


Cite this: *Analyst*, 2025, **150**, 700

Optimized machine learning approaches to combine surface-enhanced Raman scattering and infrared data for trace detection of xylazine in illicit opioids†

Rebecca R. Martens,^a Lea Gozdziński,^a Ella Newman,^a Chris Gill,^{b,a,c,e}
Bruce Wallace^{d,e} and Dennis K. Hore^{id *a,e,f}

Infrared absorption spectroscopy and surface-enhanced Raman spectroscopy were integrated into three data fusion strategies—hybrid (concatenated spectra), mid-level (extracted features from both datasets) and high-level (fusion of predictions from both models)—to enhance the predictive accuracy for xylazine detection in illicit opioid samples. Three chemometric approaches—random forest, support vector machine, and *k*-nearest neighbor algorithms—were employed and optimized using a 5-fold cross-validation grid search for all fusion strategies. Validation results identified the random forest classifier as the optimal model for all fusion strategies, achieving high sensitivity (88% for hybrid, 92% for mid-level, and 96% for high-level) and specificity (88% for hybrid, mid-level, and high-level). The enhanced performance of the high-level fusion approach (F1 score of 92%) is demonstrated, effectively leveraging the surface-enhanced Raman data with a 90% voting weight, without compromising prediction accuracy (92%) when combined with infrared spectral data. This highlights the viability of a multi-instrument approach using data fusion and random forest classification to improve the detection of various components in complex opioid samples in a point-of-care setting.

Received 1st December 2024,
Accepted 15th January 2025

DOI: 10.1039/d4an01496k

rsc.li/analyst

1 Introduction

The ongoing overdose crisis has become the leading cause of death in British Columbia, Canada, claiming over 14 000 lives since 2016.¹ Over the past decade, the illicit drug market in many communities has experienced a rapid shift toward synthetic opioids, with fentanyl and its analogues replacing

heroin.^{2–5} Recently, the illicit opioid markets have grown more complex, with frequent and unpredictable changes in drug composition across North America, including the introduction of potent co-competing sedatives adulterated into drug mixtures.^{6–12} Among these, xylazine, a veterinary tranquilizer, has become a notable concern due to its severe health risks.^{9,10,13} Xylazine has been found to reduce heart rate and breathing and when combined with opioids significantly increases the risk of respiratory depression and death.^{12–14} Notably, the effects of xylazine are unresponsive to naloxone, complicating overdose interventions and heightening risks for people who use drugs, making it a critical target for detection in harm reduction efforts.¹⁵

Harm reduction-based drug checking initiatives provide people who use drugs with essential information on sample contents, supporting informed decision-making and overdose prevention.¹⁶ Currently, many point-of-care drug checking sites rely on paper-based antibody test strips and vibrational spectroscopy methods such as infrared absorption (IR) and Raman scattering.^{17,18} While off-label use of immunoassay test strips offer increased sensitivity, they can be challenging to use correctly and lack the ability to differentiate between drug analogues, which vary in potency and effect.^{19–22} IR and Raman instruments are relatively low cost, capable of in-field

^aDepartment of Chemistry, University of Victoria, Victoria, British Columbia, V8W 3V6, Canada. E-mail: dkhore@uvic.ca

^bDepartment of Chemistry, Vancouver Island University, Nanaimo, British Columbia, V9R 5S5, Canada

^cDepartment of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA 98195, USA

^dSchool of Social Work, University of Victoria, Victoria, British Columbia, V8W 2Y2, Canada

^eCanadian Institute for Substance Use Research, University of Victoria, Victoria, British Columbia, V8W 2Y2, Canada

^fDepartment of Computer Science, University of Victoria, Victoria, British Columbia, V8W 3P6, Canada

†Electronic supplementary information (ESI) available: Information on the composition of samples; outlier detection methods; PCA feature extraction; hyperparameter tuning for all fused models; weighted voting method for high-level data fusion; standalone SERS and IR model analysis; sample composition of the testing set and all model predictions. See DOI: <https://doi.org/10.1039/d4an01496k>


analysis, and offer richer sample composition data than test strips. However, they have higher limits of detection, thereby hindering their effectiveness for detecting potent adulterants at trace levels. Additionally, street opioid samples typically have complex matrices that may include a variety of cutting agents and often dyes that absorb throughout the visible and near-infrared, presenting challenges for Raman due to strong fluorescence. Both techniques also require extensive library matching and spectral interpretation, reducing accessibility for untrained operators.^{17,23–27}

To address these challenges, researchers and drug checking facilities have investigated spectroscopic techniques that increase sensitivity and reduce fluorescence interference, with some focus on surface-enhanced Raman spectroscopy (SERS).^{28–32} SERS has also been explored for opioid detection in complex sample matrices, including toxicology screening of blood and urine^{30,33–35} and identifying fentanyl in powdered samples.^{36–42} SERS has shown promise in quantifying trace fentanyl in complex drug matrices^{39,40} and for identifying adulterants in opioid samples using machine learning techniques.^{20,43} IR spectroscopy, by contrast, has been more widely used across North America in drug checking applications,^{17,44} including in automated analysis of single and multi-component drug mixtures.²⁷ Many point-of-care drug checking services currently rely on IR spectroscopy to detect and estimate concentrations of major components, providing a comprehensive mixture analysis for substances present above its 5% w/w detection threshold.^{17,18,23,27,45,46} Integrating a simple, low-cost, and portable SERS method into routine workflows could significantly improve the detection of lower-concentration adulterants, particularly in complex opioid samples where IR falls short and other inexpensive methods, such as immunoassay test strips, fail to differentiate between substances. In drug checking services, where minimizing manual interpretation while capturing detailed sample information to identify most or all compounds in illicit mixtures is a key research focus, both IR and SERS provide valuable, yet incomplete, information when used independently.

Identifying compounds in multi-component mixtures is challenging, prompting research into automated analyses and chemometric methods to address overlapping spectral features.^{27,43,47} Supervised machine learning algorithms are particularly advantageous for analyzing complex sample matrices,^{48–50} but classifier selection can be complicated by factors such as the instrumental platform, training dataset, and hyperparameters, all of which impact performance.^{51,52} As a result, researchers have explored and compared various classifiers,^{53–58} finding that high-performing algorithms like random forest (RF), support vector machines (SVM), and *k*-nearest neighbor (KNN) demonstrate high accuracy in identifying components within complex drug mixtures.^{27,43,59–61}

Recently, data fusion has emerged as a strategy to boost model performance by combining complementary spectroscopic data, to harness the unique advantages of each technique for better detection in complex samples.^{62–65} Studies assessing early, mid, and late data fusion approaches for

Raman and IR models have shown success in differentiating watercolor ink⁶² and in multi-component quantitation.⁶³ However, research on data fusion strategies for real-world drug detection remains limited, with few studies examining fusion techniques alongside various supervised models to determine the most effective combinations. As drug checking becomes more prevalent, it is essential to recognize that no single analytical method can address all compound identification needs.^{7,18} Thus, exploring synergistic combinations of data fusion techniques and classification models to advance automated trace compound identification is a promising area of research.

In this work, we examine three data fusion strategies using SERS and IR spectral data to evaluate the predictive performance of RF, SVM, and KNN optimized models for the detection of xylazine in complex opioid samples. Given the spectral interference that hinders xylazine detection using IR,^{43,66} we developed nine supervised machine learning classifiers to assess overall improvements or reductions to performance for early, mid, and late data fusion strategies. Our findings indicate that the three random forest models built using concatenated SERS–IR spectral data, extracted features from both spectra, and a high-level weighted fusion of SERS and IR models outperform SVM and KNN models. The high sensitivity and specificity afforded to RF classification using any fusion method illustrate the potential for effectively utilizing IR data with SERS for component prediction in point-of-care drug checking.

2 Experimental

2.1 Materials and sample selection

For SERS sample preparation, 50 nm gold nanoparticle (AuNP) solution (BBI Solutions, UK), magnesium sulfate anhydrous certified powder (MgSO₄, Fisher Chemical), and deionized water (18.2 MΩ cm, Barnstead Nanopure water, Thermo Fisher Scientific) were utilized. The illicit drug samples used in this study were acquired at Substance, the Vancouver Island Drug Checking project,^{18,67} in Victoria, BC, Canada. 218 opioid samples with either fentanyl, fluorofentanyl, or heroin identified as one of the main active components were selected for analysis. Each sample was received in powdered form and was finely ground and thoroughly mixed with a spatula prior to analysis. Drug composition was determined using benzodiazepine immunoassay test strips (Rapid Response, BTNX) and paper-spray mass spectrometry (PS-MS). Sugars, cutting agents, and analogue types were identified through FTIR spectral analysis, where a drug technician used library matching of reference spectra to detect specific compounds present above 5% w/w,¹⁸ these findings were incorporated into the final sample breakdown. The selected opioid samples (*n* = 218) contained various combinations of cutting agents, opioid analogues, and adulterants. The selected opioid samples were chosen to reflect the complexity of the multi-component opioid samples currently in circulation within the illicit drug



market. A subset of 50 samples was designated as the test set to evaluate the performance of all fusion-based techniques. This subset consisted of $n = 25$ samples with xylazine (median concentration 4.35% w/w), where concentrations represent the proportion of the target drug relative to the total sample weight. The composition breakdown of the all samples analyzed in this study is provided in Table S1.†

2.2 Sample preparation and data collection

SERS measurements. SERS measurements were recorded in the 200–2000 cm^{-1} Stokes shift range using a portable Raman spectrometer (Resolve, Agilent Technologies, Santa Clara, USA) that has an 830 nm laser wavelength and a spectral resolution of 15 cm^{-1} . Details regarding the sampling method and instrument specifications can be found in previous publications.^{20,43,68} Briefly, around 1.5 mg of powdered sample was dissolved into ≈ 700 μL of deionized water, the solution was vortexed and heated to ensure homogeneity. 70 μL of the sample solution was spiked into 1.42 mL of the OD 1 50 nm AuNP solution and vortexed for 30 s. Finally, 10 μL of the aggregating agent (1 M MgSO_4) was added to make up the final 1.5 mL solution, which was vortexed again for 10 s. Opioid sample spectra were acquired at a constant laser power, collecting 10 averages, with a 2 s integration time.

FTIR measurements. A portable FTIR spectrometer (Agilent 4500a, Santa Clara, USA) equipped with a single-bounce 45° diamond internal reflective element (IRE) was used for recording attenuated total reflectance (ATR) IR absorption measurements. A small aliquot of the powdered opioid sample was placed on the IRE and the spectra were captured across the 650–4000 cm^{-1} range, with a spectral resolution of 4 cm^{-1} .

Mass spectrometry measurements. All samples in this study were analyzed using a TSQ Fortis™ triple quadrupole mass spectrometer equipped with a VeriSpray™ Paper Spray ion source (Thermo Fisher Scientific, San Jose, USA). Details on the methodology, calibration, and data processing have been previously described.^{69–72} This procedure has determined the xylazine detection limit to be 0.01% w/w. In brief, 0.5–2.1 mg of the powdered opioid sample was dissolved in 1.2 mL of methanol and vortexed, creating a solution with a concentration of approximately 1 mg mL^{-1} . A 1 μL aliquot of this solution was diluted in an internal standard cocktail to achieve a final concentration of 6 $\mu\text{g mL}^{-1}$. For mass spectrometry analysis, 10 μL of this solution was applied on a PS-MS sample strip in a VeriSpray™ sample plate to determine sample composition.

2.3 Data processing and chemometric modeling

Spectral data for samples in the training ($n = 168$) and testing ($n = 50$) sets were labeled with their corresponding sample composition information and organized into data frames using the pandas Python package.^{73,74} All statistical analyses, spectral preprocessing, and modeling was conducted in Python using the scikit-learn package.⁷⁵ These included principal component analysis (PCA), minimum covariance determinant (MCD) method, and Mahalanobis distances for outlier

detection (Fig. S1†) described in the ESI.† The derivative and normalization techniques implemented on spectral data were mean centring followed by unit variance scaling, area normalization, min–max normalization, standard normal variate (SNV), first-order derivative, and second-order derivative. Both derivatives were implemented with Savitzky–Golay smoothing (window size 5, polynomial order 2). Fig. 1 provides a representation of the mean 0th, 1st, and 2nd derivative spectra of the training set ($n = 168$) for the concatenated SERS–IR data used in this study. Additionally, classification models were developed and optimized using PCA, random forest (RF), support vector machine (SVM), and k -nearest neighbor algorithm (KNN). All models were constructed using a 5-fold cross-validation grid search on the training set and optimal hyperparameters were selected based on F1 score. Documentation of all classifier-specific parameters evaluated (Table S2†) and additional details on model construction, hyperparameter tuning, and spectral preprocessing for training and testing data are provided in the ESI.†

2.4 Data fusion strategies

This study explores three data fusion strategies: a hybrid, mid-level, and high-level approach; an overview of the procedure is presented in Fig. 2.

Hybrid data fusion. In the hybrid SERS–IR data fusion technique (Fig. 2a), all SERS and IR spectra were preprocessed using global max normalization, setting the maximum intensity value across all datasets to 1. The normalized SERS and IR spectra were concatenated, with the IR spectral values appended to the tail end of the SERS data (Fig. 1a). The fused SERS–IR dataset for training ($n = 168$) and testing ($n = 50$) comprised of 5601 variables, where the SERS and IR spectra contributed 1601 and 4000 variables, respectively. High variance samples were removed from the training set using an outlier detection method (Fig. S1†) and the final training set ($n = 154$) was used to develop the RF, SVM, and KNN models. Hyperparameter tuning of the model parameters was evaluated against individual spectral preprocessing (normalization and order of derivatives) of the SERS and IR regions in the fused dataset separately (Tables S3–S5†). The optimal combination for RF, SVM, and KNN hybrid models is detailed in the ESI† and the model performance (F1 score) on the training set is shown in Table 1.

Mid-level data fusion. For the mid-level data fusion approach (Fig. 2b), SERS and IR spectral data were each transformed into 0th, 1st, and 2nd order derivatives. These datasets were standardized about their mean and scaled to unit variance prior to PCA. The optimal order of derivative for the extracted features was determined based on the maximum F1 score achieved during hyperparameter tuning for each classifier in the training phase (Tables S6–S8†). The first-order derivative was used for RF and SVM models while no derivatives were applied for the KNN model. The Principal components (PCs) making up 95% of the cumulative variance were extracted as features for the individual SERS and IR datasets. Fig. 3 illustrates the number of principal components deter-



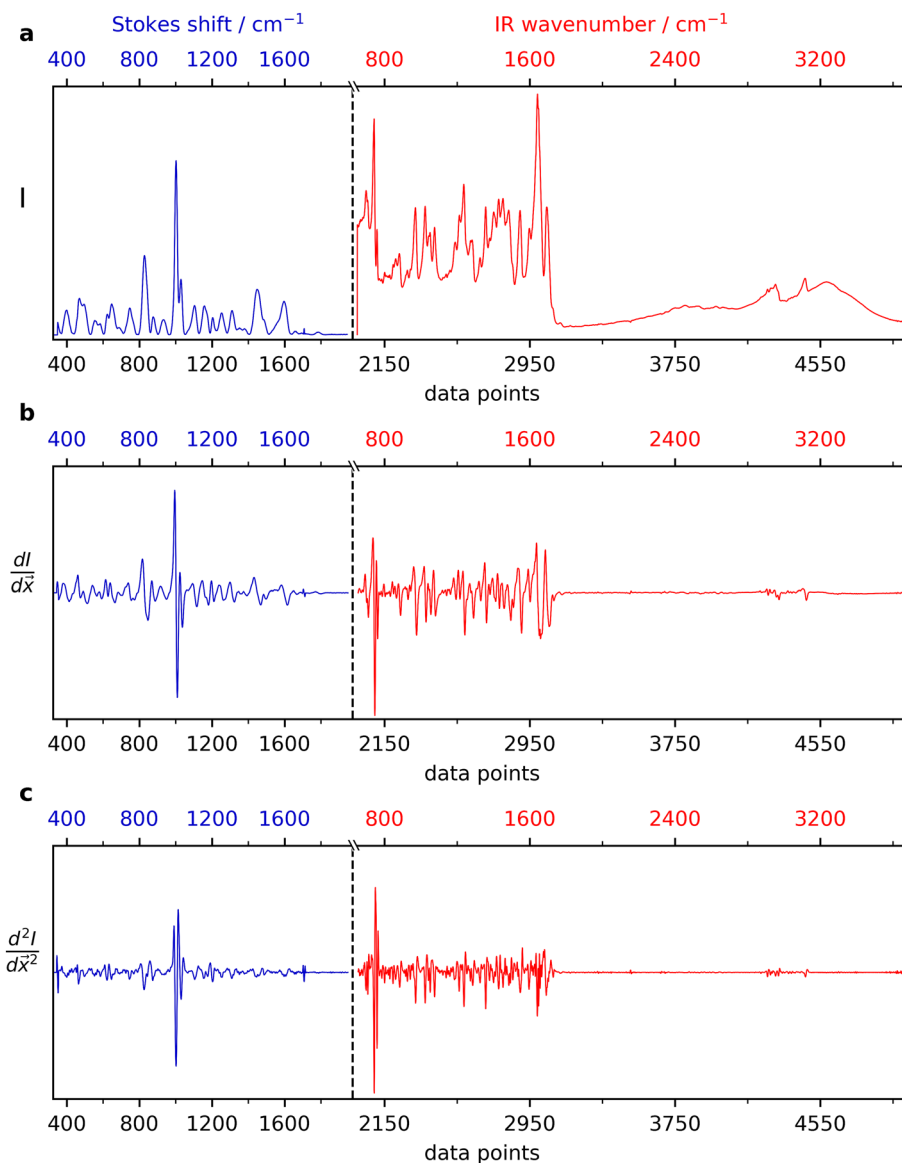


Fig. 1 Mean spectra of samples in the training set ($n = 168$) for concatenated SERS-IR fused data in the (a) 0th, (b) 1st, and (c) 2nd derivative. SERS spectra in the Stokes region of $300\text{--}2000\text{ cm}^{-1}$ is represented in blue and IR spectra in the wavenumber region of $650\text{--}4000\text{ cm}^{-1}$ is represented in red.

mined and the cumulative explained variance of the SERS (Fig. 3a) and IR (Fig. 3b) datasets for their respective order of derivatives. The extracted features were then fused together by stacking the IR features to the tail end of the SERS features in a single dataset for classification. The fused SERS-IR extracted features in the mid-level training set ($n = 168$) for the 0th derivative set included 54 PCs (46 PCs for SERS and 8 PCs for IR) and the 1st derivative set included 87 PCs (50 PCs for SERS and 37 PCs for IR) which were used to develop the RF, SVM, and KNN models, optimizing respective model hyperparameters for the PCA data (Tables S6–S8†). The optimal parameters and order of derivatives for RF, SVM, and KNN mid-level models is detailed in the ESI† and the model performance (F1 score) on the training set is shown in Table 1.

High-level data fusion. The high-level data fusion approach (Fig. 2c) combines predictions from two independently trained classification models, one built on SERS spectral data and the other on IR spectral data. High variance samples in the training set ($n = 168$) were identified for SERS ($n = 17$) and IR ($n = 15$) data using an outlier detection method (Fig. S1†) and subsequently removed, resulting in SERS ($n = 151$), IR ($n = 153$), and combined high level SERS-IR ($n = 141$) training libraries. RF, SVM, and KNN models were optimized for both SERS and IR datasets through a 5-fold cross-validation grid search, tuning model parameters and testing various spectral preprocessing techniques (Tables S9–S11†). A weighted voting classifier (sklearn.ensemble) was employed to combine predictions from the optimized standalone SERS and IR models. Weights



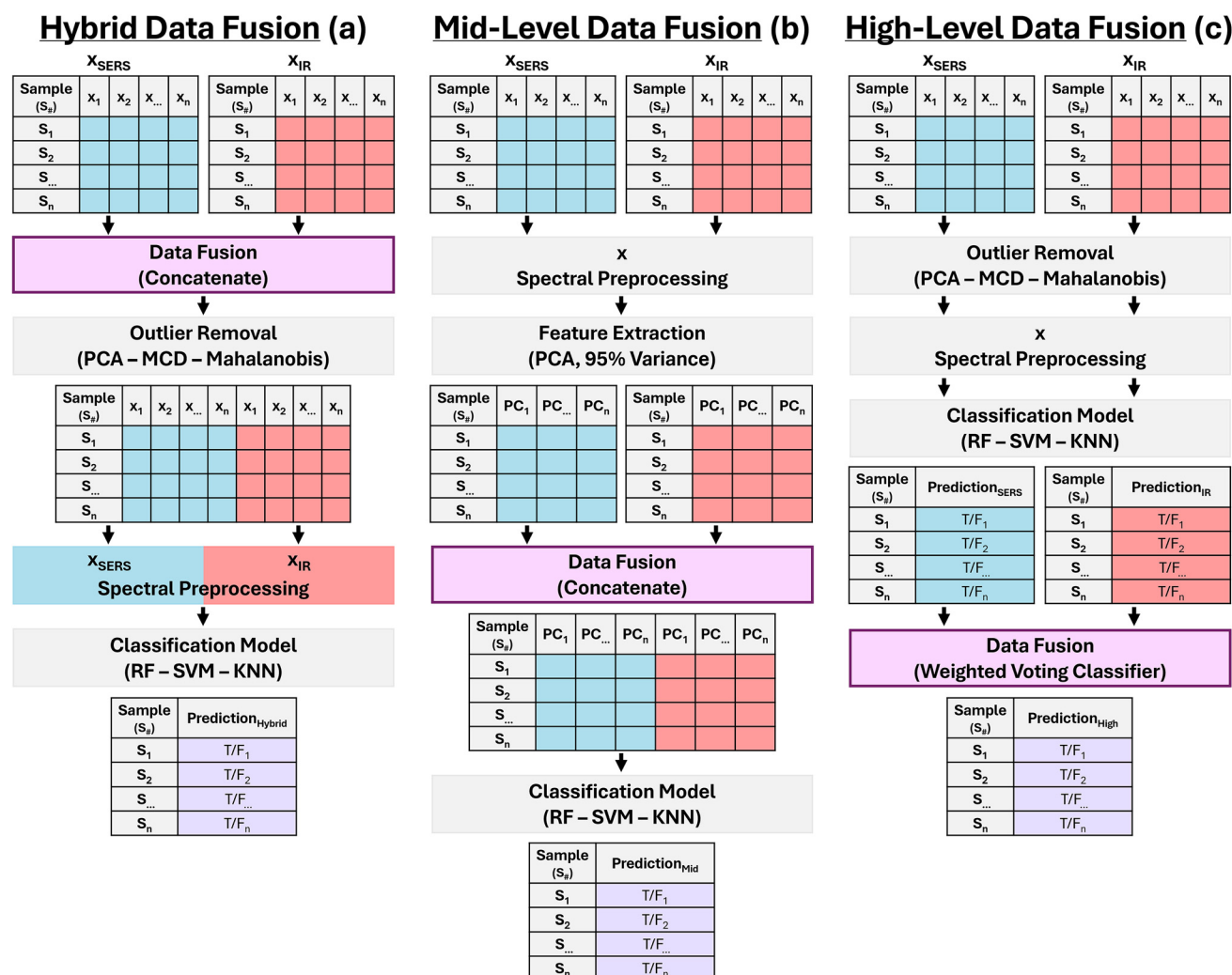


Fig. 2 Flowchart of data fusion strategies for (a) hybrid, (b) mid-level, and (c) high-level fusion approaches. The presentation style has been adapted from ref. 63.

Table 1 F1 scores for all optimized models built with SERS, IR, PCA SERS–IR and SERS–IR training data for the detection of xylazine. Optimal weights selected for SERS and IR standalone model contribution and their respective F1 scores are detailed for the construction of high-level classification models

Data fusion	Model	Training data	F1 score	Voting classifier	Weights [SERS, IR]	Ensemble F1 score
Hybrid	RF	SERS–IR	0.743	—	—	—
	SVM	SERS–IR	0.881	—	—	—
	KNN	SERS–IR	0.748	—	—	—
Mid-level	RF	PCA SERS–IR	0.616	—	—	—
	SVM	PCA SERS–IR	0.805	—	—	—
	KNN	PCA SERS–IR	0.614	—	—	—
High-level	RF	SERS	0.805	Weighted voting	[0.9, 0.1]	0.754
		IR	0.532			
	SVM	SERS	0.869	Weighted voting	[0.9, 0.1]	0.795
		IR	0.790			
	KNN	SERS	0.690	Weighted voting	[0.6, 0.4]	0.637
		IR	0.618			

summing to 1 were assigned to the predictions of the SERS and IR classification models and evaluated using a 5-fold cross-validation grid search on the high-level training set ($n =$

141). The optimal weight combination, determined based on the maximum F1 score (Tables S12–S14[†]), was applied to the predicted probability scores of the respective standalone



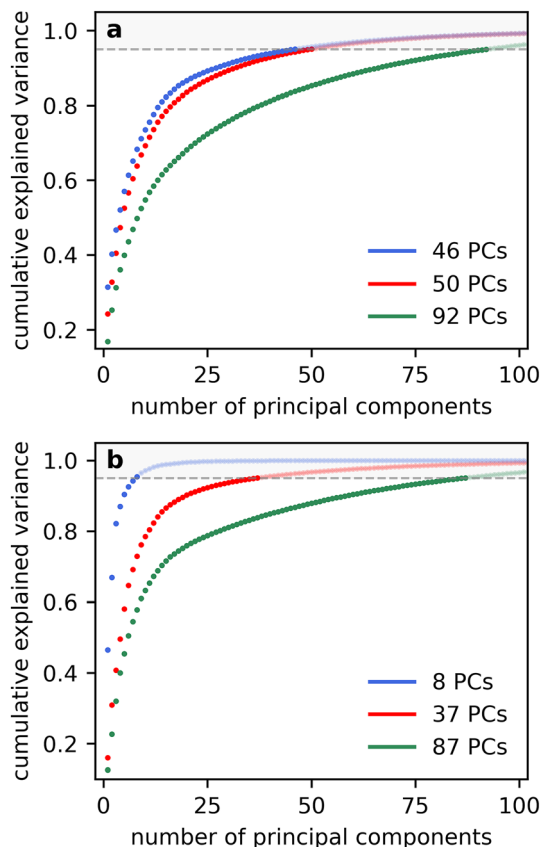


Fig. 3 The cumulative explained variance of principal components for (a) SERS and (b) IR spectral data with no derivatives (blue), first derivative (red), and second derivative (green). The number of principal components required to capture 95% of the variance is indicated for each case, with the cutoff point highlighted in grey.

models to generate the final high-level fusion predictions. Detailed parameters of the standalone models are provided in the ESI[†] while the weights of all standalone models used in the high level fusion approach, along with their training set performance (F1 score), are listed in Table 1.

3 Results and discussion

3.1 Hyperparameter optimization

SERS and IR standalone models. The initial performance of the standalone models is evaluated by examining the optimized parameters for the three classifiers built using SERS and IR spectral data (Tables S9–S11[†]).

Random forest models. The standalone RF models exhibited varied predictive performance, with F1 scores of 0.805 for SERS and 0.532 for IR during cross-validation (Table S9[†]). These results, along with the selected optimal hyperparameters, demonstrate that SERS has a stronger affinity for detecting xylazine.⁴³ The SERS RF model shows good initial performance, benefiting from a relatively high number of

decision trees and consistent normalization techniques that effectively capture complex patterns in the data. In contrast, the IR RF model, while utilizing deeper trees (maximum depth of 30), had reduced forest diversity due to the limited number of trees (30), leading to fewer averaged predictions and an increased likelihood of overfitting to the training data. Despite the increased tree depth, the poor initial performance, with an F1 score only marginally better than a baseline classifier, suggests that IR spectral data is not well-suited to capturing xylazine features within the sample matrix.

Support vector machine models. SVM models built with SERS and IR data achieved the highest individual precision and recall scores during cross-validation (Table S10[†]), with F1 scores of 0.869 for SERS and 0.790 for IR. Early performance metrics on the training set suggest both spectral devices seem compatible with SVM classification. The optimized SVM parameters for both SERS and IR data indicate a strong focus on capturing complex, non-linear relationships in the data. The use of polynomial kernels with different degrees (2 for SERS and 3 for IR) allows the models to capture interactions between features, while ‘balanced’ class weights help address class imbalance. High values of C suggest that the models are less regularized, aiming to fit the training data closely. Overall, these parameters highlight the tailored approach taken to optimize the SVM models for the specific characteristics of SERS and IR spectral data, resulting in high initial predictive performance as indicated by their respective F1 scores.

k-Nearest neighbor models. KNN models for SERS (F1 score 0.690) and IR (F1 score 0.618) showed no strong preference for either platform in detecting xylazine during cross-validation (Table S11[†]). The individual KNN models both demonstrated lower precision and recall scores compared to their hybrid counterpart (see below). The use of a single neighbour and Manhattan distance by the SERS model emphasizes its immediate data points, while the IR model applies the second derivative and Euclidean distance to capture subtle features. However, both utilize simple approaches with a single neighbour which may lead to poor generalization on unseen data, limiting performance on the validation set.

Hybrid fusion models. F1 scores of 0.743, 0.881, and 0.748 were obtained from the optimized RF, SVM, and KNN classifiers (Table 1), respectively, for predicting xylazine using concatenated SERS–IR spectral data during cross-validation on the training set. All hybrid models (Tables S3–S5[†]) demonstrated relatively high predictive performance, with the SVM model achieving the highest precision and recall (F1 score of 0.881) within the SERS–IR training set.

The optimal SVM hyperparameters (Table S4[†]) included the ‘rbf’ kernel for handling non-linear classification and the ‘scale’ kernel coefficient, which adjusts gamma based on the number of features, thereby controlling the influence of each data point according to the hybrid datasets complexity and variability. The selected preprocessing methods suggest that applying the second derivative to the IR data enhances relevant features, while the raw SERS data, normalized by area, is sufficiently informative.



For the hybrid RF model (Table S3†), the optimal spectral preprocessing techniques were the same for both SERS and IR regions, with snv normalization and no derivative, indicating a consistent approach to reducing intensity variations across the entire dataset. Similarly, the hybrid KNN model (Table S5†) selected area normalization, ensuring consistent total signal intensity across all samples. This uniformity in spectral preprocessing for the RF and KNN models suggests that the concatenated SERS–IR spectral data share underlying characteristics best captured when treated as a whole.

The selection of 50 decision trees for the hybrid RF model, indicates a balanced approach to model complexity and computational efficiency. The use of 50 trees allows the model to capture more diverse patterns related to xylazine while mitigating the risk of overfitting to the training data. In contrast, the hybrid KNN model employed a simple approach by using a single neighbour for predictions, relying on the most immediate data point's characteristics, indicating a potential for high bias to the training set.

Mid-level fusion models. The mid-level models initial predictive performance, presented in Table 1, showed that SVM had the highest F1 score (0.805), followed by RF (0.616) and KNN (0.614). SERS accounted for a greater portion of the explained variance compared to IR, likely due to its enhanced sensitivity in capturing spectral variation from low-concentration adulterants. As a result, all mid-level models incorporated more features from SERS in their training metrics for xylazine detection (Fig. 3). Consequently, hyperparameter tuning (Tables S6–S8†) was primarily focused on enhancing the predictive power of SERS data. The superior performance of SVM in the training phase is likely due to the reduced dimensionality, as the extracted features capture critical information that benefits supervised max-margin models like SVM, which can otherwise overfit in high-dimensional spaces. The chosen parameters for the SVM model (Table S7†), particularly a C value of 1, allows for a wider margin around the decision boundary, permitting some misclassifications to enhance the ability of the model to generalize to unseen data. Both RF (Table S6†) and SVM (Table S7†) favoured first-order derivative PCA (87 PCs), enhancing peak resolution and capturing subtle features related to xylazine, while KNN (Table S8†) selected the 0th derivative, opting for a less computationally intensive set of variables (54 PCs).

High-level fusion models. The contributions of the standalone models for xylazine detection and the corresponding weights applied in the fusion models (Tables S12–S14†) were used to evaluate the initial performance of the high-level data fusion approach.

Weighted voting ensemble. Despite varying F1 scores, all standalone classification models consistently showed a preference for SERS data, demonstrating its robustness in detecting xylazine compared to IR (Tables S9–S11†). This finding aligns with previous research, which highlights the superior performance of SERS-based classifiers for identifying xylazine in complex opioid matrices, while spectral overlap poses a significant challenge for IR.^{43,66}

High-level fusion of SERS and IR models evaluated F1 scores of all weight combinations summing to 1 during cross-validation on the training set (Tables S12–S14†). The fused RF, SVM, and KNN models, with optimal weights ([SERS, IR]), achieved F1 scores of 0.754, 0.795, and 0.637, respectively. These values were consistent with the initial predictive performance results for each spectral model. Optimal weights prioritized SERS information in all the final high-level fusion models, with RF and SVM high-level models weighting their SERS model contributions at 90% and KNN at 60%. The optimal weights were then applied to the predicted probability scores of the standalone SERS model and IR model on the test set and combined for all high-level models.

3.2 Performance evaluation

While the cross-validated F1 scores are a good metric to determine initial performance on the training set folds, the results of the fused models on unseen data is crucial to assessing their performance. The performance of all models is evaluated on the test set ($n = 50$) to identify the most effective data fusion strategy and classification algorithm for the detection of xylazine in complex opioid samples. A receiver operating characteristic (ROC) curve was developed using the predicted probability scores of the hybrid, mid-level and high-level models on the validation set. Area under the curve (AUC) values were calculated and used to evaluate model accuracy across the range of threshold values. ROC curves are plotted for all fused models using varying cut-off thresholds and are illustrated in Fig. 4. Optimal thresholds were determined for final prediction results of all models.

AUC values of 0.93, 0.85, and 0.68 were calculated for the hybrid RF, SVM, and KNN models, respectively (Fig. 4a). Both the RF and SVM hybrid models reflect high discriminatory power across the range of thresholds. The hybrid KNN model demonstrated a weaker performance across the range of predicted probabilities, likely due to the selected number of neighbours, where $k = 1$ results in predictions that are highly sensitive to the nearest neighbour (Table S5†). In scenarios where the nearest neighbour clearly belong to one class (0 or 1), the predicted probability of a tested sample will reflect the class and the scores for all tested samples will be either 0 or 1. Consequently, setting a threshold becomes arbitrary and no improvements can be made by changing it. Therefore, despite relatively high initial performance within the training group, the hybrid KNN model does not seem well suited to generalize to unseen data or to assign meaningful probabilities.

The mid-level AUC values of 0.95 and 0.91 for RF and SVM classifiers (Fig. 4b), respectively, represent some of the highest performing models across the range of predicted probabilities for the detection of xylazine. The mid-level KNN model however, demonstrated similar complications as its hybrid counterpart, with an AUC of 0.66, it performed the worst of all fused models on the test set.

ROC analysis, threshold determination, and final evaluation of the standalone models were conducted to compare the performance of the isolated spectral models against the



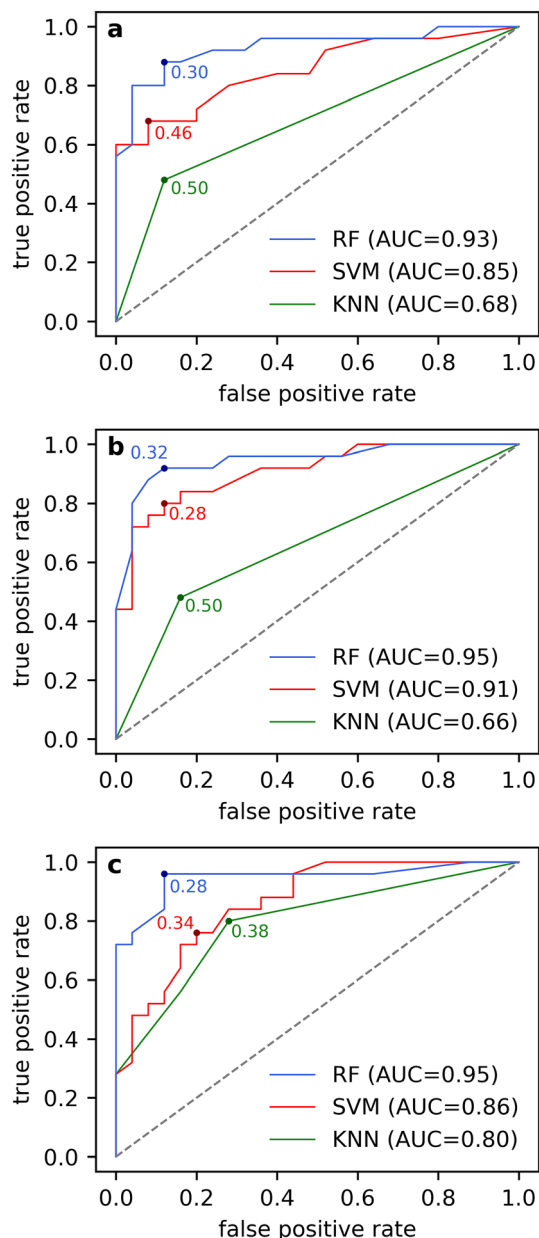


Fig. 4 A comparison between the RF model, SVM model and KNN model prediction results of xylazine on the test set ($n = 50$) for the (a) hybrid, (b) mid-level, and (c) high-level data fusion strategies. The ROC curves illustrate the true positive rate and false positive rate across various classification thresholds for xylazine detection. Optimal threshold selection for balancing sensitivity and specificity, and area under the curve (AUC) is highlighted for all models. A grey dashed line illustrates the performance of a baseline classifier.

high-level fusion approach. While the standalone models for both spectral techniques achieved relatively high AUC values (Table S15†), with the highest being the SERS RF model (AUC = 0.95), combining the weighted predictions of the standalone models in the high-level fusion approach showed increased model accuracy across the range of probability thresholds. The AUC values for the high-level RF (0.95), SVM (0.86), and KNN

(0.80) models (Fig. 4c) demonstrate improved performance across the predicted probability range compared to their hybrid counterparts. Notably, the high-level KNN model showed a significant enhancement in prediction accuracy when integrating the high-level outputs from the individual SERS and IR models.

Optimal probability thresholds for all models were determined using Youden's J statistic^{76,77} and applied to the test set. Xylazine prediction results for the hybrid, mid-level, and high-level models (Table 2) were used to evaluate the performance of the three fusion strategies and classifiers.

Of the hybrid models (Table 2), Random forest correctly identified 88% of both xylazine-positive samples and xylazine-negative samples. Showing a balanced sensitivity and specificity towards detecting xylazine. Both SVM and KNN hybrid models demonstrated a higher affinity towards identifying negative classes than they did for correctly assigning positive classes. Initial performance in the training group for SVM showed the highest precision and recall scores, but the performance on the validation set is indicative of overfitting of the hybrid SVM model to the training data.

The hybrid KNN model performs poorly in correctly assigning positive classes, likely due to an imbalance in the training dataset, where the nearest neighbour is often from the majority class (xylazine-negative). This leads to underprediction of the minority class (xylazine-positive), as the single nearest neighbour may not accurately reflect the true class distribution. KNN's effectiveness also decreases with high-dimensional data, where the distance between points becomes less informative.⁵⁶ Concatenating SERS and IR data significantly increases dimensionality, making the nearest neighbour less representative of the true class. Additionally, the computational intensity required for large data sets⁷⁸ further highlights the unsuitability of this model development strategy.

The hybrid data fusion method appears to best suit the RF binary classifier, leveraging spectral information from both platforms to achieve high accuracy (88%) and precision (88%) compared to other supervised techniques (Table 3).

Among the mid-level models (Table 2), the KNN model performs poorly, only identifying 48% of true positive samples. However, the SVM model demonstrates the highest balanced prediction success using the mid-level approach, correctly identifying 80% of positive samples and 88% of negative samples. Similarly the mid-level RF model performed well, demonstrating a higher affinity towards the positive class (92% TP) over the negative class (88% TN). The mid-level fusion approach demonstrates the highest accuracy (90%) and precision (88%) when implementing RF classification on the test set (Table 3).

Among the high-level models (Table 2), both RF and KNN fusion models outperformed their standalone counterparts when predicting xylazine on the test set (Table S16†). KNN, with the most balanced SERS weighting ratio of 0.6:0.4, showed a significant increase in identifying true positives (80% TP), at a slight expense of true negatives (72% TN), establishing high-level fusion as the optimal approach for KNN



Table 2 Xylazine prediction results of the test set ($n = 50$) for the RF, SVM, and KNN models developed using the hybrid, mid-level, and high-level SERS–IR data fusion methods. Fractional values indicate the ratio between model prediction and the number of samples in the true class ($n = 25$)

	Hybrid			Mid-level			High-level		
	RF	SVM	KNN	RF	SVM	KNN	RF	SVM	KNN
True positive	0.88 (22)	0.68 (17)	0.48 (12)	0.92 (23)	0.80 (20)	0.48 (12)	0.96 (24)	0.76 (19)	0.80 (20)
True negative	0.88 (22)	0.92 (23)	0.88 (22)	0.88 (22)	0.88 (22)	0.84 (21)	0.88 (22)	0.80 (20)	0.72 (18)
False positive	0.12 (3)	0.08 (2)	0.12 (3)	0.12 (3)	0.12 (3)	0.16 (4)	0.12 (3)	0.20 (5)	0.28 (7)
False negative	0.12 (3)	0.32 (8)	0.52 (13)	0.08 (2)	0.20 (5)	0.52 (13)	0.04 (1)	0.24 (6)	0.20 (5)

Table 3 Performance metrics summary for RF, SVM, and KNN models using all SERS–IR data fusion strategies for xylazine detection. Summary of the area under the curve (AUC), optimal threshold, accuracy, precision, sensitivity, specificity, and F1 score for opioid samples ($n = 50$) tested with RF, SVM, and KNN models

Model	AUC	Optimal threshold	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1 score (%)
Hybrid							
RF	0.93	0.30	88	88	88	88	88
SVM	0.85	0.46	80	89	68	92	77
KNN	0.68	0.50	68	80	48	88	60
Mid-level							
RF	0.95	0.32	90	88	92	88	90
SVM	0.91	0.28	84	87	80	80	83
KNN	0.66	0.50	66	75	48	84	59
High-level							
RF	0.95	0.28	92	89	96	88	92
SVM	0.86	0.34	78	79	76	80	78
KNN	0.80	0.38	76	74	80	72	77

classification over standalone, hybrid fusion, and mid-level fusion models. Both RF and SVM models prioritized SERS predictions (0.9 weighting) but benefited from increased sensitivity on the test set by incorporating IR. The SVM fusion model achieved an accuracy and F1 score of 78% (Table 3), outperforming the standalone SERS model in identifying positive classes and the standalone IR model in identifying negative classes (Table S16†).

The high-level RF model maintained strong true negative performance (88% TN) similar to the SERS model, while correctly identifying one additional true positive sample (3.17% w/w) through IR, underscoring its robustness and demonstrating effective integration of SERS and IR predictions. The fusion of high-level RF and KNN classifier outputs enhanced overall accuracy and F1 scores compared to the standalone spectral models, illustrating the compatibility of combining two techniques in fused model development. Both RF and KNN high-level models show a more meaningful balance of correctly assigned predictions, achieving the best true positive performance over their hybrid and mid-level fusion counterparts. The RF high-level model, despite misidentifying one trace (0.79% w/w) xylazine sample (Table S17†), delivered the best overall performance, correctly classifying 24 true positives and 22 true negatives, with 96% sensitivity and 88% specificity (Table 3).

The performance of RF, SVM, and KNN models varied across different data fusion strategies. The mid-level SVM achieved the highest F1 score (83%) compared to its hybrid and high-level counterparts, suggesting that margin maximiza-

tion was best suited for lower-dimensional data. In contrast, KNN consistently exhibited weaker precision and recall scores across all fusion approaches, likely due to its sensitivity to high-dimensional spaces and the increased complexity and computational demands that come with combining two spectral datasets.⁷⁸ However, the highest performing models were those using random forest classification, which showed a well-balanced trade-off between sensitivity and specificity across all fusion strategies for detecting xylazine, with robust AUC values. RF models achieved high precision and recall, with the hybrid (F1 score = 88%), mid-level (F1 score = 89%), and high-level (F1 score = 92%) models outperforming all other classifiers developed (Table 3).

RF binary classification demonstrated high predictive success for all data fusion strategies, effectively prioritizing SERS contributions for xylazine detection. In the context of drug checking, where the consequences of misclassifying a psychoactive sedative can be severe, prioritizing sensitivity is crucial. As such, the superior sensitivity (96%) and F1 score (92%) of the high-level RF model indicate that it is the most effective strategy for detecting xylazine in complex opioid samples.

3.3 Implications for point-of-care drug checking

Predictive models utilizing both SERS and IR spectroscopy in a high-level data fusion approach offer a promising solution for automated differentiation of compounds in complex, real-world opioid samples. The high sensitivity of SERS enables detection of low concentration adulterants, while incorporat-



Table 4 Performance metrics summary for SERS, IR, and high-level SERS–IR data fusion RF models for compound identification. Summary of the area under the curve (AUC), sensitivity (abbreviated as Sen.), specificity (Spe.), and F1 score for opioid samples ($n = 50$) tested with RF models. Standalone model weight contributions are included for all high-level fused RF models

Analytes	SERS				IR				SERS–IR				
	AUC	Sen. (%)	Spe. (%)	F1 (%)	AUC	Sen. (%)	Spe. (%)	F1 (%)	Weights [SERS, IR]	AUC	Sen. (%)	Spe. (%)	F1 (%)
Xylazine	0.95	92	88	90	0.82	92	56	78	[0.9, 0.1]	0.95	96	88	92
Bromazepam	0.90	84	88	86	0.83	68	92	77	[0.7, 0.3]	0.95	92	88	90
Fluorofentanyl	0.90	79	88	85	0.90	85	82	88	[0.3, 0.7]	0.92	88	82	89
Caffeine	0.43	96	20	93	1.00	100	100	100	[0.0, 1.0]	1.00	100	100	100
Erythritol	0.65	76	52	68	0.87	72	92	80	[0.5, 0.5]	0.86	80	84	82

ing IR data in a weighted fusion approach allowed for the identification of an additional xylazine-positive sample, outperforming the results of the standalone SERS RF model.

While high-level data fusion cannot fully overcome the inherent spectral limitations of IR with overlapping and low-concentration detection of xylazine in opioid mixtures,^{43,66} integrating weighted predictions of IR with SERS improved overall model sensitivity. This multi-instrumental approach leverages the strength of SERS in xylazine detection while preserving valuable compositional information from IR, creating a more comprehensive and robust model.

Reliance on a single technique often fails to meet the diverse demands of community-based drug checking, where usability, cost-effectiveness, and thorough mixture analysis are paramount. SERS excels in trace detection but faces challenges with reproducibility and bulk analysis, while IR offers reliable detection of major components but lacks sensitivity for low-concentration compounds. The high-level RF fusion model bridges these gaps, combining the sensitivity of SERS with the broader compositional analysis of IR and outperforming other machine learning models, such as SVM and KNN. Additionally, RF classification combined with high-level data fusion has the added benefit of allowing for feature importance analysis to highlight key spectral contributions, supporting intuitive data interpretation for drug checking sites and reducing the ‘black-box’ effect of many classification models.^{43,79}

As shown in Table 4, RF classification models developed using the high-level fusion approach detailed in this study effectively prioritize contributions from the higher-performing spectral-based model when differentiating several common illicit components within the test set. All RF models developed with the high-level fusion approach for xylazine, bromazepam, fluorofentanyl, caffeine, and erythritol show improved F1 scores and sensitivity over standalone models. In particular, IR aids in identifying sugars and high concentration actives, with the model leveraging IR contributions for caffeine and fluorofentanyl, while heavily relying on SERS predictions to detect low concentration adulterants like xylazine and bromazepam.

This fusion approach coupled with RF classification using SERS and IR data demonstrates a viable strategy to expanding the range of identifiable components in complex opioid

samples using field-portable technologies. By automating results, it eliminates the need for extensive spectral interpretation and thereby simplifies decision-making for both service users and providers, creating an accessible and comprehensive tool for community drug checking.

4 Conclusions

We have demonstrated the trace detection of xylazine in complex illicit opioid samples using random forest classification trained on three data fusion strategies incorporating SERS and IR spectral data. In all strategies, the performance of the random forest models surpassed that of the SVM and KNN models. Given the known interferences and challenges of identifying xylazine using IR alone, implementing a high-level data fusion technique with random forest model development maintained specificity while improving sensitivity utilizing both platforms. This approach effectively prioritized the sensitivity of SERS for trace xylazine detection while preserving valuable predictive information from IR, without compromising performance. The integration of multiple analytical techniques, supported by advanced data fusion and machine learning approaches, represents a promising strategy for advancing drug checking. This holistic approach ensures that rapid, affordable, and portable spectroscopic devices are utilized to their full potential, ultimately leading to more reliable and comprehensive compound identification in various settings.

Author contributions

Conceptualization, RRM; data collection, RRM, EN; methodology, RRM, LG, CG; analysis and interpretation, RRM; writing, RRM, DKH; funding acquisition, BW, DKH.

Data availability

Data will be made available from the corresponding author upon reasonable request.



Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This project was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) and the Canadian Institutes of Health Research (CIHR) in the form of a Collaborative Health Research Project (CHRP-549668-2020) and a New Frontiers in Research Fund (NFRFE-2022-00886). Additional support and contributions were made by the Vancouver Foundation (F0120-5607), BC Ministry of Mental Health & Addictions, and Vancouver Island Health Authority. High performance computing support and server resources were provided by the University of Victoria and the Digital Research Alliance of Canada. Allie Miskulin assisted with the acquisition of the PS-MS data. Daniel Barker-Rothschild provided valuable discussion on the topic of normalization strategies for extracted features.

References

- 1 BC Coroners Service, *Illicit Drug Toxicity Deaths in BC*, Ministry of public safety & solicitor general technical report, 2023.
- 2 J. P. Caulkins, A. Tallaksen, J. Taylor, B. Kilmer and P. Reuter, *Int. J. Drug Policy*, 2024, **124**, 104314.
- 3 P. Reuter, B. Pardo and J. Taylor, *Int. J. Drug Policy*, 2021, **94**, 103086.
- 4 L. Belzak and J. Halverson, *Health Promot. Chronic Dis. Prev. Can.*, 2018, **38**, 224–233.
- 5 S. Borden, S. R. Mercer, A. Saatchi, E. Wong, C. M. Stefan, H. Wiebe, D. K. Hore, B. Wallace and C. G. Gill, *Drug Test. Anal.*, 2023, **15**, 484–494.
- 6 A. Larnder, A. Saatchi, S. A. Borden, B. Moa, C. G. Gill, B. Wallace and D. Hore, *Drug Alcohol Depend.*, 2022, **235**, 109427.
- 7 I. Bhuiyan, S. Tobias and L. Ti, *Am. J. Drug Alcohol Abuse*, 2023, **49**, 685–690.
- 8 M. Beaulac, L. Richardson, S. Tobias, M. Lysyshyn, C. Grant and L. Ti, *Int. J. Drug Policy*, 2022, **105**, 103707.
- 9 S. Tobias, A. M. Shapiro, H. Wu and L. Ti, *Can. J. Addiction*, 2020, **11**, 28–32.
- 10 J. M. Bowles, K. McDonald, N. Maghsoudi, H. Thompson, C. Stefan, D. R. Beriault, S. Delaney, E. Wong and D. Werb, *Harm Reduct. J.*, 2021, **18**, 104.
- 11 R. A. Torruella, *Subst. Abuse Treat. Prev. Policy*, 2011, **6**, 7.
- 12 S. L. Kacinko, A. L. A. Mohr, B. K. Logan and E. J. Barbieri, *J. Anal. Toxicol.*, 2022, **46**, 911–917.
- 13 J. O'Neil and S. Kovach, *N. Engl. J. Med.*, 2023, **388**, 2274.
- 14 K. L. Sue and K. Hawk, *Addiction*, 2024, **119**, 606–608.
- 15 J. Perrone, R. Haroz, J. D'Orazio, G. Gianotti, J. Love, M. Salzman, M. Lowenstein, A. Thakrar, S. Klipp, L. Rae, M. K. Reed, E. Sisco, R. Wightman and L. S. Nelson, *Ann. Emerg. Med.*, 2024, **84**, 20–28.
- 16 B. Wallace, T. van Roode, P. Burek, B. Pauly and D. Hore, *Drugs: Educ. Prev. Policy*, 2023, **30**, 443–452.
- 17 L. Gozdziński, B. Wallace and D. Hore, *Harm Reduct. J.*, 2023, **20**, 39.
- 18 B. Wallace, R. Hills, J. Rothwell, D. Kumar, I. Garber, T. van Roode, A. Larnder, F. Pagan, J. Aasen, J. Weatherston, L. Gozdziński, M. Ramsay, P. Burek, M. S. Azam, B. Pauly, M.-A. Storey and D. K. Hore, *Drug Test. Anal.*, 2021, **13**, 734–746.
- 19 L. Gozdziński, R. Louw, C. Kielty, A. Margolese, E. Poarch, M. Sherman, F. Cameron, C. Gill, B. Wallace and D. Hore, *Harm Reduct. J.*, 2024, **21**, 63.
- 20 L. Gozdziński, A. Rowley, S. Borden, A. Saatchi, C. G. Gill, B. Wallace and D. K. Hore, *Int. J. Drug Policy*, 2022, **102**, 103611.
- 21 N. P. Weicker, J. Owczarzak, G. Urquhart, J. N. Park, S. Rouhani, R. Ling, M. Morris and S. G. Sherman, *Int. J. Drug Policy*, 2020, **84**, 102900.
- 22 M. K. Laing, L. Ti, A. Marmel, S. Tobias, A. M. Shapiro, R. Laing, M. Lysyshyn and M. E. Socias, *Int. J. Drug Policy*, 2021, **1**, 103169.
- 23 C. S. Johnson, C. R. Stansfield and V. R. Hassan, *Forensic Sci. Int.*, 2020, **313**, 110367.
- 24 T. Vankeirsbilck, A. Vercauteren, W. Baeyens, G. V. der Weken, F. Verpoort, G. Vergote and J. P. Remon, *Trends Anal. Chem.*, 2002, **21**, 869–877.
- 25 A. Lanzarotta, M. Witkowski and J. Batson, *J. Forensic Sci.*, 2020, **65**, 421–427.
- 26 R. Gonzales, K. Titier, V. Latour, A. Peyre, N. Castaing, A. Daveluy and M. Molimard, *Int. J. Drug Policy*, 2021, **88**, 103037.
- 27 L. Gozdziński, A. Hutchison, B. Wallace, C. Gill and D. Hore, *Drug Test. Anal.*, 2024, **16**, 83–92.
- 28 S. Mabbott, O. Alharbi, K. Groves and R. Goodacre, *Analyst*, 2015, **140**, 4399–4406.
- 29 D. S. Burr, W. L. Fatigante, J. A. Lartey, W. Jang, A. R. Stelmack, N. W. McClurg, J. M. Standard, J. R. Wieland, J.-H. Kim, C. C. Mulligan and J. D. Driskell, *Anal. Chem.*, 2020, **92**, 6676–6683.
- 30 A. N. Masterson, S. Hati, G. Ren, T. Liyanage, N. E. Manicke, J. V. Goodpaster and R. Sardar, *Anal. Chem.*, 2021, **93**, 2578–2588.
- 31 X. Sha, G. Fang, G. Cao, S. Li, W. Hasi and S. Han, *Analyst*, 2022, **147**, 5785–5795.
- 32 H. Segawa, T. Fukuoka, T. Itoh, Y. Imai, Y. T. Iwata, T. Yamamuro, K. Kuwayama, K. Tsijkawa, T. Kanamori and H. Inoue, *Analyst*, 2019, **144**, 2158–2165.
- 33 S. Han, C. Zhang, S. Lin, X. Sha and W. Hasi, *Spectrochim. Acta, Part A*, 2021, **251**, 119463.
- 34 C. Shende, C. Brouillette and S. Farquharson, *Analyst*, 2019, **144**, 5449–5454.
- 35 X. Su, X. Liu, Y. Xie, M. Chen, H. Zhong and M. Li, *Anal. Chem.*, 2023, **95**, 3821–3829.



- 36 T. Cooman, C. E. Ott and L. E. Arroyo, *J. Forensic Sci.*, 2023, **68**, 1520–1526.
- 37 J. Leonard, A. Haddad, O. Green, R. L. Birke, T. Kubik, A. Kocak and J. R. Lombardi, *J. Raman Spectrosc.*, 2017, **48**, 1323–1329.
- 38 M. Zhang, C. Jin, Y. Nie, Y. Ren, N. Hao, Z. Xu, L. Dong and J. X. J. Zhang, *RSC Adv.*, 2021, **11**, 11329–11337.
- 39 A. Haddad, M. A. Comanescu, O. Green, T. A. Kubic and J. R. Lombardi, *Anal. Chem.*, 2018, **90**, 12678–12685.
- 40 H. Wang, Z. Xue, Y. Wu, J. Gilmore, L. Wang and L. Fabris, *Anal. Chem.*, 2021, **93**, 9373–9382.
- 41 M. M. Kimani, A. Lanzarotta and J. S. Batson, *J. Forensic Sci.*, 2020, **66**, 491–504.
- 42 L. Gozdziński, M. Ramsay, A. Larnder, B. Wallace and D. K. Hore, *J. Raman Spectrosc.*, 2021, **52**, 1308–1316.
- 43 R. R. Martens, L. Gozdziński, E. Newman, C. Gill, B. Wallace and D. K. Hore, *Anal. Chem.*, 2024, **96**, 12277–12285.
- 44 J. N. Park, J. Tardif, E. Thompson, J. G. Rosen, J. Arrendondo, S. Lira and T. C. Green, *Int. J. Drug Policy*, 2023, **121**, 104206.
- 45 A. A. Bunaciu, H. Y. Aboul-Enein and S. Fleschin, *Appl. Spectrosc. Rev.*, 2010, **45**, 206–219.
- 46 S. Tobias, A. M. Shapiro, C. J. Grant, P. Patel, M. Lysychyn and L. Ti, *Drug Alcohol Depend.*, 2021, **218**, 108300.
- 47 L. Wang, M. O. Vendrell-Dones, C. Deriu, S. Dogruuer, P. B. Harrington and B. McCord, *Appl. Spectrosc.*, 2021, **75**, 1225–1236.
- 48 M. Guo, M. Li, H. Fu, Y. Zhang, T. Chen, H. Tang, T. Zhang and H. Li, *Spectrochim. Acta, Part A*, 2023, **287**, 122057.
- 49 S. Seifert, *Sci. Rep.*, 2020, **10**, 5436.
- 50 S. Weng, M. Qiu, R. Dong, F. Wang, L. Huang, D. Zhang and J. Zhao, *Spectrochim. Acta, Part A*, 2018, **200**, 20–25.
- 51 C. Crisci, B. Ghattas and G. Perera, *Ecol. Modell.*, 2012, **240**, 113–122.
- 52 M. Daviran, A. Maghsoudi, R. Ghezelbash and B. Pradhan, *Comput. Geosci.*, 2021, **148**, 104688.
- 53 A. Azadeh, M. Saberi, A. Kazem, V. Ebrahimipour, A. Nourmohammadzadeh and Z. Saberi, *Appl. Soft Comput.*, 2013, **13**, 1478–1485.
- 54 P. T. Noi and M. Kappas, *Sensors*, 2018, **18**, 18.
- 55 A. Murugan, S. H. Nair and K. P. S. Kumar, *J. Med. Syst.*, 2019, **43**, 269.
- 56 R. K. Halder, M. N. Uddin, A. Uddin, S. Aryal and A. Khraisat, *J. Big Data*, 2024, **11**, 113–155.
- 57 S. A. Naghibi, K. Ahmadi and A. Daneshi, *Water Resour. Manage.*, 2017, **31**, 2761–2775.
- 58 P. Pratheep Kumar, V. Mary Amala Bai and G. G. Nair, *Biomed. Signal Process. Control*, 2021, **68**, 102682.
- 59 T. Cooman, T. Trejos, A. Romero and L. Arroyo, *Chem. Phys. Lett.*, 2022, **787**, 139283.
- 60 R. Dong, S. Weng, L. Yang and J. Liu, *Anal. Chem.*, 2015, **87**, 2937–2944.
- 61 I. F. Darie, S. R. Anton and M. Praisler, *Inventions*, 2023, **8**, 56.
- 62 Y. Zou, A. Zhang, X. Wang, L. Yang and M. Ding, *J. Forensic Sci.*, 2024, **69**, 584–592.
- 63 F. Ahmmed, I. D. Fuller, D. P. Killeen, S. J. Fraser-Miller and K. C. Gordon, *ACS Food Sci. Technol.*, 2021, **1**, 570–578.
- 64 H. Liu, Y. Chen, C. Shi, X. Yang and D. Han, *LWT–Food Sci. Technol.*, 2020, **119**, 108906.
- 65 M. Zhao, M. Markiewicz-Keszycka, R. J. Beattie, M. P. Casado-Gavaldà, X. Cama-Moncunill, C. P. O'Donnell, P. J. Cullen and C. Sullivan, *Food Chem.*, 2020, **320**, 126639.
- 66 J. D. Anban, C. James, J. S. Kumar and S. Pradhan, *SN Appl. Sci.*, 2020, **2**, 1685.
- 67 B. Wallace, L. Gozdziński, A. Qbaich, S. Azam, P. Burek, A. Hutchison, T. Teal, R. Louw, C. Kiely, D. Robinson, B. Moa, M.-A. Storey, C. Gill and D. Hore, *Drugs Habits Soc. Policy*, 2022, **23**, 220–231.
- 68 M. Smith, M. Logan, M. Bazley, J. Blanchfield, R. Stokes, A. Blanco and R. McGee, *J. Forensic Sci.*, 2020, **66**, 505–519.
- 69 A. Miskulin, B. Wallace, C. Gill and D. Hore, *Drug Test. Anal.*, 2024, **16**, 1085–1093.
- 70 S. A. Borden, A. Saatchi, G. W. Vandergrift, J. Palaty, M. Lyshyshyn and C. G. Gill, *Drug Alcohol Rev.*, 2022, **41**, 410–418.
- 71 S. A. Borden, A. Saatchi, E. T. Krogh and C. G. Gill, *Anal. Sci. Adv.*, 2020, **1**, 97–108.
- 72 L. Gozdziński, J. Aasen, A. Larnder, M. Ramsay, S. A. Borden, A. Saatchi, C. G. Gill, B. Wallace and D. K. Hore, *Int. J. Drug Policy*, 2021, **97**, 103409.
- 73 T. pandas development team, pandas-dev/pandas: Pandas, 2020DOI: [10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134).
- 74 W. McKinney, *Proceedings of the 9th Python in Science Conference*, 2010, pp. 56–61.
- 75 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
- 76 K. A. Batterton and C. M. Schubert, *Stat. Med.*, 2016, **35**, 78–96.
- 77 S. J. Staffa and D. Zurakowski, *Anesthesiology*, 2021, **135**, 396–405.
- 78 J. Maillo, S. Ramirez, I. Triguero and F. Herrera, *Knowl.-Based Syst.*, 2017, **117**, 3–15.
- 79 S.-H. Luo, W.-L. Wang, Z.-F. Zhou, Y. Xie, B. Ren, G.-K. Liu and Z.-Q. Tian, *Anal. Chem.*, 2022, **94**, 10151–10158.

