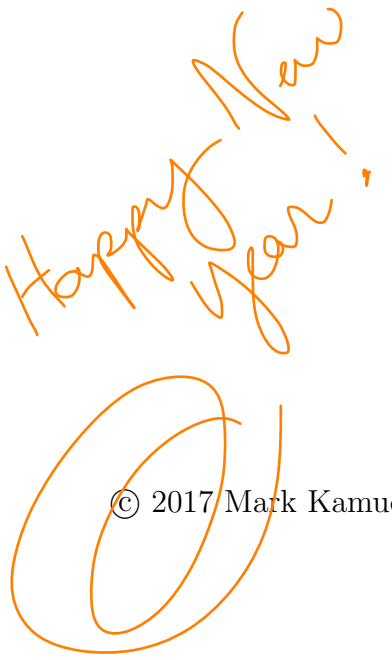


AUTOMATED ISOTOPE IDENTIFICATION ALGORITHM USING ARTIFICIAL NEURAL NETWORKS

BY

MARK KAMUDA

Preliminary Examination

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Nuclear Plasma and Radiological Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Assistant Professor Kathryn Huff, Adviser
Prof$_1$ Rizwan Uddin
Prof$_2$ Tomasz Kozlowski
Prof$_3$ Mark Hasegawa-Johnson
Dr. Clair Sullivan

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The main question addressed in this work is: Can artificial neural networks (ANN) ~~be used to solve different problems in~~ automated ~~radio~~ isotope identification and quantification ~~using simulated spectra as a training datasets~~? ANNs can incorporate abstract features of a gamma-ray spectrum in non-intuitive ways. This may give ANNs the ability to perform isotope identification and quantification for large isotope libraries practical for domestic nuclear security, operate using low-resolution NaI radiation detectors without ~~accurately~~ knowing the detector's calibration or ~~knowing the~~ background radiation.

~~The aim of this~~ dissertation ~~is to~~ demonstrate the performance of ANNs for ~~various tasks related to~~ identifying and quantifying ~~the~~ radioisotopes in low-resolution gamma-ray spectra. The low-resolution detector of interest in this work is a 2-inch by 2-inch NaI(Tl) cylindrical scintillation detector. This detector is industry standard due to its ease of use, low cost, and acceptable resolution. Tasks that will be investigated ~~are isotope identification using~~ isotopes in the ~~ANSI~~ required list ~~of isotopes for the detection and identification of radionuclides~~ [1]. The ANSI standard also requires isotope identification algorithms ~~be able~~ to operate when the radioactive material is behind shielding. ~~Because of this~~, shielding will also be incorporated into this work.

In addition to isotope identification, the ability of ANNs to quantify the count contribution from each isotope ~~will be explored~~. The neural networks ability to extract count contribution information from gamma-ray spectra is important for two reasons. The first purpose is to give more information to the user on the quantity of isotopes in an unknown spectrum. The second purpose is a possible use of neural networks in post-detonation nuclear forensics. In post-detonation nuclear forensics, a large number of radioactive fission products are created. Quantifying the amount of isotope in post-detonation debris can yield useful information

about the devices properties. Mixtures of laboratory sources will be used as a surrogate for post-detonation debris and the ability for the ANN to accurately calculate the mixture components will be ~~analyzed~~ established.

Another case where knowing the isotope quantities in a mixture of isotopes is in uranium enrichment calculations. Knowing the enrichment of uranium is important in two areas. The first are is when uranium is identified at a border crossing. Typically the spectrum would need to be given to a trained spectroscopist to quantify the enrichment, but this process could be automated on the device first used to identify it. The second case would be in treaty verification technologies. Low-resolution NaI gamma-ray detectors decrease the amount of possibly sensitive information while giving enough information to produce accurate enrichment quantification. The fact that ANNs can be taught to ignore certain patterns and only give information agreed upon by treaty signatories also makes them a good tool for treaty verification. The ability of ANNs to operate without knowing the shielding and background spectrum makes them better from a zero knowledge concept. NaI detectors also have a higher efficiency than the higher resolution HPGe. This means that the counting times for NaI are smaller than would be required to get the same number of counts using an HPGe.

CHAPTER 2

LITERATURE REVIEW

2.1 Gamma-Ray Spectroscopy for Isotope Identification

There are a number of difficult things associated with performing isotope identification. Typically this requires a trained spectroscopist. Once a spectrum of interest is identified, a spectroscopist will identify photopeaks. [2]

There are many automated radioisotope identification methods available, but few perform well given a low-resolution gamma-ray spectrum of a mixture of radioisotopes. Common methods include library comparison algorithms, region of interest (ROI) algorithms, principle component analysis (PCA), and template matching.

Library comparison algorithms attempt to match photopeak energies found in a gamma-ray spectrum with those found in a library of known isotope decay energies. Drifts and uncertainties in detector calibration can lead to misidentifying photopeaks, leading to incorrect identifications [3]. To be automated, this method needs an algorithm to extract photopeak centroids from a spectrum. Photopeak extraction algorithms face difficulties when a large number of photopeaks overlap in a spectrum, such as when a mixture of radio-isotopes are measured with a low-resolution detector [4].

ROI algorithms search for elevated counts compared to background in a region where photopeaks are expected to be for different radioisotopes. ROI algorithms may operate poorly when photopeaks of different radioisotopes overlap [3]. For this reason, large isotope libraries will preform poorly using this method. Similarly to the library comparison algorithm, calibration drift may shift photopeaks into different neighboring ROIs, leading to incorrect identification. The ROI method has been used to differentiate normally occuring radioactive material (NORM) from special nuclear material (SNM) using plastic scintillators

[5].

PCA can also be applied to radioisotope identification. The goal of PCA is to reduce the dimensionality of a dataset into uncorrelated variables [6]. Using a few of these principle components, the data may be represented in a reduced space that contains most of the information present in the original data. The transformed data can then be clustered based on isotope identity. Clustering algorithms may include K-means or Mahalanobis distance [7, 8]. PCA has been applied to isotope identification using plastic scintillators [9] and anomaly detection using both plastic scintillators and NaI detectors [10]. Despite the progress of PCA in isotope identification, there has not been significant progress in applying PCA to separating mixtures of isotopes in gamma-ray spectra.

do we know why?

Template matching algorithms find an example in a database of gamma-ray spectra that most closely matches a measured spectrum [3]. The database of spectra can contain multiple detector calibration settings, shielding materials, and source-to-detector distances. Goodness of fit can be measured using a chi-squared test, euclidean distance, or Mahalanobis distance. While a sufficient amount of example spectra can be used to identify almost any measured spectrum, the drawback of this method is the time necessary to compare a measured spectrum to the library and the computer memory necessary to store said library. This method also may have difficulty when mixtures of isotopes are considered, although work is being done to correct this [11].

2.2 Artificial Neural Networks

Artificial neural networks were first created to mimic biologic neurons. Since their creation, they have demonstrated promising results on a variety of different classification and regression tasks [12, 13, 14]. The following sections will give an overview of how ANNs learn and operate.

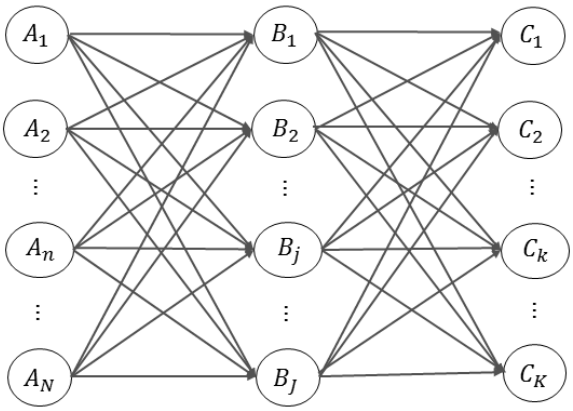
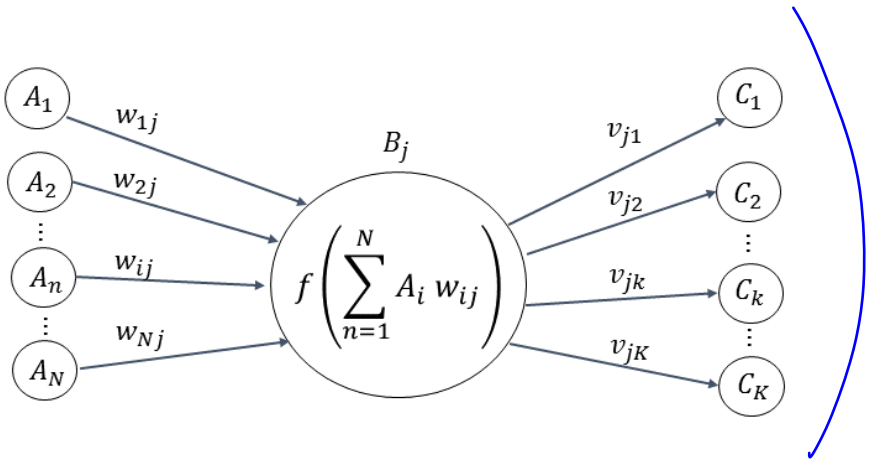


Figure 2.1: Example ANN with input layer A, hidden layer B, and output layer C.

2.2.1 Architecture and Training

ANNs work by mapping arbitrary input spaces, $\mathbb{R}^N$, to arbitrary output space, $\mathbb{R}^K$. An example one hidden layer ANN mapping $\mathbb{R}^N \mapsto \mathbb{R}^K$ is shown in Figure 2.1. Each circle represents a neuron, or node. The mathematical process governing each neuron in the ANN is shown in Figure 2.2. In Figure 2.2, the signal from the previous layer is propagated to the next by applying some function, typically sigmoidal, to the dot product of the signal from the previous layer and the weight vector going into a given node, B_j in Figure 2.2. Given a one-layer ANN with a finite number of hidden nodes, any function $\mathbb{R}^N \mapsto \mathbb{R}^K$ can be described to arbitrary precision [15]. Additional hidden layers increase the representational power of an ANN, reducing the number of nodes and computational power required to represent a function. There is no direct method to compute the number of hidden layers or nodes for a given problem. These, along with other hyperparameters, need to be optimized for a given dataset.

Artificial ANNs learn a function by changing the weights connecting the layers so that some error function is minimized for a given dataset. One popular method to update the weights is through the process of gradient descent through the backpropogation of errors [16]. The update equation for a single weight, w_j, is shown in Equations 2.1 and 2.2. In these equations, $Error$ is the given error function to be minimized (commonly mean squared error or cross entropy) and η is the learning rate of the ANN.



Are you the author of this diagram?

Figure 2.2: Summary of the operation of a single neuron.

"E"?

$$\Delta w_j = -\eta \frac{dError}{dw_j} \tag{2.1}$$

$$w_j^{new} = w_j^{old} + \Delta w_j \tag{2.2}$$

2.2.2 Hyperparameters

In general, ANNs have a tendency to memorize their training set in a process called overtraining. An overtrained ANN will tend to incorrectly identifying novel data. To prevent this, a number of hyperparameters were used to prevent overfitting and optimize performance. Unfortunately, there is currently no known method to know which hyperparameters have an impact on model performance before training. Because of this, a number of popular hyperparameters were added to the model. The following hyperparameters were used in this study: the number of neurons in each layer, the number of layers used, initial learning rate for the training algorithm, the L_2 regularization strength, and neuron dropout rate.

This is the first mention of your actual model.

Because there is no direct method to identify which hyperparameters are important or what their values should be, a random hyperparameter search can be used to find a close-to-optimum network structure and hyperparameter values. There is evidence that a random search in a given hyperparameter range finds better hyperparameters quicker than a grid search in the same range [17]. There is also a proof showing that given 60 points randomly sampled in some space with a finite minimum, the minimum of those 60 random samples is

within 5% of the true minimum with 95% probability [18].

2.3 Isotope Identification Using ANNs

There have been a number of published papers which apply ANNs to automated isotope identification. ANNs have been applied to peak fitting [19], isotope identification [20, 21], and activity estimation [20, 22]. Many of this work rely on ROI methods [23], feature extraction [24], high-resolution gamma-ray spectra as the input to the ANN [25], small libraries of isotopes, and assume perfectly calibrated detectors. ANN training methods created for high-resolution gamma-ray spectra may not perform well when trained using low-resolution spectra given the large discrepancy in resolution. ANN training that relies on ROI methods may not perform well when ROIs overlap significantly with large libraries of isotopes.

It has been shown that an ANN may be trained to perform isotope identification and quantification using low-resolution NaI gamma-ray spectrum using a library of five isotopes [26]. While promising, this study did not include complicated source mixture analysis. This study also used a library too small to be of practical use. The American National Standards Institutes (ANSI) has identified 31 gamma-ray emitting isotopes that automated isotope identification algorithms should be able to identify [1].

CHAPTER 3

PREVIOUS WORK

What work? Discuss it first...

Probably best to cite & summarize the previous work (particularly before criticizing it?)

3.1 Thesis and TNS Publication

The work I published has a lot of room for improvement. Some physics in the MCNP model were neglected, like any radiation contribution from bremsstrahlung and environmental scatter. Also the background isotope spectra were very wrong. These were generated assuming a point source of radiation. Real background is distributed in the soil. Many scattering events in soil along with skyshine contribute to a spectrum that looks very different from a point source.

What MCNP model?

There were parts of the published work that were good. The sampling method based on isotope templates is a good method to simulate lots of realistic gamma-ray spectra. One of the most difficult parts of ANN training is creating a good training set that represents reality. This method can be used to simulate most things, which is hugely useful. The results of the published work were also very promising. Despite the unrealistic physics model used in the simulation, the ANN correctly identified isotopes in a variety of simulated and real spectra.

This is extremely positive.

who?

Moving away from the MCNP model, we used GADRAS to create our template spectra. GADRAS has done all the physics heavy lifting for us, which is awesome. Also lets us simulate shielding, different scattering environments, detectors with different parameters(FWHM vs energy, nonlinear calibration, crystal dimensions), and entirely different detectors. We have demonstrated that an ANN trained with this data can accurately identify poor quality simulated spectra. Real spectra are still needed to verify these results.

The current method begins by simulating a gamma-ray spectrum dataset for a given identification and quantification task. The ANN inputs are 1024 channels of a NaI spectrum

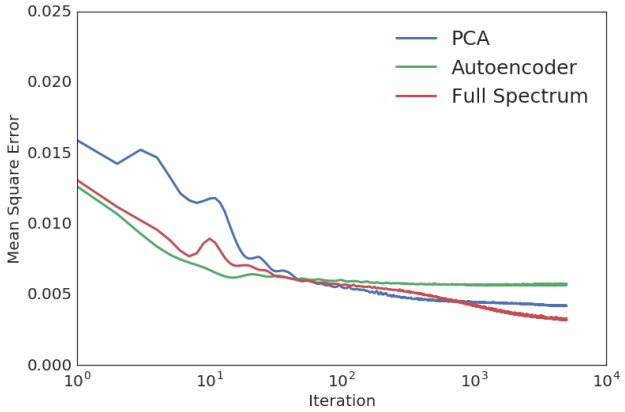


Figure 3.1: Mean square error vs training iteration for three different ANN inputs.

and the output is percent count contributions from each isotope in the library used. The number of input channels can be easily changed to accommodate other detectors.

3.2 Shielding Experiment Results

We're seeing some confusion in low-energy isotopes. Still need to validate results on real spectrum dataset. (I assume this will be fleshed out)

3.3 Uranium Enrichment Measurement Experiment

The ability for an ANN to preform isotope quantification for uranium enrichment measurements was investigated. This work also explored if dimension reduction techniques (PCA and autoencoders) improved ANN performance over using the full spectrum as ANN input. As seen in Figure 3.1, PCA and the autoencoder stopped learning before the full spectrum This indicates that using the full spectrum as ANN input may be superior to using dimension reduction techniques for this problem.

before it what?

Table 3.2 also shows that using the full spectrum as ANN input is superior to dimension reduction techniques. This table also shows that the ANN is best at identifying ^{235}U. The accuracy of U235 implies that ANN is not learning calibration drift well. The 186 keV peak from ^{235}U has no overlap, even with gain drift, with the other isotopes in the training library.

$ ^{235}U $

	Mean Square Error (10^{-3})		
Isotope	Full Spectrum	PCA	Autoencoder
U235	0.86	3.9	6.9
Pa234m	3.5	4.1	6.6
U238	4.8	5.2	7.37
U234	4.4	6.1	7.1
Th234	5.6	6.4	7.1
Th231	6.0	6.7	7.4
Average	4.2	5.4	7.1

Figure 3.2: Table showing the average mean square error from 1000 simulated enriched uranium spectra.

This allows the ANN to accurately identify it. Conversely, ^{231}Th and ^{234}Th emit similar low-energy gamma-rays. These become easily confused with gain drift. This implies that gain correction and isotope identification and quantification should be handled by a separate algorithms. Suggestions for improvements for this will be explained in later sections.

CHAPTER 4

FUTURE WORK AND PROPOSED EXPERIMENTS

Be kinder to your past work, perhaps?
Perhaps "that may extend and improve the current algorithm performance"

There are a number of experiments that may remedy deficiencies in the work that has been done. These experiments are described in this section.

4.1 Autoencoders

While autoencoders were touched on in the uranium enrichment work, they have not been explored thoroughly. There is some evidence that the autoencoder was overtrained to the A special case of denoising autoencoders will be explored for the ANSI dataset. In previous work, a single ANN had to learn to differentiate the varying background signal from the source signal. In addition to this, the ANN also had to learn gain correction. By training an autoencoder to reconstruct a background-subtracted and gain-corrected spectrum, the task of isotope identification is simplified for the ANN. This may result in more accurate identifications.

explain the meaning of this more formally first

In addition to using fully connected autoencoders, it may be more effective to use a 1-D convolutional autoencoder ANN. Because gamma-ray spectra have local structure in the form of photopeaks and Compton continua, convolutional ANN structures may be able to more easily extract information from spectra than fully connected ANNs. Fully connected ANNs do not assume there is local spatial structure in a signal, so the fully connected ANN would need to learn that there is local structure. Convolutional ANNs assume there is local structure, and the extent of this structure can be changed by changing the length of the convolutional ANNs filters.

Once the autoencoder is trained, the hidden layer and output layer (representing the reconstructed spectrum) will be used to train a separate ANN for isotope identification and

quantification. The performance of these ANNs will be compared.

4.2 Additional Simulated Detector Models

Simulating ~~addition~~ additional NaI detectors using GADRAS may help the ANN generalize to different detectors. To begin this process, two 2-inch by 2-inch NaI detectors will be modeled using GADRAS and their properties will be changed (for example: crystal dimension, calibration, scattering environment). The number of different detectors modeled will be increased and the generalization performance of the ANN will be evaluated. The generalization performance can be calculated by checking the ANNs performance on spectra produced by a detector whose properties the ANN has not seen during training.

4.3 Mixture Validation Dataset

A ~~real~~ dataset of real spectra ~~are~~ is needed to confirm the models performance. Mixtures of laboratory isotopes can be made. Mixtures will vary by signal-to-background ratio by varying source-to-detector distance and integration time. Isotopes available are: ^{60}Co, ^{137}Cs, ^{133}Ba, ^{152}Eu, ^{22}Na and others.

4.4 Bagging

Each time the ANN trains, it produces slightly different identification results and performance. Bagging (bootstrap aggregating), or the process of averaging the outputs of many ANNs, can reduce the variance in the output [27]. In addition to this, bagging more accurately displays the true performance of a given ANN structure. The number of ANNs included in the bagging process will be explored.

4.5 K-folds cross validation

K-folds cross validation will be incorporated into the model. Cross validation is the general method of determining how well a model will generalize to a dataset the model has not seen. K-folds cross validation does this by splitting the available data into k subsets. From these, k-1 subsets are used to train the ANN and the remaining subset is used as the validation dataset. This process is repeated, using each subset once as the validation dataset. Similar to bagging, cross validation will more accurately analyze how well the ANN structure will perform.

4.6 Latin Hypercube Sampling for Training Set Construction

Currently, the training set is constructed using random combinations of isotopes with random count rates. Because random sampling does not guarantee the training space is well sampled and often produces clusters, this may not be a good method to construct the training dataset. This may lead the ANN to learn a bias for random clusters in the training set. A way to reduce the chances of clusters and more uniformly sample the input space is using Latin hypercube sampling (LHS). LHS partitions each dimension of a space into N equal parts. The space is then sampled using N points, ensuring that there is only one point along each partition in each dimension.

To quantify how much LHS helps identification for a given data space, ANNs will be trained with an increasing number of samples. The ANNs performance on a real dataset will then be measured and compared. It is expected that at a certain number of samples the ANN performance will reach an asymptote. The number of samples may change for each data space (no shielding, shielding, uranium enrichment) depending on the complexity of the problem.

First few sentences should explain dropout method

4.7 Model Confidence Using the Dropout Uncertainty Method

We can exploit the dropout method used to regularized the ANN to get a confidence measure given a spectrum [28]. Once the ANN is trained, an unknown sample is presented to the model and the solution is recorded. Dropout is then used on the model using the same dropout probability that was used to train the model. The same unknown sample is then passed through the ANN and the answer recorded. This process of dropout and recording the resulting output is repeated. The variance of the outputs from this process can be used to determine the ANNs confidence in a given pattern. ?

A few results are predicted using the ANN confidence. It is also expected that high signal-to-background measurements and isotope mixtures of a few components will have a high model confidence. It is expected that low signal-to-background measurements and mixtures of many components will have a low model confidence. Additionally, the model confidence with an isotope not included in the training set should be very low. This confidence measure may be used to reduce the false alarm rate when performing isotope identification.

Additional formality may clarify this bit.

4.7.1 New Training Stopping Condition

ANNs need a condition to stop learning. This condition is typically when a certain allowable error in a validation dataset is reached or when the training error does not appreciably decrease over time. Previously in our work, the stopping condition has been based on the cost function the ANN is trying to decrease: the cross entropy. It may be better to stop training with a more useful metric. One metric could be when the maximum error in the training set reaches a certain threshold. Another metric could be to stop training when the ANNs mean squared error for a validation set stops decreasing.

CHAPTER 5

CONCLUSION

CHAPTER 6

REFERENCES

[1] *American National Standard Performance Criteria for Hand-Held Instruments for the Detection and Identification of Radionuclides, ANSI N42.34-2006*, IEEE Std., Rev. ANSI N42.34-2006, 2007.

[2] M. Rawool-Sullivan, J. Bounds, S. Brumby, L. Prasad, and J. Sullivan, "Steps toward automated gamma ray spectroscopy steps toward automated gamma ray spectroscopy: How a spectroscopist deciphers an unknown spectrum to reveal the radioactive source," 02 2010.

[3] T. Burr and M. Hamada, "Radio-isotope algorithms for NaI gamma spectra," *Algorithms*, vol. 2, pp. 339–360, March 2009.

[4] H. Xiong, "Automated wavelet analysis of low resolution gamma-ray spectra and peak area uncertainty," Master's thesis, University of Illinois at Urbana-Champaign, 2015.

[5] J. Ely, R. Kouzes, J. Schweppe, E. Siciliano, D. Strachan, and D. Weier, "The use of energy windowing to discriminate SNM from NORM in radiation portal monitors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 560, no. 2, pp. 373 – 387, 2006.

[6] J. I. T., *Principal Component Analysis*. Springer New York, 2002.

[7] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, Jul 2002.

[8] V. V. Kumari, B. R. Raju, and A. Naik, "Hybrid clustering algorithm based on mahalanobis distance and mst," *International Journal of Applied Information Systems*, vol. 3, no. 5, pp. 60–63, Jul 2012.

[9] 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, Feb 2012.

[10] R. Runkle, "Analysis of spectroscopic radiation portal monitor data using principal components analysis," *IEEE Transactions on Nuclear Science*, vol. 53, no. 3, pp. 1418–1423, 2006.

[11] J. Mattingly and D. Mitchell, "A framework for the solution of inverse radiation transport problems," *IEEE Transactions on Nuclear Science*, vol. 57, no. 6, pp. 3734–3743, 2010.

[12] S. Jeyanthia, N. Maheswaria, and R. Venkatesh, "Neural network based automatic fingerprint recognition system for overlapped latent images," *Journal of Intelligent Fuzzy Systems*, vol. 28, 2015.

[13] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Annual Conference on Neural Information Processing Systems*, pp. 1097–1105, 2012.

[14] A. Rababaah and S. D., "Integration of two different signal processing techniques with artificial neural network for stock market forecasting," *Academy of Information Management Sciences Journal*, vol. 18, pp. 63–80, 2015.

[15] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, vol. 4, pp. 251–257, 1991.

[16] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[17] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.

[18] A. Zheng, *Evaluating Machine Learning Models*. O'Reilly Media, Inc., 2015.

[19] R. Abdel-Aal, "Comparison of algorithmic and machine learning approaches for the automatic fitting of gaussian peaks," *Neural Computing and Applications*, vol. 11, no. 1, pp. 17–29, 2002.

[20] R. Abdel-Aal and M. Al-Haddad, "Determination of radioisotopes in gamma-ray spectroscopy using abductive machine learning," *Nuclear Instruments and Methods in Physics Research A*, vol. 391, pp. 275–288, 1996.

[21] M. Medhat, "Artificial intelligence methods applied for quantitative analysis of natural radioactive sources," *Annals of Nuclear Energy*, vol. 45, pp. 73 – 79, 2012.

[22] V. Vigneron, J. Morel, M. Lepy, and J. Martinez, "Statistical modelling of neural networks in y-spectrometry," *Nuclear Instruments and Methods in Physics Research A*, vol. 396, p. 642647, 1996.

[23] V. Pilato, F. Tola, J. Martinex, and M. Huver, "Application of neural networks to quantitative spectrometry analysis," *Nuclear Instruments and Methods in Physics Research A*, vol. 422, pp. 423–427, 1999.

[24] L. Chen and Y.-X. Wei, "Nuclide identification algorithm based on KL transform and neural networks," *Nuclear Instruments and Methods in Physics Research A*, vol. 598, pp. 450–453, 2009.

[25] E. Yoshida, K. Shizuma, S. Endo, and T. Oka, "Application of neural networks for the analysis of gamma-ray spectra measured with a Ge spectrometer," *Nuclear Instruments and Methods in Physics Research A*, vol. 484, pp. 557–563, 2002.

[26] P. Olmos, J. C. Diaz, J. M. Perez, P. Gomez, V. Rodellar, P. Aguayo, A. Bru, G. Garcia-Belmonte, and J. L. Pablos, "A new approach to automatic radiation spectrum analysis," *IEEE Transactions on Nuclear Science*, vol. 38, pp. 971–975, August 1991.

[27] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug 1996. [Online]. Available: https://doi.org/10.1007/BF00058655

[28] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," *International Conference on Machine Learning*, 10 2016.