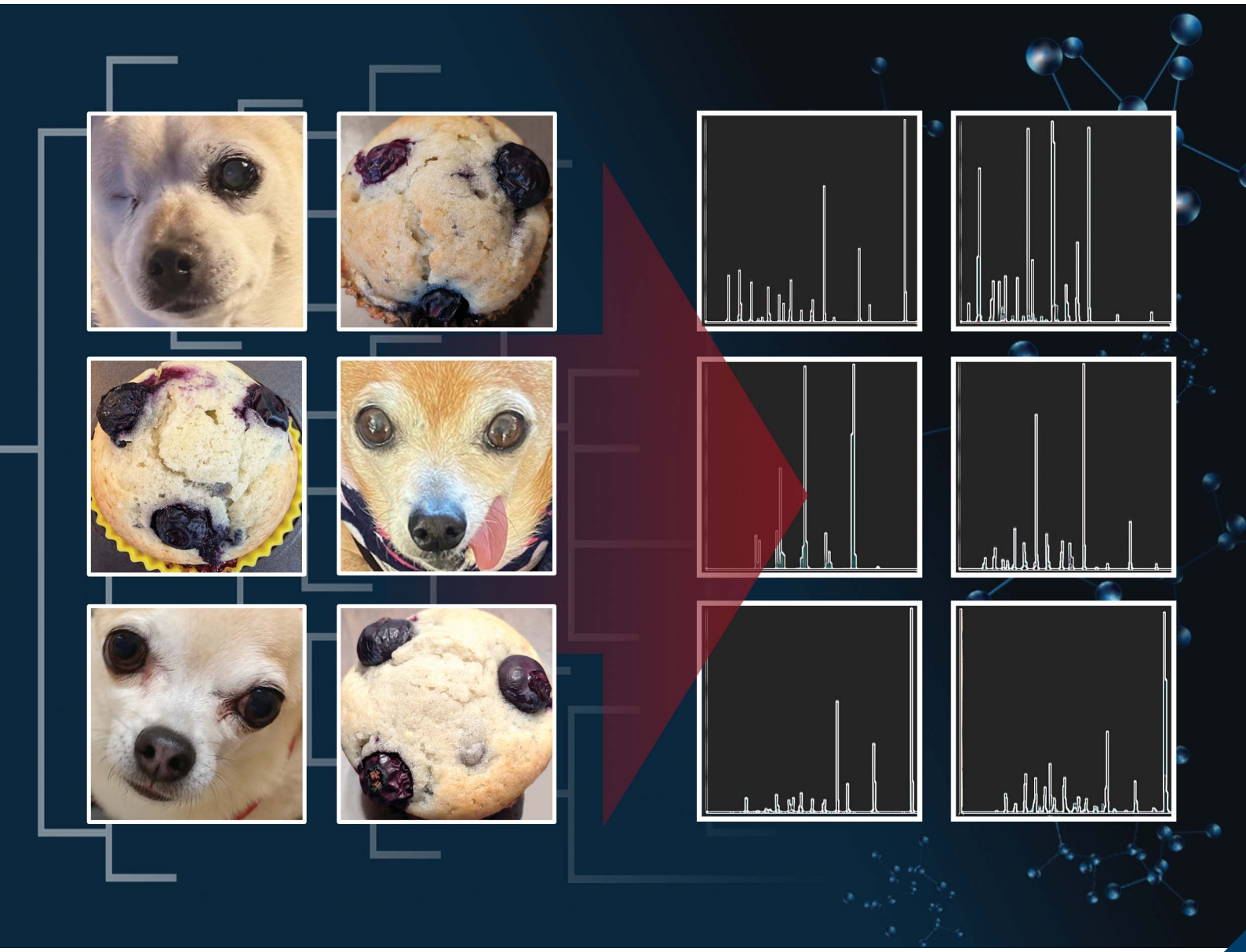
Volume 18
Number 9
5 March 2026
Pages 1759–1960

Analytical Methods

rsc.li/methods



ISSN 1759-9679

ROYAL SOCIETY OF CHEMISTRY

PAPER
Carolyn L. Fisher *et al.*
Random forest models accurately classify synthetic opioids using high-dimensionality mass spectrometry datasets

Indexed in Medline!

Analytical Methods

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Anal. Methods*, 2026, **18**, 1804

Random forest models accurately classify synthetic opioids using high-dimensionality mass spectrometry datasets

Kourosh Arasteh,[a] Steven Magana-Zook,[b] Colin V. Ponce,[c] Roald Leif,[df] Alex Vu,[ef] Mark Dreyer,[ef] Brian P. Mayer,[df] Audrey M. Williams[f] and Carolyn L. Fisher *[a]

Detection of novel threat agents presents several challenges, a principle one being the development of untargeted methods to screen an increasing number of threat chemicals whose exact structures are unknown. With the use of Machine Learning (ML) tools, we can guide the development of analytical methods for broad-spectrum detection of unbounded threat chemical families in complex mixtures. Toward this goal, we used nominal mass and high-resolution mass spectrometry data for hundreds of synthetic opioids and non-opioid compounds. We tested two ML techniques, logistic regression and random forest, to develop models towards a practical, implementable method for opioid detection. We found that of these tested ML methods, random forest models resulted in the highest validation accuracy (95+%) for both nominal mass and high-resolution classification of opioids *versus* non-opioids, with low false positive and false negative rates. The RF models were then used to successfully predict the classification of 10 compounds—five opioids and five non-opioids not part of the training and validation analysis. This application of ML is a critical step towards the development of field-deployable nominal mass spectrometers with ML-driven analyses for classification of emergent threats.

Received 7th October 2025
Accepted 29th December 2025

DOI: 10.1039/d5ay01677k

rsc.li/methods

Open Access Article. Published on 07 January 2026. Downloaded on 5/9/2026 6:40:39 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Introduction

Detection of novel and emerging chemical threats is a necessary component of nonproliferation efforts.[1] The clandestine synthesis of threat materials for illegal sale and use by state and non-state actors often aims to skirt identification which relies on existing databases through continual development of novel analogs and use of uncontrolled precursors. Additionally, combinatorial synthetic routes, such as the Ugi multicomponent reaction, result in known and unknown opioid analogs and byproducts[2] that are typically not captured in chemical reference libraries. Unfortunately, detection and identification of the resulting products often lag behind novel threat synthesis, putting forensic chemists at a perpetual disadvantage. Detection of hazardous materials relies on existing libraries and databases containing reference data (*e.g.*, mass spectra) of the chemical targets.[3] Such databases include pharmaceutical-based agents, illicit drugs, and biologically derived toxins. However, databases will always be incomplete for emergent threats.[4] Further, real-world samples may not generate mass spectra that can be adequately 'matched' to reference spectra due to minute quantities of the threat chemical relative to other chemicals, background–chemical interactions, or sheer complexity of chemicals in the collected sample. With modern, high-resolution mass spectrometers and nuclear magnetic resonance (NMR) spectrometers, *de novo* chemical structure elucidation is possible, but remains time-consuming, tedious, and dependent on subject matter expertise of the scientist.[4] In the case of chemical threat detection, rapid and reliable identification is paramount to resolving potential crises. Analytical chemistry requires new tools to allow rapid detection of known and novel threats.

The development of methods that broadly screen for known and emerging threat agents and provide their comprehensive characterization in an ever-evolving landscape would be instrumental to the field of analytical chemistry. Many of the "gold standard" fieldable detection platforms include high pressure mass spectrometers (HPMS), Fourier-Transform Infrared (FT-IR) instruments, Raman detectors, and ion mobility mass spectrometers (IMS), such as the MX908, HazMatID Elite, TruNarc, and the IONSCAN 600, respectively.[5]

[a]*Biosciences and Biotechnology Division, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. E-mail: fisher77@llnl.gov*

[b]*Computing, Global Security Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA*

[c]*Center for Applied Scientific Computing, Global Security Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA*

[d]*Nuclear and Chemical Sciences Division, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, 94550, USA*

[e]*Material Science Division, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA*

[f]*Forensic Science Center, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA*

This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 07 January 2026. Downloaded on 5/9/2026 6:40:39 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

These products were found to easily detect known opioids at high concentrations in pills and powders but performed poorly at concentrations of 10% analyte.[5] At best, these techniques still only provide presumptive evidence and confirmatory identification is still required to be done by gold-standard benchtop analytical instruments, such as liquid chromatography-mass spectrometry (LC-MS) and gas chromatography-mass spectrometry (GC-MS).[6] These lab-based analytical methods supply the definitive identification and compliance to legal standards. Despite the gold-standard status, LC-MS and GC-MS analysis is still limited to chemical identification using reference libraries.

Analytical chemists often analyze complex chemical samples by first separating components *via* chromatography, followed by electron ionization (EI) or electrospray ionization (ESI) prior to mass spectrometry analysis. Common methods include gas and liquid chromatography, which separate chemicals based on volatility and polarity, respectively,[7,8] yielding time-resolved mass spectra for each substance. Software like NIST's AMDIS,[9] Agilent's Mass Profiler Professional,[10] and Waters' Progenesis,[11] and ThermoFisher's MassFrontier[12] assist in analyzing these data through two main steps: (1) deconvolution—identifying spectral components that fluctuate together, and (2) matching—comparing spectra to a database of known standards. While this approach is valuable, it has distinct limitations. The database-centric approach focuses on individual, known chemicals; software packages are incapable of identifying novel analogs of known chemicals if they are not explicitly in the database.[4] The deconvolution step becomes problematic when chemical separation is poor or analyte concentrations are low, such as in recent years where opioids make up only a trace quantity in a street drug with many other constituents and impurities.[4,13] In these cases, deconvolution may produce highly unreliable spectra, at which point the matching is unlikely to produce meaningful results. By focusing on moiety identification, instead of chemical identification, through mass spectra analysis, we can build new analytical pipelines for chemical characterization without the same reliance on chemical reference libraries. In this context 'moiety identification' refers to ion fragments that are differentially correlated with different chemical classes and are thus distinguishing (either individually or en masse) and helpful with chemical classification of unknown chemicals not found in reference libraries. Development of data processing techniques is required to enable expedited and reliable identification of chemical classes, such as synthetic opioids, even in the case of novel compounds, poor chemical separation, and low analyte concentrations.

Machine Learning (ML), a subfield of artificial intelligence, involves the use of computational systems to learn from problem-specific training data and automate the analysis of new data within the problem space.[14] Use of ML has increased dramatically in recent years as the set of practical applications for the technology has developed.[15] Initially concerned with data-driven methods of analysis, the ML field has grown beyond pattern recognition to a widespread set of techniques useful across domains of imaging,[16] physical sciences,[17] and human social interaction,[18] among others. Along with advancement of data-driven ML techniques, improved instrumentation in academia and industry has yielded volumes of data so vast that traditional analysis methods can become infeasible.[19] Creation of 'big datasets' has led to the development of ML models that are robust to systemic noise, sensitive to latent patterns and trends, quick to provide decision support, and interpretable to human users. Despite these advances, the success of new ML approaches is heavily dependent upon having a high-quality dataset, data features easily encoded into a machine-interpretable format, and a well-defined problem statement.

ML has been applied to liquid chromatography-high resolution mass spectrometry (LC-HRMS) and GC-MS data in previous work. ML techniques have been applied to LC-HRMS data for identification of structurally similar trichothecene toxins without dependence on reference standards.[20] Other efforts have focused on predicting structural features of metabolites from raw GC-MS spectra[21] and curated features from GC-MS data[22] using decision trees.[23] A decision tree is an ML structure that recursively splits a dataset based on the values of the dataset's variables and uses this to make decisions. A Random Forest (RF) is an ensemble of many decision trees and is an established and powerful ML technique.[28] More recent works in automated GC-MS-based chemical property prediction[29,30] have utilized RF models,[28] an ML method that uses ensembles of diverse decision trees produced by training on randomized features and subsamples of a dataset (Fig. 1). RF models have exhibited robustness and high performance without extensive tuning, and their structure can be interpreted to extract the most valuable features during training.

The application of ML to GC-MS spectra of opioids, synthetic opioids, and fentanyl-related compounds in the literature has been largely focused on the subjective identification of discriminative fragments. For example, previous efforts identified potential screening fragments manually, focusing mainly on high-intensity fragment ions derived from the analysis of 'parent' chemicals with the use of computer-aided analysis.[24] However, chemical fragments can be identified more objectively using statistically driven methods and models. Statistical analysis methods were leveraged for determination of chemical attribution signatures for 3-methylfentanyl and three methods of its production.[25] More recently, atmospheric pressure solids analysis probe (ASAP) mass spectrometry was used to rapidly analyze 250 synthetic opioids, generating a high quality mass spectral library that could be utilized for ML model training to expedite data processing and screening.[26] However, not every foray into ML has been successful. For example, the METLIN team tried to predict MS fragmentation patterns using ML for an *in silico* dataset but failed to provide accurate chemical identifications in practice and was subsequently removed from the database.[27] Ultimately, the quality of the ML algorithm that is produced is directly related to the data type, quality, and quantity it is trained on. Sophisticated and complex ML models are not always necessary for producing successful algorithms if data quality and quantity is sufficiently high to address a particular problem.

Here we describe our work leveraging ML techniques to automate the classification of mass spectral data as either 'opioid' or 'non-opioid' with high accuracy (>95%). To

Open Access Article. Published on 07 January 2026. Downloaded on 5/9/2026 6:40:39 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

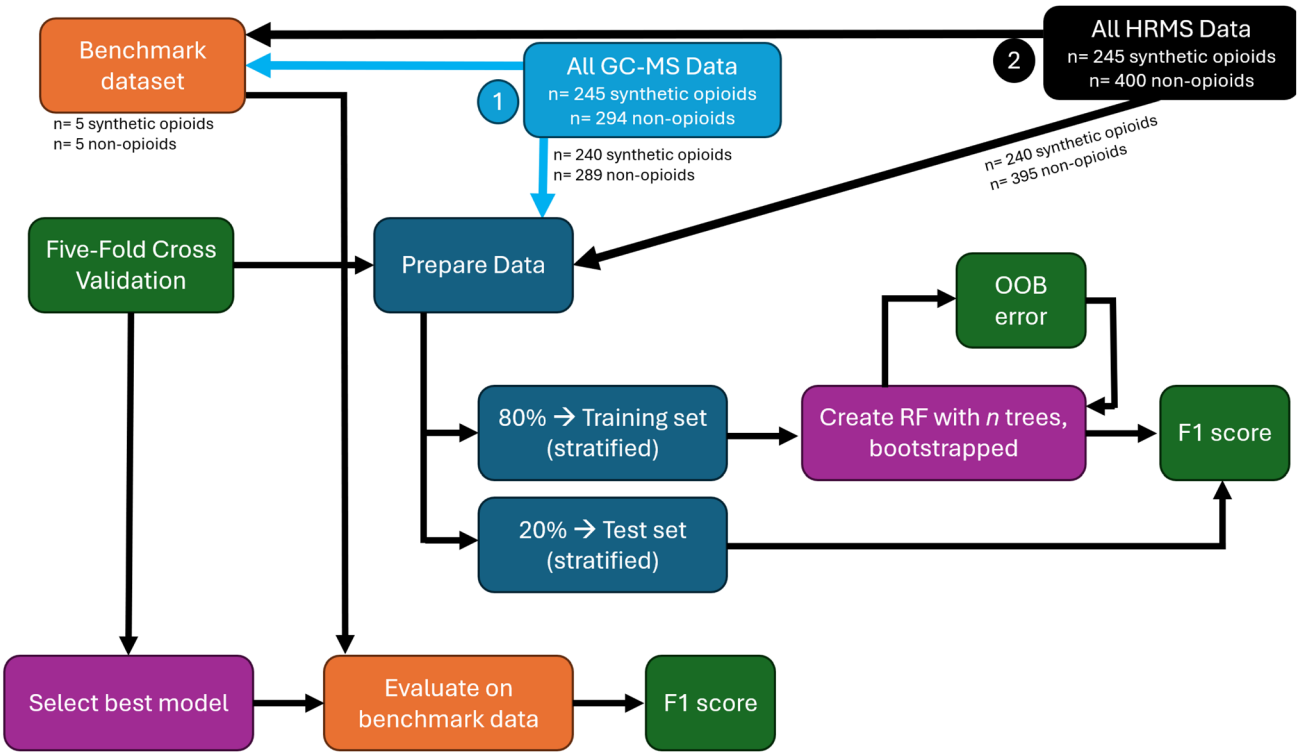


Fig. 1 Diagram summarizing the ML methods and techniques applied to the MS datasets used in this study. In the first run through of the ML-methods for model development, GC-MS data (defined as 245 synthetic opioids and 294 non-opioids) have $n = 10$ benchmark data removed before preparation of the remaining data for the stratified 80/20 train/test split and subsequent steps. Likewise, in the second (2) run through of the ML-methods above, HRMS data (defined as 245 synthetic opioids and 400 non-opioids) have $n = 10$ benchmark data removed before preparation of the data and the remaining steps in the method. Orange refers to the benchmark dataset; green refers to the various methods of calculating error; blue refers to the dataset split into training and testing sets; purple refers to steps involving the ML model development or application.

determine the best ML model for this specific binary classification task, we compared the performance of random forest (RF) and logistic regression (LR) models trained on GC-MS data and high-resolution MS/MS (HR MS/MS) data. The GC-MS and HR MS/MS data were generated on a non-exhaustive list of ~250 synthetic opioids and ~480 non-opioid chemicals. Unlike Monroy *et al.* 2025,[31] we collected data from two different chemical groups using two different analytical techniques, with different levels of resolution (*i.e.*, nominal mass and high accuracy mass). Through these additional data collection and curation efforts, we were able to assess the value for both electron and electrospray ionization (EI and ESI), low and high mass accuracy data, and binary classification for two chemical groups. Due to differences in the analytical methods used, not all compounds resulted in usable data by both GC-MS and HR MS/MS analytical methods. Still, we found that nominal mass GC-MS data and HR MS/MS data both resulted in highly effective ML models (validation accuracy >95%) for binary classification of synthetic opioids from non-opioid compounds. Surprisingly, the ML models based on the nominal GC-MS mass data performed at the same level of efficacy as the ML models based on HR MS/MS data. These results demonstrate the utility of ML applications on even low resolution (*e.g.*, nominal mass) data sets. Further development of ML classification techniques will enable expedited and reliable identification of chemical groups without available reference spectra.

Materials and methods

Materials and sample preparation

All solvents were purchased from Fisher-Scientific (Hampton, NH, USA); acetonitrile and methanol (99.9% minimum) were Optima LC-MS grade reagents, while dichloromethane (99.8% minimum) and ethyl acetate (99.9% minimum) were GC-MS grade reagents. Ultrapure water was produced using a Milli-Q IQ 7000 with QPOD dispenser. A Fentanyl Analog Screening (FAS) kit and FAS Emergent Panels 1–4 were acquired from Cayman Chemical (Ann Arbor, MI) in collaboration with the Centers for Disease Control and Prevention. Each compound of the opioid kit was reconstituted and dissolved using LC-MS grade methanol producing a nominal 400 μg mL^{-1} solution. Further dilution in methanol was carried out to produce a 10 μg mL^{-1} standard amendable to GC-MS analysis. Additionally, serial dilutions were performed using a mixture of ultrapure water and methanol to prepare standards for HR MS/MS analysis. Similarly, a "non-opioid" set of chemical compounds were compiled from multiple sources and diluted with ethyl acetate to 10 μg mL^{-1}. This included a set of ISO-certified multi-

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 07 January 2026. Downloaded on 5/9/2026 6:40:39 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

component pesticide reference materials that were obtained from Thermo Scientific (Waltham, MA). These standards were available as part of the Thermo Scientific Pesticide Explorer Collection. Additional non-opioid reference materials were obtained from AccuStandard (New Haven, CT), in a custom multi-component analyte mix (S-22329) and as single-component solutions. These substances were selected to represent compounds from the following classes of chemicals: organochlorine, organobromine, organophosphorus, organosulfur, nitroaromatic, polycyclic aromatic hydrocarbon, *N*-heterocyclic aromatic, alkane hydrocarbon, fatty acid methyl ester, and phthalate ester. All chemicals used in this study are summarized in SI Table 1 and 2.

Gas chromatography-mass spectrometry data collection

Chemicals were analyzed on an Agilent 7890A GC coupled to an Agilent 5975C MS detector. The GC column used for the analysis was an Agilent JW DB-5ht capillary column (30 m × 0.25 mm id × 0.10 μm film thickness). Ultra-high purity helium, at 1.0 mL min^{-1}, served as the carrier gas. The inlet was operated in pulsed splitless mode (25 psi for 1 minute, followed by a 50 mL min^{-1} purge flow), with the injector temperature set at 265 °C and injection volumes of 1 μL. The oven temperature program was as follows: 50 °C, held for 1 min, increased at 30 °C min^{-1} to 335 °C, held for 4.5 min. The MS ion source and quadrupole temperatures were 230 and 150 °C, respectively. Electron ionization was used with an ionization energy of 70 eV. The MS was scanned from m/z 29 to 600 in 0.4s, with a solvent delay of 5 min. Through the use of chromatographic separation and the Automated Mass Spectral Deconvolution and Identification System (AMDIS) software program developed by the National Institute of Standards and Technology (NIST), an individual mass spectrum was exported for each chemical (regardless of whether it was in a mixture or a pure analytical standard). This was done using an automated process to extract consistent and unbiased mass spectral data for every compound in the GC-MS files. The tabulated mass spectral data (nominal mass) for every chemical were normalized to the base peak in each mass spectrum and exported as a text file for ML applications.

High-resolution orbitrap mass spectrometry data collection

High-resolution accurate mass MS/MS data were collected on a ThermoFisher Scientific Q Exactive HF-X Orbitrap mass spectrometer equipped with a heated electrospray source in positive polarity. Material was directly injected into the mass spectrometry using direct infusion, bypassing any liquid chromatographic separation. Data from 10 μg mL^{-1} 50 : 50 methanol:water dilutions of all compounds were collected over a range of 20 collision energies (CE) repeated six times. To increase the breadth of fragments generated by each chemical and to test a wide range of CEs to generate ML models optimized for opioid detection, both synthetic opioid and non-opioid chemical precursors were evaluated using both linearly spaced (*i.e.*, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200) and multiplicatively spaced (or 'nonlinearly spaced'; *i.e.*, 10, 12, 14, 16, 19, 22, 26, 30, 35, 41, 48, 57, 66, 78, 91, 106, 125, 146, 171, 200) CE series. The RAW files generated for each directly infused compound were converted to processable text files for use in the Python3 environment using the RawConverter program developed previously by He *et al.* (2015).[32]

Machine learning methods

Random Forest (RF) and Linear Regression (LR) models for both GC-MS and HR MS/MS datasets were evaluated. First, five randomly selected synthetic opioid and five non-opioid mass spectra (from both GC-MS and HR MS/MS datasets) were removed to create a benchmark dataset and to be used to test the best model (Fig. 1); the mass spectra from the benchmark dataset were not previously used in the training or testing of any ML model. For the remaining spectra in both the GC-MS and HR MS/MS datasets, a standard ML holdout procedure (in which 20% of the data is uniformly at random selected to be 'held out' for testing purposes and not used for training) was applied. Data were stratified to maintain equal ratios of synthetic opioids *versus* non-opioids for all ML models tested. We used the 'f1_score' function from the scikit-learn metrics library, where the F1 score is calculated as the harmonic mean of precision and recall for the model's predictions on the validation set (*i.e.*, test set). This randomized method of training, or "bootstrapping" samples to build the trees, also provides a valuable estimate of generalizability and error of a model. The data that are not used for the construction of a given tree are considered an "out-of-bag" (OOB) observation (additional explanation below in note† and summarized in Fig. 2). Additionally, we applied five-fold cross validation which divides the dataset equally into five equally sized sets and then trains and tests the model five times. We validated models using three modes of calculating error: validation accuracy for each test set, cross validation accuracy summarized as an F1-score, and OOB error. All three are standard, well-accepted methods of measuring performance[23] and calculating error using orthogonal methods is a best practice in ML model assessment and validation.[23]

GC-MS model selection, training, and testing

Using the GC-MS data, LR and RF models for classifying opioids from non-opioids were assessed for performance. The performance of RF models with forest sizes with 1–200 trees were evaluated with bootstrapped samples to evaluate out-of-bag (OOB) error. Due to a slight imbalance in number of members in each of the two classes, class weighting during training was employed to avoid known issues with RF models trained on imbalanced data.[35] The RF model with the lowest tree count and

† For any given observation, there are several trees in a random forest that have not "seen" this observation (or in our case, mass spectra for a specific chemical) during training. We can use only those trees that have not trained on this observation to attempt to predict the observation's class, and in this way we evaluate prediction error. A classification is thus generated for all such observations, and when compared to the true classification of these observations, an error rate can be estimated. This error rate is known as OOB error.

Open Access Article. Published on 07 January 2026. Downloaded on 5/9/2026 6:40:39 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

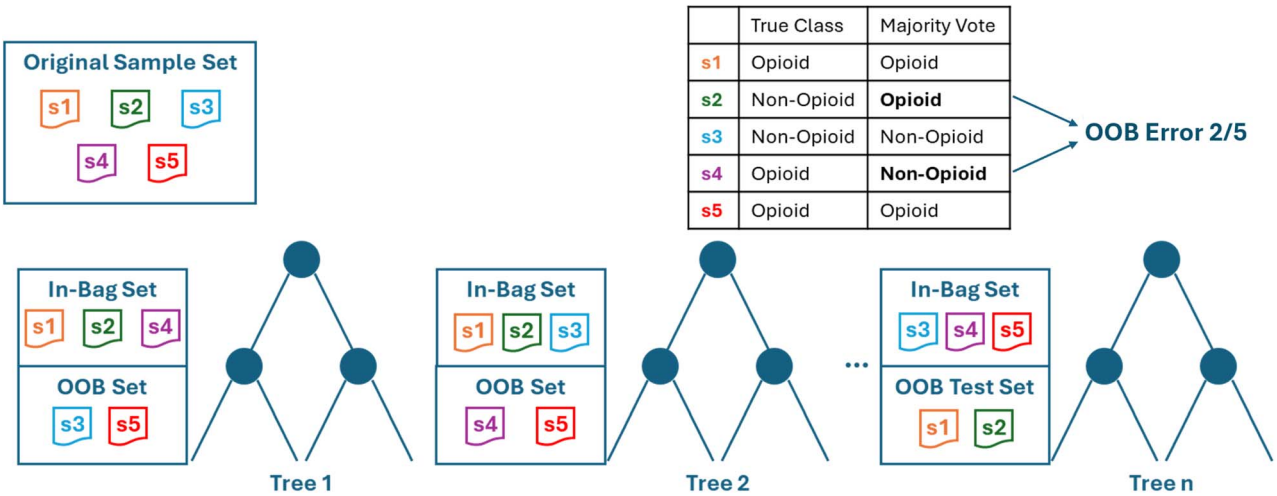


Fig. 2 Diagram of training by bootstrap aggregation (bagging) during RF model. For each round of training, the original training set of spectra is subsampled with-replacement to create an "in-bag set" for each decision tree in the random forest to be trained on. Each tree is evaluated on the complementary set of samples for each in-bag set, known as the "OOB set", and the consensus of the forest on these OOB samples produces the OOB error metric recorded over the training of the random forest model.

a stable, relatively low OOB error were further evaluated. This trained RF model was used on the benchmark dataset (*e.g.*, the 10 compounds pulled before the 80/20 split) and evaluated classification accuracy.

A binary, single-class LR model was utilized to serve as a baseline measurement of linear model efficacy on the GC-MS classification task. The LR model employed class-weighting similar to the RF models. To explore hyperparameter settings, a grid search was used to evaluate penalty norm (L1 *vs.* L2) and regularization strength (20 values from 10^{-4} to 10^{4} spaced on a log scale). All possible combinations of these hyperparameters were trained *via* stratified five-fold cross-validation. The best hyperparameter tuning was determined by finding the model with the highest accuracy on the left-out data in the cross-validation sets. The highest-scoring hyperparameters were used to create a new LR model, which was then trained on the training set, tested using the benchmark dataset, and evaluated for accuracy. As in the case of RF, a high accuracy on the test set would suggest the model will generalize to previously unseen data. The LR model that performed best used the L1 norm for loss and had an inverse regularization strength of $C = 4.28133$.

HR MS/MS model selection, training, and testing

As with the GC-MS data, the approach for applying RF models to HR MS/MS data utilized class weighting during training with bootstrapped samples. OOB error was used for quantifying model prediction errors when evaluating different forest sizes. For HR MS/MS data, RF models were trained with all forest sizes with 1–200 trees and RF model with minimal tree count that had a stable, relatively low OOB error was selected. This trained RF model was used on the benchmark dataset (*e.g.*, the 10 compounds pulled before the 80/20 split) and evaluated accuracy. After training, highest performing HR MS/MS RF model on the validation set was queried for the highest importance fragment ions. Due to the nature of higher-precision HR MS/MS data being reported in accurate mass, rather than nominal mass, fragment ions were aggregated to the nearest nominal mass (nearest integer) to compare feature importance with those found in the GC-MS case. (Fragment ions were not aggregated during the RF model training and validation; only during the process to generate Fig. 3B).

As with the GC-MS data set, a binary, single-class LR model was also used to produce a baseline expectation of model efficacy on the HRMS classification task. Class weighting was used and explored both L1 and L2 norms but modified the set of inverse regularization strengths to 10 values between 10^{-20} to 10^{10} spaced on a log scale. This modification was done to explore higher degrees of regularization to ensure convergence of the classifier. To compare performance with the LR model for GC-MS data, the highest-scoring hyperparameters were used to create a new LR model, which was then trained on the HRMS training set, tested using the benchmark dataset, and evaluated for accuracy. As in the case of RF, a high accuracy on the test set would suggest the model will generalize to previously unseen data.

Benchmark evaluation testing for GC-MS and HR MS/MS ML models

The best-performing, highest-accuracy ML models (defined by the validation accuracy percentage) were used to predict the classification of 5 synthetic opioids and 5 non-opioids that were randomly selected and previously set aside in the benchmark dataset.

Results and discussion

Mass spectral data obtained from synthetic opioids and non-opioid chemicals were collected on two instruments: (1) an electron ionization GC-MS for nominal mass data and (2) a HR MS with positive mode electrospray ionization for HR MS/MS

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 07 January 2026. Downloaded on 5/9/2026 6:40:39 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

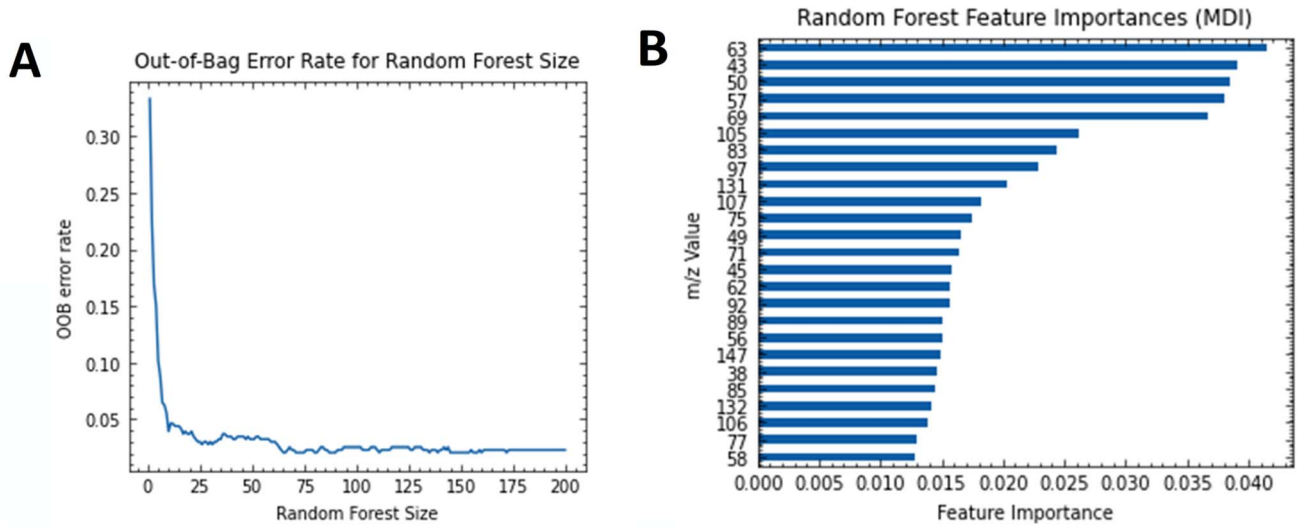


Fig. 3 (A) Out-of-bag (OOB) error during bootstrap selection of GC-MS training data for random forest models with 1–200 trees. (B) GC-MS feature importance scores determined from mean decrease in impurity (MDI) for the highest accuracy RF model we evaluated (75 trees).

data. Due to differences inherent to the GC-MS and HR MS/MS analytical techniques and instrumentation, only 245 of the synthetic opioids were analyzed by GC-MS and a different set of 245 were analyzed by HR MS/MS, with 242 synthetic opioids analyzed by both techniques (SI. Table 1). Of the ~480 non-opioid chemicals, 294 were analyzed by GC-MS, 400 were analyzed by HR MS/MS, and 216 chemicals were analyzed by both techniques (SI. Table 2). The total number of synthetic opioids and non-opioids that were analyzed by GC-MS, HR MS/MS, and both techniques is summarized in Table 1. Our goal was to develop ML models for classification with at least 95%+ accuracy in the validation data set, and we found this goal to be easily achievable for most LR and RF models for both the GC-MS and HR MS/MS datasets. As we utilized class-weighting and our data were relatively balanced, a high accuracy on the validation set would suggest the model will generalize to data not used to create the model.

Applying ML models to GC-MS data

Using the GC-MS data, random forest (RF) and linear regression (LR) models for classifying opioids from non-opioids were assessed. The best LR model had a mean training accuracy of 94.9% over five folds of cross-validation and a 96.2% accuracy on the validation test set (Table 2). The RF models for the GC-MS data generated higher training and validation accuracies, 97.9% and 99.1% respectively, than the LR models (Table 2). Thus, we next tested a range of RF decision tree ensembles up to 200 trees. Our results indicated that the model's out-of-bag (OOB) error rate stabilized at approximately 2.5% when the ensemble size reached around 75 trees, suggesting that further increases in the number of trees provided diminishing returns in terms of model performance (Fig. 3A). As described above, RFs are an ensemble of decision trees and RF models become more powerful as more trees are added to the ensemble. Thus, as all ensemble methods (including RF) have accuracies that

Table 1 Total number of synthetic opioids and non-opioids analyzed by GC-MS, HR MS/MS, or both analytical techniques (see SI Tables 1 and 2 for chemicals analyzed by each analytical technique)

	Number of synthetic opioids analyzed (out of 250)	Number of non-opioids analyzed (out of 478)
GC-MS	245	294
HR MS/MS	245	400
Both GC-MS and HR MS/MS	242	216

Table 2 Performance of models trained on GC-MS and HR MS/MS data

Model	Training accuracy	Validation accuracy	Validation F1-score
GC-MS RF (75 trees)	0.979	0.991	0.991
GC-MS LR (best model)	0.949	0.962	0.963
HR MS/MS RF (125 trees)	0.955	0.959	0.957
HR MS/MS LR (best model)	0.812	0.813	0.813

Open Access Article. Published on 07 January 2026. Downloaded on 5/9/2026 6:40:39 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

eventually plateau as the ensemble grows. 75 trees were not required for a high-validation accuracy RF model; we found that only 11 trees were needed to generate an RF model with a lower 4.6% OOB error (Fig. 3A). However, we chose to utilize the 75 tree RF model as this forest size was the smallest for which OOB error appeared to stabilize. In predicting the classes of the validation test set, the 75-tree RF model exhibits a validation F1-score of 99.1%.

After this RF model with 75 trees in the ensemble was generated, its structure was queried for the highest importance fragment ions that could be utilized for identifying the most discriminative fragment ions to separate the synthetic opioids from the non-opioids (Fig. 3B). We used both impurity-based feature importance (*i.e.*, "impurity profiling") and permutation importance to evaluate the highest importance fragment ions in differentiating between the opioids and non-opioids in our dataset. Impurity-based feature importance is computed from the structure of the RF model and can indicate how well a feature can be used to split the dataset cleanly when the feature's importance is measured across all trees in the model. Impurity-based feature importance tends to highlight features with many different values, such as m/z ratio, that vary widely between compounds. For example, if you have both m/z ratios and compound types (such as opioid or non-opioid), the model may focus more on the m/z ratio because it has more possible values. This measure shows which features help the model sort the training data, but it does not always mean those exact features will be useful for future samples, such as in the test dataset. Despite this caveat, it is interesting to note the model's resultant highly "important" features discovered experimentally *via* impurity-based calculations often corresponded to known fragments that are characteristic of the synthetic opioid class.

Permutation importance is a more computationally expensive metric that overcomes drawbacks of impurity based importance. This metric is calculated by randomly shuffling, or permuting, the values of a given feature, and computes the change in model performance. Permutation importance is less useful in cases where features are highly correlated. For example, we observed adjacent ions to be highly correlated, which can be expected for ^{13}C fragment isotope patterns. Nevertheless, we see some features selected *via* permutation importance that correspond to known fragmentation of opioids. Fig. 3B shows the top 25 fragment ions ranked in descending order of impurity-based feature importance for the best RF model (75 trees, 99.1% accuracy, Table 2). Specifically, these listed ions are not to be interpreted as necessary for defining the synthetic opioid class, but rather, these ions are the most important for distinguishing the opioids from non-opioids in our training set. Despite this caveat, one might expect some overlap between "defining" ions and "distinguishing" ions; indeed, ions 57, 77 and 105 all made the list, and these ions can be easily traced back to structural components (*i.e.*, fragments ions) of many fentanyls.[24,33] However, many other ions are of similarly high importance but cannot be directly traced back to a fentanyl moiety (*e.g.*, 63, 43, 50). Since the task assigned to the ML model was to distinguish two classes of chemicals, it is possible that these distinguishing ions are not descriptive for the opioid class but are present in the non-opioid class. Ultimately, depending on the classification task and data used, the feature importances will vary.

It is important to note that while decision trees generally have an ease of interpretability, other ML techniques do not. Random forests of decision trees introduce stochastic subsampling and tree construction that confound this property. However, RF models do allow for some understanding of feature importance during the training process and the decrease in accuracy that can result from removing certain features. For the problem of MS classification, these feature understanding methods can be used to derive fragments that are more diagnostic than others in classifying spectral signatures. This is useful in developing more targeted screens for better identification of opioids from trace samples or samples with a complex background, though future work is necessary to validate this as a path forward.

Testing the predictive power of the GC-MS RF model through evaluation of benchmark dataset

Our final validation for the GC-MS RF model was the classification of 10 compounds simulated to be of 'unknown' origin (*i.e.*, outside of the training and validation datasets). The 10 compounds, shown in Table 4, were randomly selected from both the synthetic opioid and non-opioid data sets to use for this benchmark dataset. This prediction step was also used to understand classification and feature selection processing time on a typical computer (4 cores @ 1.9 GHz). Table 4 includes the prediction (1/0) and the prediction score based on this RF model. Although this is only a small sample size of synthetic opioids ($n = 5$) and non-opioid chemicals ($n = 5$), the GC-MS RF model classified all ten compounds in a total of 0.037 seconds, or a rate of 270.02 compounds/second, with high RF scores, resulting in accurate classification assignments for all 10 simulated 'unknown' chemicals.

Applying ML to the HR MS/MS data to generate a high accuracy model

All GC-MS data was acquired at the conventionally accepted 70 eV, which provides adequate fragmentation for most compounds. The HR MS/MS data, which uses the soft ionization technique of ESI, relies on collisional dissociation for fragmentation. Since no single CE will generate all possible fragments, we applied linearly spaced and nonlinearly spaced collision energies (CEs) between 10–200 to maximize the number of fragments available to the ML models. The wide range of CEs was necessary to increase the full range of possible fragments that could be important for higher accuracy chemical classification. For small molecules, determining the appropriate CE to use for optimal precursor and fragment detection is impossible without prior efforts. The types of chemical bonds found in small molecules are too variable and selecting for only a single CE is usually a compromise between the detection of larger and smaller fragments. Furthermore, the appropriate CE selection for an unknown chemical is near impossible to know or predict, even if information about CEs for the chemical class

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 07 January 2026. Downloaded on 5/9/2026 6:40:39 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

is known. Therefore, trying a wide range of CEs and evaluating which ones have an impact on the model makes the most sense when training ML models to classify unknown chemicals. Traditionally, analytical chemists utilize linearly spaced CEs, yet we found that the nonlinearly spaced CEs at lower voltage provided a comprehensive detection of mid-range fragments and were preferable for downstream ML models (Table 3). Notably, while a CE of 35 created the highest accuracy RF models with only 25 trees and the highest mean accuracy of 96%, a CE of 10 was able to use only 7 trees to create an RF model with an 84% mean accuracy (Table 3). Beyond a CE of 41, mean accuracy drops from 96% to 91%, indicating that higher CEs are a bit less distinguishing between the two chemical classes than lower CEs. However, despite the fact the CE of 200 is incredibly high and likely not the most optimal choice for any chemical analyzed in this dataset, this result showing that even a CE of 200 can generate a highly accurate ML model with a RF of 35 trees and 91% accuracy (Table 3) in distinguishing synthetic opioids from non-opioids is nonetheless noteworthy. Our results show that even a sub-optimal CE for two entire classes of chemicals still produced an ML model that effectively distinguished the two classes. This result is important to the field of ML in MS as new ML approaches continue to be explored, tested, and optimized.The HR MS/MS RF model with the highest testing accuracy, and included all CEs, was 95.5% for an ensemble of 125 trees with the lowest OOB error (Table 2 and Fig. 4A). For HR MS/MS data, OOB error decreases linearly until around 125 trees, perhaps due to higher cardinality of ion features. It is interesting to note that both nominal mass GC-MS data and the HR MS/MS data both resulted in RF models with very high accuracy (>95%); one dataset did not distinctly outperform the other. We expected the HR MS/MS data–which included comprehensive coverage (20 nonlinearly spaced CEs) of high mass accuracy fragmentation data for all compounds–would generate models with much higher validation accuracies than the GC-MS models. Instead, we found that the best HR MS/MS RF model had a validation accuracy of 95.9% and the GC-MS RF model had 99.1% (Table 1). Additionally, the best LR model for the HR MS/MS data only had 81.3% validation accuracy, about 15 percentage points behind the GC-MS LR model with 96.2% validation accuracy (Table 2).

Table 3 Number of trees in each Random Forest (RF) ensemble at specific Collision Energies (CEs) with the highest corresponding mean accuracy as determined by five-fold cross validation (CV5)

	Lowest number of RF trees in an ensemble with the highest mean accuracy (%) using CV5	
HRMS collision energy used	Number of trees	Mean accuracy (%)
10	7	84
12	35	89
14	45	92
16	45	94
19	25	95
22	45	96
26	45	96
30	45	96
35	25	96
41	40	96
48	40	95
57	45	95
66	35	94
78	25	94
106	45	94
125	25	93
146	25	93
171	35	92
200	35	91

We report the top 25 fragment ions aggregated to the nearest nominal mass for easy comparison to the features in the GC-MS analysis (Fig. 4B). Viewing the top 25 fragment ions, we note two of the same characteristic ions of fentanyl (77 and 105) as highly important to the model along with two other ions (91 and 188) as well, all of which are in alignment with previously published work.[24–26] Such ions are characteristic of many (but not all) fentanyl analogs [summarized in 33]. The dependence of both the GC-MS and HR MS/MS trained RF models on some of the same *m*/*z* ions (especially 105 and 77) suggests the two RF models might have some similar structure, despite different datasets and data structures. As such, the ions that are denoted with high feature importance by the models do not need to be chemically characterized or identified to be of use for distinguishing between chemical classes using an ML model or in the development of a targeted MS experiment. While this might seem counter-intuitive to a chemist, the chemical identity of fragment ions is of no consequence to the ML models developed in this work.

It is well understood that inherent biases exist for all ML methods due to differences in datasets, and the GC-MS and HR MS/MS training data that were used to generate these ML models were indeed different (see SI Tables 1 and 2). However, we also believe that the differences in the models could also potentially be because of the accurate mass readings of HR MS/MS. The diffuse peak intensities across several features can be more difficult for the RF model to analyze. For example, *m*/*z* fragments 188.1434, 188.1435, and 188.1436 were present within the HR MS/MS dataset for several chemicals in the synthetic opioid datasets (data not shown) yet all three are within the 3-ppm mass error for the instrument. Thus, from the perspective of the instrument, all represent the same chemical fragment ion. However, the machine learning model we developed and employed did not bin any ions as a signal feature. When binning is not performed, the model treats 188.1434, 188.1435, 188.1436 as three separate features (instead of one chemical identified by the instrument), and thus further increases the complexity of the dataset. In ML, complexity is hard to learn and can inhibit model development, refinement, and validation. We attempted one method of binning by "clipping off" the last decimal point for detected ions to lower the complexity of the dataset (*e.g.*, the 188.1434, 188.1435, and 188.1436 ions would all coalesce into a single "188.143" *m*/*z* ion and a single feature for the ML model to learn). By doing so, the ML models accuracies increased for the HR MS/MS data (data not shown), but this method was overly simplistic and needs to

Open Access Article. Published on 07 January 2026. Downloaded on 5/9/2026 6:40:39 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

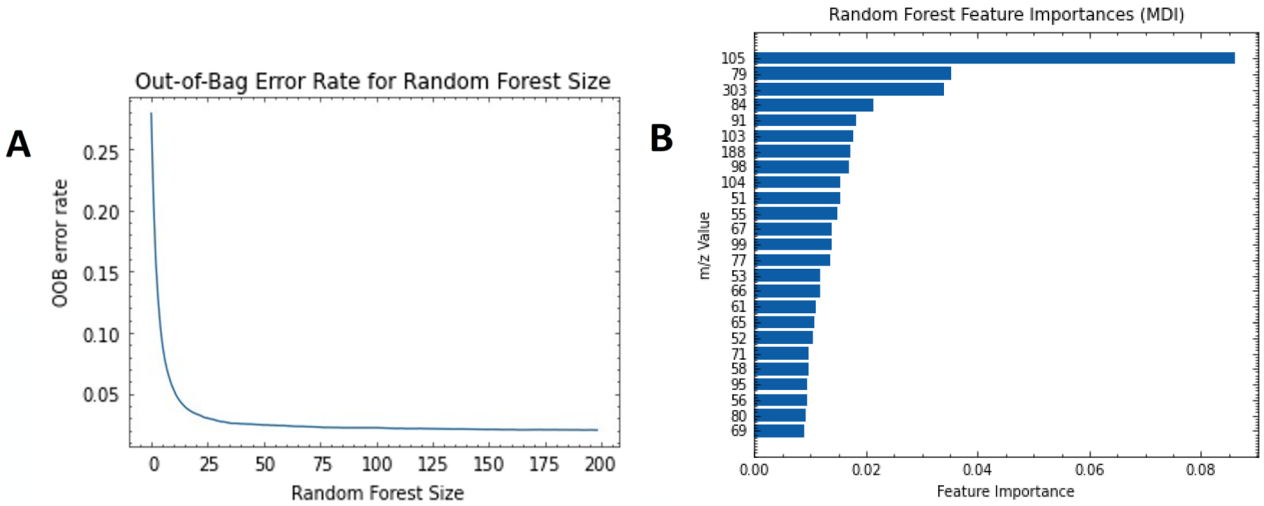


Fig. 4 (A) Out-of-bag (OOB) error during bootstrap selection of HR MS/MS training data for random forest models with 1–200 trees. (B) Feature importance scores determined from mean decrease in impurity (MDI) for the highest accuracy RF model we evaluated (125 trees) on HR MS/MS data. Testing the predictive power of the HR MS/MS RF model through evaluation of benchmark dataset.

Table 4 GC-MS and HR MS/MS RF predictions during ML model development and testing were generated with each prediction score as a probability that the model would generate the true label of "1", where a synthetic opioid = 1 and non-opioid = 0

'Unknown' Chemical	GC-MS RF score	HR MS/MS RF score	True label
alpha-methylacetylfentanyl	0.97	0.85	1
despropionyl *ortho*-fluorofentanyl	0.89	0.92	1
Ethiofencarb	0.11	0.07	0
N-6-APDB fentanyl	0.95	0.82	1
Phenylfentanyl	0.80	0.90	1
Pirimiphos ethyl	0.08	0.25	0
Quizalofop ethyl	0.03	0.07	0
Rotenone	0.13	0.22	0
Thiazopyr	0.05	0.17	0
Thiofentanyl	0.95	0.69	1

be further refined. As such, binning is an important consideration for future ML models using MS datasets, especially with high resolution and multi-dimensional datasets. For this current work, the HR MS/MS ML models are correlating all dispersive ions (*e.g.*, 188.1434, 188.1435, 188.1436, *etc*) for a fragment into the correct classification with high accuracy (Table 2). By correctly binning fragment ions, the classification accuracies will increase beyond these already high values. We suggest further work toward developing a more nuanced representation of groups of ions as features, rather than treating each accurate mass as a separate feature for some applications, to yield better results with learning methods that use features as decision boundaries.

We performed the same final prediction test on the best HR MS/MS RF model (125 trees, 95.9% accuracy) through classification of the same 10 compounds in the benchmark dataset (Table 4). As with the GC-MS RF model, the HR MS/MS RF model classified all ten compounds in a total of 0.169 seconds, or a rate of 59.0 compounds/second, with high RF scores and accurate chemical classifications. The prediction scores between the GC-MS and HR MS/MS RF models varied, confirming that different RF models were generated based on the different datasets used. However, both models were still successful in their prediction task for binary classification of "unknown" chemicals.

Conclusions

Opioids are a significant chemical threat category in the landscape of the 21st century, and the ML methodologies developed through this work can eventually be applied to real-world sample analyses and rapid field-deployable screening tools to counter this threat. We used ML to develop a binary classification model to separately identify pure chemical standards of a non-exhaustive list of synthetic opioids and non-opioids (*i.e.*, structurally diverse pesticides, pollutants and hydrocarbons). MS data obtained from both nominal mass GC-MS and HR MS/MS instruments were used to develop models for binary classification to separate synthetic opioids and non-opioid chemicals with 95–99%+ accuracy. Interestingly, the nominal mass GC-MS data generated ML models with similar very high validation accuracy as the HR MS/MS ML models. While there are

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 07 January 2026. Downloaded on 5/9/2026 6:40:39 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

several reasons this could be the case (such as different instrumentation biases that resulted in different usable datasets), we hypothesize that high resolution accurate mass readings are simply not required for the RF model to effectively distinguish the studied chemical classes. Notably, both the GC-MS RF model and the HR MS/MS RF model generated predictions with RF scores that matched the true-label identification for each of the 10 chemicals in the benchmark dataset (Table 4). These results support that with high quality mass spectral data and optimized ML models, classification of unknown chemicals is achievable with high accuracy. Future work to increase the number of chemical classes, the complexity of the sample (*e.g.*, concentration of analyte, matrix conditions, *etc*), and the amount of data in the mass spectral databases leveraged for ML model development are necessary next steps before this technology can be implementable for real-world samples.

Our results support that traditional nominal mass GC-MS instruments can generate high quality data for building highly effective ML models. Such a result is noteworthy for future potential field-deployable operations when an HRMS instrument might not be a feasible option for data collection. Currently ion-trap mass spectrometer with soft electrospray ionization and MS/MS capabilities have been successfully transported and deployed in the field for rapid drug screening analysis.[34] Coupling such current capabilities with a ML-based analytical workflow would further expedite screening and data processing. Additionally, our work shows that by training ML models to classify chemical groups, reference-free classification of unknown chemicals is possible. Current field-deployable and benchtop mass spectrometers all rely heavily on reference libraries and mass spectral databases for identification. An impactful future workflow to screen samples on mass spectrometers and then leverage ML models for fast analysis and classification of chemicals into groups of interest would cut down on the number of missed unknown chemicals and data analysis time. This overall workflow, which includes the data curation and processing, ML model application and testing, and prediction testing and method error analysis, represents a broadly applicable approach that could be applied to screening and identifying other classes of threat compounds including novel biotoxins and explosives. Furthermore, ML models and importance feature determination could be used in developing a targeted data analysis method for opioid identification. Overall, this work supports the development of science and technology tools and capabilities to meet challenges in emerging threat identification.

Author contributions

Kourosh Arasteh: writing – original draft preparation, formal analysis, methodology, software, validation, writing – review and editing. Stephen Magana-Zook: investigation, formal analysis, methodology, software, validation. Colin V. Ponce: conceptualization, formal analysis, validation, writing – original draft preparation, writing – review and editing, funding acquisition. Roald Leif: data curation, formal analysis, investigation, writing – review and editing. Alex Vu: data curation, investigation. Mark Dreyer: data curation, formal analysis, investigation, writing – review and editing. Brian Mayer: conceptualization, data curation, writing – review and editing, funding acquisition. Audrey M. Williams: conceptualization, writing – review and editing, funding acquisition. Carolyn L. Fisher: conceptualization, formal analysis, methodology, software, validation, writing – original draft preparation, writing – review and editing, funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Data availability

Data for this article, including the source code, GC-MS, and HRMS datasets, are available at GitHub at https://github.com/kouarasteh/FentaML.

Supplementary information (SI): supplemental Table 1: list of synthetic opioid standards used to generate GC-MS and/or HRMS/MS data. Supplemental Table 2: list of non-opioid standards used to generate GC-MS and/or HRMS/MS data. See DOI: https://doi.org/10.1039/d5ay01677k.

Acknowledgements

C.L.F. acknowledges Lawrence Livermore National Laboratory support from Laboratory Directed Research and Development projects GS-21-FS-038, for the funding and execution of this project, and 23-LW-029, for the fine-tuning of the ML models and the writing and revising of the manuscript. The authors would also like to thank Deon Anex for productive and helpful conversations; Philip Paul (may he rest in peace) for early conversations in using mass spectrometry data for machine learning models; Michelle Rubin for her professional editing and refining of this manuscript; and Ian McGovern at the Data Science Institute at LLNL for his review, suggestions, and helpful comments for the manuscript. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, (LLNL-JRNL-2012291).

References

1 United Nations Counter-Terrorism Centre (UNCCT), *Ensuring Effective Interagency Interoperability and Coordinated Communication in Case of Chemical And/or Biological Attacks*, United Nations Counter-Terrorism Centre, New York, 2017.

2 A. Varadi, T. C. Palmer, N. Naselton, D. Afonin, J. J. Subrath, V. Le Rouzic, A. Hunkele, G. W. Pasternak, G. F. Marrone, A. Borics and S. Majumdar, *ACS Chem. Neurosci.*, 2015, **6**(9), 1570–1577, DOI: 10.1021/acschemneuro.5b00137.

3 A. Kabir and K. G. Furton, in *Handbooks in Separation Science, Gas Chromatography*, ed. C. F. Poole, Elsevier, Cambridge, MA, 2021, edn 2nd, pp. 745–791.

Open Access Article. Published on 07 January 2026. Downloaded on 5/9/2026 6:40:39 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

4 M. Joshi and E. Sisco, *WIREs Forensic Sci.*, 2023, **5**, 5–e1486, DOI: 10.1002/wfs2.1486.
5 A. Bradly, K. F. Mo, A. Melville, D. Saunders, S. Jensen, V. Lin, J. Mobberley, C. Doll, D. Orton and A. Heredia-Langner, Pacific Northwest National Laboratory, Performance assessment of field-portable instruments and assays for fentanyl and fentanyl-related compounds, 2023, https://www.pnnl.gov/sites/default/files/media/file/ST_PNNL_Fentanyl_Reference_Spectra_Final_Report_CLEARED_Public_Release.pdf, accessed August 22, 2025.
6 S. A. Borden, J. Palaty, V. Termopoli, G. Famiglini, A. Cappiello, C. G. Gill and P. Palma, *Mass Spectrom. Rev.*, 2020, **39**, 703–744, DOI: 10.1002/mas.21624.
7 V. L. McGuffin, in *Chromatography*, ed. E. Heftmann, Elsevier, Amsterdam, 2004, edn 6th, vol. 69, pp. 1–93, DOI: DOI: 10.1016/S0301-4770(04)80007-1.
8 O. Coskun, *North Clin. Istanb.*, 2016, **3**, 156–160, DOI: 10.14744/nci.2016.32757.
9 National Institute of Standards and Technology (NIST), Automated mass spectral deconvolution and identification system (AMDIS), 2019, https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:amdis, accessed August 22, 2025.
10 *Agilent Technologies, Mass Profiler Professional Software, Version 15.0*, Agilent Technologies Inc., Santa Clara, CA, 2015, https://www.agilent.com/en/product/software-informatics/mass-spectrometry-software/data-analysis/mass-profiler-professional-software, accessed August 22, 2025.
11 Waters Corporation, *Progenesis QI Software*, Waters Corporation, Milford, MA, https://www.waters.com/nextgen/us/en/products/informatics-and-software/mass-spectrometry-software/progenesis-qi-software/progenesis-qi.html, accessed August 22, 2025.
12 Thermo Fisher Scientific, *MassFrontier Spectral Interpretation Software, Version 8.0*, Thermo Fisher Scientific, Waltham, MA, https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/mass-frontier-spectral-interpretation-software.html, accessed August 22, 2025.
13 Drug Enforcement Administration (DEA), *DEA Safety Alert: DEA Laboratory Testing Reveals that 6 Out of 10 Fentanyl-Laced Fake Prescription Pills Now Contain a Potentially Lethal Dose of Fentanyl*, DEA, Washington, DC, https://www.dea.gov/alert/dea-laboratory-testing-reveals-6-out-10-fentanyl-laced-fake-prescription-pills-now-contain, accessed August 22, 2025.
14 C. Janiesch, P. Zschech and K. Heinrich, *Electron. Mark.*, 2021, **31**, 685–695, DOI: 10.1007/s12525-021-00475-2.
15 C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2007.
16 A. Voulodimos, N. Doulamis, A. Doulamis and E. Protopapadakis, *Comput. Intell. Neurosci.*, 2018, 7068349, DOI: 10.1155/2018/7068349.
17 G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto and L. Zdeborová, *Rev. Mod. Phys.*, 2019, **91**, 045002, DOI: 10.1103/RevModPhys.91.045002.
18 W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang and D. Yin, in *Proceedings of the 2019 World Wide Web Conference (WWW '19)*, ACM, New York, NY, USA, 2019, pp. 11, DOI: DOI: 10.1145/3308558.3313488.
19 C. W. Tsai, C. F. Lai, H. C. Chao and A. V. Vasilakos, *J. Big Data*, 2015, **2**, 21, DOI: 10.1186/s40537-015-0030-3.
20 B. P. Mayer, M. L. Dreyer, M. C. P. Conaway, C. A. Valdez, T. Corzett, R. Leif and A. M. Williams, *Anal. Chem.*, 2023, **95**(35), 13064–13072, DOI: 10.1021/acs.analchem.3c01474.
21 J. Hummel, N. Strehmel, J. Selbig, D. Walther and J. Kopka, *Metabolomics*, 2010, **6**, 322–333, DOI: 10.1007/s11306-010-0198-7.
22 K. Vancampenhout, K. Wouters, B. De Vos, P. Buurman, R. Swennen and J. Deckers, *Soil Biol. Biochem.*, 2009, **41**, 568–579, DOI: 10.1016/j.soilbio.2008.12.023.
23 L. Breiman, J. Friedman, C. J. Stone and R. A. Olshen, *Classification and Regression Trees*, CRC Press, Boca Raton, FL, 1984, DOI: DOI: 10.1201/9781315139470.
24 Q. Nan, W. Hejian, X. Ping, S. Baohua, Z. Junbo, D. Hongxiao, Q. Huosheng, S. Fenyun and S. Yan, *J. Am. Soc. Mass Spectrom.*, 2020, **31**, 277–291, DOI: 10.1021/jasms.9b00112.
25 B. P. Mayer, C. A. Valdez, A. J. DeHope, P. E. Spackman and A. M. Williams, *Talanta*, 2018, **186**, 645–654, DOI: 10.1016/j.talanta.2018.02.026.
26 K. A. Reyes Monroy, R. Koerber and G. F. Verbeck, *Rapid Commun. Mass Spectrom.*, 2025, **39**, 9, DOI: 10.1002/rcm.9994.
27 G. Siuzdak, The Analytical Scientist, 2025, https://theanalyticalscientist.com/issues/2025/articles/jan/the-xcms-metlin-story/, accessed August 22, 2025.
28 L. Breiman, *Mach. Learn.*, 2001, **45**, 5–32, DOI: 10.1023/A:1010933404324.
29 L. S. Whitmore, R. W. Davis, R. L. McCormick, J. M. Gladden, B. A. Simmons, A. George and C. M. Hudson, *Energy Fuels*, 2016, **30**, 8410–8418, DOI: 10.1021/acs.energyfuels.6b01952.
30 S. Qiu and J. Wang, *J. Food Sci.*, 2015, **80**, S2296–S2304, DOI: 10.1111/1750-3841.13012.
31 K. A. Monroy, R. McCrary, I. Parry, C. Webber, T. D. Golden and G. F. Verbeck, *J. Am. Soc. Mass Spectrom.*, 2025, **36**(3), 587–600, DOI: 10.1021/jasms.4c00455.
32 L. He, J. Diedrich, Y. Y. Chu and J. R. Yates III, *Anal. Chem.*, 2015, **87**(22), 11361–11367, DOI: 10.1021/acs.analchem.5b02721.
33 Cayman Chemical, Laboratory guide for fentanyl identification, naming, and metabolism, *Cayman Chemical*, Ann Arbor, MI, https://cdn2.caymanchem.com/cdn/cms/caymanchem/LiteratureCMS/800205.pdf, accessed August 22, 2025.
34 A. N. Couch, C. Chang and J. T. Davidson, *Forensic Sci. Int.*, 2025, **367**, 112381, DOI: 10.1016/j.forsciint.2025.112381.
35 C. Chen, A. Liaw and L. Breiman, *Using Random Forest to Learn Imbalanced Data*, University of California, Department of Statistics, Berkeley, CA, 2004, https://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf, accessed August 22, 2025.

This journal is © The Royal Society of Chemistry 2026