Proceedings of the INMM & ESARDA Joint Virtual Annual Meeting
August 23-26 & August 30-September 1, 2021

FIELD TEST OF A MACHINE LEARNING TECHNIQUE FOR SAFEGUARDS ASSESSMENT AT A FUEL FABRICATION FACILITY

Author
Sean Branney, Ph.D.
Savannah River National Laboratory

Author
Samaria Muhammad, Ph.D.
Savannah River National Laboratory

Author
Ahmad Rushdi, Ph.D.
Sandia National Laboratory

Author
Matthew Sternat, Ph.D.
Sandia National Laboratory

Author
Jesse Jones, Ph.D.
Sandia National Laboratory

ABSTRACT

Many safeguards systems used by the IAEA are extremely expensive because of the need to authenticate the data produced by those systems. Their prohibitive costs limit both the number of systems available to support effective safeguards as well as the current scope of unattended monitoring. Therefore, as an increasing number of facilities require safeguards, inexpensive unattended systems will become necessary. To help pursue this end, we are conducting an ongoing field trial of a machine learning algorithm to draw safeguards conclusions from an array of inexpensive sensors deployed at a fuel fabrication site in the United States. Specifically, we are testing the machine learning method's ability to identify activities at a fuel fabricator facility and unusual movements that are inconsistent with their expected pattern of behaviors. This small-scale trial is a follow-up to a proof-of-concept previously conducted on this method. If successful, the conclusion of this test will encourage a follow-up, larger-scale trial in which additional sensor types (e.g., radiation detection devices, scales, etc.) are incorporated into the machine learning method for a full-scale, "real world" application of this method. This trial will help us assess the potential for machine learning technological advancements to advance nuclear safeguards in an inexpensive, accessible way. Findings will have direct and practical implications for international safeguards technology and IAEA operations.

INTRODUCTION

The landscape of nuclear power is ever-changing. Countries are pursuing nuclear energy in increasing numbers, as the international community is attempting to move towards more sustainable and eco-friendly ways to meet energy needs. However, because of the dual-use nature of many nuclear-related commodities and activities, it is necessary to effectively implement safeguards for existing and burgeoning nuclear programs. It is also important that these safeguards are cost-effective and efficient, given the increasing number of resources that will be required to match the heightened number of countries pursuing nuclear programs.

Despite their significant advances in recent history, current safeguards measures can be costly, inefficient, and in some assessments, invasive (ex., Morrison 2006). Spectroscopic radiation detectors, for example, which are highly effective, can become prohibitively expensive when implemented in large numbers within a facility. However, with the aid of machine learning, it

may be possible to use non-spectroscopic detectors to draw safeguards conclusions. Likewise, in-person inspections, which can also serve as an invaluable tool for validating the peaceful use of a nuclear program, have been critiqued as invasive, because inspectors can probe and must be hosted (*ibid*). This is particularly true when inspectors tour sensitive military facilities (*ibid*). Inspectors also have inefficiencies that can be optimized by well-strategized misuse or diversion activities. Safeguards measures that ameliorate some of these methods' disadvantages would be particularly useful in the modern day, as more countries pursue domestic programs.

As the nuclear landscape changes, so do the capabilities with which to monitor it. The advancement of artificial intelligence (AI) and machine learning (ML) technologies, which have flourished in recent years, provide an opportunity to create more cost-effective, efficient safeguards. Machine learning models can be created with readily available computer programming toolkits, and developers are constantly refining and optimizing tools for these models. Algorithms' ability to collate and process various disaggregated information remotely and with relatively little financial burden could help improve safeguard measures by making them more accessible and efficient. Likewise, the cameras and sensors that could be used to collect data for machine learning models to classify can aggregate various anomalies that a human might fail to detect, but which could also be symptomatic of facility misuse or materials diversion (Rushdi et al. 2019). The IAEA already uses some of this technology in complex fuel cycles for nuclear materials accountancy (*ibid*). However, expanding this method into simpler fuel cycle processes and to also detect anomalistic behaviors (rather than primarily for material accountancy) could further advance safeguards in a cost-efficient and effective way.

Because of this potential, it is timely to test the machine learning models' potential application on nuclear safeguards. To this end, researchers at Savannah River National Laboratory (SRNL) and Sandia National Laboratory (SNL) are conducting an ongoing field test of machine learning methods' abilities to assess nuclear safeguards at a fuel fabrication facility. The goal of this paper is to explain the machine learning method being implemented, discuss its effectiveness in a completed proof-of-concept, and explore practical takeaways that have emerged in the field test thus far for practitioners who might hope to implement a similar method and take it to scale in the future.

A DESCRIPTION OF THE METHOD

The authors are currently conducting a field test to examine the feasibility and effectiveness of collecting data from a fuel fabrication facility, applying machine learning algorithms to examine those data, and drawing safeguards conclusions about the facility. For this field study, the data relates to 30B cylinders at a fuel fabrication facility. Data consists of "pattern-of-life" information, such as the locations of cylinders, their everyday movements, etc. After training machine learning algorithms to (1) identify cylinders through image classification methods and (2) identify typical patterns of life, the machine learning model will be asked to classify cylinders' activities as usual or unusual. The goal of this technique is to help safeguards monitors of the potential misuse of diversion of nuclear material at a fraction of the cost of other effective safeguards measures.

Abnormal behaviors could include a cylinder leaving a facility at an unusual place or time, or cylinder movements to and from the cylinder yard that indicate unusual behavior in feeding the pellet lines, for example. When automated, if the algorithm detects an abnormality, it will be able to send an alert to the relevant monitoring practitioners. The practitioners can then use this prompt to help them assess the potential for misuse or diversion of nuclear materials. In the event this information cannot be transmitted offsite, the information will be stored locally for review by an inspector when they are next at the facility. If taken to scale, additional information from other safeguards a given facility might also have (ex., scale weights) could be incorporated into the model to further improve its remote assessment capabilities.

In practice, it could take a considerable amount of time to collect enough data to identify patterns of life sufficiently enough to decrease false alarms and maximize true positives. However, after adequately training the machine learning algorithms, they will require very little input and provide considerable return on investment. Therefore, though the method requires time and data on the front end, its inputs decrease in demand over time.

PROOF-OF-CONCEPT

The first step to test this method was conducting a proof-of-concept in which researchers from SRNL and SNL conducted an analysis to help ensure this machine learning method would be appropriate for safeguards monitoring (see Rushdi et al. 2019 for proof-of-concept). In some cases, machine learning models can be difficult to train adequately enough to provide a satisfactory ratio of true to false positives and negatives. The proof-of-concept was undertaken to address these concerns.

The research team trained an algorithm against a fabricated dataset of 50+ sensors the team was monitoring. The algorithm was trained to predict whether the "pseudo-facility" that the data would have come from was engaged in activities that were indicative of "normal" behaviors ("N"), "misuse" of the facility ("M"), or "diversion" of material ("D"). The team's goal was not to provide live accountancy of materials, as some IAEA safeguards already do with similar technology, but to provide indicators that material or operational quantity anomalies (e.g., wrongfully high enrichment) might be present.

To test this technique, the team created a dataset describing normal patterns of life within the pseudo-facility that would have been picked up by the sensors' signals. Variables from the signals included material weights, time on scales, enrichment levels, ID matches, temperatures, and strain gauges' rack statuses. The team then injected synthetic data that would mimic symptoms of misuse or diversion (e.g., mass changes, product quantity changes, flow duration changes). The synthetic data simulated various paths to misuse or divert materials. The team ensured that variables followed a normal probability distribution, so that most observations were normal, with a minority not so (see Figure 1 for distributions).

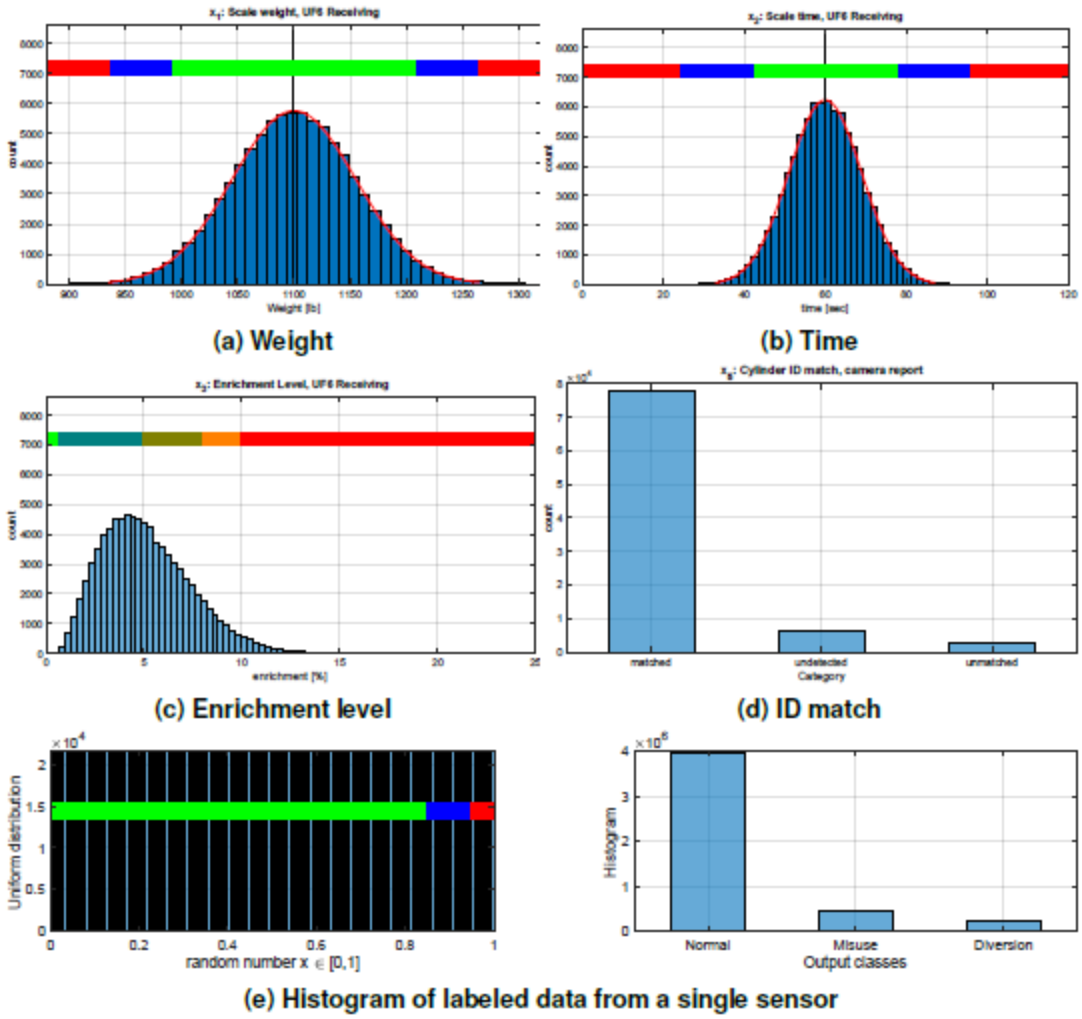

Figure 1. Probability Distribution of Sensor Signals (Source: Rushdi et al. [2019])

To test the algorithm, the team created rules they hoped the algorithm would detect. If the sum of reported signals was lower than a given number, the model should assign it the status of normal ("N"). If the sum was larger than or equal to that number, but also smaller than another greater threshold, it should be assigned the status of misuse ("M"). If the sum exceeded or was equal to the final threshold, it should be assigned the status of diversion ("D"). Extra weight was given to the most critical sensors, such as enrichment detectors. These rules are described in Table 1 below:

Table 1. Rules of Classification (Source: Rushdi et al. [2019])

Case	A	B	C
$R_1 = A$	$\mathcal{N} : S < \alpha_1 N_s$ $\mathcal{M} : \alpha_1 N_s < S < \alpha_2 N_s$ $\mathcal{D} : S > \alpha_2 N_s$		
$R_2 = A \vert B$	$\mathcal{N} : S < \alpha_1 N_s$ $\mathcal{M} : \alpha_1 N_s < S < \alpha_2 N_s$ $\mathcal{D} : S > \alpha_2 N_s$	$\mathcal{M} : E_2 < \beta_2 E, E_1 > \beta_1 E$ $\mathcal{D} : E_2 > \beta_2 E$	
$R_3 = A \vert B \vert C$	$\mathcal{N} : S < \alpha_1 N_s$ $\mathcal{M} : \alpha_1 N_s < S < \alpha_2 N_s$ $\mathcal{D} : S > \alpha_2 N_s$	$\mathcal{M} : E_2 < \beta_2 E, E_1 > \beta_1 E$ $\mathcal{D} : E_2 > \beta_2 E$	$\mathcal{M} : W_2 < \gamma_2 W, W_1 > \gamma_1 W$ $\mathcal{D} : W_2 > \gamma_2 W$

where R_1 is rule one, R_2 is rule two, R_3 is rule three, N_s is the number of sensors, S is the sum of sensor signals, E_1 is the number of enrichment sensors reporting misuse, E_2 is the number of enrichment detectors reporting diversion, W_1 is the number of scales reporting misuse, and W_2 is the number of scales reporting diversion. The same rules can be visualized in Figure 2 below:

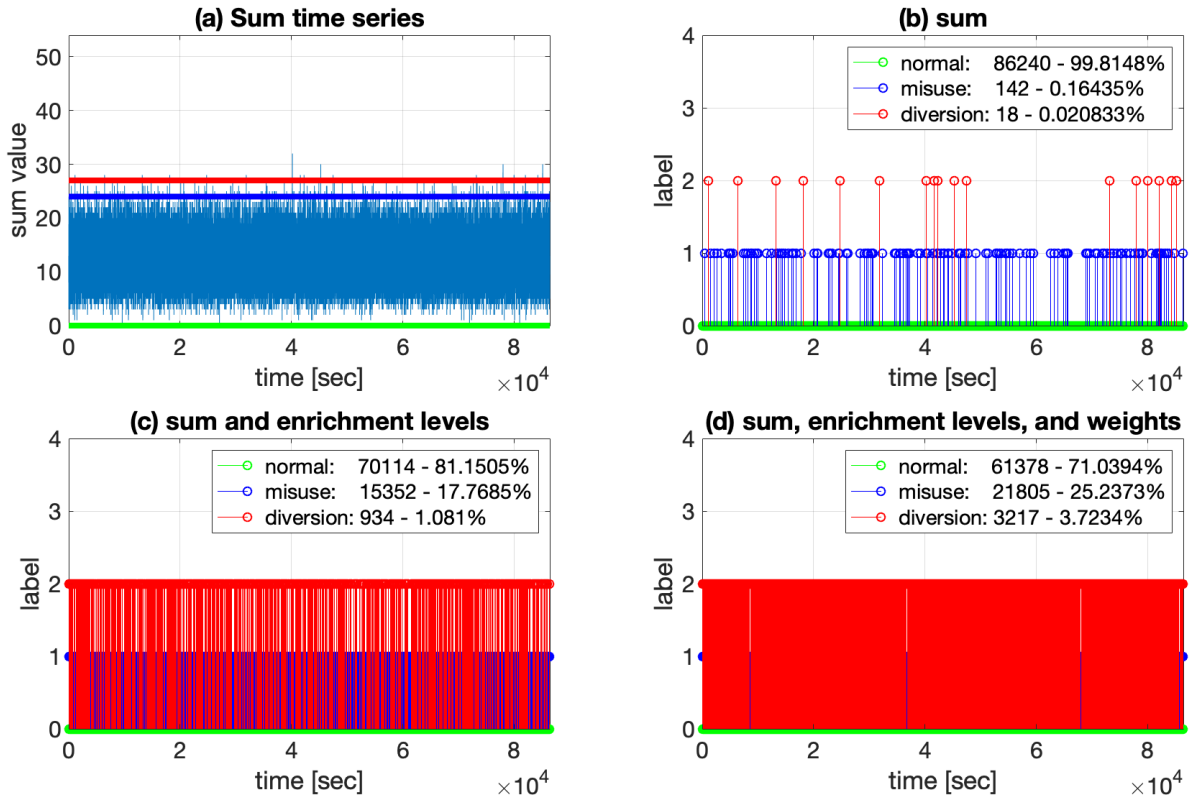

Figure 2. Visualized Rules of Classification (Source: Rushdi et al. [2019])

After assembling the dataset, the team tested various machine learning models' (e.g., logistic regression, naïve Bayes, k-nearest neighbors, etc.) ability to assign labels of normalcy, misuse, and diversion. Because most data should appear normal, a model could boast a misleadingly high

accuracy rate, in which the only correct predictions are true normals, but no misuse or diversion labels were correctly assigned. Therefore, the team considered models' abilities to predict R_2 and R_3 (see Table 1 for definitions), which include misuse and diversion, as well.

As can be seen in Tables 2, 3, and 4 (below), the models had considerable success classifying behaviors that were indicative of normalcy, misuse, and diversion. Accuracy rates were upwards of 99%. As documented in Tables 3 and 4, Gradient boosting improved scores on R_2 and R_3; though, it had the disadvantage of a comparatively longer training time.

Table 2. Train and Test Scores using R_1 (Source: Rushdi et al. [2019])

	classifier	train_score	test_score	train_time
0	Logistic Regression	0.999304	0.999270	0.686506
3	Neural Net	0.999238	0.999270	39.128846
4	Nearest Neighbors	0.999238	0.999270	8.891924
5	Linear SVM	0.999238	0.999270	4.042764
7	Random Forest	1.000000	0.999270	43.508844
1	Decision Tree	1.000000	0.998617	0.987446
6	Gradient Boosting Classifier	0.997416	0.996082	1675.279600
2	Naive Bayes	0.610629	0.615801	0.312202

Table 3. Train and Test Scores using R_2 (Source: Rushdi et al. [2019])

	classifier	train_score	test_score	train_time
6	Gradient Boosting Classifier	0.999653	0.998227	1968.456724
7	Random Forest	1.000000	0.963493	81.406278
3	Neural Net	0.961197	0.962529	50.938250
5	Linear SVM	0.961214	0.962336	93.580762
4	Nearest Neighbors	0.961495	0.962259	12.997938
0	Logistic Regression	0.960949	0.961642	0.662952
1	Decision Tree	1.000000	0.953662	1.583436
2	Naive Bayes	0.925538	0.927718	0.331022

Table 4. Train and Test Scores using R_3 (Source: Rushdi et al. [2019])

	classifier	train_score	test_score	train_time
6	Gradient Boosting Classifier	0.999687	0.997169	1872.820478
7	Random Forest	1.000000	0.983594	77.830042
1	Decision Tree	1.000000	0.980840	1.632468
3	Neural Net	0.978504	0.975449	156.648988
5	Linear SVM	0.921837	0.912733	253.030408
0	Logistic Regression	0.874738	0.870380	0.877348
4	Nearest Neighbors	0.866176	0.845131	11.414636
2	Naive Bayes	0.770543	0.766358	0.543786

The proof-of-concept provided a promising basis to continue developing this method for scaled application of safeguards monitoring. However, testing the concept in a real fuel fabrication facility and under different conditions is important to further improving it. It would be advantageous to better understand how different sensors correlate with each other, how the models perform with and without various tools (for example, without enrichment detectors), and

how the models handle data coming in at different rates. For these reasons, it is helpful to assess the technique in practice.

THE CURRENT STUDY

The current study builds upon the proof-of-concept. It is an ongoing study in which data is gathered from a fuel fabrication facility using remote sensors, and is then used to train a machine learning algorithm to 1) identify cylinders and 2) identify unusual activities.

At the time of writing, the team is training the algorithm to positively identify different kinds of cylinders that are used in the fabrication facility. The team is also working on gathering data this will be used to train an algorithm to detects patterns in cylinders' activities. In an automated process, which is the ultimate goal of this series of studies, the algorithm would need to first positively identify cylinders, then recognize their patterns of life, identify unusual behaviors, and alert the analyst.

The cylinders' behavioral data will include variables such as the time of day, the direction a cylinder is moving, etc. In practice, these inputs would require little besides cameras and a computer program, providing an economically-feasible monitoring option for analysts. The data used in this project could be combined with other data facilities have available, such as radiation detectors, etc. As the team gathers data and trains the models, they will note models' performance and the prospects for applying the technique in practice.

As the team conducts this field test, several considerations have emerged that could help future researchers and practitioners aiming to apply similar techniques to safeguards. First, the security concerns associated with facilities that operate as part of the nuclear fuel cycle, such as fuel fabrication facilities, affects how this technique can be applied. For example, in the current study, cameras used to collect data from the facility could not have Wi-Fi capabilities, but still needed to meet certain technological standards (ex., having a certain frame rate). Relatedly, assuaging facility operators' concerns regarding the incorporation of any cameras at all into their facilities was an obstacle the team faced and could be an important factor for future researchers and practitioners to consider, as this process can delay or completely halt the ability to monitor. Some facility operators might prefer remote cameras to physically present people, who must be escorted and could affect everyday operations when present; however, others might consider a camera's constant presence to be invasive. Practitioners must recognize these factors when considering the implementation of this safeguards technique.

Another consideration is the length of time required to train a machine learning model to 1) successfully classify images and 2) learn and label routine versus non-routine activities. To produce a highly accurate algorithm with appropriate levels of sensitivity to changes in behavioral patterns, huge swaths of data would need to be incorporated into a dataset. Therefore, when taking this technique up to scale, additional time needs to be factored in to train the model before it can become operable. However, as noted previously, this front-end input should result in considerable back-end payoffs in terms of labor and money dedicated to upkeep.

CONCLUSIONS

Machine learning algorithms can help remote safeguards practitioners recognize and respond to potential facility misuse or materials diversion. They can also aggregate various pieces of information that could be symptomatic of misuse or diversion, but which could also be easy to overlook by humans. The IAEA has applied machine learning to complex fuel cycles and for material accountancy purposes; however, it has not yet expanded these capabilities to include simpler fuel cycles or material/operational anomalies (Rushdi et al. 2019).

As more countries pursue and express interest in nuclear power programs, the ability to incorporate inexpensive, effective safeguards is becoming imperative. Using machine learning capabilities to remotely classify fuel cycle activities is a potential path forward in advancing safeguards and providing feasible evaluation options. The authors' field test of this method is meant to provide additional insight into how the effectiveness and practicality of this method.

The proof-of-concept conducted provides promising results for the prospects of applying the machine learning technique. Its accuracy rates with realistically-distributed data were upwards of 99%. The field test of a fuel fabrication facility is ongoing, with the researchers currently training the image classification algorithm to identify cylinders and working on data collection. Thus far, security concerns and selecting cameras that can withstand rigorous security measures but also be technologically sufficient have created obstacles that can provide lessons to practitioners aiming to take this method to scale. Additionally, while this study is focused on assessing the performance and feasibility of machine learning methods to detect unusual activities, connecting those assessments to an automated alerting system is a next step that will be useful in providing a model tool to take to scale.

Machine learning techniques are a promising approach to effectively meet the need to safeguard nuclear power programs, particularly as countries pursue the programs in higher numbers. As these models are further studied and developed, it is possible that they can become a useful tool for the international safeguards community to effectively carry out its goals.

ACKNOWLEDGEMENTS

We would like to acknowledge and sincerely thank the collaborating fuel fabrication facility's staff for their cooperation in conducting this field test. We would also like to thank our counterparts at other US national laboratories for their significant contributions to this project, and for fostering a collaborative environment between the US national laboratories.

REFERENCES

Morrison, D. (2006). "Brazil's Nuclear Ambitions, Past and Present." *Nuclear Threat Initiative.* Retrieved from https://www.nti.org/analysis/articles/brazils-nuclear-ambitions/.

Rushdi, A.; Sternat, M.; Branney, S. (2019). "Machine Learning for Activity Prediction in Nuclear Safeguards Using a Mass Array of Unauthenticated Sensors." *US Department of Energy*. Report.