Anomaly Detection in Gamma Ray Spectra: A Machine Learning Perspective

Shiven Sharma, Colin Bellinger and Nathalie Japkowicz
School of Electrical Engineering and Computer Science
University of Ottawa
Email: {sshar009, cbell052, nat}@uottawa.ca
Rodney Berg and Kurt Ungar
Radiation Protection Bureau Health Canada
Email: {rodney.berg, kurt.ungar }@hc-sc.gc.ca

Abstract—With Canadian security and the safety of the general public in mind, physicists at Health Canada (HC) have begun to develop techniques to identify persons concealing radioactive material that may represent a threat to attendees at public gatherings, such as political proceedings and sporting events. To this end, Health Canada has initiated field trials that include the deployment of gamma-ray spectrometers. In particular, a series of these detectors, which take measurements every minute and produce 1,024 channel gamma-ray spectrum, were deployed during the Vancouver 2010 olympics. Simple computerized statistics and human expertise were used as the primary line of defence. More specifically, if a measured spectrum deviated significantly from the background, an internal alarm was sounded and an HC physicist undertook further analysis into the nature of the alarming spectrum. This strategy, however, lead to a significant number of costly and time consuming false positives. This research applies sophisticated machine learning algorithms to reduce the number of false positives to an acceptable level, the results of which are detailed in this paper. In addition, we emphasize the primary findings of our work and highlight avenues available to further improve upon our current results.

I. Introduction

Health Canada's role in monitoring the levels of radiation in the atmosphere surrounding nuclear power plants is long standing. More recently, their involvement, with respect to the health and safety of Canadians, has been expanded into monitoring public events, the objective being to undertake research into automating the identification of concealed radioactive materials that have a potential to pose a significant threat to attendees at large public gatherings.

Machine learning is a vast and ever expanding area of artificial intelligence that presents considerable potential in the domain of radioactive threat detection. In order to facilitate the development and testing of algorithms for this purpose, the NARNIA (New Algorithms for Radiological Network Inferential Alarming) project was created by Health Canada. NARNIA involved the deployment of a series of gamma-ray spectrometers around the various venues at the Vancouver 2010 Olympics for the duration of the games. The result was a repository of datasets that continue to be used in the development of a multi-categorical alarm system. We have subsequently been collaborating on the development of such a system.

Our research into the applicability of Machine Learning solutions to the problem of anomaly detection in gamma-ray spectra led to the observation that spectral readings taken during heavy rain events arise from a probability distribution that is distinct from those coming during the absence of rain, or light to moderate rainfall. As a result, we propose a two-tier system that first separates spectra based on rain and non-rain events using a standard binary classification algorithm, and then detects anomalous readings using the Mahalanobis Distance as an anomaly detection algorithm.

The subsequent sections elaborate on our research and the resulting architecture along with the results. Specifically, Section II provides insight into the datasets used during our research, along with our procedures for converting them into a form suitable for use by Machine Learning algorithms, and the subsequent analysis of the data. In Section III, we discuss the various machine learning paradigms, the challenges posed by their nature with respect to the machine learning algorithms, and with the implications of heavy rain events, as heavy rain elevates levels of radiation due to radon daughter washout from it. Given that Anomaly Detection is at the core of this research, we present the algorithms considered for inclusion in the system in Section IV. The resulting system is discussed in Section V. The experiments undertaken to validate the system are presented in Section VI, and their results are discussed in Section VII. Finally, we conclude the report in Section VIII, with a discussion of possible avenues for future research within this problem domain.

II. Gamma-Ray Spectrometer Data

A gamma-ray spectrometer is an instrument that measures the distribution of the intensity of gamma radiation versus the energy of each photon. The resulting measurements can be plotted as in Fig. 1.

This realization of the data is in log form, which is the standard method for viewing and analyzing such data. Fig. 1 *(i)* and Fig. 1 *(ii)* correspond to a pure background measurement and background plus Technicium, respectively. Energy is represented in terms of channels on the x-axis and the counts, which indicate the intensity, are recorded on the y-axis. The lower channels in the plots correspond with lower
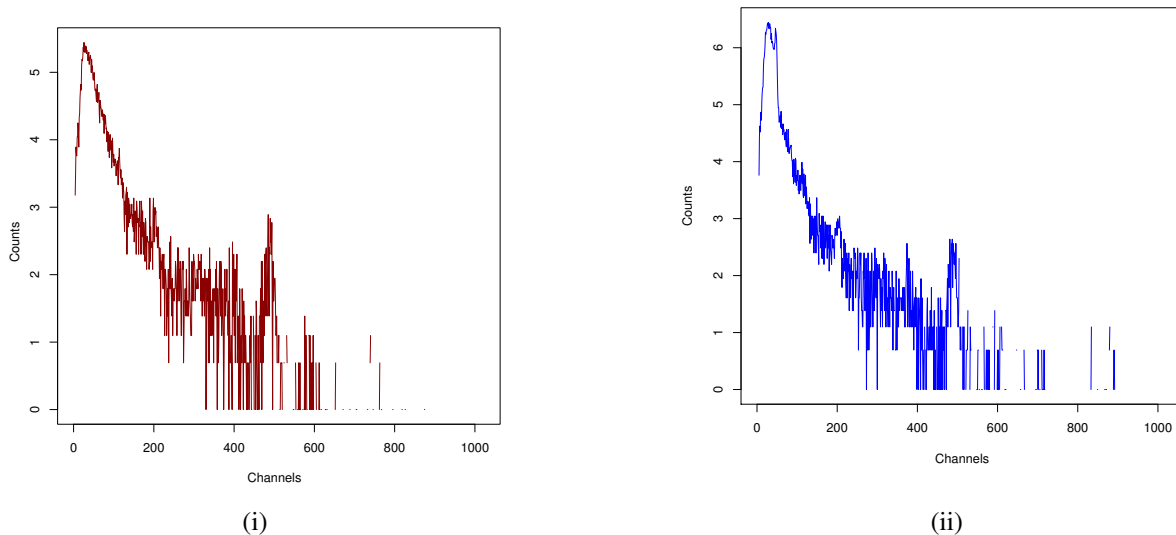
(i)   (ii)

Fig. 1. These figures contain randomly selected examples from the HC dataset. Plotted on the log-scale, sub-figure *(i)* depicts a background instance and sub-figure *(ii)* depicts and instance containing the medical isotope Technicium.

energy photons; the energy increases with the channels. The isotopes of interest in our experiments all peak well below channel 600, and thus, to minimize the affects of the so-called 'curse-of-dimensionality' our algorithms only utilize the first 600 channels (which corresponds to 1800keV).

Health Canada deployed NaI detectors during the Vancouver Winter Olympic Games in 2010 thereby producing a catalogue dataset that can be used for the development and testing of Machine Learning approaches to multi-category alarm systems. The data was collected in one-minute intervals during the one-month period surrounding the games, thereby producing a dataset of 708,480 spectra readings. Upon installation, the detectors provided the means to collect spectral data which could then be utilized for detecting radioactive isotopes of interest.

Data from three stations was utilized in the experiments described below. Apart from the plethora of background measurements recorded at these sites, the readings contained spectra from three medical isotopes, namely Iodine, Thallium and Technicium (one of the stations also had readings for Caesium, which were the result of a check-source). In addition to the medical isotopes that were measured and identified as a result of people passing by the gamma-ray spectrometers, artificial Cobalt spectra at varying signal strengths was generated for use in the development of the system. These were subsequently incorporated into the data.

Finally, to aid in the development of a strong classification system, the data was labelled as belonging to a rain or non-rain class. This weather information was acquired from the Environment Canada website for the period of the Vancouver Olympic Games, and by observing the increase in radiation by examining the spectra.

III. MACHINE LEARNING ALGORITHMS

Given that the crux of our work involves the utilization of machine learning tools to improve the detection capabilities of the system employed by Health Canada, it is prudent to provide a comprehensive overview of machine learning with respect to past work done in related fields, the various paradigms and their associated algorithms and the algorithms chosen for the final system, along with the challenges presented by the given spectral data.

A. Related Work

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, based on models learned from training data.

Olmos *et al.*, in [1] and [2], 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 in [3], when applying *abductive* machine learning to identify a small set of radioisotopes in gamma-ray spectroscopy. More recently, Kangas *et al.* reported the results of applying *multilayer perceptron neural networks* in [4] 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.*, in [5], to determine $^{235}U/U_{total}$ ratios, and Yoshida *et al.* for radionuclide detection in uranium ore [6].

While this study suggest that machine learning algorithms offer the possibility of improved efficiency and detection, the studies are far from comprehensive, and generally rely on

synthetic or overly simplistic laboratory data. In addition, the heavy reliance on static artificial neural network-based solution shows good promise, but leave open the question of dynamic and non-dynamic artificial neural network solutions.

B. Machine Learning: Directions Forward

The research area of machine learning is vast, and includes a range of algorithms that offer the possibility of greatly improving isotope and novel event detection. Off-the-Shelf classification algorithms can often improve accuracy while reducing the need for human intervention. The choice between one-class classification (OCC) (alternatively referred to as anomaly detection) and multi-class learning comes down to the practicalities of the data. While multi-class learning is largely accepted as the stronger of the two learning techniques, its performance is contingent on the available data. For multi-class learners to reach their full potential, a representative set of data drawn from each class is required. In domains containing a large number of features, or events that are particularly difficult to measure, accessing sufficient data from all classes can be prohibitive. In such circumstances, as is the case here, one-class learning becomes a more viable alternative. We provide a discussion on the various one-class learning algorithms considered by us in the subsequent section.

IV. Anomaly Detection

One-class learning and Anomaly detection are synonymous; both rely on the detection of unusual, or anomalous, events. Given the importance that this paradigm plays within our proposed system, we explicitly highlight the four anomaly detection algorithms we considered for inclusion into the system.

The four techniques considered for analyzing the probabilistic likelihood of instances belonging to a particular class are the autoassociator (AA), support vector machines (SVM), the Mahalanobis distance (MD) and the Variance in Angle Spectrum (VAS). For our purposes, as we are only interested in the ranking of the instances, and therefore, we do not apply a threshold for classification purposes; we simply order the instances based on the specific metric of the methods, the aim being to discover any groups that form within this ordering, and then examine the majority class present within these groups.

The Autoassociator: A complete description of the autoassociator is beyond the scope of this paper, but we direct the interested reader to [7] for past work and analysis. Briefly, the autoassociator is a form of neural network in which the number of output nodes equals the number of input nodes. In other words, one can visualize it as a sort of a *compression-decompression* network, the task of the network being to be able to, with sufficiently low error, reconstruct the input at the output, by first compressing it within the hidden layer, and thereafter decompressing it at the output nodes. The data used for training such a network comes from a single class.

During the application of a standard AA classifier the Reconstruction Error (RE) between the input and output nodes is calculated. By applying a predefined threshold to this RE,
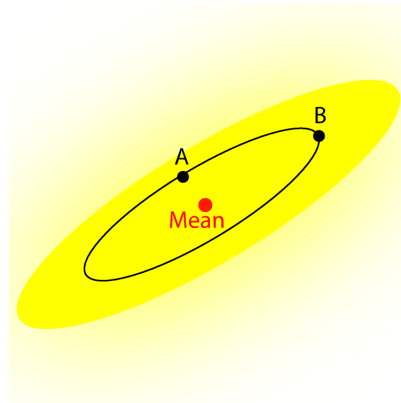

Fig. 2. Illustration of the Mahalanobis Distance between two points **A** and **B** from the mean. Both points have the same Mahalanobis distance, but different Euclidean distances from the mean.

instances can be classified as belonging to either the class on which the network was originally trained (if the error falls below the threshold), or an outlier class (if the error falls above the threshold). However, as thresholding is impractical here, we ordered the instances based on decreasing RE, and use the rankings to identify instances of interest.

Support Vector Machine: Schölkopf *et al.*, in [8], proposed a one-class extension to the existing support vector techniques, for the estimation of support in high-dimensional spaces. In general terms, their approach maps the training data into a dot product feature space, and inserts a hyperplane in a manner that separates the origin from the data with maximal margin. Since ranking is of interest, we can rely on the distance between a novel instance in the test set and the maximal margin hyperplane as the score upon which ranking occurs. Thus, instances with a strong similarity to the background have larger positive distances, instances on the edge of the background distribution obtain values close to zero, and, finally, those that are dissimilar to the background distribution obtain negative distances from the maximum margin hyperplane.

Mahalanobis Distance: The Euclidean distance offers no information with regards to where points lie within the distribution of interest. To this end, we employ the Mahalanobis Distance (MD), a parametric distance-measure used for Gaussian distributions. Intuitively, if one visualizes a Gaussian distribution as being composed of hyperelliptical contours, the MD gives the distance with respect to the position of instances within these contours; specifically, it discriminates between instances lying in different density regions of the distribution. Two points that might be within the same Euclidean distance from the mean might have a very different MD, as one may lie in a low density region, and another may lie in a high density region, and vice versa. Figure 2 illustrates the idea of two points having the same MD from the mean (*i.e.* they lie on the same hyperelliptical contour representing a region of equivalent density) but different Euclidean distance.

The calculation of the MD involves knowing the mean μ and the covariance matrix Σ of the distribution. Once these are known, the MD of an instance x from the mean μ is calculated

as:

$$MD(x, \mu) = (x - \mu)^T \Sigma^{-1} (x - \mu)$$

We rank the instances using their MD from the mean of the distribution underlying the background class, using the distance as a measure of their likelihood of belonging to this distribution.

Variance in Angle Spectrum: The Variance in Angle Spectrum (VAS) [9] uses the variance in the cosine of the angles between an instance p and all other instances $x \in D$, where D is the set of all instances. An outlier is identified as having a low variance. This is based on the rationale that an outlier will most likely lie in a border region of the distribution, and therefore most points will lie in the same direction. A non-outlier point, on the other hand, will be in the middle of the distribution, and thus most points will surround it with equal density, thereby giving a high variance in the angle. The variance for every point $p \in D$ is calculated as:

$$VAS(p) = var_{x,y \in D} \left(\frac{\langle \vec{xp} \cdot \vec{xp} \rangle}{\|\vec{xp}\| \cdot \|\vec{yp}\|} \right)$$

A low value of $VAS(p)$ will imply that p is most likely an outlier, and vice versa.

V. PROPOSED SYSTEM

In this section, we discuss the system that has been devised to detect events of interest in the NARNIA dataset. To date, the system operates primarily as an anomaly detection system, with the onus put on the user to decide upon the most appropriate threshold, along with the determination of the specific isotope, or isotopes, causing the deviation from the norm. However, future work is planned to address both of these issues.

A. Overview

Preliminary work has demonstrated that spectral data of interest, particularly those measured in the presence of medical isotopes (MIs), and those synthetically augmented with isotopes of a greater threat, can be successfully identified with modern machine learning (ML) tools. This ability is of specific importance as it reduces the reliance on human experts, and presents the possibility of increased accuracy. The significant challenge to date has been the result of measurements taken unknowingly during heavy rain events. In particular, when the training spectra contains a series of unlabelled rain and non-rain events, the performance of the ML algorithms is severely hindered. Alternatively, if we have *a-priori* knowledge of the presence of heavy rain, our algorithms can produce strong rankings.

Based on our previous results, we can conclude that if we know in advance that the spectrum of interest was measured during heavy rain, we can accurately attribute it to its appropriate class. Thus, an ensemble classification strategy appears to be of great significance to the success of any ML solution within this domain. We test the hypothesis that standard binary classifiers can be applied to separate spectra recorded during heavy rain events from all others (for simplicity we refer to
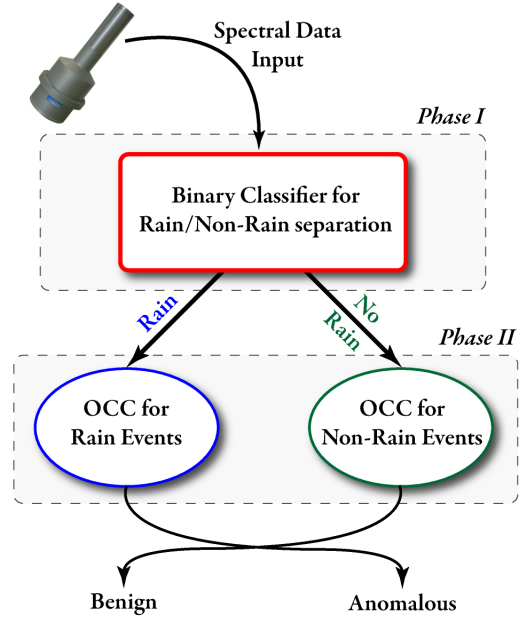

Fig. 3. The proposed architecture for anomaly detection in gamma-rays.

such events as *rain* and *non-rain* events regardless of the fact that the non-rain class may contain some moderate rain events) in the first phase of the ensemble method. Subsequently, spectra of interest can be sent to a classifier that was built either purely from data recorded during heavy rain events, or non-rain events. Fig. 3 depicts the architecture of the ensemble classification strategy.

B. Phase I: Rain separation

In this phase, we rely on the availability of labelled training data to build a classifier that can recognize the presence of a heavy rain event. We are specifically interested in heavy rain events as they perturb the background distribution in such a way that the identification of spectral events of interest becomes significantly more difficult.

A variety of learning strategies may be appropriate for this phase of the system. In theory, there is sufficient rain data in the NARNIA data set to apply binary learning. This is ideal, as binary classifiers are known to produce very strong results. As a result, we have applied a set of binary learners to this problem; the specifics of which are discussed in the following section.

Unfortunately, the labelling processes that is required to produce a training set is extremely time consuming and prone to inaccuracies. Thus, standard binary learning may not achieve its full potential. The identification of background instances measured in the absence of rain, on the other hand, is a much simpler matter. Thus, we hypothesize that one-class learning may produce a better model for separating rain from non-rain events. In addition, methods, such as clustering and expectation maximization are worth considering in the future, along with the range of methods that incorporate temporal context.

TABLE I
AUC VALUES FOR ALL STATIONS, OVER ALL BINARY CLASSIFIERS, FOR BOTH RAIN AND NON-RAIN ANOMALY DETECTION SYSTEMS.

Results over Rain Classifications from Phase I								
Station 6			Station 12			Station 13		
IBK	J48	NBC	IBK	J48	NBC	IBK	J48	NBC
0.993	0.991	0.994	0.998	0.998	0.998	0.999	0.999	0.999
Results over Non-Rain Classifications from Phase I								
Station 6			Station 12			Station 13		
IBK	J48	NBC	IBK	J48	NBC	IBK	J48	NBC
0.992	0.993	0.994	0.999	0.999	1.000	1.000	0.999	1.000

C. Phase II: Isotope/Anomaly Detection

Subsequent to rain classification, the instance is passed to the appropriate rain or non-rain model, where it is assessed for its normality. Specifically, is this another background instance, or is it deserving of further consideration?

Pursuant to our previous discussion regarding the imbalance resulting from the scarcity of spectra of isotopes of interest, this phase relies solely on an anomaly detection algorithm. We have thus far described four different algorithms worthy of consideration. Our initial experiments on obtaining rankings from each of these algorithms have led us to conclude that the Mahalanobis Distance (MD) offers the best results, as opposed to either the Autoassociator (AA), support vector machine (SVM) or the Variance in Angle Spectrum (VAS). Both the AA and VAS have extremely long training times, which increase drastically, proportional to the dimensionality of the data being trained. But perhaps crucially the greatest advantage that the MD has over AA and VAS is the near-Gaussian nature of the spectral data. MD calculates the distance explicitly assuming a Gaussian distribution; given the nature of the data provided, this allows MD to be remarkably accurate, thereby giving rankings that are as close to being probabilistically accurate. It is for this reason that we chose the MD over the others for the second phase of our system.

VI. EXPERIMENTS

This section details the experimental framework undertaken to validate the performance of our two-tier system for anomaly detection. We begin by providing a detailed description of data pre-processing, followed by the training and testing procedures used for validating the system.

A. Data Pre-processing

The initial phases of pre-processing have been described in Section II. Data from the measuring stations was split into two parts, namely background data and medical isotope data. Health Canada had also provided us with artificially generated Cobalt data, which was divided into two parts, rain and non-rain. Each of these contained Cobalt spectra of varying intensities, ranging from 50 (*i.e.* the weakest strength) to 500 (*i.e.* the strongest strength). In accordance with our discussions with the experts at Health Canada, we augmented this data with the medical isotope data from each of the three stations; it was made clear to us that the Cobalt data was station independent, and thus could be used in conjunction with the medical isotopes from all stations.

The issue of high dimensionality was resolved by considering only the energy ranges from 0 to 1800keV, as the isotopes that interest Health Canada all lie below 1800keV.

The background data from each station is further divided into two sets, one for rain events and one for non-rain events. The labelling of the background data into rain and non-rain events is done based on information provided by Health Canada, and the weather information for Vancouver provided by Environment Canada. The purpose of this explicit labelling is to facilitate binary learning for Phase I. Both sets are combined, and then divided into 10 folds, each fold being 90% of the cardinality of the combined set. The remainder of the set is used for validation purposes. This process is called 10-fold cross-validation. The validation set from the background data is augmented with the medical isotope data to produce our final testing set.

B. Training and Testing Procedures

As detailed in previous sections, our system is a two-tiered model. Phase I is comprised of a binary learning algorithm, which aims to induce a discriminant function that can separate normal spectra from spectra observed during heavy rain events. We use three different binary learning algorithms, namely Decision Trees, Naïve Bayes and an Instance Based Learner. The training set extracted from the background data, as described in the aforementioned sub-section, is passed on to the binary learner of Phase I. Once the classifier has been induced, we can label the testing data, comprised of the remainder of the background data along with the medical isotope data, and proceed to Phase II.

Phase II consists of two anomaly detection systems, one for rain events and another for non-rain events. Specifically, we use the Mahalanobis Distance (MD). The Mahalanobis distance relies on the calculation of a mean and a covariance matrix. These are calculated using the training data; the matrices for the rain system are calculated using the rain events from the background training set, and conversely, the matrices for the non-rain system are calculated using the non-rain events from the background training data. Note that we only use the background data for calculating the matrices, since the anomalies we expect to find are relative to the background.

Once we have our matrices, the rain and non-rain datasets of the testing data resulting from Phase I are passed to the appropriate anomaly detection system, and once the MD has been calculated, we obtain our final rankings. It is expected that the spectra displaying the greatest anomaly to the mean would have the largest distances, and vice versa.
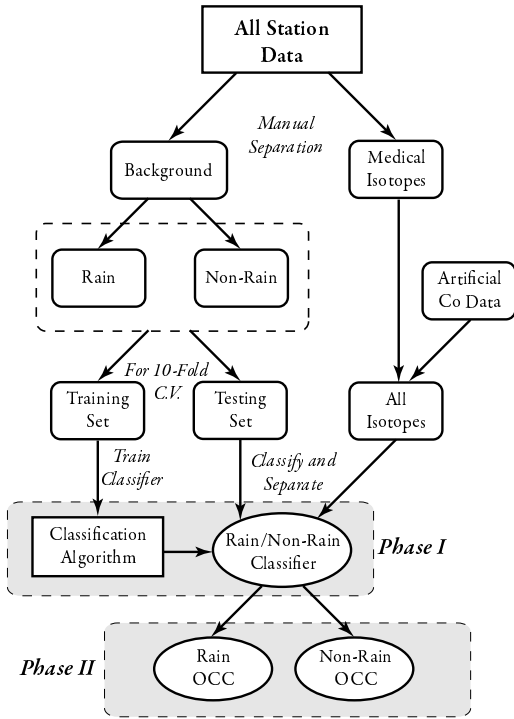
Fig. 4. This illustrates the complete experimental framework undertaken to validate our proposed architecture.

Fig. 4 graphically illustrates the entire experimental framework. The following section will present the results obtained based on this framework.

VII. RESULTS

The results obtained from the experiments conducted, as described in the previous section, are presented here. We first produce the Area Under the ROC Curve (AUC) values for each of the three binary classifiers, for each station. This is followed by the ROC plot for NBC on station 6. In the interest of brevity, we have included only the ROC plot for the best classifier on the most difficult station, station 6.

In order to verify whether the classifier induced during Phase I does indeed provide an improvement over the final rankings obtained from Phase II, we also present the AUC values obtained in the absence of Phase I. Specifically, we build the matrices for MD using both rain and non-rain background data, and produce distance rankings using a testing set comprised of both rain and non-rain background and medical isotope data. We conclude the section with a discussion elaborating on the conclusions derived from these results.

A. Results from the Two-Tier Anomaly Detection System

Table I presents the AUC values obtained from the rankings produced by the system over all stations, using each of the three binary classifiers, during Phase I.

The ROC curve for the results obtained by Health Canada for station 6 is given in Fig. 5. For our architecture, in all cases, the Naïve Bayes Classifier (NBC) produces the best results, and thus, we display the associated ROC curve for
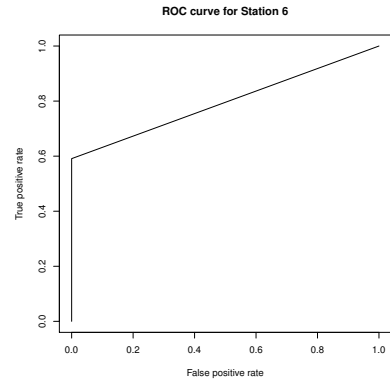

Fig. 5. The ROC plot for Station 6. This plot was generated using the results obtained by utilizing the techniques employed at Health Canada.

NBC at stations 6, for both rain and non-rain events in Fig. 6. It can clearly be seen that incorporating Machine Learning for anomaly detection results in a marked improvement in performance. Moreover, this is, indeed, the case for the data collected at the other stations as well.

B. Results from no rain separation

Here we display the results obtained when no binary classifier is induced in order to separate rain spectra from non-rain spectra and pass it to the appropriate anomaly detection system. Consequently, only a single anomaly detection system is used. Table II shows the AUC values for this system.

TABLE II
AUC VALUES FROM NO RAIN SEPARATION.

Station	AUC Value
Station 6	0.991
Station 12	0.999
Station 13	0.999

C. Discussion

The AUC values for NBC in Table I demonstrate the spectacular performance of the two-tier system. Indeed, an AUC value of 1 is one that is the ultimate goal of every classifier; it implies a perfect ranking, devoid of any classification errors. This can be attributed primarily to the nature of the probabilistic distribution of the individual channels of the spectral data, a fact alluded to in Section V-C. The near-Gaussian distribution of each channel makes the feature space of the domain exceptionally conducive to the application of the Mahalanobis Distance (MD). This in turn produces the results that we see here.

It is interesting to note that even without using a binary classifier to split testing data into rain and non-rain classes, the MD based anomaly detection system still produces exceptional AUC values, as seen in Table II. In comparison with results from Table I, however, we note that the two-tier system outperforms the single-tiered system over all stations. Even though the differences in the results may appear to be minuscule, it is prudent to be wary of the fact that a minute difference
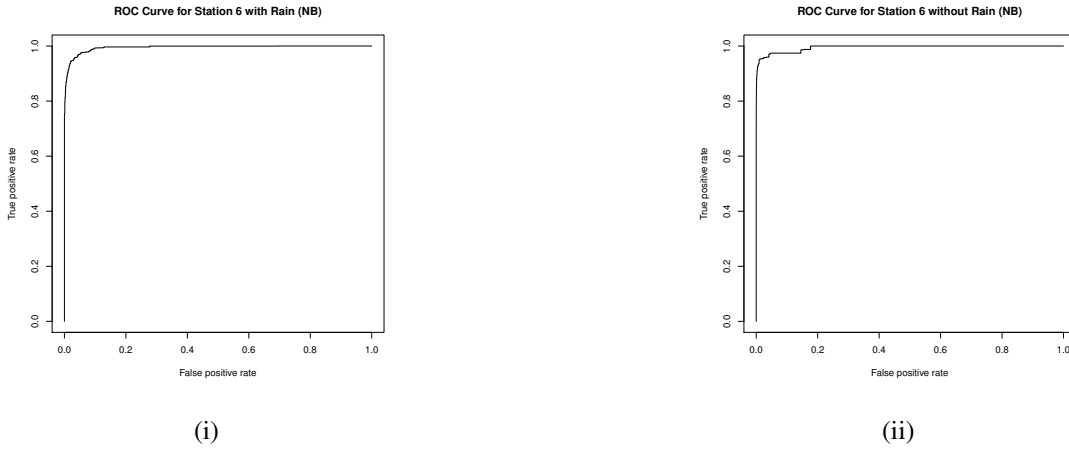
(i)

(ii)

Fig. 6. The ROC curves for Station 6 for both rain and non-rain events, using the Naïve Bayes Classifier and the Mahalanobis Distance..

TABLE III
TOTAL NUMBER OF TRUE POSITIVES (TPS) AND FALSE POSITIVES (FPS)
INCURRED WITH (2-TIER) AND WITHOUT (1-TIER) RAIN SEPARATION.

Station 6			
1-Tier		*2-Tier*	
TPs	FPs	TPs	FPs
3758	336	3950	144
Station 12			
1-Tier		*2-Tier*	
TPs	FPs	TPs	FPs
3856	98	3918	36
Station 13			
1-Tier		*2-Tier*	
TPs	FPs	TPs	FPs
3900	24	3914	10

in AUC values can still imply a large difference in correct classifications. In other words, the single-tiered system ranks more background instances over medical isotopes, and vice versa, as compared to the two-tier system. This is illustrated in Table III, where we can see that without inducing a rain separating classifier, the single-tier system gives more than twice as many false alarms as the two-tier system. Note that the true positive and false positive values are given for the Naïve Bayes classifier based two-tier system, since this classifier provided the best AUC values overall.

VIII. CONCLUSION AND FUTURE WORK

The primary objective of the project undertaken was to devise an automated system that could identify and alarm appropriately on anomalous gamma-ray spectral readings, with little, if any, intervention on the part of the human personnel involved. Machine Learning tools were considered in order to achieve the stated objectives, and the result was a two-tier system for spectral anomaly detection. The two-tier design was devised based on the observation that spectral readings acquired during heavy rain events had been identified as being detrimental to effective anomaly detection, and therefore would need to be filtered out *a priori*. Thus, in Phase I of the two-tier system, we induce a binary classifier that learns to discriminate between spectral readings occurring during heavy rain events, and readings occurring during all other conditions. Spectra passed to this classifier would be classified into either case, and passed on to the appropriate anomaly detection systems (*i.e.* one for either rain or non-rain events), which comprise Phase II of the two-tier design. The Mahalanobis Distance is used as the anomaly detector in this phase, as the Gaussian assumption inherent in its calculations conforms exceedingly well with the probability distributions of the individual energy channels of the data. The final product of the system is a ranking of spectral readings based on their Mahalanobis Distance from the appropriate background means, with higher values indicating a larger anomalous nature.

The results obtained from our experiments validate the employment of a two-tier design for automating anomaly detection. AUC values obtained by using the Naïve Bayes classifier during Phase I indicate that an overwhelming majority of medical isotope spectra are ranked above the background spectra. Of particular importance is the fact that the spectra for Cobalt, specifically spectra with weak signal strengths, are also ranked above the background spectra. This is a vast improvement over the systems employed at Health Canada, as we were informed that their systems could not identify spectra with strengths below 200 (the strengths ranged from 500, being the strongest, to 50, being the weakest). This further attests to the strength and accuracy of our anomaly detection system.

However, there still remains scope for further research into improving the efficiency of our system. The output of the system, as of now, is a list of rankings based on the Mahalanobis distance. Ideally, what would be required is a threshold on the distance that would split the output into two separate classes, one on which an alarm would be required, and the other which is benign. The selection of an appropriate threshold, then, is an area that we will further look into.

A major limiting factor in Machine Learning research, especially in domains similar to spectral anomaly detection, is the scarcity of data from particular classes, which, ironically, tend to be classes we are most interested in identifying. This has been alluded to previously in this paper. Specifically, the isotopes we are most interested in identifying have no

representative spectra in the data provided. This naturally complicates accurate identification of said isotopes. One possible method to alleviate the problem of data scarcity is to generate artificial data that would best reflect real data, if it were available. However, as can be expected, the quality of artificial data would greatly affect the resulting classifier induced through it. We are currently working with the experts at Health Canada in obtaining data relating to isotopes of major interest that reflects genuine spectra as accurately as possible. Upon obtaining these spectra, we will conduct further experiments, using them in conjunction with the isotopes already available, and evaluate the efficiency of our system, modifying it appropriately to conform with the new data.

To conclude, we have developed a two-tier system for ranking spectra based on how anomalous they are from the background spectra. The rankings produced are significantly better than those produced by manual inspection, particularly in the case for Cobalt. Furthermore, the inclusion of a classifier, prior to anomaly detection, to separate spectra based on whether or not they were obtained during heavy rain events or not, greatly reduces the number of false positives, as opposed to the case where *a-priori* separation is not undertaken. We have identified several areas which are deemed worthy of future research that have the potential to lead to further improvements within our spectral anomaly detection system.

REFERENCES

[1] 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.

[2] 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, 1992.

[3] 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 Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 391, no. 2, pp. 275–288, 1997.

[4] L. J. Kangasa, P. E. Kellera, E. R. Sicilianoa, R. T. Kouzes, and J. H. Elya, "The use of artificial neural networks in pvt-based radiation portal monitors," *Nuclear Instruments and Methods in Physics Research*, vol. 587, pp. 398–412, 2008.

[5] V. Vigneron, J. Morel, M. Lepy, and J. 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.

[6] 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.

[7] N. Japkowicz, C. Myers, and M. Gluck, "A novelty detection approach to classification," in *International Joint Conference on Artificial Intelligence*, vol. 14, 1995, pp. 518–523.

[8] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," *Advances in neural information processing systems*, vol. 12, pp. 582–588, 2000.

[9] H. P. Kriegel, M. Schubert, and Z. Arthur, "Angle-based outlier detection in high-dimensional data," in *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '08. New York, NY, USA: ACM, 2008, pp. 444–452. [Online]. Available: http://doi.acm.org/10.1145/1401890.1401946