Supplementary Information (SI) for Analyst.

Supplementary information for

Optimized Machine Learning Approaches to Combining Surface-Enhanced Raman Scattering and Infrared Data for Trace Detection of Xylazine in Illicit Opioids

Rebecca R. Martens,[1] Lea Gozdzialski,[1] Ella Newman,[1] Chris Gill,[2,1,3,5] Bruce Wallace,[4,5] Dennis K. Hore[1,5,6,*]

[1]*Department of Chemistry, University of Victoria, Victoria, British Columbia, V8W 3V6, Canada*

[2] *Department of Chemistry, Vancouver Island University, Nanaimo, British Columbia, V9R 5S5, Canada*

[3] *Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA 98195, United States*

[4]*School of Social Work, University of Victoria, Victoria, British Columbia, V8W 2Y2, Canada*

[5]*Canadian Institute for Substance Use Research, University of Victoria, Victoria, British Columbia, V8W 2Y2, Canada*

[6]*Department of Computer Science, University of Victoria, Victoria, British Columbia, V8W 3P6, Canada*

e-mail: dkhore@uvic.ca

Contents

S1. Sample Component Information

The breakdown of components and their concentrations in the samples used in the study ($n = 218$) as identified by paper-spray mass spectrometry, benzodiazepine test strips, and infrared spectroscopy is indicated in Table S1.

S2. Outlier Detection Method

To improve model performance and limit variance in the training data, high variance samples were removed from the hybrid and high-level spectral training sets. The spectral data for all samples in the training set ($n = 168$) were first mean centred and scaled to unit variance. Principal component analysis (PCA) was then applied reduce the dimensionality and visualize sample variance in the first 2 principal component space. Outliers were selected using the minimum covariance determinant (MCD) method, and Mahalanobis distances were calculated. A threshold based on the χ^2 distribution was used to flag outliers. Figure S1 illustrates the outlier detection process, showing the first two principal components and robust Mahalanobis distances for SERS, IR, and concatenated SERS–IR data. This method ensured that the training libraries consisted solely of complex multi-component samples that clustered within the first two principal components. The outliers identified by the process were removed, establishing the final training spectral datasets of SERS ($n = 151$), IR ($n = 153$), and concatenated SERS–IR ($n = 154$) which were used for hyperparameter optimization and model development.

S3. PCA Feature Extraction

To extract the relevant features for the mid-level data fusion approach, the IR and SERS datasets ($n = 168$ for each) were processed into 0th, 1st, and 2nd derivatives, resulting in three datasets for both IR and SERS. All datasets were standardized using StandardScalar normalization, and PCA was applied to extract features capturing 95% of the variance. For the 0th derivative, 46 PCs from SERS and 8 from IR were retained, for the 1st derivative, 50 PCs from SERS and 37 from IR, and for the 2nd derivative, 92 PCs from SERS and 87 from IR. The resulting PC scores from IR were fused to the end of the SERS PC scores in a single extracted feature matrix using hstack, producing fused SERS–IR PC datasets. These fused datasets were used to perform hyperparameter tuning for random forest (RF), support vector machine (SVM), and k-nearest neighbor (KNN) models, with the optimal model parameters and derivative combination determined based on the F1 score.

The validation datasets for both SERS ($n = 50$) and IR ($n = 50$) were preprocessed with the respective optimal order of derivatives for the given model and were standardized using the same StandardScaler models that were fitted on the training data. The PCA transformations were then applied to the standardized test data using the pre-trained PCA models. The transformed PCA features from both the SERS and IR validation datasets are then concatenated to form a test feature SERS–IR matrix before assessing model performance.

Compound	Concentration Range	Median
	(% w/w)	
Caffeine (n = 188)	Identified (No Quantity)	–
Fentanyl (n = 180)	0.01 – 71.68	8.71
Fluorofentanyl (n = 139)	0.15 – 78.48	5.50
Erythritol (n = 93)	Identified (No Quantity)	–
Bromazolam (n = 90)	0.01 – 52.56	2.01
Xylazine (n = 66)	0.05 – 32.15	1.52
ANPP (n = 29)	0.41 – 28.69	0.92
Chloroisobutyryl Fentanyl (n = 16)	0.10 – 4.29	0.15
Undifferentiated Benzodiazepine (n = 11)	Identified (No Quantity)	–
Etizolam (n = 9)	0.29 – 61.89	1.37
Flubromazepam (n = 7)	0.18 – 2.80	0.65
Heroin (n = 6)	2.05 – 8.39	2.55
Flualprazolam (n = 6)	0.57 – 1.90	0.69
Isobutyryl Fentanyl (n = 5)	3.38 – 5.17	4.15
Acetylfentanyl (n = 4)	0.22 – 3.22	1.72
Flubromazolam (n = 3)	1.94 – 7.33	2.15
Acetylcodeine (n = 2)	0.62 – 0.72	–
Carfentanil (n = 2)	0.58 – 0.96	–
Metonitazene (n = 1)	Identified (No Quantity)	–
N-desethyl Isotonitazene (n = 1)	18.78	–
Morphine (n = 1)	1.18	–
Acetylmorphine (n = 1)	1.93	–
4-Anilino-boc-piperidine (n = 1)	Identified (No Quantity)	–

Table S1: Compounds identified by paper-spray mass spectrometry, benzodiazepine immunoassay test strips, and infrared spectroscopy in all samples ($n = 218$). Concentration ranges and median concentrations are listed for all compounds.

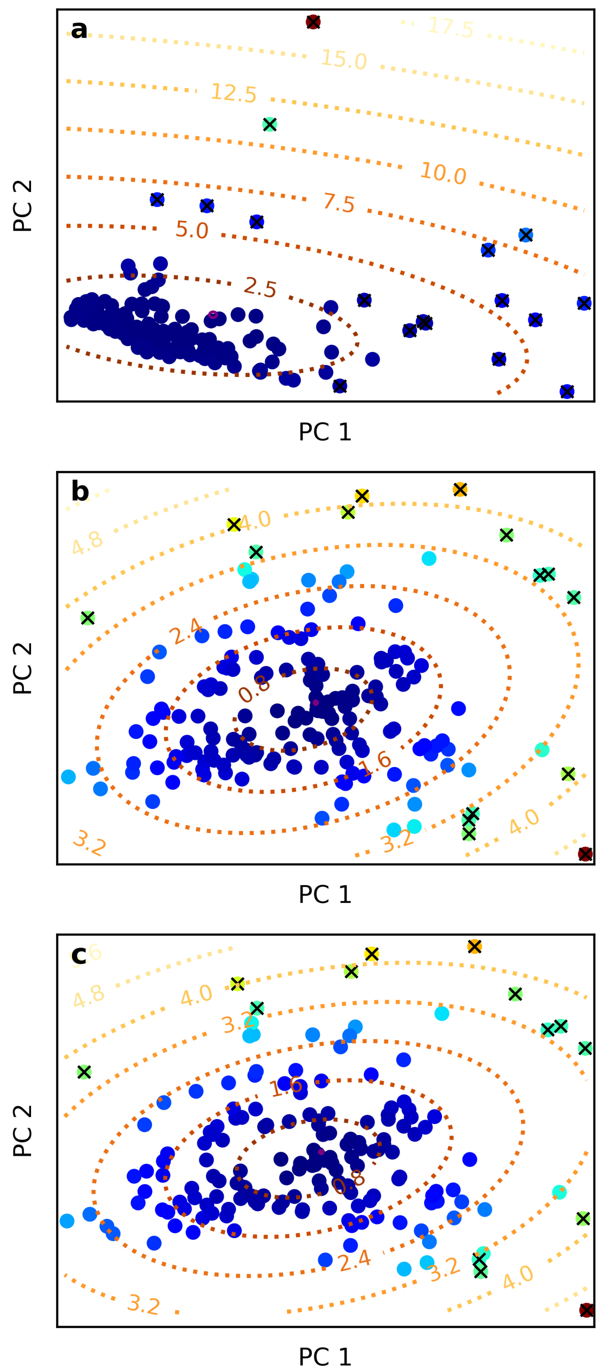

Figure S1: Outlier detection using PCA and robust Mahalanobis distance for (a) SERS, (b) IR, and (c) concatenated SERS–IR fused spectral data. Each subplot illustrates the scatter plot of the first two principal components, colour-coded by Mahalanobis distance. Overlaid contour plots represent the Mahalanobis distances, with identified outliers marked with x. Gradient colours of the data points range from dark blue (low variance) to red (high variance).

S4. Hyperparameter Tuning and Fused Model Development

S4.1. Hyperparameter Tuning

Binary classification for xylazine prediction in the hybrid, mid-level, and high-level training sets was performed using random forest (RF), support vector machine (SVM) with C-support vector classification (SVC), and k-nearest neighbor (KNN) models (scikit-learn). Default settings from the scikit-learn package were applied to all model parameters not adjusted during hyperparameter tuning.

All fusion models were optimized using a 5-fold cross-validated grid search to identify the best combination of parameters and spectral preprocessing techniques. Optimal values were selected based on the highest F1 score from the grid search. The spectral preprocessing techniques applied to the training data varied depending on the fusion method. Hyperparameters were tailored to each classifier and remained consistent across all fusion approaches. Table S2 summarizes the parameters tested for each classifier type.

S4.2. Hybrid Model Development

The concatenated SERS–IR training set ($n = 154$) was used to construct the hybrid RF, SVM, and KNN models. Model parameters were tested in combination with spectral preprocessing methods for the SERS and IR regions seperately using a pipeline. The preprocessing pipeline was designed to transform the raw SERS and IR spectral data before model training. The pipeline first restricted the data to the relevant spectral regions (1601 variables for SERS and 4000 for IR). Next, various preprocessing techniques were applied, including normalization (SNV, min-max, area, and none) and derivative computation (0, 1, and 2) to enhance spectral features for SERS and IR regions in the same matrix. These preprocessing steps were systematically optimized using grid search and cross-validation to identify the best combination of preprocessing methods and classification model parameters for all hybrid models. Results for the optimized hybrid model parameters for concatenated SERS–IR spectral data are shown for the RF (Table S3), SVM (Table S4), and KNN (Table S5) classifiers for the detection of xylazine.

S4.3. Mid-Level Model Development

The SERS–IR PC training sets ($n = 168$) in the 0th (54 PCs), 1st (87 PCs), and 2nd (179 PCs) derivative were used to construct the mid-level RF, SVM, and KNN models. Model parameters were tested for all derivative training sets using grid search and cross-validation. Results for the optimal combination of model parameters and order of derivatives selected for the SERS–IR PC training set are shown for the RF (Table S6), SVM (Table S7), and KNN (Table S8) mid-level classifiers for the detection of xylazine.

Model Optimization

Parameter	Description	Grid Search Values
Random Forest		
n_estimators	The number of decision trees in the forest	5, 10, 20, 30, 40, 50, 100, 150, 200, 300, 400, 500, 800, 1000, 2000
max_depth	The depth of each tree	None, 1, 2, 3, 4, 5, 10, 20, 30, 50
min_samples_leaf	The minimum number of samples needed to form a leaf node	1, 2, 3, 4, 5, 10, 20, 50
SVM		
C	The regularization parameter	1, 10, 100
kernel	Defines the kernel type used in the algorithm	'linear', 'rbf', 'poly', 'sigmoid'
gamma	The kernel coefficient for 'rbf', 'poly', and 'sigmoid' kernels	'auto', 'scale'
degree	The degree of the polynomial ('poly') kernel function	1, 2, 3, 4, 5
coef0	Independent term in the kernel function (relevant only for 'poly' and 'sigmoid' kernels)	0.0, 0.1, 0.5, 1.0, 2.0
class_weight	The weights associated with classes	None, 'balanced'
KNN		
n_neighbors	The number of neighbours used for k-neighbors queries	1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 35, 37, 39
weights	The weight function used in prediction	'uniform', 'distance'
algorithm	The algorithm used to calculate the nearest neighbours	'auto', 'ball_tree', 'kd_tree', 'brute'
metric	How the algorithm calculates the distance between data points	'euclidean', 'manhattan', 'minkowski'

Table S2: Investigation of hyperparameters using a 5-fold cross-validation grid search for development of random forest (RF), support vector machine (SVM), and *k*-nearest neighbor (KNN) classification models. Showing hyperparameter descriptions and the parameter combinations tested for all models.

Random Forest (RF)

Parameter Name	Description	Grid Search Values	Hybrid SERS—IR Data Fusion	
			Best Parameter	F1 Score
n_estimators	The number of decision trees in the forest	5, 10, 20, 30, 40, 50, 100, 150, 200, 300, 400, 500, 800, 1000, 2000	50	0.743
max_depth	The depth of each tree	None, 1, 2, 3, 4, 5, 10, 20, 30, 50	None	
min_samples_leaf	The minimum number of samples needed to form a leaf node	1, 2, 3, 4, 5, 10, 20, 50	1	
deriv_sers	The order of derivatives of the SERS spectral data	0, 1, 2	0	
deriv_ir	The order of derivatives of the IR spectral data	0, 1, 2	0	
norm_sers	The normalization process of the SERS spectral data	‘min-max', 'snv', 'none', 'area'	‘snv’	
norm_ir	The normalization process of the IR spectral data	‘min-max', 'snv', 'none', 'area'	'snv’	

Table S3: Investigation of RF hyperparameters and spectral preprocessing techniques using a 5-fold cross-validation grid search for the detection of xylazine using the hybrid SERS–IR data fusion method. Showing the combinations tested and the optimal set of parameters selected based on F1 score.

Support Vector Machine (SVM)

Parameter Name	Description	Grid Search Values	Hybrid SERS—IR Data Fusion	
			Best Parameter	F1 Score
C	The regularization parameter	1, 10, 100	10	0.881
kernel	Defines the kernel type used in the algorithm	'linear', 'rbf', 'poly', 'sigmoid'	‘rbf’	
gamma	The kernel coefficient for 'rbf', 'poly', and 'sigmoid' kernels	'auto’, ‘scale'	‘scale’	
degree	The degree of the polynomial ('poly') kernel function	1, 2, 3, 4, 5	N/A	
coef0	Independent term in the kernel function (relevant only for ‘poly’ and ‘sigmoid’ kernels)	0.0, 0.1, 0.5, 1.0, 2.0	N/A	
class_weight	The weights associated with classes	None, ‘balanced’	‘balanced’	
deriv_sers	The order of derivatives of the SERS spectral data	0, 1, 2	0	
deriv_ir	The order of derivatives of the IR spectral data	0, 1, 2	2	
norm_sers	The normalization process of the SERS spectral data	‘min-max', 'snv', 'none', 'area'	‘area’	
norm_ir	The normalization process of the IR spectral data	‘min-max', 'snv', 'none', 'area'	'none’	

Table S4: Investigation of SVM hyperparameters and spectral preprocessing techniques using a 5-fold cross-validation grid search for the detection of xylazine using the hybrid SERS–IR data fusion method. Showing the combinations tested and the optimal set of parameters selected based on F1 score.

k-Nearest Neighbors (KNN)

Parameter Name	Description	Grid Search Values	Hybrid SERS—IR Data Fusion	
			Best Parameter	**F1 Score**
n_neighbors	The number of neighbours used for k-neighbors queries	1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 35, 37, 39	1	0.748
weights	The weight function used in prediction	'uniform', 'distance'	'uniform'	
algorithm	The algorithm used to calculate the nearest neighbours	'auto', 'ball_tree', 'kd_tree', 'brute'	'auto'	
metric	How the algorithm calculates the distance between data points	'euclidean', 'manhattan', 'minkowski'	'manhattan'	
deriv_sers	The order of derivatives of the SERS spectral data	0, 1, 2	0	
deriv_ir	The order of derivatives of the IR spectral data	0, 1, 2	0	
norm_sers	The normalization process of the SERS spectral data	'min-max', 'snv', 'none', 'area'	'area'	
norm_ir	The normalization process of the IR spectral data	'min-max', 'snv', 'none', 'area'	'area'	

Table S5: Investigation of KNN hyperparameters and spectral preprocessing techniques using a 5-fold cross-validation grid search for the detection of xylazine using the hybrid SERS–IR data fusion method. Showing the combinations tested and the optimal set of parameters selected based on F1 score.

Random Forest (RF)

Parameter Name	Description	Grid Search Values	Mid-Level SERS—IR Data Fusion	
			Best Parameter	**F1 Score**
n_estimators	The number of decision trees in the forest	5, 10, 20, 30, 40, 50, 100, 150, 200, 300, 400, 500, 800, 1000, 2000	500	0.616
max_depth	The depth of each tree	None, 1, 2, 3, 4, 5, 10, 20, 30, 50	10	
min_samples_leaf	The minimum number of samples needed to form a leaf node	1, 2, 3, 4, 5, 10, 20, 50	1	
deriv	The order of derivatives of the spectral data for PCA feature extraction	0, 1, 2	1	

Table S6: Investigation of RF hyperparameters using a 5-fold cross-validation grid search for the detection of xylazine using the mid-level SERS–IR data fusion method. Showing the combinations tested and the optimal set of parameters selected based on F1 score.

Support Vector Machine (SVM)

Parameter Name	Description	Grid Search Values	Mid-Level SERS—IR Data Fusion	
			Best Parameter	F1 Score
C	The regularization parameter	1, 10, 100	1	0.805
kernel	Defines the kernel typer used in the algorithm	'linear', 'rbf', 'poly', 'sigmoid'	'sigmoid'	
gamma	The kernel coefficient for 'rbf', 'poly', and 'sigmoid' kernels	'auto', 'scale'	'auto'	
degree	The degree of the polynomial ('poly') kernel function	1, 2, 3, 4, 5	N/A	
coef0	Independent term in the kernel function (relevant only for 'poly' and 'sigmoid' kernels)	0.0, 0.1, 0.5, 1.0, 2.0	0.1	
class_weight	The weights associated with classes	None, 'balanced'	'balanced'	
deriv	The order of derivatives of the spectral data for PCA feature extraction	0, 1, 2	1	

Table S7: Investigation of SVM hyperparameters using a 5-fold cross-validation grid search for the detection of xylazine using the mid-level SERS–IR data fusion method. Showing the combinations tested and the optimal set of parameters selected based on F1 score.

k-Nearest Neighbors (KNN)

Parameter Name	Description	Grid Search Values	Mid-Level SERS—IR Data Fusion	
			Best Parameter	F1 Score
n_neighbors	The number of neighbours used for k-neighbors queries	1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 35, 37, 39	1	0.614
weights	The weight function used in prediction	'uniform', 'distance'	'uniform'	
algorithm	The algorithm used to calculate the nearest neighbours	'auto', 'ball_tree', 'kd_tree', 'brute'	'auto'	
metric	How the algorithm calculates the distance between data points	'euclidean', 'manhattan', 'minkowski'	'manhattan'	
deriv	The order of derivatives of the spectral data for PCA feature extraction	0, 1, 2	0	

Table S8: Investigation of KNN hyperparameters using a 5-fold cross-validation grid search for the detection of xylazine using the mid-level SERS–IR data fusion method. Showing the combinations tested and the optimal set of parameters selected based on F1 score.

Random Forest (RF)

Parameter Name	Description	Grid Search Values	SERS		IR	
			Best Parameter	F1 Score	Best Parameter	F1 Score
n_estimators	The number of decision trees in the forest	5, 10, 20, 30, 40, 50, 100, 150, 200, 300, 400, 500, 800, 1000, 2000	40	0.805	30	0.532
max_depth	The depth of each tree	None, 1, 2, 3, 4, 5, 10, 20, 30, 50	3		30	
min_samples_leaf	The minimum number of samples needed to form a leaf node	1, 2, 3, 4, 5, 10, 20, 50	4		4	
deriv	The order of derivatives of the spectral data	0, 1, 2	0		0	
norm	The normalization process of the spectral data	'min-max', 'snv', 'none', 'area'	'snv'		'snv'	

Table S9: Investigation of RF hyperparameters and spectral preprocessing techniques using a 5-fold cross-validation grid search for the detection of xylazine using SERS and IR data. Showing the combinations tested and the optimal set of parameters selected based on F1 score.

S4.4. High-Level Model Development

The SERS ($n = 151$) and IR ($n = 153$) spectral training sets were each used to construct their respective RF, SVM, and KNN standalone models. Classifier-specific hyperparamters were tested in combination with various spectral preprocessing methods (normalization and derivatives) to optimize the parameters for the standalone SERS and IR models. The optimized parameters for detecting xylazine were identified for RF (Table S9), SVM (Table S10), and KNN (Table S5) models across both spectroscopic platforms.

S5. Voting Classifier (Weighted)

A weighted voting classifier (sklearn.ensemble) was constructed to combine the predictions from both SERS and IR models in the high-level data fusion approach. The high variance samples identified for both standalone SERS ($n = 17$) and IR ($n = 15$) training data were removed to establish the high-level training library ($n = 141$) for assessing weight combinations. All combinations of weights for the SERS and IR standalone models, summing to 1, were evaluated using a 5-fold cross-validation grid search on the training data. The voting classifier was initialized in 'soft' voting mode, enabling probability-based weighting. The optimal weight combination for the SERS and IR models was selected based on the highest F1 score. The tested weight combinations and their respective results are shown for the high-level RF (Table S12), SVM (Table S13), and KNN (Table S14) fused models. The best performing weights were applied to the predicted probability scores of the standalone models on the test set and added together for all the high-level fused model predictions.

S6. Analysis of SERS and IR Standalone Models

All SERS and IR standalone models developed for the high-level data fusion method were evaluated independently on the validation set ($n = 50$) to assess xylazine prediction results

Support Vector Machine (SVM)

Parameter Name	Description	Grid Search Values	SERS		IR	
			Best Parameter	F1 Score	Best Parameter	F1 Score
C	The regularization parameter	1, 10, 100	100	0.869	100	0.790
kernel	Defines the kernel type used in the algorithm	'linear', 'rbf', 'poly', 'sigmoid'	'poly'		'poly'	
gamma	The kernel coefficient for 'rbf', 'poly', and 'sigmoid' kernels	'auto', 'scale'	'auto'		'scale'	
degree	The degree of the polynomial ('poly') kernel function	1, 2, 3, 4, 5	2		3	
coef0	Independent term in the kernel function (relevant only for 'poly' and 'sigmoid' kernels)	0.0, 0.1, 0.5, 1.0, 2.0	0.5		0.1	
class_weight	The weights associated with classes	None, 'balanced'	'balanced'		'balanced'	
deriv	The order of derivatives of the spectral data	0, 1, 2	0		1	
norm	The normalization process of the spectral data	'min-max', 'snv', 'none', 'area'	'snv'		'snv'	

Table S10: Investigation of SVM hyperparameters and spectral preprocessing techniques using a 5-fold cross-validation grid search for the detection of xylazine using SERS and IR data. Showing the combinations tested and the optimal set of parameters selected based on F1 score.

k-Nearest Neighbors (KNN)

Parameter Name	Description	Grid Search Values	SERS		IR	
			Best Parameter	F1 Score	Best Parameter	F1 Score
n_neighbors	The number of neighbours used for k-neighbors queries	1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 35, 37, 39	1	0.690	1	0.618
weights	The weight function used in prediction	'uniform', 'distance'	'uniform'		'uniform'	
algorithm	The algorithm used to calculate the nearest neighbours	'auto', 'ball_tree', 'kd_tree', 'brute'	'auto'		'auto'	
metric	How the algorithm calculates the distance between data points	'euclidean', 'manhattan', 'minkowski'	'manhattan'		'euclidean'	
deriv	The order of derivatives of the spectral data	0, 1, 2	0		2	
norm	The normalization process of the spectral data	'min-max', 'snv', 'none', 'area'	'area'		'min-max'	

Table S11: Investigation of KNN hyperparameters and spectral preprocessing techniques using a 5-fold cross-validation grid search for the detection of xylazine using SERS and IR data. Showing the combinations tested and the optimal set of parameters selected based on F1 score.

High-Level RF Model	
Weights [SERS, IR]	**F1 Score**
[0.0, 1.0]	0.442
[0.1, 0.9]	0.409
[0.2, 0.8]	0.419
[0.3, 0.7]	0.448
[0.4, 0.6]	0.577
[0.5, 0.5]	0.618
[0.6, 0.4]	0.674
[0.7, 0.3]	0.727
[0.8, 0.2]	0.753
[0.9, 0.1]	0.754
[1.0, 0.0]	0.742

Table S12: Investigation of weights tested for standalone SERS and IR RF model predictions using a 5-fold cross-validation grid search on the high-level training set ($n = 141$) for the classification of xylazine. The table shows the weight combinations tested and their corresponding F1 scores. The highlighted row indicates the optimal weight combination selected for the high-level RF model.

High-Level SVM Model	
Weights [SERS, IR]	**F1 Score**
[0.0, 1.0]	0.476
[0.1, 0.9]	0.482
[0.2, 0.8]	0.482
[0.3, 0.7]	0.626
[0.4, 0.6]	0.694
[0.5, 0.5]	0.717
[0.6, 0.4]	0.693
[0.7, 0.3]	0.722
[0.8, 0.2]	0.755
[0.9, 0.1]	0.795
[1.0, 0.0]	0.771

Table S13: Investigation of weights tested for standalone SERS and IR SVM model predictions using a 5-fold cross-validation grid search on the high-level training set ($n = 141$) for the classification of xylazine. The table shows the weight combinations tested and their corresponding F1 scores. The highlighted row indicates the optimal weight combination selected for the high-level SVM model.

High-Level KNN Model	
Weights [SERS, IR]	**F1 Score**
[0.0, 1.0]	0.552
[0.1, 0.9]	0.552
[0.2, 0.8]	0.552
[0.3, 0.7]	0.552
[0.4, 0.6]	0.552
[0.5, 0.5]	0.505
[0.6, 0.4]	0.637
[0.7, 0.3]	0.637
[0.8, 0.2]	0.637
[0.9, 0.1]	0.637
[1.0, 0.0]	0.637

Table S14: Investigation of weights tested for standalone SERS and IR KNN model predictions using a 5-fold cross-validation grid search on the high-level training set ($n =$ 141) for the classification of xylazine. The table shows the weight combinations tested and their corresponding F1 scores. The highlighted row indicates the optimal weight combination selected for the high-level KNN model.

of the spectroscopic platforms. ROC analysis was performed on all SERS and IR models across the range of predicted probabilities to determine the AUC and the cut-off threshold. Optimal thresholds were calculated using Youden's J statistic and applied to the test set to evaluate the xylazine prediction results of all individual models. The summary of AUC, optimal threshold, and all performance metrics for SERS and IR standalone models using RF, SVM, and KNN classifiers is illustrated in Table S15. A detailed breakdown of sample composition and xylazine prediction results of the RF, SVM, and KNN classifiers using SERS, IR, and high-level SERS–IR data on the validation set ($n = 50$) is presented in Table S16.

	Model	AUC	Optimal Threshold	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1 Score (%)
RF								
	SERS	0.95	0.26	90	88	92	88	90
	IR	0.82	0.14	74	68	92	56	78
SVM								
	SERS	0.85	0.48	78	82	72	84	77
	IR	0.84	0.18	80	78	84	76	81
KNN								
	SERS	0.70	0.50	70	78	56	84	65
	IR	0.70	0.50	70	81	52	88	63

Table S15: Performance metrics summary for RF, SVM, and KNN models using SERS and IR spectral data 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.

S7. Validation Makeup

The complete breakdown of sample composition and the xylazine prediction results of all fused models on the validation set ($n = 50$) is illustrated in Table S17.

AERL	Sample Composition	Xylazine									
		Concentration (% w/w)	RF Model Results			SVM Model Results			KNN Model Results		
			SERS	IR	High-Level	SERS	IR	High-Level	SERS	IR	High-Level
5184	Fentanyl (6.95%) + Etizolam (1.07%) + Flubromazepam (0.73%) + Xylazine (0.5%) + Fluorofentanyl (0.46%) + Caffeine (nan%)	0.5	TP	TP	TP	TP	TP	TP	TP	FN	TP
5404	Fentanyl (4.36%) + Bromazolam (0.99%) + Xylazine (0.64%) + Caffeine (nan%)	0.64	TP	TP	TP	TP	TP	TP	FN	FN	FN
5405	Fentanyl (3.68%) + Bromazolam (0.95%) + Xylazine (0.5%) + Caffeine (nan%)	0.5	TP	TP	TP	FN	TP	FN	FN	TP	TP
5416	Fentanyl (17.62%) + Fluorofentanyl (0.37%) + Xylazine (0.96%) + Erythritol (nan%) + Caffeine (nan%)	0.96	TP	TP	TP	FN	TP	FN	FN	TP	TP
5417	Fentanyl (6.45%) + ANPP (0.46%) + Fluorofentanyl (0.65%) + Xylazine (1.74%) + Caffeine (nan%)	1.74	TP	TP	TP	TP	TP	TP	TP	TP	TP
5424	Fentanyl (6.21%) + Fluorofentanyl (0.71%) + Xylazine (1.64%) + Benzodiazepine (undifferentiated) (nan%) + Caffeine (nan%) + Erythritol (nan%)	1.64	TP	TP	TP	TP	TP	TP	TP	TP	TP
5452	Fentanyl (15.48%) + Etizolam (1.06%) + Flubromazepam (2.8%) + Xylazine (4.57%) + Caffeine (nan%)	4.57	TP	FN	TP	TP	FN	TP	TP	FN	TP
5453	Fentanyl (10.53%) + Etizolam (1.06%) + Xylazine (0.88%) + Caffeine (nan%)	0.88	TP	TP	TP	TP	TP	TP	TP	TP	TP
5464	Fentanyl (4.4%) + Fluorofentanyl (0.49%) + Bromazolam (4.32%) + Xylazine (0.98%) + Caffeine (nan%)	0.98	TP	TP	TP	TP	TP	TP	FN	TP	TP
5493	Fentanyl (11.18%) + Fluorofentanyl (0.65%) + Xylazine (4.12%) + Benzodiazepine (undifferentiated) (nan%) + Caffeine (nan%)	4.12	TP	TP	TP	TP	TP	TP	TP	FN	TP
5779	Fentanyl (6.63%) + Xylazine (14.88%) + Caffeine (nan%)	14.88	TP	TP	TP	FN	TP	FN	FN	FN	FN
5934	Fentanyl (6.48%) + Xylazine (1.15%) + Fluorofentanyl (0.29%) + Benzodiazepine (undifferentiated) (nan%) + Caffeine (nan%)	1.15	TP	TP	TP	TP	FN	TP	FN	FN	FN
10073	Fentanyl (16.21%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	TN	TN	TN	FP	TN	TN	TN	TN
10543	Fentanyl (1.36%) + Fluorofentanyl (7.87%) + Bromazolam (3.15%) + Xylazine (3.17%) + Caffeine (nan%) + Erythritol (nan%)	3.17	FN	TP	TP	FN	TP	FN	FN	TP	TP
10549	Fentanyl (20.46%) + Fluorofentanyl (0.46%) + Erythritol (nan%) + Caffeine (nan%)	0	FP	FP	FP	FP	TN	FP	TN	TN	TN
10866	Fentanyl (17.97%) + Caffeine (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
10876	Fentanyl (12.02%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
10880	Fentanyl (9.14%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	FP	TN	TN	TN	TN	TN	TN	TN
10898	Fentanyl (10.96%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
12816	Fluorofentanyl Base (71.48%) + Fentanyl (0.75%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
12992	Fentanyl (2.21%) + Fluorofentanyl (7.94%) + Erythritol (nan%) + Caffeine (nan%) + ANPP (1.66%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
13141	Fluorofentanyl (nan%) + ANPP (1.08%) + Fentanyl (71.68%) + 4-Anilino-boc-piperidine (nan%)	0	TN	TN	TN	TN	TN	TN	TN	FP	FP
13163	Fentanyl (0.16%) + Fluorofentanyl (12.13%) + Bromazolam (0.82%) + Xylazine (8.47%) + Caffeine (nan%) + Erythritol (nan%)	8.47	TP	TP	TP	TP	TP	TP	TP	TP	TP
13167	Fluorofentanyl (19.7%) + Xylazine (0.1%) + Caffeine (nan%) + Erythritol (nan%)	0.1	TP	TP	TP	TP	TP	TP	TP	TP	TP
13830	Fluorofentanyl (37.02%) + Fentanyl (0.29%) + Caffeine (nan%)	0	TN	TN	TN	TN	TN	TN	TN	FP	FP
13833	Fluorofentanyl (5.77%) + Xylazine (5.43%) + Caffeine (nan%) + Erythritol (nan%)	5.43	TP	TP	TP	TP	TP	TP	TP	FN	TP
14414	Fluorofentanyl (19.24%) + Fentanyl (0.16%) + Bromazolam (1.29%) + Xylazine (13.06%) + Erythritol (nan%) + Caffeine (nan%)	13.06	TP	TP	TP	TP	TP	TP	TP	TP	TP
14793	Fentanyl (5.57%) + Fluorofentanyl (3.74%) + Bromazolam (0.9%) + Xylazine (1.84%) + Caffeine (nan%) + Erythritol (nan%)	1.84	TP	TP	TP	TP	TP	TP	FN	FN	FN
14799	Fentanyl (11.63%) + Fluorofentanyl (4.35%) + Bromazolam (0.96%) + Xylazine (0.79%) + Caffeine (nan%) + Erythritol (nan%)	0.79	FN	FN	FN	FN	FN	FN	FN	FN	FN
15080	Fluorofentanyl (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
15083	Fluorofentanyl (21.16%) + Fentanyl (1.69%) + Bromazolam (10.89%) + Caffeine (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
15095	Fluorofentanyl (3.49%) + Fentanyl (2.46%) + Bromazolam (1.7%) + Caffeine (nan%)	0	TN	FP	TN	TN	TN	TN	TN	TN	TN
15096	Fentanyl (25.53%) + Fluorofentanyl (1.89%) + Xylazine (0.47%) + Benzodiazepine (undifferentiated) (nan%) + Caffeine (nan%)	0.47	TP	TP	TP	TP	TP	TP	TP	FN	TP
15106	Fentanyl (8.94%) + Fluorofentanyl (0.83%) + Bromazolam (3.01%) + ANPP (1.3%) + Chloroisobutyryl fentanyl (0.14%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	TN	TN	FP	TN	FP	FP	TN	FP
15107	Fentanyl (4.6%) + Fluorofentanyl (2.29%) + Bromazolam (2.01%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	FP	TN	TN	FP	TN	TN	TN	TN
15108	Fentanyl (12.51%) + Chloroisobutyryl fentanyl (1.1%) + ANPP (3.9%) + Acetylfentanyl (0.29%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	FP	TN	TN	TN	FP	FP	TN	FP
15265	Fentanyl (13.24%) + Isobutyryl fentanyl (4.54%) + Heroin (3.16%) + Bromazolam (9.01%) + Caffeine (nan%) + Erythritol (nan%)	0	FP	TN	FP	TN	FP	TN	TN	FP	FP
15269	ANPP (0.57%) + Fentanyl (14.89%) + Fluorofentanyl (0.63%) + Bromazolam (0.37%) + Caffeine (nan%)	0	TN	FP	TN	TN	FP	TN	TN	TN	TN
15481	Fluorofentanyl (21.0%) + Caffeine (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
15482	Fentanyl (0.4%) + Fluorofentanyl (3.59%) + Bromazolam (1.46%) + Xylazine (4.46%) + Erythritol (nan%) + Caffeine (nan%)	4.46	TP	TP	TP	TP	TP	TP	TP	TP	TP
15483	ANPP (0.59%) + Fentanyl (34.6%) + Bromazolam (35.69%) + Flualprazolam (1.06%) + Xylazine (7.48%) + Erythritol (nan%)	7.48	TP	TP	TP	TP	TP	TP	FN	TP	TP
15487	Heroin (2.47%) + Fentanyl (11.32%) + Bromazolam (1.68%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	FP	TN	TN	FP	TN	TN	TN	TN
15489	Fentanyl (1.76%) + Fluorofentanyl (0.38%) + Bromazolam (0.32%) + Flubromazolam (2.15%) + Xylazine (0.15%) + Caffeine (nan%)	0.15	TP	TP	TP	FN	TP	TP	TP	FN	TP
15498	ANPP (0.65%) + Fentanyl (12.62%) + Isobutyryl fentanyl (3.73%) + Bromazolam (12.38%) + Caffeine (nan%)	0	TN	TN	TN	TN	TN	TN	FP	TN	FP
15501	ANPP (0.5%) + Fentanyl (12.92%) + Fluorofentanyl (0.55%) + Bromazolam (0.43%) + Caffeine (nan%)	0	TN	FP	TN	FP	FP	FP	FP	TN	FP
15503	Fentanyl (1.68%) + Fluorofentanyl (2.0%) + Bromazolam (5.88%) + Caffeine (nan%)	0	TN	FP	TN	TN	TN	TN	TN	TN	TN
15504	Fentanyl (9.83%) + Fluorofentanyl (15.42%) + Bromazolam (5.57%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	FP	TN	TN	TN	TN	TN	TN	TN
15505	Fentanyl (11.5%) + Fluorofentanyl (24.25%) + Bromazolam (3.11%) + Caffeine (nan%) + Erythritol (nan%)	0	FP	FP	FP	FP	TN	FP	TN	TN	TN
15550	Fentanyl (18.35%) + Bromazolam (8.23%) + Xylazine (15.3%) + Flubromazolam (1.94%) + Caffeine (nan%) + Erythritol (nan%)	15.3	TP	TP	TP	FN	TP	FN	FN	TP	TP
15617	Fentanyl (58.05%) + Bromazolam (nan%) + Xylazine (15.36%) + ANPP (0.97%) + Flualprazolam (1.90%)	15.36	TP	TP	TP	TP	FN	TP	TP	FN	TP

Table S16: Complete breakdown of validation set ($n = 50$) sample composition as identified by paper-spray mass spectrometry, benzodiazepine immunoassay test strips, and infrared spectroscopy. Concentration values of xylazine (blue) are described next to true positive (TP), true negative (TN), false positive (FP), and false negative (FN) prediction results of the SERS, IR, and high-level SERS–IR data fusion models. Correct predictions by the respective RF, SVM, and KNN models are shown in white, while incorrect predictions are in grey.

AERL	Sample Composition	Xylazine									
		Concentration (% w/w)	Hybrid Fused Model Results			Mid-Level Fused Model Results			High-Level Fused Model Results		
			RF	SVM	KNN	RF	SVM	KNN	RF	SVM	KNN
5184	Fentanyl (6.95%) + Etizolam (1.07%) + Flubromazepam (0.73%) + Xylazine (0.5%) + Fluorofentanyl (0.46%) + Caffeine (nan%)	0.5	TP	TP	TP	TP	TP	TP	TP	TP	TP
5404	Fentanyl (4.36%) + Bromazolam (0.99%) + Xylazine (0.64%) + Caffeine (nan%)	0.64	TP	FN	FN	TP	TP	TP	TP	TP	FN
5405	Fentanyl (3.68%) + Bromazolam (0.95%) + Xylazine (0.5%) + Caffeine (nan%)	0.5	FN	TP	FN	TP	FN	FN	TP	FN	TP
5416	Fentanyl (17.62%) + Fluorofentanyl (0.37%) + Xylazine (0.96%) + Erythritol (nan%) + Caffeine (nan%)	0.96	TP	TP	FN	TP	TP	FN	TP	FN	TP
5417	Fentanyl (6.45%) + ANPP (0.46%) + Fluorofentanyl (0.65%) + Xylazine (1.74%) + Caffeine (nan%)	1.74	TP	TP	FN	TP	TP	FN	TP	TP	TP
5424	Fentanyl (6.21%) + Fluorofentanyl (0.71%) + Xylazine (1.64%) + Benzodiazepine (undifferentiated) (nan%) + Caffeine (nan%) + Erythritol (nan%)	1.64	TP	TP	TP	TP	TP	TP	TP	TP	TP
5452	Fentanyl (15.48%) + Etizolam (1.06%) + Flubromazepam (2.8%) + Xylazine (4.57%) + Caffeine (nan%)	4.57	TP	FN	TP	TP	TP	FN	TP	TP	TP
5453	Fentanyl (10.53%) + Etizolam (1.06%) + Xylazine (0.88%) + Caffeine (nan%)	0.88	TP	TP	FN	TP	TP	FN	TP	TP	TP
5464	Fentanyl (4.4%) + Fluorofentanyl (0.49%) + Bromazolam (4.32%) + Xylazine (0.98%) + Caffeine (nan%)	0.98	TP	FN	FN	TP	TP	TP	TP	TP	TP
5493	Fentanyl (11.18%) + Fluorofentanyl (0.65%) + Xylazine (4.12%) + Benzodiazepine (undifferentiated) (nan%) + Caffeine (nan%)	4.12	TP	TP	FN	TP	TP	FN	TP	TP	TP
5779	Fentanyl (6.63%) + Xylazine (14.88%) + Caffeine (nan%)	14.88	TP	TP	FN	TP	TP	FN	TP	FN	FN
5934	Fentanyl (6.48%) + Xylazine (1.15%) + Fluorofentanyl (0.29%) + Benzodiazepine (undifferentiated) (nan%) + Caffeine (nan%)	1.15	FN	FN	FN	TP	TP	FN	TP	TP	FN
10073	Fentanyl (16.21%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	FP	TN	FP	FP	TN	TN	TN	TN
10543	Fentanyl (1.36%) + Fluorofentanyl (7.87%) + Bromazolam (3.15%) + Xylazine (3.17%) + Caffeine (nan%) + Erythritol (nan%)	3.17	TP	TP	FN	TP	TP	FN	TP	FN	TP
10549	Fentanyl (20.46%) + Fluorofentanyl (0.46%) + Erythritol (nan%) + Caffeine (nan%)	0	FP	TN	FP	TN	TN	TN	FP	FP	TN
10866	Fentanyl (17.97%) + Caffeine (nan%)	0	TN	TN	TN	FP	FP	TN	TN	TN	TN
10876	Fentanyl (12.02%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
10880	Fentanyl (9.14%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
10898	Fentanyl (10.96%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
12816	Fluorofentanyl Base (71.48%) + Fentanyl (0.75%)	0	TN	TN	TN	FP	TN	TN	TN	TN	TN
12992	Fentanyl (2.21%) + Fluorofentanyl (7.94%) + Erythritol (nan%) + Caffeine (nan%) + ANPP (1.66%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
13141	Fluorofentanyl (nan%) + ANPP (1.08%) + Fentanyl (71.68%) + 4-Anilino-boc-piperidine (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	FP
13163	Fentanyl (0.16%) + Fluorofentanyl (12.13%) + Bromazolam (0.82%) + Xylazine (8.47%) + Caffeine (nan%) + Erythritol (nan%)	8.47	TP	TP	TP	TP	TP	TP	TP	TP	TP
13167	Fluorofentanyl (19.7%) + Xylazine (0.1%) + Caffeine (nan%) + Erythritol (nan%)	0.1	TP	TP	TP	TP	TP	TP	TP	TP	TP
13830	Fluorofentanyl (37.02%) + Fentanyl (0.29%) + Caffeine (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	FP
13833	Fluorofentanyl (5.77%) + Xylazine (5.43%) + Caffeine (nan%) + Erythritol (nan%)	5.43	TP	TP	TP	TP	TP	TP	TP	TP	TP
14414	Fluorofentanyl (19.24%) + Fentanyl (0.16%) + Bromazolam (1.29%) + Xylazine (13.06%) + Erythritol (nan%) + Caffeine (nan%)	13.06	TP	TP	TP	TP	TP	TP	TP	TP	TP
14793	Fentanyl (5.57%) + Fluorofentanyl (3.74%) + Bromazolam (0.9%) + Xylazine (1.84%) + Caffeine (nan%) + Erythritol (nan%)	1.84	TP	FN	FN	TP	TP	TP	TP	TP	FN
14799	Fentanyl (11.63%) + Fluorofentanyl (4.35%) + Bromazolam (0.96%) + Xylazine (0.79%) + Caffeine (nan%) + Erythritol (nan%)	0.79	FN	FN	FN	FN	FN	TP	FN	FN	FN
15080	Fluorofentanyl (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
15083	Fluorofentanyl (21.16%) + Fentanyl (1.69%) + Bromazolam (10.89%) + Caffeine (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
15095	Fluorofentanyl (3.49%) + Fentanyl (2.46%) + Bromazolam (1.7%) + Caffeine (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
15096	Fentanyl (25.53%) + Fluorofentanyl (1.89%) + Xylazine (0.47%) + Benzodiazepine (undifferentiated) (nan%) + Caffeine (nan%)	0.47	TP	TP	TP	TP	FN	TP	TP	TP	TP
15106	Fentanyl (8.94%) + Fluorofentanyl (0.83%) + Bromazolam (3.01%) + ANPP (1.3%) + Chloroisobutyryl fentanyl (0.14%) + Caffeine (nan%) + Erythritol (nan%)	0	FP	TN	TN	TN	TN	FP	TN	FP	FP
15107	Fentanyl (4.6%) + Fluorofentanyl (2.29%) + Bromazolam (2.01%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
15108	Fentanyl (12.51%) + Chloroisobutyryl fentanyl (1.1%) + ANPP (3.9%) + Acetylfentanyl (0.29%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	TN	TN	TN	TN	TN	TN	FP	FP
15265	Fentanyl (13.24%) + Isobutyryl fentanyl (4.54%) + Heroin (3.16%) + Bromazolam (9.01%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	FP	TN	TN	TN	TN	FP	TN	FP
15269	ANPP (0.57%) + Fentanyl (14.89%) + Fluorofentanyl (0.63%) + Bromazolam (0.37%) + Caffeine (nan%)	0	TN	TN	TN	TN	TN	TN	TN	TN	TN
15481	Fluorofentanyl (21.0%) + Caffeine (nan%)	0	TN	TN	TN	TN	FP	TN	TN	TN	TN
15482	Fentanyl (0.4%) + Fluorofentanyl (3.59%) + Bromazolam (1.46%) + Xylazine (4.46%) + Erythritol (nan%) + Caffeine (nan%)	4.46	TP	TP	TP	TP	TP	TP	TP	TP	TP
15483	ANPP (0.59%) + Fentanyl (34.6%) + Bromazolam (35.69%) + Flualprazolam (1.06%) + Xylazine (7.48%) + Erythritol (nan%)	7.48	TP	TP	TP	FN	TP	FN	TP	TP	TP
15487	Heroin (2.47%) + Fentanyl (11.32%) + Bromazolam (1.68%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	TN	TN	TN	TN	FP	TN	TN	TN
15489	Fentanyl (1.76%) + Fluorofentanyl (0.38%) + Bromazolam (0.32%) + Flubromazolam (2.15%) + Xylazine (0.15%) + Caffeine (nan%)	0.15	TP	FN	TP	TP	FN	FN	TP	TP	TP
15498	ANPP (0.65%) + Fentanyl (12.62%) + Isobutyryl fentanyl (3.73%) + Bromazolam (12.38%) + Caffeine (nan%)	0	TN	TN	FP	TN	TN	TN	TN	TN	FP
15501	ANPP (0.5%) + Fentanyl (12.92%) + Fluorofentanyl (0.55%) + Bromazolam (0.43%) + Caffeine (nan%)	0	TN	TN	TN	TN	TN	TN	TN	FP	FP
15503	Fentanyl (1.68%) + Fluorofentanyl (2.0%) + Bromazolam (5.88%) + Caffeine (nan%)	0	TN	TN	FP	TN	TN	FP	TN	TN	TN
15504	Fentanyl (9.83%) + Fluorofentanyl (15.42%) + Bromazolam (5.57%) + Caffeine (nan%) + Erythritol (nan%)	0	FP	TN	TN	TN	TN	FP	TN	TN	TN
15505	Fentanyl (11.5%) + Fluorofentanyl (24.25%) + Bromazolam (3.11%) + Caffeine (nan%) + Erythritol (nan%)	0	TN	TN	TN	TN	TN	TN	FP	FP	TN
15550	Fentanyl (18.35%) + Bromazolam (8.23%) + Xylazine (15.3%) + Flubromazolam (1.94%) + Caffeine (nan%) + Erythritol (nan%)	15.3	TP	TP	TP	TP	FN	FN	TP	FN	TP
15617	Fentanyl (58.05%) + Bromazolam (nan%) + Xylazine (15.36%) + ANPP (0.97%) + Flualprazolam (1.90%)	15.36	TP	FN	FN	TP	TP	FN	TP	TP	TP

Table S17: Complete breakdown of validation set ($n = 50$) sample composition as identified by paper-spray mass spectrometry, benzodiazepine immunoassay test strips, and infrared spectroscopy. Concentration values of xylazine (blue) are described next to true positive (TP), true negative (TN), false positive (FP), and false negative (FN) prediction results of the hybrid, mid-level, and high-level SERS–IR data fusion models. Correct predictions by the respective RF, SVM, and KNN models are shown in white, while incorrect predictions are in grey.