

Self-Simulated Learning Artificial Intelligence for the Detection of High Explosives in Soil by Mid Infrared Laser Spectroscopy

Reynaldo Villareal-Gonzalez¹, Nataly Galán-Freyre¹, Leonardo Pacheco-Londoño¹, and Samuel Hernández-Rivera²

¹Universidad Simon Bolivar

²University of Puerto Rico Mayaguez

May 5, 2020

Abstract

A tunable mid-infrared (MIR) laser (quantum cascade laser, QCL) was used for the detection of TNT and RDX in soil samples at a concentration range from 0 to ~20% w/w. This type of sensing is complicated due to the complexity of the matrix, i.e., the diversity of compounds contained in soil. Thus, the high explosives (HE) detection in soil by QCL was assisted with an Artificial Intelligence (AI) system. AI managed to predict these HE in seven kinds of soils using minimum information Machine Learning (ML). The models were generated only from neat HE and soil spectra, without necessity using experimental spectra of the mixes. AI used these neat spectra to simulate the spectra of HEs/soil mixes. The simulated data was used to train the ML models and then were tested with real spectra of HEs/soils mixes. The method was designated as “Self-Simulated Learning Artificial Intelligence” (SSLAI). This methodology has advantages for applications in field scenarios where the matrices are unknown because SSLAI models do not need to be trained with real samples a priori. Models would only have to be fed with spectra for the neat components to train itself. The methodology was tested with mixes of seven soils and two explosives. Test samples were classified into three concentrations ranges: high concentration test ($\text{Test_H} > 10\%$ w/w), medium concentration test ($10\% \text{ w/w} > \text{Test_M} > 3\%$ w/w), and low concentration test ($\text{Test_L} < 3\%$ w/w). The results show that it is possible to correctly predict these two HE/soil mixes from the simulated data. Specifically, for TNT and RDX, SSLAI achieved a high precision in the prediction for the high and medium concentration tests (Test_H and Test_M). However, for both samples with concentrations below 3% w/w (Test_L), the number of false positives increased, and the precision was reduced.

1 Introduction

The field detection of explosives has received considerable attention globally for many years due to its importance in forensic applications, defense and security, and environmental control. High explosives (HEs) have been used for decades in military practice and mining excavation. These are considered as a contamination source which can generate poisoning of population humans and animals, and producing genetic diseases that threaten human health. ^[1,2] In order to monitor the quality of HE-contaminated soils, this study focuses on the detection of HE in soil using mid-infrared (MIR) laser spectroscopy with a Quantum Cascade Laser (QCL) source as a remote method of analysis. ^[3-13] QCL spectroscopy first demonstrated in 1994 by Faist *et al.* ^[14] offers several benefits over conventional or thermal source MIR spectroscopy, such as room temperature operation, small beam sizes, long lifetimes, low energy consumption, long-term power stability, and fine-tuning of the output frequency. ^[15] Several groups have demonstrated the capability of remote sensing of HE and others analytes using MIR laser spectroscopy with QCLs. ^[14, 16-27] The typical methods used to detect HE is destructive, require sampling, transferring samples to the lab, and performing a proper treatment of the sample for later detection. These methods include protocols based on gas chromatography-mass

spectroscopy (GC-MS), gas chromatography-chemiluminescence (GC-CL), ion mobility spectrometry (IMS), [28] immunosensors, [29] electrophoresis, [30] fluorescence, [31] electrochemical methods, [30,32] high-pressure liquid chromatography (HPLC), [33,34] and HPLC-MS. [33] However, *in situ* detection of HE in the soil is not easy due to the presence of solid interfering materials such as organic and inorganic compounds, which vary for each type of soil [35-37] which makes this detection a challenge for the analyst. Other studies conducted by this research group involved the characterization [38-40] interactions, [41-43] and detection [44-45] of HE in soil using Raman and FT-IR spectroscopy. [46,47] In all of these cases, the detection was marginally possible because the crystalline particles of HE had to be found in the solid matrix by microscopy to achieve the detection.

Today, AI methods are becoming more popular because they have demonstrated to be a rapidly evolving research area that offers sophisticated and advanced approaches capable of addressing complicated and challenging problems. Besides, AI-based systems have a variety of applications in different sectors, such as engineering, economics, medicine, military, marine sciences, and others. [48] Therefore, AI allows the transfer of human knowledge to machines through analytical models and learning from the data. This is a task that can be accomplished by soft-computing methodologies. [49]

AI uses minimum information (spectra of neat explosives and clean soil) for the development of the Machine Learning (ML) models without the necessity of experimental data of the mixes. AI: Self-Simulated Learning Artificial Intelligence (SSLAI) models were tested with real spectra of experimental mixes of HE/soils. SSLAI models do not need to be trained with real contaminated soil samples or real mixes of HE/soils. The model would only have to be fed with spectra of neat HE and soils for the model to train itself. This provides the possibility of HE detection in field applications with the advantages that the natural solid matrices could be unknown.

2 Methodology

2.1 Reagents

HE used as analytes in this study were cyclotrimethylenetrinitramine (RDX), and 2,4,6-trinitrotoluene (TNT). TNT was purchased from ChemService (West Chester, PA, USA), and RDX was synthesized in the lab. Seven types of natural soils were used: first natural (Soil-1) and second natural soil (Soil-2) were obtained from suburban sites located near the municipality of Mayaguez, PR, USA (coordinates: 18°13'25.7"N 67°07'51.2"W); third natural soil (Soil-3) and fourth natural soil (Soil-4) from (coordinates: 18°9'36"N 67°6'40"W); fifth natural soil (Soil-5), sixth natural soil (Soil-6), and seventh natural soil (Soil-7) were obtained from suburban sites located near the municipality of Ponce, PR, USA (coordinates: 18°00'39.9"N 66°36'50.6"W).

2.2 Samples Preparation

Calibration samples did not require complicated preparation steps. Neat samples consisted of 0.20 g of neat HEs or neat soils. Samples of mixes also consisted of approximately 0.20 g of HE/soil with concentrations of 0% to 20% w/w for each HE. The test samples were prepared by grinding the explosive into a fine powder using a mortar and pestle, followed by mixing in a mini vortex mixer for approximately 10 s at 3000 rpm. The mixed samples were ground again and mixed in the mini vortex mixer for a second time. The test samples, were classified into three concentrations ranges; high concentration test (Test_H > 10% w/w), medium concentration test (10% w/w > Test_M > 3% w/w) and low concentration test (Test_L < 3% w/w).

2.3 Data acquisition and QCL system

Before measuring the QCL spectra, a background spectrum of a roughened gold substrate was obtained. This background spectrum provided a good and smooth reference trace, due to the lack of MIR signals from the employed solid matrix. The samples were placed in the wells of metal holders (2.54 cm in diam., 10 mm deep). The sample surface was flattened to create a smooth surface for accurate measurements. Duplicate

spectra were collected at ten (10) different locations on the surface, resulting in a total of twenty (20) spectra per sample. These were used for the calibration and test analyses. This process was used for the neat samples, and each concentration in the test mixes samples. The spectra were obtained in reflectance (R) mode at a distance of approximately 15 cm, using a LaserScan MIR pre-dispersive spectrometer (Block Engineering, Marlborough, MA, USA) equipped with three tunable MIR lasers with a tuning range from 990 to 1111, 1111 to 1178 and 1178 to 1600 cm^{-1} , the scan time was 0.5 s approximately on each laser for a total time scan of 1.5 s. The average power typically varied between 0.5 to 10 mW across the 600 cm^{-1} total tuning range with 100:1 transverse electromagnetic polarization (TEM_{00}) and beam divergence of < 2.5 mrad in the x-axis and < 5 mrad in the y-axis. The spectrometer had a ZnSe lens of 3 in. in diameter, which was used to focus the MIR beam, to collect the reflected light, and to focus the light onto a thermoelectrically cooled mercury-cadmium-telluride (MCT) detector. The wavelength accuracy and precision were 0.5 cm^{-1} and 0.2 cm^{-1} , respectively. The spectroscopic system worked best at a laser head to target distance of 15 \pm 3 cm. Each laser diode formed an elliptical spot of 4 x 2 mm^2 at the focal plane of the ZnSe lens due to the difference of beam divergence in the axes.

2.4 Self-Simulated Learning Artificial Intelligence (SSLAI) Analysis

A classifier algorithm used to compare machine learning (ML) methods for classification was developed in Python 3 using the library of sklearn 3.2. ^[50] Ten ML methods for classification were employed: K-Neighbors Classifier, Support Vector Machine (SVM), Nu-Support Vector Classification (NuSVC), Decision Tree Classifier, Random Forest Classifier, AdaBoost Classifier, Gradient Boosting Classifier, Gaussian Naive Bayes, Linear Discriminant Analysis, and Quadratic Discriminant Analysis. A basic description of each ML method used is included in **Table 1**.

The most efficient ML method for the prediction of the data was selected by the algorithm, considering the highest accuracy value and the lowest Log Loss value (defined later). To accomplish this, it was necessary to provide the algorithm with the simulated training data and the real test data. The algorithm trained each of the ten methods, calculating the accuracy and the log-loss, and comparing the results.

The package of libraries in scikit-learn ^[50] provides a set of open-source software of efficient AI techniques for the Python programming language. They are accessible to non-experts of ML and apply to various scientific disciplines. The calibration model was developed with simulated training data. The simulated training data are linear combinations of each of the spectra of neat HE with each of the spectra of neat soil considering their compensation of intensities in percentages. External evaluation of the predictions was accomplished using HE/soil mixes. These data were not used in the calibration model. A schematic representation of SSLAI analysis is shown in **Fig. 1**.

The parameters that were used to evaluate the performance of the classification model developed were: recall, Log Loss, precision, f1-score, weighted average, support, and accuracy. In binary classification, recall of the positive class is also known as “sensitivity,” and recall of the negative class is “specificity.” The Log Loss function is used in (multinomial) logistic regression and extensions of it, such as Neural Networks, defined as the negative log-likelihood of the true labels given a probabilistic classifier’s predictions. The Log Loss is only defined for two or more labels. For a single sample with true label y_t in $\{0,1\}$ and estimated probability y_p that $y_t = 1$, the Log Loss is (**Eq. (1)**):

$$-\log P(y_t|y_p) = -(y_t) \log(y_p) + (1 - y_t) \log(1 - y_p). \quad (1)$$

The precision is also called the positive predicted value, and it is the ratio: $\text{TP} / (\text{TP} + \text{FP})$ where TP is the number of true positives, and FP is the number of false positives. The precision (intuitively) is the ability of the classifier of not labeling as positive a sample that is negative. The f1-score is also known as the balanced f-score or the f-measure. The f1-score can be interpreted as a weighted average of the precision and recall, where the f1-score best value is 1, and the worst is 0. The relative contributions of precision and recall to the f1-score are equal. The formula for the f1-score is shown in **Eq. 2**:

$$\text{f1-score} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall}). \quad (2)$$

The f1-score in multi-class and multi-label cases is the average of the f1-score for each class, with weighting that depends on the average parameter. The support is the number of records used, i.e., the numbers of spectra for each class. The accuracy classification score in multi-label classification computes the accuracy subset: the set of labels predicted for a sample must exactly match the corresponding set of labels in y_true . The reported averages include the macro average (averaging the unweighted mean for each label) and the weighted average (averaging the support-weighted mean for each label).

Other parameters for the model evaluations were obtained using areas under the receiver operator curve (ROC) plots. This was generated using the ML method selected to evaluate the “sensor” performance. The ROC plots allow inspection of the fundamental trade-off in the models between TP and FP. This provides much more information than a straightforward accuracy calculation. When comparing two models, the ROC plots clearly show that a curve that is entirely over another represents a model with better results regardless of the threshold used. The area under a ROC plot is equal to the probability that a randomly selected positive case will receive a higher score than a randomly chosen negative case. In other words, it is the probability of sensing the HE.

3 Results

3.1 QCL MIR spectra

Reflectance spectra for TNT and RDX were acquired. **Fig. 2(a)** shows the characteristic spectroscopic signals for each HE. For TNT, the prominent signal located at 1567 cm^{-1} ($-\text{NO}_2$ asymmetric stretch), 1472 cm^{-1} (CH_3 deformation), 1445 cm^{-1} (C-C ring stretch), 1359 cm^{-1} (ring stretch), 1199 cm^{-1} (CH in-plane ring bend), 1173 cm^{-1} (CH in-plane ring bend scissoring), 1089 cm^{-1} (CH_3 bend twisting), and 1025 cm^{-1} (CH in-plane ring bend rocking; CH_3 deformation) were observed. ^[52] For RDX, the important band at 1321 cm^{-1} for $-\text{N}-\text{NO}_2$ symmetric stretch and the signal at 1593 cm^{-1} for the $-\text{N}-\text{NO}_2$ asymmetric stretch were observed. Other signals that stand out include the band at 1234 cm^{-1} and the band at 1034 cm^{-1} for $-\text{N}-\text{C}-\text{N}$ stretch; the signals at 1016 cm^{-1} for the combination $-\text{N}-\text{NO}_2$ stretch and in-plane C-H bending. ^[53-55]

Fig. 2(b) shows the reflectance spectra of the seven soil samples, identified by Soil-1, Soil-2, Soil-3, Soil-4, Soil 5, Soil-6, and Soil-7. A Principal Components Analysis (PCA) was carried out with the spectra of standard samples of clays and sand, and it was found that PC2 could differentiate between the classes clay and sand. The PC model enabled to conclude that Soils 1-4 have a higher clay proportion (60-65%) in comparison with Soils 5-7, with a clay proportion of 22-29%, considered mostly sandy soils (data not shown). These percentages were calculated from the distances of the PC2 scores with respect to the average of the classes (sand and clay).

Figs. 2(c) and **2(d)** show a comparison of the simulated spectra with the experimental spectra considering the mixes at high concentrations and with the same type of soil (Soil 1). The characteristic fingerprint signatures were observed for both TNT and RDX with Soil 1, respectively, **Figs. 2(c)** and **2(d)**. However, some differences in the spectral patterns between the simulated and experimental spectra are observed.

3.2 Comparison of ML methods

A classifier algorithm to compare ten ML methods was done. **Fig. 3** shows a graphical representation of the relationship between % Accuracy and Log Loss values generated by the classifier algorithm, for each of the tests performed with the ten ML methods selected previously. In this figure, the marker color is associated with the name of the ML method, and the marker shape with the concentration range of the test.

Moreover, from these results, the most efficient methods were selected. Four ML methods with the lower value of Log Loss (closer to zero) and a higher value of % Accuracy (ML methods with % Accuracy values close to 100%) were selected. The selected methods were: RFC, SVM, NuSVC, and GBC. Next, these four ML methods were evaluated by a ROC analysis, obtaining as a result that the best ML model for

classification is the one carried out by the RFC algorithm, demonstrating that RFC has a better probability of sensing.

Another reason for which RFC was chosen is that when there is binary discrimination, the precision parameter is decisive. Since the RFC model was the one with the highest precision value in the different tests, it was selected as the optimum model. This means that the number of false positives using this model is low.

3.3 Random Forest Classifier Analysis

The performance of the RFC model was evaluated with the test samples using the parameters observed in **Table 2**. Excellent results were obtained in the precision parameters, with values close to 1.00 for high and medium concentration tests, Test_H, and Test_M, respectively. However, for the tests of low concentrations (Test_L), the precision parameter remained at 1.00 for TNT but decreased to 0.75 for RDX. This decrease in the precision parameter for RDX is because some samples at low concentrations of TNT (3 samples, see confusion matrix) were predicted as RDX. However, it is essential to highlight that despite that the RFC model did not correctly predict most of the samples with TNT and RDX for the low concentration test, the model maintained a good accuracy. This is because at these concentrations ($< 3\%$), soil particle size generates a lack of homogeneity in the samples, creating explosive clusters that make detection difficult since the interrogation area for the QCL spot is only $4 \times 2 \text{ mm}^2$. This spot size generates a certain probability of not finding HE particles when this sensing takes place on the sample surface. A possible solution to this problem is to increase the interrogation area of the QCL system to generate an adequate average of the spectral acquisition.

The recall values in the low concentration test for both explosives were poor because most of the samples that had HE were classified as having no HE (none). According to f1-score values, this parameter measures the model in a general way, considering the recall and precision values, throwing poor values for the low concentration test.

Fig. 4 shows the analysis of the ROC curves for each of the Tests using the model generated by the RFC method. The ROC curve evaluates the model with different decision thresholds and measures the probability of sensing, which is calculated as the integration of the area under the ROC curve. The probability of sensing for the medium and high concentration test was excellent, with values close to 1.00. Although the probability of sensing for the low concentration test was expected to be poor, it presented a moderate probability of sensing.

4 Conclusion

This research establishes a strategic method for the detection of HE (TNT, RDX) in soil samples. The benefits of the methodology “Self-Simulated Learning Artificial Intelligence” (SSLAI) is that simulated data generated from spectra of neat HE and clean soil were used to create the training data set used to generate ML models. SSLAI was validated using 3 tests with experimental spectra of HE/soils mixtures and obtained favorable results. The remote detection was achieved using the benefits of MIR laser spectroscopy, which can be applied in field applications.

The detection of threat chemicals in solid matrices such as natural soil is complicated, specifically in field applications where the detection time is essential, and where often these solid matrices are unknown. This methodology permits to generate the model at the moment with insufficient information for training, surpassing previous methodologies where the volume of experimental data necessary to generate good prediction models takes time to collect and analyze the data.

The synergy between QCL technology and SSLAI can be a promising strategy for the detection of weapons of mass destruction in the fields of defense and security because solid matrices such as soil contain many interferences and varieties from the multiple components that they contain and also because they are sub-

strates/media with low reflectivity. Therefore, the detection of analytes in soil matrices is considered a challenge.

Disclosures

The authors declare no conflicts of interest.

Acknowledgments

This material is based upon work supported by the U.S. Department of Homeland Security, Science and Technology Directorate, Office of University Programs, under Grant Award 2013-ST-061-ED0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

References

1. T. Frische, H. Hoper, *Chemosphere* **2003** , 50, 415-427.
2. S. N. Correa-Torres, L. C. Pacheco-Londono, E. A. Espinosa-Fuentes, L. Rodriguez, F. A. Souto-Bachiller, S. P.Hernandez-Rivera, *J. Environ. Monit.* **2012** , 14, 30.
3. A. Pettersson, S. Wallin, H. Ostmark, A. Ehlerding, I. Johansson, M. Nordberg, H. Ellis and A. Al-Khalili, *Proc. SPIE* **2010** , 7664(76641K).
4. C. S.-C. Yang, E. E. Brown, U. Hommerich, F. Jin, S. B. Trivedi, A. C. Samuels, A. P. Snyder, *Appl. Spectrosc.* **2012** , **66** , 1397.
5. A. K. Misra, S. K. Sharma, T. E. Acosta, J. N. Potter, D. E. Bates, *Appl. Spectrosc.* **2012** , 66, 1279.
6. J. L. Gottfried, F. C. De Lucia, C. A. Munson, A. W. Miziolek, *Appl. Spectrosc.* **2008** , 62, 353.
7. J. R. Castro-Suarez, L. C. Pacheco-Londono, M. Velez-Reyes, M. Diem, T. J. Tague, S. P. Hernandez-Rivera, *Appl. Spectrosc.* **2013** , 67, 181.
8. J. C. Carter, S. M. Angel, M. Lawrence-Snyder, J. Scaffidi, R. E. Whipple, J. G. Reynolds, *Appl. Spectrosc.* **2005** , 59, 769.
9. L. A. Averett, P. R. Griffiths, *Appl. Spectrosc.* **2008** , 62, 383.
10. L. Pacheco-Londono, W. Ortiz-Rivera, O. M. Primera-Pedrozo, S. P. Hernandez-Rivera, *Anal. Bioanal. Chem.* **2009** , 395, 323.
11. C. W. Van Neste, L. R. Senesac, T. Thundat, *Anal. Chem.* **2009** , 81(5), 1952.
12. J. J. Moros, J. A. Lorenzo, P. Lucena, L. M. Tobaría, J. J. Laserna, *Anal. Chem.* **2010** , 82, 1389.
13. J. Moros, J. J. Laserna, *Anal. Chem.* **2011** , 83, 6275.
14. J. Faist, F. Capasso, DL Sivco, C. Sirtori, AL Hutchinson, AY Cho, *Science* **1994** , 264, 553.
15. L. Hvozdar, N. Pennington, M. Kraft, M. Karlowatz, B. Mizaikoff, *Vibrational Spectroscopy* **2002** , 30, 53.
16. J. L. Ruiz-Caballero L. A. Blanco-Riveiro, I. A. Ramirez-Marrero, L. A. Perez-Almodovar, A. M. Colon-Mercado, J. R. Castro-Suarez, L. C. Pacheco-Londono, S. P. Hernandez-Rivera, *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, 519, 012007.
17. L. C. Pacheco-Londono, N. J. Galan-Freyte, A. M. Figueroa-Navedo, R. Infante-Castillo, J. L. Ruiz-Caballero, S. P. Hernandez-Rivera, *J. Chemom.* **2019** , 33:e3167.
18. L. C. Pacheco-Londono, J. A. Aparicio-Bolano, N. J. Galan-Freyte, A. D. Roman-Ospino, S. P. Hernandez-Rivera, *Appl. Spectrosc.* **2019** , 73, 17.
19. L. C. Pacheco-Londono, J. L. Ruiz-Caballero, M. L. Ramirez-Cedeno, R. Infante-Castillo, N. J. Galan-Freyte, S. P. Hernandez-Rivera, *Molecules* **2019** , 24, 3494.
20. L. C. Pacheco-Londono J. R. Castro-Suarez, N. J. Galan-Freyte, A. M. Figueroa-Navedo, J. L. Ruiz-Caballero, R. Infante-Castillo, S. P. Hernandez-Rivera, "Mid-Infrared Laser Spectroscopy Applications I: Detection of Traces of High Explosives on Reflective and Matte Substrates", pp. 11-34, in "Infrared

Spectroscopy: Principles, Advances, and Applications” M. El-Azary, ed. IntechOpen, London, UK, 2019.

21. A. Ruther, M. Pfeifer, V. A. Lorenz-Fonfria, S. Ludeke, *J. Phy. Chem B* **2014** , 118, 3941.
22. S. Ludeke, M. Pfeifer, P. Fischer, *J. Amer. Chem. Soc.* **2011** , 133, 5704.
23. Q. Shi, D. D. Nelson, J. B. McManus, M. S. Zahniser, M. E. Parrish, R. E. Baren, K. H. Shafer, C. N. Harward, *Anal. Chem.* **2003** , 75, 5180.
24. K. Worle, F. Seichter, A. Wilk, C. Armacost, T. Day, M. Godejohann, U. Wachte, J. Vogt, P. Radermacher, B. Mizaikoff, *Anal. Chem.* **2013** , 85, 2697.
25. N. J. Galan-Freyre, L. C. Pacheco-Londono, A. D. Roman-Ospino, S. P. Hernandez-Rivera, *Appl. Spectrosc.* **2016** , 70, 1511.
26. A. C. Padilla-Jimenez, W. Ortiz-Rivera, C. Rios-Velazquez, I. Vazquez-Ayala, S.P. Hernandez-Rivera, *Opt. Eng.* **2014** , 53, 061611.
27. M. C. Phillips, B. E. Bernacki, *Opt. Eng.* **2012** , 52, 061302.
28. J. Hildenbrand, J. Herbst, J. Wollenstein, A. Lambrecht, *Proc. of SPIE* **2009** , 7222, 72220B.
29. U. Narang, P. R. Gauger, F. S. Ligler, *Anal. Chem.* **1997** , 69, 2779.
30. A. Hilmi, J. H. T. Luong, *Environ. Sci. Tech.* **2000**, 34, 3046.
31. S. Kumar, N. Venkatramaiah, S. Patil, *J. Phys. Chem. C* **2013** , 117, 7236.
32. R. L. Marple, W. R. LaCourse, *Anal. Chem.* **2005** , 77, 6709.
33. T. W. Sheremata, A. Halasz, L. Paquet, S. Thiboutot, G. Ampleman, J. Hawari, *Environ. Sci. Technol.* **2001** , **35** , 1037.
34. S. L. Larson, W. A. Martin, B. L. Escalon, M. Thompson, *Environ. Sci. Technol.* **2008** , 42, 786.
35. N. B. Gallagher, T. A. Blake, P. L. Gassman, *J. Chemom.* **2005** , 19, 271.
36. M. Forouzangohar, D. Cozzolino, R. S. Kookana, R. J. Smernik, S. T. Forrester, D. J. Chittleborough, *Environ. Sci. Technol.* **2008** , 42, 3283.
37. N. B. Gallagher, P. L. Gassman, T. A. Blake, *Environ. Sci. Technol.* **2008** , 42, 5700.
38. M. D. Hernandez, I. Santiago, I. Y. Padilla, *Proc. of SPIE* **2006** , 6217, 621736.
39. B. Baez, S. N. Correa, S. P. Hernandez-Rivera, M. de Jesus, Castro, M.E., Mina, N., Briano, J.G., *Proc. of SPIE* **2004** , 5415, 1389.
40. A. Torres, I. Padilla, S. Hwang, *Proc. of SPIE* , 2005, 6553, 65531Q.
41. G. M. Herrera-Sandoval, L. M. Ballesteros, N. Mina, J. Briano, M. E. Castro, S. P. Hernandez-Rivera, *Proc. of SPIE* **2005** , 5794, 1245-1253.
42. A. Blanco, N. Mina, M. E. Castro, J. Castillo-Chara, S. P. Hernandez-Rivera, *Proc. of SPIE* **2005** , 5794, 1281.
43. L. M. Ballesteros, G. M. Herrera, M. E. Castro, J. Briano, N. Mina, S. P. Hernandez-Rivera, *Proc. of SPIE* **2005** , 5794, 1254.
44. S. P. Hernandez-Rivera, C. A. Manrique-Bastidas, A. Blanco, O. M. Primera, L. C. Pacheco, J. Castillo-Chara, M. E. Castro, N. Mina, N., *Proc. of SPIE* **2004** , 5415, 474.
45. C. Osorio, L. M. Gomez, S. P. Hernandez-Rivera, M. E. Castro, *Proc. of SPIE* **2005** , 5794, 803.
46. C. A. Manrique-Bastidas, N. Mina, M. E. Castro, S. P. Hernandez-Rivera, *Proc. of SPIE* **2005** , 5794, 1358.
47. A. Blanco, L. C. Pacheco-Londono, A. J., Pena-Quevedo, S. P. Hernandez-Rivera, *Proc. of SPIE* , **2006** , 6217, 621737.
48. C. W. Chan, G. H. Huang, *Engineering Applications of Artificial Intelligence* **2003** , 16, 75.
49. L. Iliadis, C. Jayne, "Engineering Applications of Neural Networks," A. J. Turner, Ed., *12th INNS EANN-SIG International Conference, EANN2011* , **2011** .
50. L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, Andreas Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt, G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," *European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases* , Prague, Czech Republic, **2013** .
51. J. Zhu, S. Rosset, J. Zhu, H. Zou, *Statistics and its interface* **2009** . 2, 349.

52. L. F. Alzate, C. M., Ramos, N. M. Hernandez, S. P. Hernandez, N. Mina, *Vib. Spectrosc* **2006** , 42, 357.
 53. R. Infante-Castillo, L. C. Pacheco-Londono, S. P. Hernandez-Rivera, *J. Mol. Struct* **2010** , 970, 51-58.
 54. L. C. Pacheco-Londono, J. L. Ruiz-Caballero, M. L. Ramirez-Cedeno, R. Infante-Castillo, N. J. Galan-Freyre, S. P. Hernandez-Rivera, *Molecules* **2019** , 24, 3494.
 55. R. Infante-Castillo, L. Pacheco-Londono, and S. P. Hernandez-Rivera, *Spectrochim. Acta, Part A: Molecular and Biomolecular Spectroscopy* 2010, 76, 137.

GRAPHICAL ABSTRACT

A Machine Learning (ML) strategy was used to enhance mid-infrared (MIR) laser reflectance spectra based detection of high explosives (HE) in mixes with soil samples. The Artificial Intelligence (AI) methodology termed “Self-Simulated Learning Artificial Intelligence” (SSLAI) demonstrated that it is possible to detect a chemical threat in solid matrices such as natural soil. There is an excellent possibility to transfer the methodology to field applications where the detection time is essential, and where often these solid matrices are unknown.

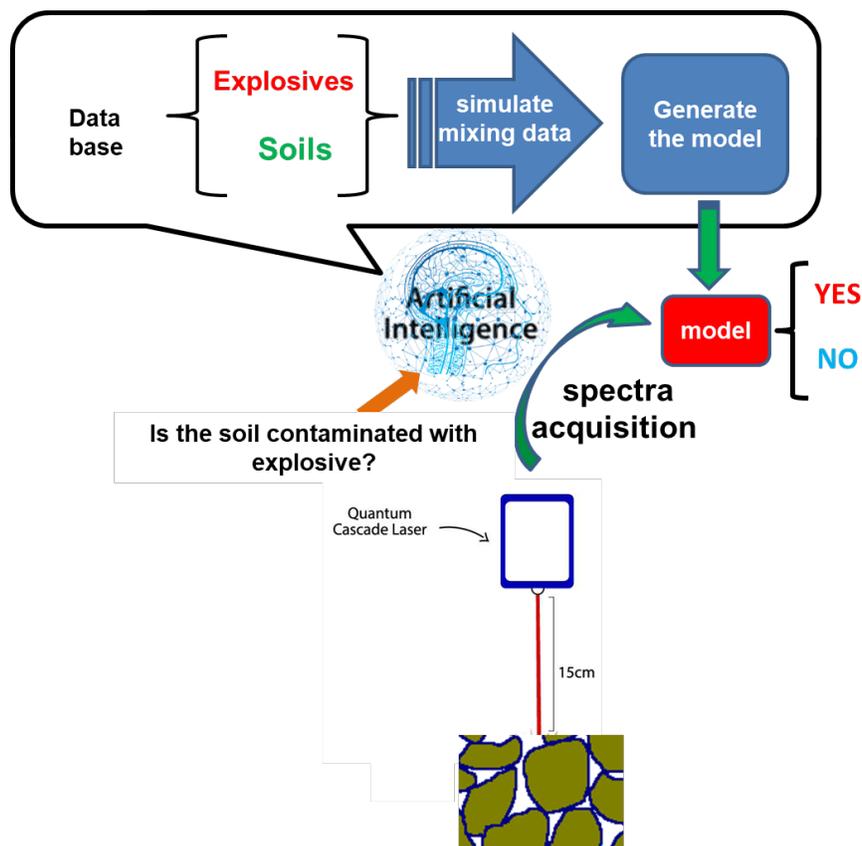


Table 1. Basic description of machine learning methods for classification ^[50]

ML METHODS	DESCRIPTION
K-Neighbors Classifier (KNC)	K-Neighbors Classifier is a neighbors-based classification where k is an integer value.
SVM	Support vector machines (SVM) and NuSVM are algorithms capable of performing classification.
NuSVM	
Decision Tree Classifier (DTC)	Decision Tree Classifier is a non-parametric supervised learning method. It is an algorithm for classification and regression.
Random Forest Classifier (RFC)	Random Forests Classifier is an ensemble learning method for classification, that uses multiple decision trees.

ML METHODS	DESCRIPTION
AdaBoost Classifier (ABC)	An AdaBoost (51) classifier is a meta-estimator that begins by fitting a classifier
Gradient Boosting Classifier (GBC)	Gradient Boosting Classifier builds an additive model in a forward stage-wise fas
Gaussian Naive Bayes (GNB)	In the Gaussian Naive Bayes, the likelihood of the features is assumed to be Gau
Linear Discriminant Analysis (LDA)	Linear Discriminant Analysis is a classifier with a linear decision boundary gener
Quadratic Discriminant Analysis (QDA)	Quadratic Discriminant Analysis, it is a classifier with a quadratic decision bound

Table 2. Performance evaluation parameters for the RFC model

	Precision	Recall	f1-score
	Test_H > 10% w/w	Test_H > 10% w/w	Test_H > 10% w/w
NONE	0.98	1.00	0.99
TNT	1.00	1.00	1.00
RDX	1.00	0.93	0.96
	10 % w/w > Test_M > 3% w/w	10 % w/w > Test_M > 3% w/w	10 % w/w > Test_M > 3% w/w
NONE	0.96	1.00	0.98
TNT	1.00	0.82	0.90
RDX	1.00	0.81	0.89
	Test_L < 3% w/w	Test_L < 3% w/w	Test_L < 3% w/w
NONE	0.84	1.00	0.91
TNT	1.00	0.11	0.20
RDX	0.75	0.15	0.24

Hosted file

AIL-20 AI QCL-SOIL R VILLAREAL FIGURE 1 2-MAR-20.tif available at <https://authorea.com/users/301976/articles/431924-self-simulated-learning-artificial-intelligence-for-the-detection-of-high-explosives-in-soil-by-mid-infrared-laser-spectroscopy>

Hosted file

AIL-20 AI QCL-SOIL R VILLAREAL FIGURE 2 2-MAR-20.tif available at <https://authorea.com/users/301976/articles/431924-self-simulated-learning-artificial-intelligence-for-the-detection-of-high-explosives-in-soil-by-mid-infrared-laser-spectroscopy>

Hosted file

AIL-20 AI QCL-SOIL R VILLAREAL FIGURE 3 2-MAR-20.tif available at <https://authorea.com/users/301976/articles/431924-self-simulated-learning-artificial-intelligence-for-the-detection-of-high-explosives-in-soil-by-mid-infrared-laser-spectroscopy>

Hosted file

AIL-20 AI QCL-SOIL R VILLAREAL FIGURE 4 2-MAR-20.tif available at <https://authorea.com/users/301976/articles/431924-self-simulated-learning-artificial-intelligence-for-the-detection-of-high-explosives-in-soil-by-mid-infrared-laser-spectroscopy>