# Feature Importance and Classification from Disparate Data Streams for Facility Pattern-of-Life Characterization in the Context of Nuclear Safeguards

Emily Casleton, Paul Mendoza, Rosalyn Rael, Jonathan Woodring, and Vlad Henzl
*Los Alamos National Laboratory, Los Alamos, NM, USA*

**Abstract**
Current practices for safeguards verification are largely based first on manual and independent interpretation of data streams from separate sensors followed by expert driven contextual interpretation. The work presented here develops a framework to integrate and utilize the wealth of information contained within large and heterogeneous datasets. The approach is demonstrated on persistent, disparate data collected from a nuclear training facility at Los Alamos National Laboratory. Features are first extracted from the individual data streams, then various feature-importance metrics are employed to down-select the feature matrix to a subset that is best able to characterize the facility operations, i.e. its "pattern of life". The resulting features are then input into supervised learning methods to classify modes of facility operations. The fusion of these disparate data streams yields a more accurate characterization of facility operations than any data stream individually and with a rather high degree of confidence. In addition, through the criterion of feature importance we are able to rank the sensor modalities with respect to the information they provide to characterize facility operations. This approach can be useful in a limited sensor deployment scenario or in planning stages of safeguards measures implementation. The developed framework can be adapted and applied to any other type of facility or sensor, and the associated data streams.

## Introduction

Current practices for safeguards verification are largely based on manual and independent interpretation of data streams from separate sensors followed by expert driven contextual interpretation. The work presented here integrates a wealth of information from many different modalities collated within large and heterogeneous datasets. The approach is demonstrated on persistent and event-based disparate data collected from a nuclear training facility at Los Alamos National Laboratory (LANL) and aims to classify activities of interest at the facility. The framework presented here could be adapted and applied to any other type of facility or sensors and the associated data streams.

To answer the question of what type of activity is occurring within the facility under scrutiny, a large set of features is extracted from the individual data streams of eight different and diverse modalities, ranging from temperature and humidity to nuclear material accountability tracking records. A feature is deemed `important' if it is informative for classifying modes of facility operations, determined through an analysis with a *random forest [4]*. After an initial screening, two well-established feature-importance metrics for random forests, the *Gini impurity index* and *permutation importance-based methods*, are applied to the features. In addition, we present a novel feature importance metric based on operator spoofing, apply it to the features and compare to the other metrics. This spoofing method is similar to the permutation-based

approach, but addresses the question of which features are robust to mislabeled data. A random forest is trained on intentionally spoofed data, and then features are ranked based on which are able to recover the 'true' classification.

## The Facility

The facility under scrutiny within this study is a Safeguards Training Facility located at LANL at Technical Area 66 (TA-66) [1]. It consists of a single building (see Figure 1), which is classified as a Category 3 nuclear facility, and includes offices, conference rooms, and a laboratory for nuclear material nondestructive assay (NDA) training courses. These courses involve the measurement of both uranium and plutonium, and the students consist of IAEA inspectors or other nuclear safeguards (both domestic and international) professionals. Other operations that take place at the facility include general office work, seminars and meetings in the conference rooms, courses that do not involve measuring nuclear material (e.g., statistics), individual course preparation, and testing instrumentation in the lab. The presence of certain uranium and plutonium bearing items during some of the training courses yield certain safeguards-relevant operations such as material accounting and control, including escorting of course participants by designated LANL employees.

*Figure 1-Safeguards Training Facility Located at TA-66.*

Overall, there are only a few and fairly simple operational modes at TA-66, certainly in comparison to the operational complexity of commercial nuclear power plants or Category 1 nuclear facilities. On the other hand, a majority of nuclear facilities operate within a finite set of well-defined states, e.g., a nuclear power plant is usually either operating or refueling, and diversion or misuse scenarios are distinct from those states within the normal operation set, and are relatively straight forward when compared to facilities that handle bulk nuclear material, such as production facilities.

## Data and Feature Extraction

The TA-66 facility was monitored using multiple, continuously collected, disparate data streams. Persistently collected data is recorded with seven inexpensive sensor platforms that integrate sensors of light, temperature, and humidity, measured at 1 Hz (i.e., one measurement

per second) and tri-axial vibration, measured with a 4 Hz frequency (i.e., four measurements per second). A series of simple calibration tests, using items such as a space heater and flashlight, were conducted to ensure sensors were working as expected and to test the sensitivity of the sensor platforms. Neutron counts are also persistently collected with a Portable Handheld Neutron Counter (PHNC) [2] stationed within the laboratory. The neutron data consists of singles and doubles rates, with one measurement recorded every 3 seconds. The approximate location of the persistent sensors can be seen in Figure 2.

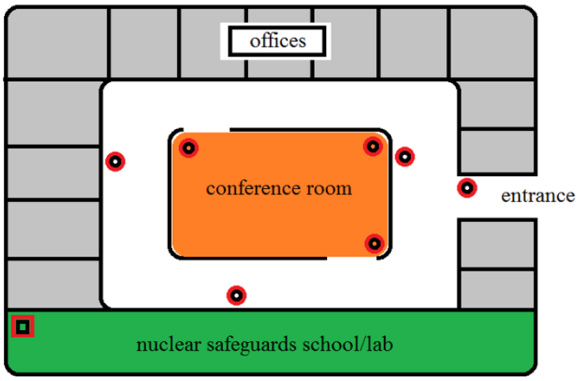


*Figure 2-Simplified floor plan of the facility. Red circles represent sensor platform location and red square represent the neutron detector. Actual floor plan restricted.*

Event based data was also collected and incorporated into the analysis. This data includes materials movement logs, through a nuclear data accountability system called Source Tracker, as well as badge reader data indicating entry and exits into the laboratory and the building, and room scheduling, which involves reserving some of the common areas, such as the conference room, for a meeting or presentation, but does not indicate whether the room was actually used during that time.

The large volume and variety of data necessitated a thoughtful data management plan. The data is currently collected via a Wi-Fi network to a central location, cleaned up and stored in a database. The dataset itself is a critical element in advancing new methods development and testing, as there is currently no facility data open source to the research community due to confidentiality concerns.

In order to resolve the persistent and event-based data and differing sampling rates onto a common time scale, features were extracted from all data streams [3]. This results in a new dataset consisting of a feature matrix that is aligned in time. These features represent a summary of the activity observed by the data stream over each time interval. Features considered include summaries of empirical distributions (e.g., moments, extreme values), counts, fractions of counts, entropy, and tailored features derived from knowledge about a specific modality or a combination between modalities. After various checks and data cleaning, the resulting feature matrix consists of 473 hourly features extracted from raw data collected from July 18, 2018 through February 23, 2020; spanning 1 year, 7 months, and 5 days of data. The breakdown of modality from which the features were extracted is shown in Table 1. Note that the 'Combo' feature modality is actually a combination of features from two different modalities.

| Neutron | Light | Humid | Temp | Vibration | Room Schedule | Badge Reader | Source Tracker | Combo |
|---|---|---|---|---|---|---|---|---|
| 24 | 44 | 44 | 44 | 126 | 5 | 1 | 20 | 165 |

*Table 1-Number of features extracted from each modality. Note that light, humidity, temperature, and vibration are combined for all seven sensors.*

The response variable of interest is a categorical variable indicating the current operating state of the facility. There are five possible categories for each hour: standard working operations (i.e., weekday between 7:00 am and 6:00 pm), course day, preparation for course day, standard non-working operations, and holidays. Course days include those that involve nuclear material as well as those that do not, and preparation days are defined as 7 calendar days prior to a course, excluding weekends and holidays.

## Feature Selection Methodologies

Because the question of interest is classifying the operation state of the facility, i.e., the response variable of interest is a categorical variable, the analysis methodology to be used is classification. We will focus on the methodology of random forests, as they have become a very popular data analysis tool due to their successful application to classification problems in various application areas. A random forest is an ensemble of decorrelated decision trees, which, by using a variety and large number of trees, increases the stability of the predictions from the model [4]. Random forests are particularly suited to handle the "large $p$, small $n$" problem, where there are many potential predictor variables (large $p$), but much fewer observations (small $n$).

As an initial screening of features, we determine which features are correlated with the response variable. Those that are uncorrelated will likely not have any predictability power, thus we aim to select features whose distributions are distinct for different response variable categories. However, because our response variable is categorical, measures of correlation such as the Pearson correlation coefficient are not appropriate. Instead, we will use the results of the Kruskal-Wallis test to initially eliminate uninformative features. This statistical test is a non-parametric alternative to the one-way analysis of variance (ANOVA) and tests the assumption that all categories have been drawn from the same population against the alternative that at least one is different, through a ranked sum process [5].

Two popular methodologies for assessing variable importance within an analysis performed via random forests are the Gini index and permutation methods. The Gini index is based on node impurity and is often used as the optimization criteria in creating the splits within the decision trees that make up the random forest [6]. As a variable selection method, the Gini index averages the decrease in node impurity over all splits for a given variable. This method has been shown to be biased with both categorical and continuous predictors [7]; however, all predictor variables considered in this work are continuous.

Alternatively, the permutation method of variable importance is based on the premise that randomly permuting the values of a predictor is analogous to removing it from the model. Unlike the Gini index, the permutation methods are applicable for methodologies other than random forests. For each variable, the model's prediction ability is compared when its values are in order against when the values are randomly shuffled [6]. A large decrease in prediction ability indicates high variable importance. This approach uses out-of-bag observations for prediction and is more computationally expensive than the Gini index [8].

In this work we will also present a novel variable importance metric similar to the permutation variable importance metric that specifically addresses a question of interest within the project, that of operator misdeclaration. Thus, we are interested in which variables are robust with respect to prediction ability when trained with data that has been intentionally spoofed. In other words, when a declaration is not correct and an operator tries to fool us, the variables are still able to recover the true classification label.

This approach, named *Spoofed*, will have a similar form to the permutation-based method, but with a few modifications. The algorithm is detailed below:

1. *Change the response variable labels for a few observations, i.e., emulate receiving intentionally spoofed data from an operator.*
2. *Fit a random forest using the spoofed data.*
3. *Predict the spoofed days and compute the prediction accuracy using the true variable labels.*
4. *For each variable of interest:*
    a. *Permute the variable values.*
    b. *Predict the spoofed days.*
    c. *Compare prediction accuracy from unpermuted.*

There are a few differences between this method and the original, permutation methods. First, we are only interested in predicting the observations for which we have intentionally spoofed. Second, the prediction accuracy used in this method will represent training error, because we will be predicting observations that were used to train the model, in contrast to the traditional permutation methods that uses out-of-bag sample predictions and thus test error. However, if the spoofed observations were left out of the training and only used for prediction, it would not address the question of interest. Although the error will be biased high, the ranking of the variables will not be affected because the bias will be consistent across variables, and we are not interested in the raw prediction ability, only the ability to rank.

## Results

The three feature selection methods were applied to the 473 features extracted from the 1 year, 7 months, and 5 days of data from eight modalities collected from the TA-66 Safeguards training facility. The results of each, through the features selected from each modality, are compared in Figure 3 and Figure 4. The first figure shows the raw number of features selected by each method from each modality. Note that the Kruskal-Wallis initial method of down selection

was applied before the Gini, Permutation, or Spoofed methods, thus each of these approaches will have equal or fewer numbers per modality than the After KW or Original approaches.

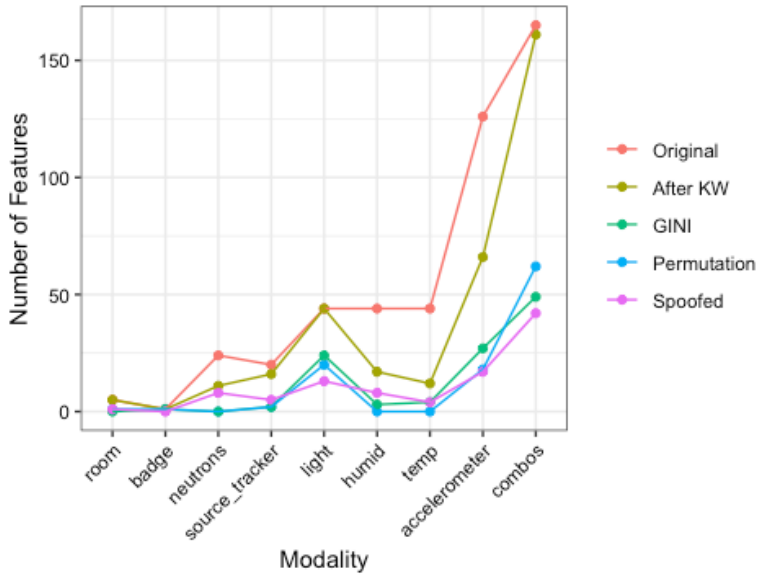


*Figure 3-Number of features for each modality for the various feature selection methods.*

Figure 4 displays the same information as Figure 3, but in terms of percentages. These percentages are calculated based on the number of features presented to the method, thus the percentages for the "After KW" method are based on the original number of features and the percentages for the Gini, Permutation, and Spoofed methods are based on the "After KW" number of features.

The initial screening using the Kruskal-Wallis test cut the number of features from 473 to 333. As can be seen in the results figures, the method chose to keep a majority of the combination features, but selected a small percentage of the humidity, temperature and accelerometer features. It also selected less than half of the features from the neutron detector. This is because these modalities look similar for all classification levels; i.e., the natural background of neutrons, humidity, temperature, and vibration dominate most of the extracted features and any discernable difference due to facility operations will not significantly contribute to the feature values. However, the fact that it did not completely eliminate any modality indicates that there were some features of each modality that are able to inform the facility operations.

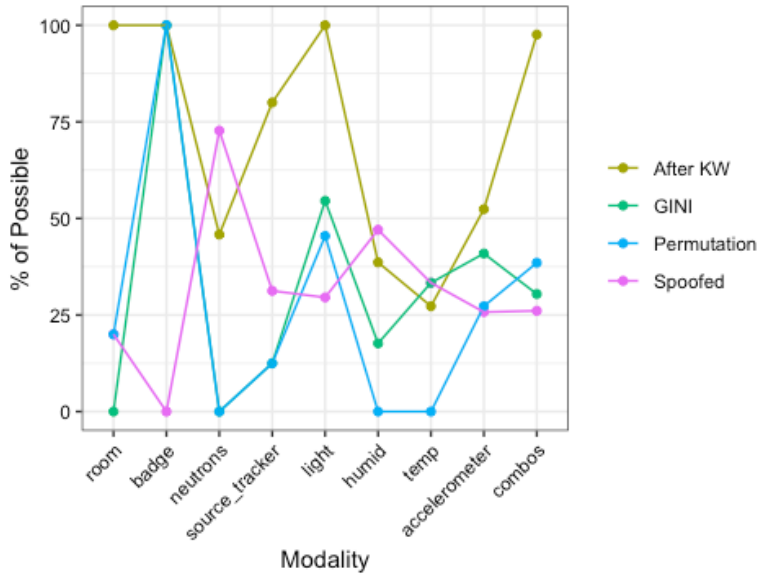


*Figure 4-Percentage of the features selected for each modality for the various feature selection methods.*

The first feature selection method applied was the permutation-based importance using the *cforest* function from the R package *party* [9]. This method results in an importance value for each feature. After plotting the values of the importance metric, it was determined that a cutoff of 0.0012 allowed for a natural separation between two groups of variables and resulted in 110 features. The application of the Gini index was performed with the *randomForest* function in the R package by the same name [10]. The application of the Gini index also results in a score for each variable, so again the top 110 features were chosen.

For the application of the Spoofed method to the TA-66 data, two courses, or ten days of "course day" observations were changed to have a label of "standard non-working operations". The two courses for which the data were modified include an "*Advanced Neutron NDA*" course and a "*Fundamentals of NDA*" course. A random forest was fit to the spoofed data. Then, each of the 333 features values were permuted one at a time. After each permutation, the spoofed days were predicted, and prediction accuracy was computed based on the true, i.e., not spoofed, observation label. Any feature for which the prediction accuracy decreased as compared with the unpermuted values was selected. This resulted in selection 98 of the features.

The results shown in Figure 3 and Figure 4 indicate the different feature modalities that the different feature selection methods favored. Neither the permutation nor the Gini methods chose any neutron features. The permutation method also did not choose any humidity or temperature features, and favored more of the combination features than the Gini method. The spoofed method was the only method that did not choose any of the badge reader features, but did select features from the neutron modality. This could be because the badge reader data is easily tricked and the neutron data is not, where the other features do not change with the mislabeled data.

Despite minor details (e.g., training versus test error), the spoofed method is most analogous to the permutation method, where the main difference is the data on which the method was trained. Figure 4 shows that these two methods produce almost the opposite in terms of features selected, specifically in terms of the badge reader, neutron, humidity, and temperature modalities. This implies that if the data is spoofed or correctly labeled can lead to a noticeably different mix of variables that are chosen as important.

In a realistic safeguards application, we will not know whether the data has been spoofed or not. However, the results in Figure 4 indicate that the selected features could be an indicator of spoofed data. If features and operator logs are received over time, and variable selection is performed sequentially, assuming that the status of the facility operations has remained relatively consistent, a large shift in feature importance could indicate the data has been tampered. In addition, the results of the spoofed feature selection methodology can be used as a design tool to indicate the data streams and modalities that can best inform when the data has been corrupted. This approach can be useful in a limited sensor deployment scenario or in the planning stages of safeguards measures implementation.

## Conclusions

The data collected from the persistent sensors and the extracted features have been used to identify features that are important for the purpose of classifying facility operation mode. The ultimate goal of this analysis would be to identify drastic changes in facility cadence, implying that a change has occurred within the facility, leading to increased awareness of operations either by operators or inspectors.

A natural question to ask is which feature importance methodology is 'best'; however, the answer to this question is application specific. For example, although not presented here, but already in [11], the criterion of feature importance can be used to rank the sensor modalities with respect to the information they provide to characterize facility operations. We have presented a method of ranking variable importance based on those that answer a specific question, specifically, correctly predicting data that has been intentionally spoofed. This framework could be extended to answering other questions of interest, such as, which features are important for predicting only courses, or which features are important when the facility dramatically, but intentionally, modifies its typical cadence of activities, (e.g., in response to a pandemic). Choosing which feature importance method to use should be informed by the question of interest and on which method most accurately addresses that issue.

For future work, we will further explore the features selected by the different methods, specifically examining which modalities compose the combination features. In addition, the idea of spoofing has been demonstrated to rank the features, but we have not quantitively addressed the details of detecting spoofed data.

## Acknowledgements

This work was originally supported internally by LANL Laboratory Directed Research and Development program and is currently sponsored by the U.S. Department of Energy (DOE), Office of Defense Nuclear Nonproliferation's Nonproliferation Research and Development (NA-22). Approved for public release LA-UR-21-28133.

## References

[1] NPR Morning Edition, "How do you find plutonium? Go to nuclear inspector school" (19 October 2015). https://www.npr.org/2015/10/19/449031762/how-do-you-find-plutonium-go-to-nuclear-inspector-school

[2] H. O. Menlove, "Manual for the Portable Handheld Neutron Counter (PHNC) for Neutron Survey and the Measurement of Plutonium Samples," Los Alamos National Laboratory, Los Alamos, NM, 2005.

[3] D. L. Hall and J. Llinas. "An Introduction to Multisensor Data Fusion." *Proceedings of the IEEE,* 85(1):6-23, 1997.

[4] G. James, D.Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*. Vol. 112. New York: springer, 2013.

[5] M. Hollander, D. A. Wolfe, and E.Chicken. *Nonparametric statistical methods*. Vol. 751. John Wiley & Sons, 2013.

[6] C. Strobl, A. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. "Conditional variable importance for random forests." *BMC bioinformatics*, 9(1):1-11, 2008.

[7] C. Strobl, A. Boulesteix, A. Zeileis, and T. Hothorn. "Bias in random forest variable importance measures: Illustrations, sources and a solution." *BMC bioinformatics,* 8(1): 1-21, 2007.

[8] K. J. Archer, and R. V. Kimes. "Empirical characterization of random forest variable importance measures." *Computational statistics & data analysis*, 52(4): 2249-2260, 2008.

[9] T. Hothorn, P.Buehlmann, S. Dudoit, A. Molinaro, and M. Van Der Laan. "Survival Ensembles". *Biostatistics*, 7(3): 355—373, 2006.

[10] A. Liaw and M. Wiener. "Classification and Regression by randomForest". R News, 2(3): 18—22, 2002.

[11] R. Rael, E. Casleton, V. Henzl, P. Mendoza, J. L. Woodring. "Pattern Of Life Analysis Of A Facility Under Nuclear Safeguards", Paper presented at INMM Annual meeting, held virtually, July 12-16, 2020.