2015 IEEE 14th International Conference on Machine Learning and Applications

# Synthetic Oversampling for Advanced Radioactive Threat Detection

Colin Bellinger
School of Electrical Engineering and Computer Science
University of Ottawa
Email: colin.bellinger@gmail.com

Nathalie Japkowicz
School of Electrical Engineering and Computer Science
University of Ottawa
Email: nat@eecs.uottawa.ca

Christopher Drummond
National Research Council Canada
Gouvernement du Canada / Government of Canada
Ottawa, Canada
Email:Christopher.Drummond@nrc-cnrc.gc.ca

***Abstract*—Gamma-ray spectral classification requires the automatic identification of a large background class and a small minority class composed of instances that may pose a risk to humans and the environment. Accurate classification of such instances is required in a variety of domains, spanning event and port security to national monitoring for failures at industrial nuclear facilities. This work proposes a novel form of synthetic oversampling based on artificial neural network architecture and empirically demonstrates that it is superior to the state-of-the-art in synthetic oversampling on the target domain. In particular, we utilize gamma-ray spectral data collected for security purposes at the Vancouver 2010 winter Olympics and on a node of Health Canada's national monitoring networks.**

## I. Introduction

Given recent events, such as the nuclear accident in Japan and the threat of terrorism, monitoring and the early detection of radioactive threats is of significant importance. A particularly common form of monitoring involves the use of gamma-ray spectroscopy. The threat of terrorism has necessitated the monitoring of ports of entry to ensure that nuclear material is not being secretly moved across borders [1]. This process typically involves the scanning of individual shipping containers in order to identify and prevent the illicit movement of nuclear materials. Similarly, the entry points of certain events, such as the G8 meetings and the Vancouver 2010 Olympics, may represent high profile targets, and are thus monitored to ensure security from radioactive threats [2]. More generally, it is considered prudent to monitor radiation at a national level with particular points of interest in mind, such as downwind from nuclear power plants and isotope production facilities [3], the goal being to instil confidence that national industries are safe, and to facilitate fast detection and reaction should a malfunction occur.

This work focuses on the real-time monitoring that takes place in the national monitoring network developed by the *Radiation Protection Bureau of Health Canada* and their work in securing high profile events. In both scenarios, a large volume of data is collected, and the resulting gamma-ray spectra must be quickly analyzed in order to ensure public safety. Though highly proficient at identifying troubling spectra, the data volume dictates that reliance on human experts alone is not feasible. Rather, a combination of computer technology for identifying spectra affected by isotopes of interest, in conjunction with the post-*hoc* analysis of an expert, is necessary. Thus, the objective is to ensure that all spectra affected by isotopes of interest are reviewed by a physicist, with as few benign spectra being analyzed as possible.

Machine learning algorithms have been applied to this task [2, 3] and have succeeded in improving performance beyond the more traditional peak fitting algorithms that look to identify specific isotopes. The false positive rate, however, remains a concern. One approach to dealing with the accuracy of positive predictions is to manage the significant class imbalance that exists in this domain. The standard requirement in machine learning is that a representative set of examples is available for the purpose of training. This learning task, however, is complicated by the fact that the isotopes of interest in the data occur much less frequently than the background class. As a result, significantly fewer non-background examples are available for training the classifier.

Three paradigms exist for coping with imbalance:

- *one-class classification*: is ideal when examples are extremely rare or non-existent;
- *cost-sensitive methods*: are often appropriate when specific costs are known and extensive choice in the classification method is not required; and,
- *Sampling methods*: are generally preferable when costs are not well known and classifier independence is desirable.

It has been shown that the benefit of one-class classification quickly erodes as the number of training samples is increased [4]. From the perspective of imbalanced classification, it has been shown that sampling and cost-sensitive methods are often comparable in terms of their effect on performance [5, 6]. Thus, we focus on sampling for its simplicity in implementation and applicability to all binary classification algorithms without modification.

Bagged random undersampling (BRUS) and SMOTE have empirically been shown to be superior methods on different imbalanced domains [7, 8]. Both of these methods, however, have known weaknesses that affect their performance on our target domain. In particular, the degree of imbalance demands that too much majority information is discarded by BRUS, whereas SMOTE's use of distance measures impacts its performance on the high-dimensional gamma-ray spectral

978-1-5090-0287-0/15 $31.00 © 2015 IEEE
DOI 10.1109/ICMLA.2015.58

data. In addition, the kNN framework applied by SMOTE to generate synthetic instances causes them to be placed inside the convex-hull; this can negatively impact classification by either over- or under-generalizing, depending on the properties of the domain. This has necessitated the development of post-hoc cleaning techniques, such as the removal of Tomek links [9] and, conversely, means of adding additional diversity to the synthesized set [10, 11].

As an alternative, we propose a novel method of synthetically oversampling the minority class that is based on the induction of an artificial neural network. Specifically, we introduce a method of DEnoising Autoencoder-based Generative Oversampling (DEAGO) for modelling the minority class and synthesizing new instances to balance the training set. By inducing a neural network model, we are able to avoid the short-comings of distance-based solutions on the high-dimensional spectral data. Moreover, through the training process, the autoencoder model learns to approximate the shape of the minority distribution rather than naively placing synthetic points between nearest neighbours. This ensures that the samples generated by DEAGO represent the global structure of the class. Our results show that DEAGO is superior to SMOTE and bagged-random undersampling on the target domain in terms of the AUC on both gamma-ray spectral test domains.

## II. Gamma-Ray Spectra Data

In this section, we describe the datasets acquired from Health Canada for the development of the classification system.

### A. Collection

The data utilized in this study was collected as part of two separate, but related, health and safety programs. In particular, the VAN dataset was collected at three Vancouver 2010 Winter Olympics venues and the SAAN data was collected as part of the national monitoring network at Saanich, British Columbia. Both datasets were recorded with sodium iodine (NaI) gamma-ray spectrometers (GRS) which measure photon counts with respect to energy in KeV (see Figure 1 for an example of the VAN data). Four months of data is used in the experiments from the national monitoring network, with each sample collected over a 15-minute interval. Due to time constraints, specifically with respect to the need for fast detection and reaction, noisier, 1-minute samples were collected for security at the Olympic venues.

### B. Processing

Annotated data files from a Health Canada database were provided for the experiments. The annotations specified the presence of isotopes of interest; this task was conducted in advance by government physicists. These files were converted to *csv* and imported to R[1] for analysis and cleaning. Each subsequent experiment was conducted in R.

The data from the national monitoring network was sampled according to 512 spectral channels. Based on the channels that are affected by the isotopes of interest, the final dataset included only the first 250 channels. The Olympics data was sampled with 1024 channels and, once again, we reduced the feature-space of our dataset to half of this based on the energy levels at which the isotopes of interest are expected to occur.

[1] http://cran.r-project.org

### C. Properties

The VAN dataset has 39,000 background instances and 39 instances marked as isotopes of interest; thus, the dataset is heavily imbalanced. The SAAN data contains 11,500 background instances and 29 technicium-99 instances that form the class of interest; therefore, this learning task also includes a significant degree of absolute imbalance.

Two examples of the data are plotted in log-form in Figure 1; these correspond to a background reading and a sample that has been affected by Technicium-99. In this, data energy is represented in terms of channels on the $x$-axis, and the counts, which indicate the intensity, are recorded on the $y$-axis. Unshielded and poorly shielded isotopes produce specific and identifiable peaks in the background distribution, as the signals are not degraded by shielding materials. In the case of Technicium-99 instance, we clearly see this in the low energy range of the spectrum. In other cases, we might expect a much more muted signal.

## III. Related Work

This section presents the existing work related to isotope detection, machine learning applied to such tasks, and class imbalance.

### A. Hypothesis Testing for Isotope Detection

The hypothesis testing algorithm for isotope detection is the preferred method resulting from Health Canada's experiments, and is analogous to a multi-class classifier. It is based on a set of one or more user-defined *regions of interest*, each associated with an isotope of interest. The advantage of this is that an alarm can be triggered for more than one isotope at any given time.

In terms of application, when presented with a novel spectrum, one confidence value is produced for each region of interest (ROI), based on the area under that region. In order for an alarm to be generated, the confidence value must exceed the associated alarm threshold for the region, the definition of which is not tied to the algorithm in general. The specific thresholds were set according to data sampled from the background and the intuition of the analysts at Health Canada.

Despite having the advantage of being able to detect multiple isotopes, it's fundamental weakness is it's lack of sensitivity. This motivates the use of machine learning for detecting isotopes.

### B. Machine Learning and Gamma-Ray Spectroscopy

Machine learning algorithms first appeared in the context of spectral analysis in the early 1990s. These initial studies were, in essence, laboratory experiments aimed at determining the effectiveness of machine learning algorithms at identifying the presence of a single isotope of interest in test spectra.

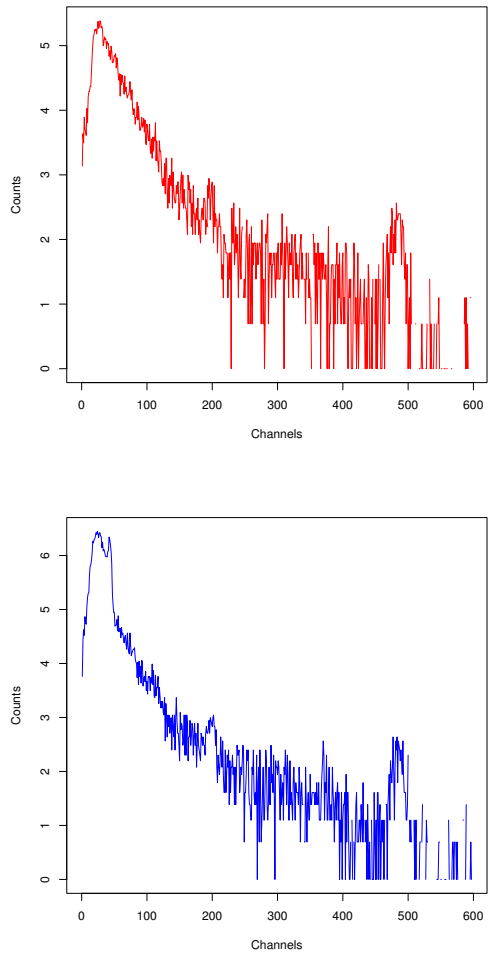


Fig. 1. These figures contain randomly selected examples from the VAN dataset. Plotted on top in log-scale is a background instance. Below this is an instance containing Technicium-99. As can be seen, the high counts in the Tc-99 spectrum below channel 100 represent are indicative of a weakly shielded Tc-99 sample.

Olmos *et al.*, in [12] and [13], for example, applied a *linear associative memory neural network* to identify the presence of $^{133}Ba$ and $^{57}Co$ in spectra recorded during laboratory experiments, and found that the results represented an improvement over traditional 'peak-fitting' strategies. Abdel-Aal and Al-Haddad reported improved results [14], when applying *abductive* machine learning to identify a small set of radioisotopes in gamma-ray spectra. More recently, Kangas *et al.* reported the results of applying *multilayer perceptrons* [15] to analyze the shape of low resolution polyvinyl toluene spectra data acquired from port monitoring technology. Multilayer perceptrons were also applied by Vignerson *et al.*, [16], to determine $^{235}U/U_{total}$ ratios, and Yoshida *et al.* for radionuclide detection in uranium ore [17].

While these studies suggest that machine learning algorithms offer the possibility of improved efficiency and detection, they are far from comprehensive. In this work, and in previous work [2], we seek a deeper understanding of the effectiveness of machine learning for the detection of isotopes of interest, and compare them to those obtained by deployed systems.

### C. Class Imbalance

In gamma-ray spectral classification, the class of interest is rare and the data is high-dimensional. Previously, problems of class imbalance have been dealt with in a variety of ways, including cost-sensitive methods, random undersampling and oversampling, and informed sampling methods. Sampling and cost-sensitive methods have been shown to be competitive and, in some cases, theoretically related [5, 6]. These methods have their strengths and weaknesses [18]; however, they do not add new information to the learning process. Instead, they only adjust the classification bias relative to the existing training data. When the minority training set is small, these approaches risk overfitting the minority class. To prevent this, additional examples are required.

Synthetic oversampling methods apply a generative bias to the minority training instances to synthesize additional training instances. Rifia *et al.*, [19] accurately articulate the objective of modelling from finite training sets as one of deciding upon how to redistribute the probability mass associated with the training instances.

SMOTE [8] assumes that the best place to insert probability mass is between nearest neighbours. Depending on the distribution of the training instances, this can create small clusters of high-density that spread in directions unassociated with the latent distribution. Such a spread is one reason that expensive post-hoc "cleaning" processes, such as the removal of Tomek links [9], must often be applied. In addition to being slow, cleaning processes insert their own biases that are difficult to validate, and utilize distances measures that are generally inaccurate in high-dimensional spaces.

## IV. Data Synthesization

This section presents the two data synthesization methods and demonstrates their functionality over an artificial domain.

### A. SMOTE

SMOTE has shown considerable success [8] in alleviating the class imbalance problem. SMOTE works by generating synthetic minority instances between existing instances of the minority class. Thus, its performance is dependent upon the availability of instances in the minority training set. More specifically, SMOTE finds the $k$-nearest neighbours of each instance, $\mathbf{x}_i$, and generates a synthetic point between $\mathbf{x}_i$ and a randomly selected instance, $\mathbf{x}_j$, in the nearest neighbour set. Formally, the new point $\mathbf{x}_{new}$ is generated as:

$$\mathbf{x}_{new} = \mathbf{x}_i + (\mathbf{x}_j - \mathbf{x}_i) \times \delta \quad (1)$$

where $\delta = [0, 1]$ causes $\mathbf{x}_{new}$ to be placed at a random distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. The direct result of this is that vectors connecting nearest neighbours become populated with uniformly distributed, randomly generated synthetic instances. This is clearly seen in Figure 2 where the synthetic distribution formed by the blue circles spans nearest neighbours in the

training set (red triangles). The result is that the synthetic instances do not accurately represent the target distribution shown in grey. Some of the worst examples cover regions completely outside the target distribution; however, this does not necessarily mean that the impact on performance will be poor.

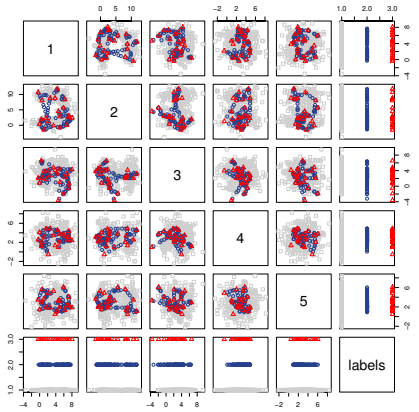


Fig. 2. The synthetic samples resulting from SMOTE method with 20 training instances.

### B. Denoising Autoencoder-Based Minority Oversampling

As an alternative to SMOTE, we present a synthetic sampling technique (DEAGO) that is based on the modelling and reconstructive capabilities of denoising autoencoders. Denoising autoencoders are neural networks that learn to reconstruct clean versions of the network input at the output layer [20].

The general algorithm for modelling and sampling is presented in Algorithm 1. The algorithm takes the training instances $\mathcal{X}$ as an input and performs normalization prior to training the denoising autoencoder, $dea$. $dea$ is then trained with $h$ hidden units for $e$ epochs via gradient descent with back propagation. Subsequent to the creation and training of the network, the normalized sample initiation set is mapped via $dea$ to the induced distribution.

**Algorithm 1** daego$(\mathcal{X}, e, h, \sigma, \mathcal{X}_{init})$

**Input:**

i) $\mathcal{X}$, the instances of the class training set.
ii) $e$, the number of training epochs.
iii) $h$, the number of hidden units in the autoencoder.
iv) $\sigma$, the training noise.
v) $\mathcal{X}_{init}$, the sample initiation set.

**Output:**

i) $\mathcal{Y}$, the synthetic samples.

**Method:**

1: Normalize the training dataset $\mathcal{X}$ into $[\mathcal{X}_{norm}, norm\_params]$.
2: Apply the normalization parameters $norm\_params$ to the sample initiation set $\mathcal{X}_{init}$.
3: Create a denoising autoencoder network $dea$ from the parameter set $e$, $h$, $\sigma$.
4: Train the network on the normalized training data $x\_norm$.
5: Map $\mathcal{X}_{init}$ to the induced manifold via $dea$.
6: Denormalize the mapped synthetic instances $\mathcal{Y}$ with the normalization parameters $norm\_params$.
7: Return $\mathcal{Y}$

**End Algorithm**

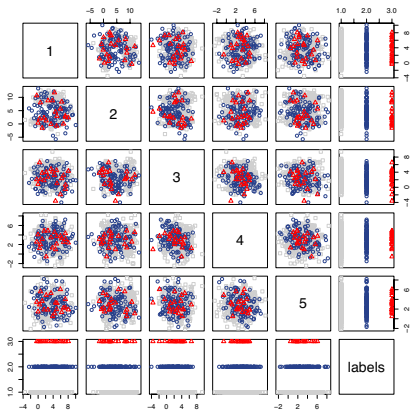


Fig. 3. DEAGO synthetic samples generated the training instances with Gaussian noise added for the sample initiation points on the cData with 20 training instances. This synthetic set has the lowest density difference relative to the target set.

The sample initiation points themselves can have an impact on the synthetic set depending on how they are selected. This is due to the fact that the induced function $g(f(\tilde{x}))$ maps the input $\tilde{x}$ learn distribution. Thus, if the sample initiation set is distributed in a small region relative to the learnt distribution, the resulting samples will populate the nearby region of the induced distribution. This can be an effective way to bias the induced classifier in favour of the minority class.

The output, $\mathcal{Y}$, of the DEAGO algorithm is denormalized and returned for use in inflating the minority training set. Our empirical results suggest the use of $\mathcal{X}_{init} = \mathcal{X} + \mathcal{N}(\cdot)$ for a synthetic set that covers all regions of the induced distribution. Alternatively, the synthetic set can be biased towards the majority class by utilizing the majority class instances in $\mathcal{X}_{init}$. Finally, the complexity and generality of the function induced by DEAGO is controlled by the number of hidden units $u$ and training noise $\sigma$.

Figure 3 presents the generative result of DEAGO in the same context that SMOTE was previously reviewed. It is clear from this that, unlike SMOTE, DEAGO induced a global model based on the training set that appears as red triangles, and that, as a result, the synthetic samples (blue circles) better represent target distribution shown in grey.

## V. Experimental Methodology

As previously described, the gamma-ray classification performed in this study is intended to identify isotopes of interest for the purpose of maintaining health and security. The volume of data produced makes it impractical to rely solely on human analysis. Thus, machine learning, particularly classification, has the potential to have significant impact on this domain.

TABLE I. THE MEAN 5X2-FOLD CROSS VALIDATED AUC RESULTS FOR SAAN AND VAN.

| | SAAN | | VAN | |
|---|---|---|---|---|
| | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ |
| DEAGO | 0.894 | 0.049 | 0.989 | 0.002 |
| SMOTE | 0.818 | 0.086 | 0.939 | 0.032 |
| BRUS | 0.742 | 0.048 | 0.836 | 0.072 |
| MLP | 0.786 | 0.149 | 0.840 | 0.087 |

This section briefly describes the methodology applied to evaluate the performance of each method on the gamma-ray spectral classification tasks.

In order to validate the performance of each method, we run 5x2-fold cross validated (CV) experiments on the SAAN and VAN datasets. The AUC for each combination of imbalance correction technique and classifier is recorded for each iteration of CV and the mean and standard deviations are reported. The AUC provides an excellent general assessment of the classifier as it is not affected by the degree of imbalance in the test set, and 5x2-fold CV is used in place of the more common 10-fold CV as it has been observed that 5x2 CV has a lower probability of issuing a Type I error as compared to k-fold CV [21]. Furthermore, the amount of test data available in 5x2-fold CV is larger in comparison to 10-fold CV.

We present results based on the SMOTE algorithm and a novel form of synthetic oversampling that is based on autoencoders, which we have denoted DEAGO. For completeness, we also include the results of BRUS, as this method has seen success on some imbalanced domains. In each case, subsequent to the rebalancing phase, we apply a multilayer perceptron classifier (MLP). Future work will consider the performance of other state-of-the-art classifiers.

## VI. RESULTS

This section reports the results of our experiments on gamma-ray classification. The 5x2-fold cv tests on the SAAN and VAN data are performed using MLPs trained on the natural distribution, and with data balanced via BRUS, SMOTE and DEAGO. Finally, the results of the top two methods (DEAGO and SMOTE) are tested for statistical significance using the t-test. The columns of the table specify the corresponding means and standard deviations for each dataset.

In both cases, BRUS decreases the AUC produced by the MLP classifier. This is likely due to the degree of imbalance in the datasets, which necessitates that a large number of majority class instances be randomly removed in order to achieve a relative balance. Alternatively, both SMOTE and DEAGO, which add novel instances to the training process in order to balance the training set, lead to improvements in performance.

Statistical analysis via the paired t-test shows that the mean AUC of $0.894$ on the SAAN data produced by DEAGO is significantly better than the score of $0.818$ produced by SMOTE. Similarly, the mean AUC of $0.989$ for VAN produced by DEAGO is also a significant improvement over the AUC of $0.939$ produced by SMOTE.

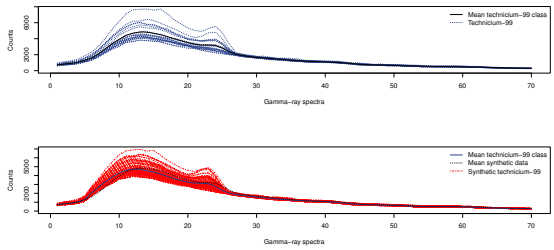


Fig. 4. SMOTE synthetic spectra produced from train 20 examples.

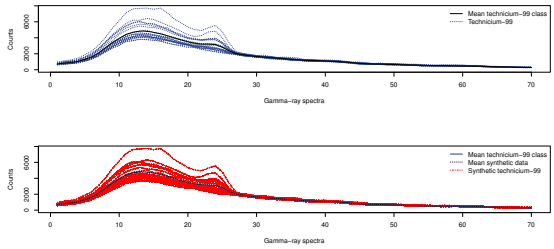


Fig. 5. DEAGO synthetic spectra produced from train 20 examples.

### A. Synthesized Instances

In order to further understand the results of our experiments, we proceed to examine gamma-ray spectra synthesized by DEAGO and SMOTE. Figure 4 and Figure 5 depicts the *isotope of interest* class as blue dotted lines, the mean by the solid line and the synthesized instances as red dashed lines for SMOTE and DEAGO, respectively.

Whilst SMOTE generated many reasonable synthetic instances, it is clear that the distribution of these instances is less representative of the latent distributions. Specifically, the density of the synthetic instances inside the convex-hull make the distributions appear unimodal; we can see, however, from the training instances that the spectra are multimodal below channel 30. Alternatively, DEAGO does a better job of maintaining the latent distribution while adding novel instances.

## VII. DISCUSSION

The results on both gamma-ray spectral datasets demonstrate the considerable promise of DEAGO on gamma-ray spectral domains. Indeed, DEAGO produces statistically significantly better results on VAN and SAAN, with an AUC that is nearly $0.08$ higher on SAAN and $0.05$ higher on VAN. In addition, DEAGO proves to be very consistent in its performance. It produces a standard deviation of $0.002$ and $0.049$ for VAN and SAAN respectively, whereas SMOTE has standard deviations of $0.032$ and $0.086$ and BRUS produces standard deviations of $0.072$ and $0.048$. The consistency of DEAGO is very noteworthy, as many imbalanced classification systems, including those tested here, tend to vary greatly over multiple runs when the minority class is as small as it is here. Once again, this demonstrates that DEAGO is capable of inducing a good model from a synthetic oversampling perspective from small minority training sets.

Finally, although neural networks are sometimes slow to train, the training time is much shorter for this application due to the fact that DEAGO is trained on the small minority training set. In addition, it benefits from not requiring that multiple classifiers be trained, as is the case with BRUS, and no post-hoc cleaning is needed, such as with SMOTE.

## VIII. Conclusion

Given the risks associated with nuclear materials, be they from malfunctions at an industrial facility or terrorist threats, it is of considerable importance that we develop the means to monitor and detect the presence of isotopes of interest. In particular, this work focuses on the identification of isotopes of interest based on gamma-ray spectra sampled from NaI spectrometers on the national monitoring network and those used to secure entrance to the Vancouver 2010 Olympics. Our results demonstrate that by synthetically oversampling the minority class that is composed of the isotopes of interest, we can mitigate the negative effects of imbalance in the domain, thereby improving the classification results. Moreover, our experiments show that DEAGO produces statistically significantly better AUC results than SMOTE on both datasets.

## References


[1] S. R. Dalal and B. Han, "Detection of radioactive material entering national ports: A Bayesian approach to radiation portal data," *Annals of Applied Statistics*, vol. 4, pp. 1256–1271, 2010.

[2] S. Sharma, C. Bellinger, and N. Japkowicz, "Clustering based one-class classification for the comprehensive nuclear test-ban treaty," in *Advances in Artificial Intelligence, Lecture Notes in Computer Science*, D. Kosseim, Leila and Inkpen, Ed., vol. 7310. Springer Berlin / Heidelberg, 2012, pp. 181–193.

[3] V. Barnabe-Lortie, C. Bellinger, and N. Japkowicz, "Smoothing gamma ray spectra to improve outlier detection," in *Computational Intelligence for Security and Defense Applications (CISDA), 2014 Seventh IEEE Symposium on*, 2014, pp. 1–8.

[4] C. Bellinger, S. Sharma, and N. Japkowicz, "One-Class versus Binary Classification: Which and When?" pp. 102–106, Dec. 2012.

[5] M. A. Maloof, "Learning when data sets are imbalanced and when costs are unequal and unknown," in *ICML-2003 workshop on learning from imbalanced data sets II*, 2003, pp. 1–2.

[6] K. McCarthy, B. Zabar, and G. Weiss, "Does cost-sensitive learning beat sampling for classifying rare classes?" in *Proceedings of the 1st international workshop on Utility-based data mining*, 2005, pp. 69–77.

[7] B. C. Wallace, K. Small, C. E. Brodley, and T. a. Trikalinos, "Class Imbalance, Redux," *2011 IEEE 11th International Conference on Data Mining*, pp. 754–763, Dec. 2011.

[8] N. Chawla, K. Bowyer, L. Hall, and K. W.P., "SMOTE: Synthetic Minority Over-Sampling Technique," *J. Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[9] I. Tomek, "Modifications of CNN," *IEEE Trans. System, Man, Cybernetics*, vol. 6, no. 11, pp. 769–772, 1976.

[10] H. Han, W.-y. Wang, and B.-h. Mao, "Borderline-SMOTE : A New Over-Sampling Method in Imbalanced Data Sets Learning," *Advances in intelligent computing*, pp. 878–887, 2005.

[11] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, no. 3, pp. 1322–1328, Jun. 2008.

[12] P. Olmos, J. Diaz, J. Perez, P. Gomez, V. Rodellar, P. Aguayo, A. Bru, G. Garcia-Belmonte, and J. de Pablos, "A new approach to automatic radiation spectrum analysis," *Nuclear Science, IEEE Transactions on*, vol. 38, no. 4, pp. 971–975, 1991.

[13] P. Olmos, J. Diaz, J. Perez, G. Garcia-Belmonte, P. Gomez, and V. Rodellar, "Application of neural network techniques in gamma spectroscopy," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 312, no. 1-2, pp. 167–173, Feb. 1992. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/016890029290148W

[14] R. E. Abdel-Aal and M. N. Al-Haddad, "Determination of radioisotopes in gamma-ray spectroscopy using abductive machine learning," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 391, no. 2, pp. 275–288, 1997.

[15] L. J. Kangas, P. E. Keller, E. R. Siciliano, R. T. Kouzes, and J. H. Ely, "The use of artificial neural networks in PVT-based radiation portal monitors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 587, no. 2-3, pp. 398–412, Mar. 2008.

[16] V. Vigneron, J. Morel, M. C. Lépy, and J. M. Martinez, "Statistical modelling of neural networks in gamma-spectrometry," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 369, no. 2-3, pp. 642–647, 1996.

[17] 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 Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 484, no. 1-3, pp. 557–563, 2002.

[18] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[19] S. Rifai, Y. Bengio, Y. Dauphin, and P. Vincent, "A Generative Process for Sampling Contractive Auto-Encoders," in *International Conference on Machine Learning*, no. 1, 2012.

[20] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked Denoising Autoencoders : Learning Useful Representations in a Deep Network with a Local Denoising Criterion," *The Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.

[21] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural computation*, vol. 10, no. 7, pp. 1895–1923, 1998.