Classification Analytics of ^{239}Pu and ^{235}U Source Signatures Using Gamma Spectral Regions

Nageswara S. V. Rao, David Abrecht, David A. Hooper, Jennifer Ladd-Lively,
Oak Ridge National Laboratory
David Meier, Pacific Northwest National Laboratory

Abstract

Machine learning detection methods using gamma signatures from spectral measurements of low-intensity ^{239}Pu and ^{235}U sources are studied. NaI detectors located at different distances from the source have been used to collect the training and independent testing data sets. The source is introduced via a shielded conduit into the facility where it is surrounded by 21 NaI detectors deployed over 6 x 6 meters area in the formation of two concentric circles and a spiral. The counts in gamma spectral regions associated with these two sources are estimated at 1 second intervals for each NaI detector, and are used as classifier features for detecting the source presence. Eight different classifiers with five basic properties — namely, smooth, non-smooth, statistical, structural, and hyper-parameter tuning — are trained and tested using the background and source measurements collected over multiple experimental runs. While the overall classifier performance improved as detectors closer to the source are used, some identically produced detectors underperformed but differently between two sources. Some classifiers achieved lower training error but their testing error based on independent measurements is higher for both sources. Overall, these results indicate significant over-fitting by these methods, and illustrate the complexity of training and selecting the machine learning methods to solve these detection problems.

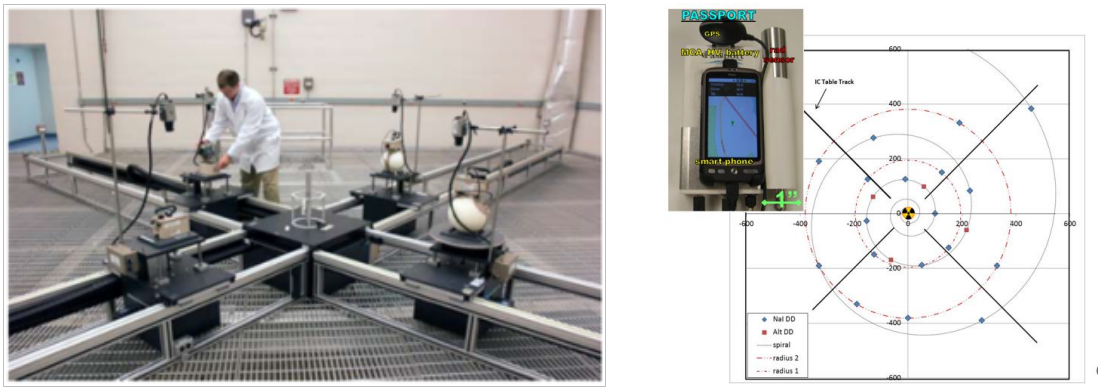
Primary Area

Nuclear Security & Physical Protection Technologies: Nuclear and Radiological Detection Technologies

1 Introduction

Signatures of low-intensity radioactive sources are important for detection tasks in nuclear safeguards, non-proliferation and security areas [1, 2]. We study the signatures of ^{239}Pu and ^{235}U

Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).



(a) *LSI faciity* (b) *detector formation*

Figure 1: *Gamma spectra of* ^{239}*Pu and* ^{235}*U collected using 21 Passport NaI detectors.*

sources for detection analytics using gamma spectra collected by portable NaI detectors. In general, low-intensity ^{239}Pu and ^{235}U sources may be encountered in different forms during their fuel cycle [3]. For example, signatures of ^{239}Pu may be found during the production steps involving target irradiation and radiochemical dissolution, and those of both sources may be found during the post-production transport [3]. The detection of such sources using gamma spectra in these contexts constitutes a specific measurement-driven subclass of inverse problems studied in nuclear forensics and related areas [4]. The detection of these low-intensity sources is particularly challenging in these environments since the Poisson distribution of gamma measurements makes it hard to distinguish them from the background. The problem of using gamma measurements to detect, localize and measure the strength of a radiation source, has been studied under various formulations using a wide variety of techniques including, maximum likelihood estimation, Bayesian estimation, sequential probability ratio test, particle filters, machine learning, and others. In particular, the machine learning (ML) methods are receiving increased attention due to the availability of software and data sets [5, 6, 7].

Our objective is to gain insights into ML methods for detecting these sources, referred to as classifiers, using hand held gamma spectra detectors that can be easily deployed in the field, for example, NaI detector by Passport Systems and CsI detector by Kromek. The application of ML methods to these problems is complex often requiring judicious selection and use of training data sets and classifier methods [5, 8]. In this paper, we study these aspects for ^{239}Pu and ^{235}U sources using data sets from structured experiments. We utilize the gamma spectra collected at the Low Scatter Irradiator (LSI) facility at Savannah River National Laboratory (SRNL) using 21 NaI detectors from Passport Systems, shown in Fig. 1. The activity levels in spectral regions associated with ^{235}U and ^{239}Pu sources are estimated as counts at 1 second intervals, and are used as features to train the classifiers for detecting the presence of a source. Previous results on ^{235}U source indicated complex patterns in terms of variations in the quality of detector spectra and the errors of classifiers expressed as a function of detector distance from the source [9, 10]. In this paper, we study the signatures of ^{239}Pu and compare them to ^{235}U results, with the addition of a hyper-parameter tuning ML method. The eight different ML methods are chosen to represent their diversity of design, namely smooth and non-smooth, statistical and structural methods, and one composite method that tunes their hyper parameters and selects among them [11, 12, 13]. The *error profile* of a ML classifier to detect a source is its classification error (including false alarms and missed detection) expressed as a function of the detector distance from the source.

Our results show that the error profile of ML classifiers improves overall but not strictly monotonically as measurements from the detectors closer to source are used. Spectra from some de-

Bin #	Lower Bound (keV)	Upper Bound (keV)	ISOTOPE(s)
01	42	86	^{241}Am (Americium 241) ^{210}Tl (Thallium 210)
02	64	103	^{133}Ba (Barium 133) ^{109}Cd (Cadmium 109)
03	105	145	^{57}Co (Cobalt 57) ^{239}Pu (Plutonium 239)
04	123	160	^{99m}Tc (Technetium 99m) ^{235}U (Uranium 235)
05	166	203	^{235}U (Uranium 235)
09	384	442	^{239}Pu (Plutonium 239)

Table 1: Spectral bins used for estimating the counts for the sources.

tectors resulted in lower quality signatures although all detectors are identically produced and configured. These under-performing detectors are different for the two sources. Among the ML methods, three classifiers achieved the lower training errors. However, their testing error based on independent measurements is higher in a similar way for both sources, indicating significant over-fitting by these methods.

The organization of this paper is as follows. The data sets are briefly described in Section 2, and the detector spectral measurements and features are described in Section 3. The ML error profiles and the performance of the detectors are described in Section 4. The performance of three groups of detectors based on their distance to the source are described in Section 5. A summary of our contributions and directions for future work are described in Section 6.

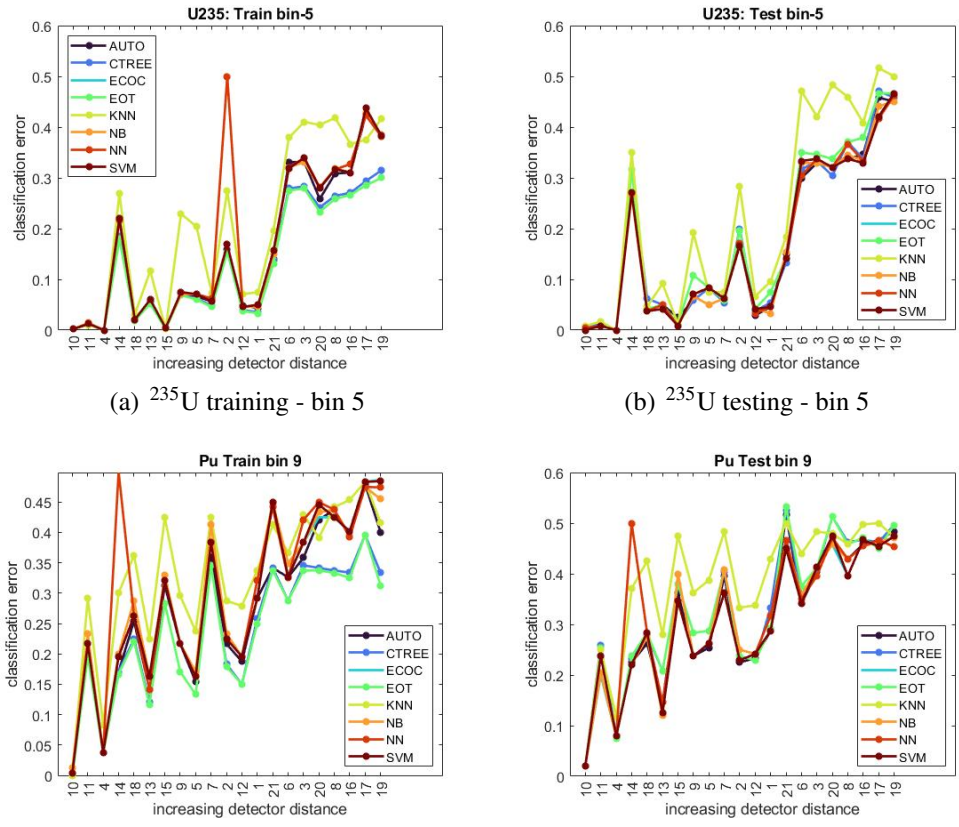
2 IRSS Measurements and Data sets

Indoor and outdoor NaI detector configurations are used with multiple source strengths and types, different background profiles, and various types of source and detector movements to conduct structured experiments under the Domestic Nuclear Detection Office's (DNDO) Intelligent Radiation Sensor System (IRSS) program [7]. In this paper, we utilize measurements collected using ^{235}U source of 191 uCi strength and ^{239}Pu source of 800 uCi strength during these experiments conducted at the SRNL LSI facility. In both cases, the source is introduced at the center via a shielded conduit, and is stationary during the tests, and the detectors are arranged in the fixed geometric pattern shown in Fig. 1(b), which consisted of two rings with radii 2 and 4 meters with 4 and 5 detectors, respectively, and a spiral of 12 detectors. These detectors are identically produced and configured, and the variations in their spectral quality in terms of derived source signatures are attributed to their NaI sensor material. For each source, multiple runs were executed some of which are used for training, and others for independent testing. In addition, several background runs were carried out with no source present. Gamma spectrum from each detector is collected every second, and each run using the source lasted 120 seconds and each background run lasted 60 seconds. We utilize six runs with ^{235}U and ^{239}Pu sources and ten background runs with no source. The detectors form three groups of seven: the *inner group* with seven nearest to source, the *middle group* with next seven nearest, and the *outer group* with rest of detectors.

3 Features and Scenarios

The spectrum from each detector is mapped into 21 spectral bins corresponding to a set of isotopes and the detector's energy resolution [7], and a partial set corresponding to eight isotopes (including ^{235}U and ^{239}Pu) with the corresponding bin numbers and energy bounds are shown in Table 1. The counts in 21 spectral bin are estimated at 1 second intervals, and the appropriate ones are used as features to train classifiers for detecting the presence of a source. For ^{235}U source, counts in bin 5 that corresponds to 166 - 203 keV range are used as a primary classifier feature, and those in bin 4 that corresponds to 123 - 160 keV may additionally include those from ^{99m}Tc source. Similarly, for ^{239}Pu source, bin 9 that correspond to 384 - 442 keV is a primary classifier feature, whereas bin 3 that corresponds to 105 - 145 keV may additionally include those from ^{57}Co source. More details about measurements and count estimates from different detectors can be found in [9]. Eight classifiers are trained and tested using the background and source measurements collected over multiple experimental runs under two scenarios, wherein each background runs consists of 60 spectra and each source run consists of 120 spectra. They are collected at the rate of 1 per second, and the counts in bins are estimated for each spectrum, and used in two scenarios:

- **Scenario S1:** Training with 240 counts from 2 background runs and 1 source run, and testing with 240 counts from different sets of 2 background runs and 1 source run.
- **Scenario S2:** Training with 480 counts from 4 background runs and 2 source runs, and testing with 840 counts consists of different sets of 6 background runs and 4 source runs.



(a) ^{235}U training - bin 5

(b) ^{235}U testing - bin 5

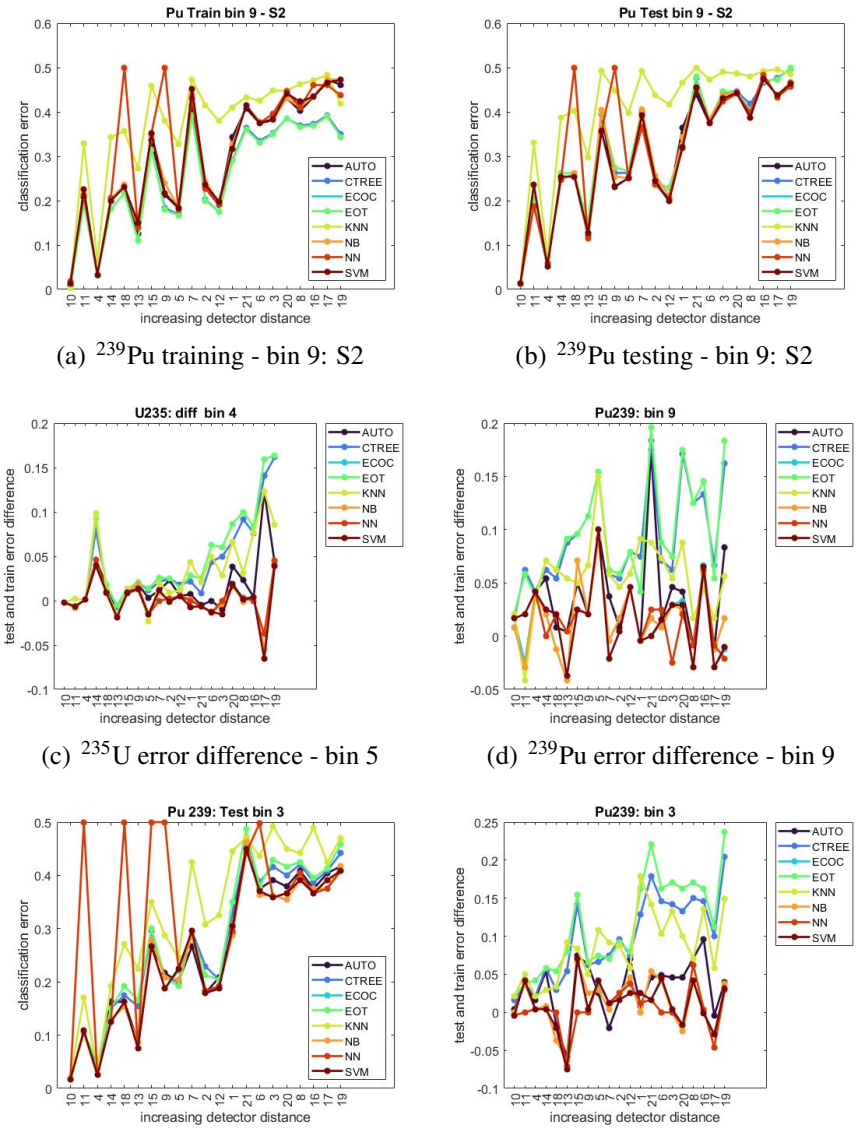
(c) ^{239}Pu training - bin 9

(d) ^{239}Pu testing - bin 9

Figure 2: *Training and testing error profiles of ML classifiers for detecting ^{235}U and ^{239}Pu expressed as a function of increasing detector distance.*

4 Detection and Device Performance

Eight ML methods or classifiers are tested using the data sets from scenarios S1 and S2, and their performance is assessed using the error profiles plotted as a function of detector distance from the source, as shown in Figs. 2 and 3.



(a) ^{239}Pu training - bin 9: S2

(b) ^{239}Pu testing - bin 9: S2

(c) ^{235}U error difference - bin 5

(d) ^{239}Pu error difference - bin 9

(e) ^{239}Pu testing - bin 3

(f) ^{239}Pu error difference - bin 3

Figure 3: *Testing and training error profiles and their differences for ML classifiers for detecting ^{235}U and ^{239}Pu expressed as a function of increasing detector distance.*

4.1 ML Error Profiles

Seven different classification methods and a composite method, provided by the Matlab ML toolkit, are tested, and brief descriptions of the following six classifiers are provided in [5] (more detailed descriptions and analytical aspects of classifiers can be found in several references, for example [14, 15, 16]).

Classification Trees (CTREE)	Error Correcting Output Codes (ECOC)
Ensemble of Trees (EOT)	k-Nearest Neighbors (KNN)
Naive Bayes (NB)	Support Vector Machine (SVM)
Neural Network (NN)	Auto Tuning and Selection method (AUTO)

The seven individual classifier methods represent a variety of designs: CT and EOT are tree-based with non-smooth classification functions; kNN is based on the nearness concept in the feature space; NN, SVM and ECOC utilize smooth underlying functions; and NB is based on statistical principles. In addition, AUTO uses the hyper-parameter searching of CT, EOT, KNN, NB, and SVM methods and choose one of them based on training data set.

The training and testing errors improve overall for both sources as measurements from the detectors closer to source are used, as indicated by the error profiles in Fig. 2. The KNN method has higher error for most detectors for both sources, and NN has much higher error but only for fewer detectors. The disparity between training and testing errors is higher in CTREE, ECOC, and EOT methods for both sources — they consistently have lower training errors at longer distances, as indicated in Figs. 2(a) and (c) for ^{235}U and ^{239}Pu, respectively for scenario S1. The corresponding testing errors indicate that SVM and NB methods achieve lower or comparable errors as indicated in Figs. 2(b) and (d) for ^{235}U and ^{239}Pu, respectively for scenario S1. For ^{239}Pu, the larger data sets of scenario S2 resulted in a similar phenomenon as shown in Figs. 3(a) and (b), but with less variation and somewhat lower error compared to scenario S1; similar results for ^{235}U are in [9].

The errors of ^{239}Pu are higher than those of ^{235}U for both training and testing as shown Fig. 2, even though their intensities are 800 uCi and 191 uCi, respectively. Specifically, the testing errors of 13 out of 14 detectors are below 20% for ^{235}U, whereas only 3 out 14 detectors have such error for ^{239}Pu.

The difference between testing and training error profiles for ^{235}U and ^{239}Pu are mainly positive for CTREE, ECOC, EOT and KNN as shown in Figs. 3(c) and (d), respectively, which indicates over-fitting to the training data by them. The utilization of bin 3 counts as an alternative classifier feature for ^{239}Pu resulted in similar performance as indicated in Figs. 3(e) and (f); the performance is similar for ^{235}U with the use of bin 4 feature in Fig. 3(c) and using both features as described in [5].

4.2 Detector Device Performance

The training and testing error profiles of classifiers for ^{235}U shown in Figs. 2(a) and (b) have an overall increasing trend except for two outlier detectors 14 and 2, in the inner and middle groups, respectively. The errors of almost all of ML methods of these detectors are higher than others in the corresponding group, which indicates the property of the device rather than classifier method. For ^{239}Pu, similar performance is observed, namely overall increase in classification error with detector distance as shown in Fig. 2(c) and (d), but there is much more variation across the detectors. Detectors 15 has the highest error for 6 out of 8 classifiers in the inner group for ^{239}Pu versus detector 14 for ^{235}U. And, detector 7 has the highest error in the middle group for ^{239}Pu versus detector 2 for ^{235}U. Since all detectors use the same codes for bin count estimation and ML, these differences are mainly attributable to NaI material used in them. Furthermore, the under-performing detectors are different between the two sources.

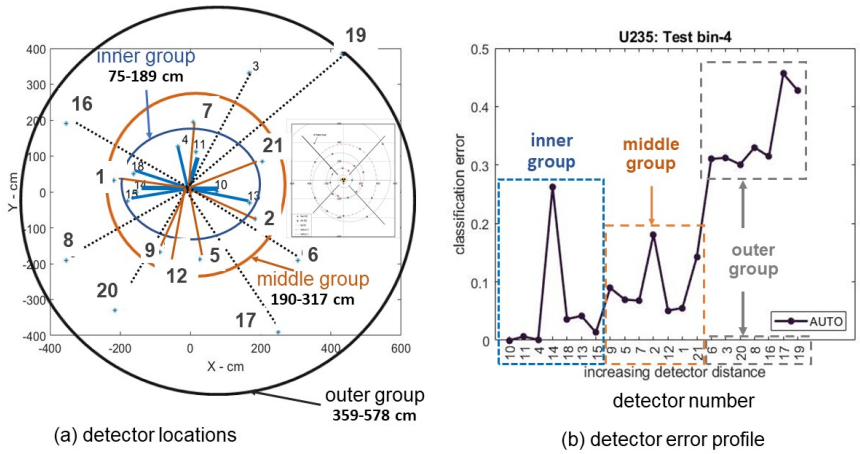


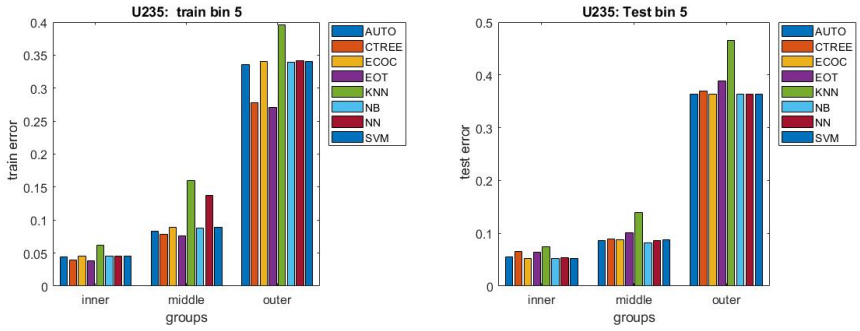
Figure 4: *Inner, middle and outer groups of detectors represent increasing detector distance.*

5 Inner, Middle and Outer Groups

The detectors of three groups of seven, namely, inner (75-189 cm), middle (190-317 cm) and outer (359-578 cm), are shown in Fig. 4. We consider the errors averaged over each group to assess the overall performance of classifiers for ^{235}U and ^{239}Pu in Figs. 5 and 6, respectively. They average out the variations due to individual detector devices within each group and illustrate the overall effects of increased detector distances. There is an increasing trend of training and testing errors from inner to outer groups for ^{235}U as shown in Figs. 5(a) and (b), respectively. There is a similar trend for ^{239}Pu, but the errors are higher and have higher variation among the classifiers within each group as shown in Figs. 6(a) and (b); similar trend is noted for error profiles based on individual detectors in Section 4.1 and also in [9] for ^{235}U. The difference between the testing and training errors is shown for ^{235}U and ^{239}Pu in Figs. 5(c) and 6(c), respectively. Here, the positive values indicate over-fitting, but these difference values need to be interpreted within the context of actual errors; for example, the negative values for NN (Fig. 5(d)) are associated with high training error, which still results in overall high testing error. For ^{239}Pu, these differences are positive (Fig. 6(d)), thereby indicating over-fitting across ML methods and groups. The performance of individual ML methods can be visualized using the corresponding stacked plots for ^{235}U and ^{239}Pu in Figs. 5 and 6, respectively, for training and testing errors. They show lower testing errors for ECOC, NB and SVM for both sources, even though the training errors are lower for CTREE and EOT methods, thereby showing the over-fitting by the latter.

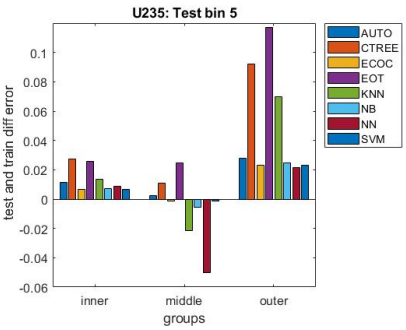
6 Conclusions

We studied the signatures of low-intensity ^{235}U and ^{239}Pu sources from a detection analytics perspective. The detection performance improves as distance from detector to source is decreased, but is not strictly monotonic; indeed, a few devices had very high classification errors despite being close to the source independent of which ML method is used for detection. There are complex patterns even under the controlled environment under which these data sets were collected, namely, clean background, and the stationary source and detectors; they were previously observed for ^{235}U in [9], and this paper confirms them for ^{239}Pu. Several methods provided near-optimal detection of nearby detectors, and their differences became prominent at larger distances, and the error is in general higher for ^{239}Pu compared to^{235}U. For both sources, the classifiers that achieved lowest

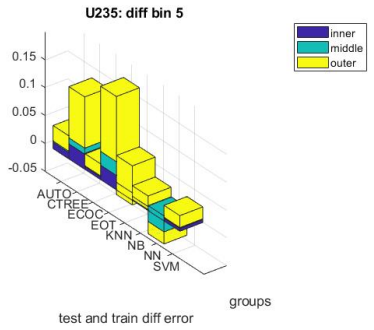


(a) ^{235}U training - bin 5

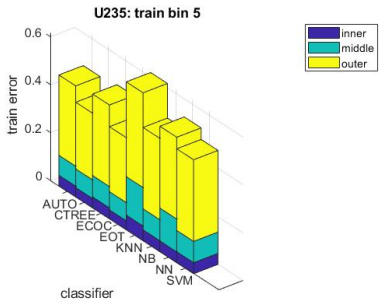
(b) ^{235}U testing - bin 5



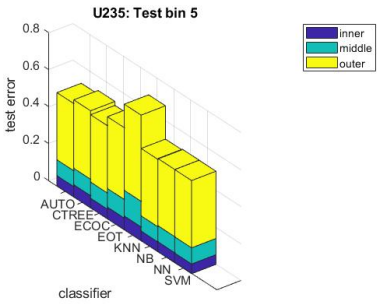
(c) ^{235}U error difference - bin 5



(d) ^{235}U error difference - bin 5



(e) ^{235}U training - bin 5

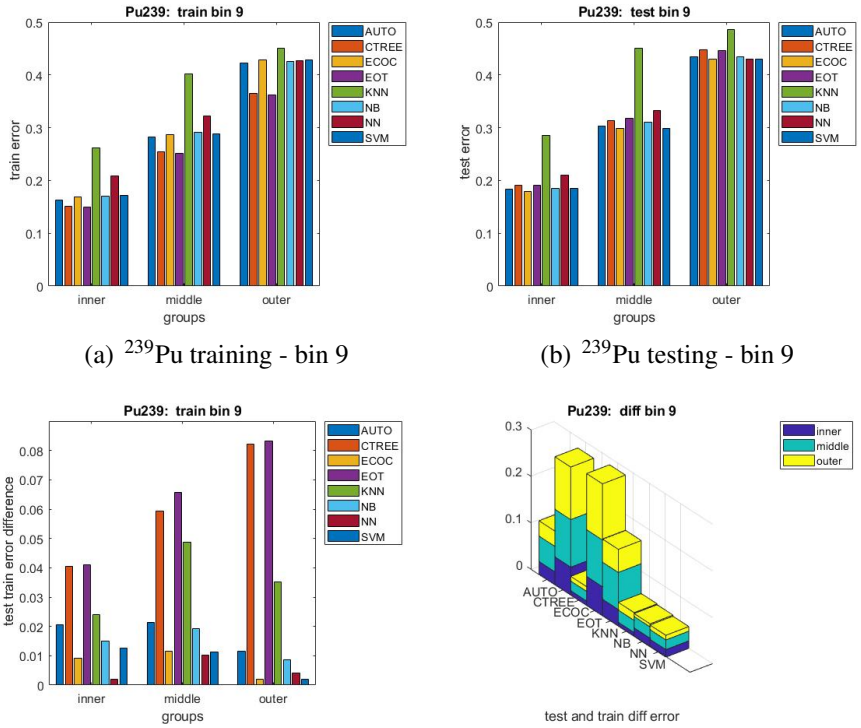


(f) ^{235}U testing - bin 5

Figure 5: *Training and testing errors, and their differences of inner, middle and outer groups for ^{235}U source.*

training error did not result in lowest testing error, which is an indication of their over-fitting to training data.

Future research directions include studying the effects of complex background, particularly in outdoor and field scenarios. It would be of future interest to study the fusion of classifiers to assess conditions under which they outperform individual classification methods for ^{239}Pu. Such approaches, however, have been shown to be vulnerable to over-fitting for ^{235}U sources [10]. Another future direction includes utilizing multiple detectors for ^{239}Pu, which were shown to outperform single detectors in detecting ^{235}U [10] source other radioactive sources [17], and also using analytical methods [18, 19]. Physics-based explanations for the performance ML methods is a topic of future investigation. The detection problem is independently studied in this paper as a binary classification problem of ^{239}Pu and ^{235}U separately, and it would be of future interest to combine them in a multi-class framework.



(a) ^{239}Pu training - bin 9

(b) ^{239}Pu testing - bin 9

(c) ^{239}Pu error difference - bin 9

(d) ^{239}Pu error difference - bin 9

(e) ^{239}Pu training - bin 9

(f) ^{239}Pu testing - bin 9

Figure 6: *Training and testing errors, and their differences of inner, middle and outer groups for* ^{239}*Pu source.*

Acknowledgments

This research was supported by the U.S. Department of Energy, National Nuclear Security Administration, Office of Defense Nuclear Nonproliferation Research and Development (DNN R&D) and was performed at Oak Ridge National Laboratory managed by UT-Battelle LLC for the U.S. Department of Energy under Contract DE-AC05-00OR22725.

References

[1] M. J. Kristo and S. J. Tumey, "The state of nuclear forensics," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 294, pp. 656–661, 2013. Proceedings of the Twelfth International Conference on Accelerator Mass Spectrometry, Wellington, New Zealand, 20-25 March 2011.

[2] K. J. Moody, P. M. Grant, and I. D. Hutcheon, *Nuclear Forensic Analysis*. CRC Press, 2005.
[3] E. L. Connolly and P. G. Martin, "Current and prospective radiation detection systems, screening infrastructure and interpretive algorithms for the non-intrusive screening of shipping container cargo: A review," *Journal of Nuclear Engineering*, vol. 2, no. 3, pp. 246–280, 2021.
[4] J. R. Lewis, A. Zhang, and C. M. Anderson-Cook, "Comparing multiple statistical methods for inverse prediction in nuclear forensics applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 175, pp. 116–129, 2018.
[5] N. S. V. Rao, C. Greulich, S. Sen, J. Hite, K. J. Dayman, A. D. Nicholson, D. E. Archer, M. J. Willis, I. G. R. D. Hunley, J. Johnson, A. J. Rowe, I. R. Stewart, and J. M. Ghawaly, "Classification of dissolution events using fusion of effluents measurements and classifiers," in *Institute of Nuclear Materials Management Annual Meeting*, 2020.
[6] C. Drummond, "An incremental machine learning algorithm for nuclear forensics," in *Advances in Artificial Intelligence* (E. Bagheri and J. C. Cheung, eds.), (Cham), pp. 194–207, Springer International Publishing, 2018.
[7] N. S. V. Rao, S. Sen, M. L. Berry, C. Q. Wu, K. M. Grieme, R. R. Brooks, and G. Cordone, "Datasets for radiation network algorithm development and testing," in *2016 IEEE Nuclear Science Symposium*, 2016.
[8] N. S. V. Rao, C. Redding, D. Hooper, and J. Ladd-Lively, "Selection of training sets for ^{235}u source detection classifiers using gamma signatures," in *Symposium on Radiation Measurements and Applications*, 2023.
[9] N. S. V. Rao, D. Hooper, and J. Ladd-Lively, "Study of classifiers for u-235 source signatures using gamma spectral measurements," in *Institute of Nuclear Materials Management Annual Meeting*, 2022.
[10] N. S. V. Rao, D. Hooper, and J. Ladd-Lively, "On feature, classifier and detector fusers for u-235 signatures using gamma spectral counts," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2022.
[11] E. Alpaydin, *Introduction to Machine Learning*. MIT Press, 2020. fourth edition.
[12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley and Sons, Inc., 2001. Second Edition.
[13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, 2001.
[14] K. P. Murphy, *Probabilistic Machine Learning: An introduction*. MIT Press, 2022.
[15] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. The MIT Press, 2018. second edition.
[16] L. Devroye, L. Gyorfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New York, 1996.
[17] N. S. V. Rao, S. Sen, N. J. Prins, D. A. Cooper, R. J. Ledoux, J. B. Costales, K. Kamieniecki, S. E. Korbly, J. K. Thompson, J. Batcheler, R. R. Brooks, and C. Q. Wu, "Network algorithms for detection of radiation sources," *Nuclear Instruments and Methods Phy. Res. A*, vol. 784, pp. 326–331, June 2015.
[18] N. S. V. Rao, J. C. Chin, D. K. Y. Yau, C. Y. T. Ma, and R. N. Madan, "Cyber-physical tradeoffs in distributed detection networks," in *Internation Conference on Multisensor Fusion and Integration*, 2010.
[19] N. S. V. Rao, C. Y. T. Ma, and D. K. Y. Yau, "On performance of individual, collective and network detection of propagative sources," in *International Conference on Information Fusion*, 2013.