinho HFR, Vieira A, Nené NR, Barradas NP (2005) Artificial neural network analysis of multiple IBA spectra. Nucl Instrum Method Phys Res B 228:383–387. https://doi.org/10.1016/j.nimb.2004.10.075
34. Gui J, Sun Z, Wen Y, Tao D, Ye J (2021) A review on generative adversarial networks: algorithms, theory, and applications. IEEE Trans Knowl Data Eng. https://doi.org/10.1109/TKDE.2021.3130191
35. Lebese T, Ruan X (2022) The use of generative adversarial networks to characterise new physics in multi-lepton final states at the LHC. https://arxiv.org/pdf/2105.14933.pdf
36. IAEA-Physics-neutrons (2022) NAA-ANN-1: proof of concept of application of artificial neural networks to neutron activation analysis data. https://github.com/IAEA-Physics-neutrons/NAA-ANN-1