



OPEN

High accuracy machine learning identification of fentanyl-relevant molecular compound classification via constituent functional group analysis

Mengyu Xu^{1,4}, Chun-Hung Wang², Anthony C. Terracciano^{3,4}, Artem E. Masunov^{2,5,6,7,8} & Subith S. Vasu^{3,4}✉

Fentanyl is an anesthetic with a high bioavailability and is the leading cause of drug overdose death in the U.S. Fentanyl and its derivatives have a low lethal dose and street drugs which contain such compounds may lead to death of the user and simultaneously pose hazards for first responders. Rapid identification methods of both known and emerging opioid fentanyl substances is crucial. In this effort, machine learning (ML) is applied in a systematic manner to identify fentanyl-related functional groups in such compounds based on their observed spectral properties. In our study, accurate infrared (IR) spectra of common organic molecules which contain functional groups that are constituents of fentanyl is determined by investigating the structure–property relationship. The average accuracy rate of correctly identifying the functional groups of interest is 92.5% on our testing data. All the IR spectra of 632 organic molecules are from National Institute of Standards and Technology (NIST) database as the training set and are assessed. Results from this work will provide Artificial Intelligence (AI) based tools and algorithms increased confidence, which serves as a basis to detect fentanyl and its derivatives.

Fentanyl (*N*-phenyl-*N*-[1-(2-phenylethyl)piperidin-4-yl]propionanilide)¹ is a synthetic opioid with extensive medical applications in acute analgesia. Fentanyl is an agonist for the μ -opioid receptor with higher potency than morphine, exhibiting an effective dose (ED₉₅) of 0.45–0.60 $\mu\text{g}/\text{kg}$ in dilation and curettage owe able in part to its 92% bioavailability^{2,3}. Once in the human body, metabolizing fentanyl begins with *N*-oxidative dealkylation and is eventually excreted within the urine^{4,5}. Fentanyl can be administered via inhalation, ingestion, oral exposure, injection, or transdermal means⁶. Fentanyls which are primarily powdered crystalline substances, typically white, obtained outside of a pharmaceutical setting may exist as a white powder form mixed with heroin and cutting agents, or pressed into counterfeit opioid prescription pills⁷. Fentanyl may even be weaponized and formed into an aerosol⁸. Such widespread use of fentanyl as either a primary compound chemical weapon or adulterating substance within other drugs poses a significant threat for officers the public at large, officers responding to incidents, and others given its fentanyls ample means of exposure⁹.

¹Statistics and Data Science, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816, USA. ²NanoScience Technology Center, University of Central Florida, 12424 Research Parkway, Orlando, FL 32826, USA. ³Mechanical and Aerospace Engineering, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816, USA. ⁴Center for Advanced Turbomachinery and Energy Research, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816, USA. ⁵School of Modeling, Simulation, and Training, University of Central Florida, 3100 Technology Parkway, Orlando, FL 32816, USA. ⁶Department of Chemistry, University of Central Florida, 4111 Libra Dr., Orlando, FL 32816, USA. ⁷South Ural State University, Lenin pr. 76, Chelyabinsk 454080, Russia. ⁸National Research Nuclear University MEPhI, Kashirskoye shosse 31, Moscow 115409, Russia. ✉email: subith@ucf.edu

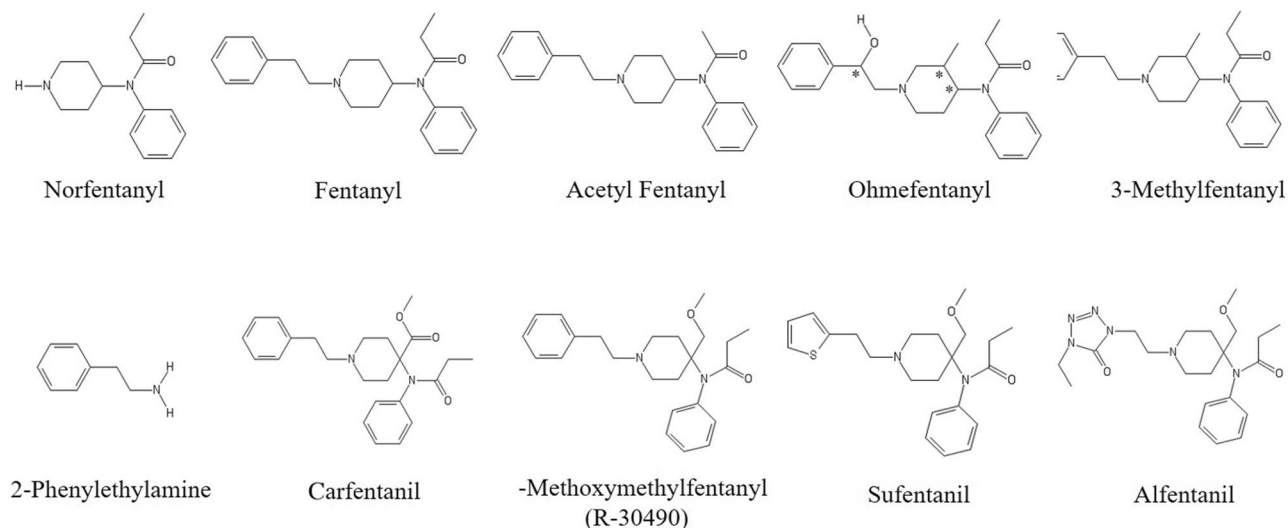


Figure 1. Fentanyl and selected analogues. * denotes chiral center of ohmfentanyl, which is a stereoisomer. Images generated using ChemDraw tool (PerkinElmer).

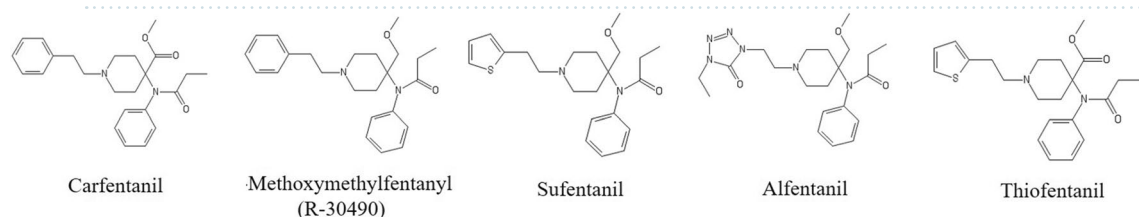


Figure 2. Carfentanil and selected analogues. Images generated using ChemDraw tool (PerkinElmer).

A comprehensive review of fentanyl and its analogs has been done in by Armenian et al.¹⁰. Carfentanil¹¹, sufentanil¹², alfentanil¹³, and remifentanil¹⁴ are fentanyl analogs with more potency than fentanyl for medicinal use. These and other fentanyls all exhibit some common key functional groups.

One promising method of identifying fentanyl and its analogues is through the fingerprinting of their unique infrared (IR) spectral properties (reflectivity, absorptivity, or transmissivity) as this technique can reveal characteristics of the numerous intramolecular bonds¹⁵. However, identifying fentanyls in practice is difficult. Clandestine production often exhibits variations in products between batches and as previously stated there are numerous formulations of fentanyl and carfentanil. As can be seen in Figs. 1 and 2, fentanyl, carfentanil, and their analogues exhibit varying functional groups and simultaneously may have conformers for the same composition which are denoted by a * at a particular bond. These structural variations can result in significantly different infrared spectral properties. Despite these differences, there are structural similarities enabling the compound to bind to receptors within the body and perform a similar function.

In analytical chemistry, IR spectroscopy tables¹⁶ had been widely used to identify compounds based on the correlations between IR absorption and common types of molecular bonds and functional groups. In this work, we aimed to answer the following question: can we identify constituent functional groups of fentanyl and its analogues from the IR absorption data?

Here we use a database containing 632 molecules, 591 of which have at least one of the functional groups found in the parent compound of fentanyl. The rescaled IR absorption spectra of these molecules are shown in Fig. 3. From these data we construct machine learning algorithms to identify molecules which are compromised from at least one of these functional groups.

A main challenge of learning from the spectral data is that the number of variables is the number of wavenumbers where absorptive data is obtained. The analysis suffers from a curse of dimensionality. There has been a growing body of literature in chemometrics that study the identification of compounds from their spectral properties via data-driven algorithms. In most of the work, spectral properties measured at each wavenumber is treated as a predictor, and widely-used classification or clustering machine learning methods has been applied for analysis directly or after dimension reduction with methods such as principal component analysis, see Refs.^{17–24} among others. In this work, we treat the underlying IR spectra as smooth functions and apply functional principal component analysis by approximating the data with a basis of a few orthogonal smoothed eigenfunction and then perform classification via functional generalized linear models. Comparing to the high-dimensional multivariate classification, the functional analysis is more interpretable. It associates the functional groups with features in the IR spectra in terms of the functional components along the whole range such as peaks and troughs. Results

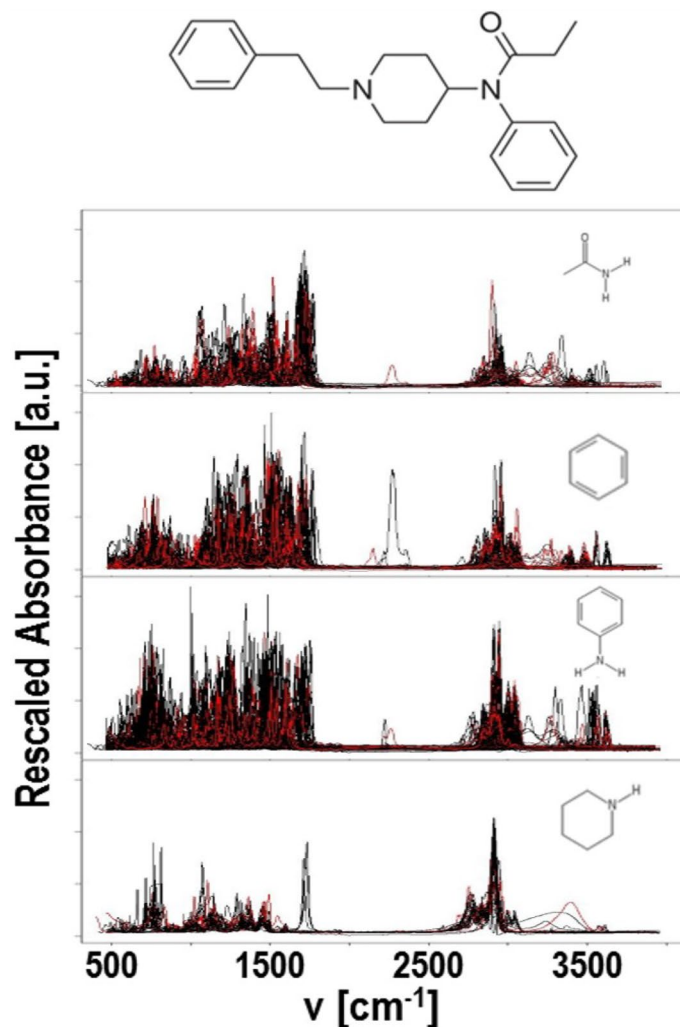


Figure 3. IR spectral curves of compounds with identified functional groups within fentanyl. Black: compounds in the training set; red: compounds in the test set. Image generated using ChemDraw tool (PerkinElmer).

from this work will provide Artificial Intelligence (AI) based tools and algorithms increased confidence, which serves as a basis to detect fentanyl and its derivatives.

Methodology

Let $\alpha_1, \alpha_2, \dots, \alpha_p$ be the wavenumbers at which absorbance are recorded, $v_i(\alpha)$ be the underlying absorbance at any wavenumber α for the i th molecule and $v_{i,j} = v_i(\alpha_j) + \epsilon_{i,j}$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$ be the observed absorbance. The goal of this study is to distinguish if a given molecule contains a certain functional group such as amide. Let $y_i = 1$ (molecule i contains the functional group), $i = 1, 2, \dots, n$ be the binary observations. We establish a classification model with $\{v_{i,j}\}_{1 \leq i \leq n, 1 \leq j \leq p}$ as the predictor and $\{y_i\}_{1 \leq i \leq n}$ as the response, via a functional logistic regression with functional principal component basis.

Dataset. The data set for both training and testing was obtained from the public database of National Institute of Standards and Technology (<https://www.nist.gov/>), which provides spectral information of the selected 632 compound molecules considered in this work. Such compounds can be categorized into one of the following eight groups: (i) amide only, (ii) aniline only, (iii) benzene only, (iv) piperidine only, (v) amide and aniline simultaneously, (vi) amide and benzene simultaneously (vii) distinct aniline and benzene simultaneously, and (viii) none of the above constituent functional groups. The group information of the data is listed in Table 1 with the names of the compounds listed in the Supplementary Appendix A.

For each molecule, the discretization of the IR spectra is recorded with unaligned maximum wavenumbers ranges of 243 cm^{-1} to $4,000.7 \text{ cm}^{-1}$ (7.28–119.94 THz), with the majority of the wavenumbers ranging from 500 to $4,000.7 \text{ cm}^{-1}$ (14.99–119.94 THz). Preprocessing of the discrete raw data consists of rescaling each absorbance function such that it has unit standard deviation and converting it to regularly spaced measurement via interpolation at the common wavenumbers $500\text{--}4,000 \text{ cm}^{-1}$ (14.99–119.92 THz) with step size 4 cm^{-1} (0.12 THz),

	Amide only	Aniline only	Benzene only	Piperidine only	Amide and aniline	Amide and Benzene	Aniline and Benzene	Benzene and Piperidine	None	Total
Training set	30	139	217	18	24	22	15	8	33	506
Testing set	7	35	54	5	6	5	4	2	8	126
Total	37	174	271	23	30	27	19	10	41	632

Table 1. Number of molecules with the functional groups.

yielding a total of $p = 876$ discrete points. For compounds with a slope in the IR spectra, we apply a linear baseline correction to remove the linear trend between the lowest point and the right end in the IR spectra. No subsequent smoothing of the data is performed in preprocessing. It is assumed that spectra obtained from the NIST database are close enough to standard temperature and pressure to neglect broadening and thermal effects. The preprocessed dataset is decomposed into distinct training and testing sets, consisting of 506 and 126 molecules, respectively, as shown in Table 1.

Functional principal component analysis (functional PCA). Denote $\bar{v}(\alpha) = n^{-1} \sum_{i=1}^n v_i(\alpha)$ as the mean function and $\sigma(s, t) = (n-1)^{-1} \sum_{i=1}^n [v_i(s) - \bar{v}(s)][v_i(t) - \bar{v}(t)]$ as the covariance function of the sample. Functional PCA decomposes the underlying absorbance functional $v_i(\alpha)$ as

$$v(\alpha) = \bar{v}(\alpha) + \sum_{k=1}^{\infty} f_{ik} \phi_k(\alpha), \quad i = 1, 2, \dots, n, \quad (1)$$

where $\phi_k(\alpha)$, $k = 1, 2, \dots, K$ are the functional principal components (PCs), and f_{ik} is the score of molecule i for principal component k . Cutting off at a finite integer K and estimate $v_i(\alpha)$ according to

$$\hat{v}_i(\alpha) = \bar{v}(\alpha) + \sum_{k=1}^K f_{ik} \phi_k(\alpha), \quad i = 1, 2, \dots, n. \quad (2)$$

The functional PCs are the eigenfunctions of the covariance function, $\sigma(s, t)$, where $\phi_k(\alpha)$ corresponds to the k th largest eigenvalue λ_k . The corresponding PC scores are $f_{ik} = \int \phi_k(t)[v_i(t) - \bar{v}(t)] dt$, $k = 1, 2, \dots, K$. The number of PCs K can be selected such that the majority, e.g. 90%, of the variance is explained by (2).

The functional PCs are orthonormal, i.e., $\int \phi_j(\alpha) \phi_k(\alpha) d\alpha = 0$ if $j \neq k$ and 1 if $j = k$, and they explain the most proportion of variance in the data in descending order. In other words, the first PC score f_{i1} , $i = 1, 2, \dots, n$ maximizes the sample variance among the inner products between the data and all functionals ξ 's subject to the constraint $\int \xi(\alpha)^2 d\alpha = 1$; the second PC score f_{i2} , $i = 1, 2, \dots, n$ maximizes the remaining variance among the inner products of the data and all functionals ξ 's subject to the constraint $\int \xi(\alpha)^2 d\alpha = 1$ and $\int \xi(\alpha) \phi_1(\alpha) d\alpha = 0$, and so on. The proportion of variance that is explained by f_{ik} , $1 \leq i \leq n$ is $\lambda_k / \sum_{k=1}^{\infty} \lambda_k$. In the estimation, we regularize the functional PCs and impose roughness penalty to the sample covariance function, measured by the norm of second-order derivative of the function. Details of functional PCA can be found in Ref.²⁵.

Functional generalized linear model (functional GLM). Assume the binary response $y_i \sim \text{Bernoulli}(p_i)$, where p_i is the probability that the molecule i contains the functional group. Denote the logit function of p_i as $l_i = \log\left(\frac{p_i}{1-p_i}\right)$ and assume that l_i is a linear function of $v_i(\alpha)$. That is

$$l_i = \beta_0 + \int \beta(\alpha) [\hat{v}_i(\alpha) - \bar{v}(\alpha)] d\alpha,$$

where the intercept β_0 and the functional parameter $\beta(\alpha)$ are unknown quantity and coefficient functionals of interest. Represent $\beta(\alpha) = \sum_{k=1}^{\infty} \beta_k \phi_k(\alpha)$. According to (1), the above linear model can be written as

$$l_i = \beta_0 + \sum_{k=1}^K \beta_k f_{ik}, \quad (3)$$

where the unknown coefficients $\beta_0, \beta_1, \dots, \beta_K$ can be estimated by the maximum likelihood estimation (MLE) for the generalized linear model²⁶. The above logistic regression can be extended to multiclass cases. Details are omitted and interested readers are referred to Ref.²⁷.

For a new molecule with IR spectrum $v^*(\alpha)$, the probability p^* can be estimated by $\hat{p}^* = \frac{\exp(\hat{l}^*)}{1 + \exp(\hat{l}^*)}$, where $\hat{l}^* = \hat{\beta}_0 + \sum_{k=1}^K \hat{\beta}_k f_{ik}^*$, where $\hat{\beta}_k$ is the MLE of β_k , and f_{ik}^* are the PC scores of the new molecule, $k = 1, 2, \dots, K$. If \hat{p}^* exceeds a given threshold such as 0.5, the molecule is classified as having the functional group.

Comparing to the high-dimensional multivariate classification with $p = 876$ predictors, the functional GLM represents the absorbance by a small number of orthogonal basis functions enabling statistical and computational efficiency. More importantly, the GLM associates the functional groups with features in the IR spectra in terms of the functional components along the whole range, while the multivariate classification treats absorbance at each wavenumber as a predictor and trains the model by assuming conditions such as sparsity.

	Ground truth		
		True	False
Prediction	True	True positive (TP)	False positive (FP)
	False	False negative (FN)	True negative (TN)

Table 2. Confusion matrix of a classification problem.

Classification performance metrics. In evaluating the performance of classification results we obtain the confusion matrix as shown in Table 2 by comparing the predicted labels against the ground truth from the test set.

Let

$$FPR = \frac{FP}{FP + TN}, \quad FNR = \frac{FN}{FN + TP}.$$

The Receiver Operating Characteristic Curve (ROC) is a curve plotting $1-FNR$ (a.k.a., sensitivity) against the FPR (a.k.a., 1-specificity). For each classification model we report the Area Under the Curve (AUC) and use AUC as an evaluation metric of the algorithm performance.

Results and discussion

In this study, we train functional GLM models and evaluate its capability to fingerprint the functional groups including amide, aniline, benzene and piperidine in the IR spectrum of a molecule. Parameters for the GLM are estimated from the IR spectral data and functional groups information of the training set to train the model. The constructed model is then used to predict the functional groups from only IR spectra data of a testing dataset. We evaluate the accuracy of the models by comparing the prediction with the ground truth in the testing set. The functional data analysis is implemented using the R package *fda.usc*²⁸ and the ROC curve and AUC are obtained from R package *pROC*²⁹.

We investigate two scenarios. Scenario one considers potential interactions which may arise from the simultaneous existence of multiple functional groups reflected by distinct IR spectral patterns. We classify the testing molecules into 8 nonoverlapping groups. Whereas in scenario two, the model is trained to predict if a molecule contains the constituent functional groups, where compounds with simultaneous presence of multiple functional groups is also counted.

Feature representation: functional PCA. We first construct the regularized functional PCs basis from the training set. The mean function and the leading PCs are presented in Fig. 4.

From Fig. 4, one can see the following. First, as shown in Fig. 3, the IR spectral data in each group is very diverse and has various patterns. Accordingly, in panel (c) of Fig. 4, the proportion of variance explained by the functional PCs increases slowly as more PCs are considered, where the first functional PC explains nearly 20% of the overall variance, with the second describing 10%. Over 80% of the total variance may be considered with the first 22 PCs. In addition, in panel (d) to (f) we present the leading 4 PC scores of compounds with no more than one functional groups of interest. The first four PC scores alone does not partition the four groups; however, we can observe a general trend in which, compounds with amide tend to have small scores of PC1, PC2 and PC4, and high scores in PC 3; anilines tend to have high scores in PC2, while piperidine containing compounds are clustered within a narrow range of PC2 and PC4 with high PC1 scores.

Prediction for appearance of functional groups from IR spectra. Four functional GLMs are constructed to predict the following responses from the IR spectra data and are assigned as a specific model respectively.

- Model 1 whether the molecule has an *amide* functional group.
- Model 2 whether the molecule has an *aniline* functional group.
- Model 3 whether the molecule has a *benzene* functional group.
- Model 4 whether the molecule has a *piperidine* functional group.

In constructing the functional GLM models, we take $K = 22$, which is the minimum number of PCs such that at least 80% of variance is explained. In Model 4, because the response is imbalanced with only 10 piperidine out of 478 training molecules, we perform an oversampling adjustment before estimating the model. That is, the 10 observations in the training set are resampled and reused such that there are 100 molecules in the training set are piperidine. The ROC curves for Model 1 to Model 4 are presented in the area under the ROC curve on Fig. 5 are shown in the Table 5.

We present the confusion matrices from the testing set of Models 1–4 (AUCs presented in Table 3) in Tables 4, 5, 6, 7 respectively, with a threshold probability of 0.5. The rates of correctly identifying the corresponding functional groups are 96.03%, 90.48%, 87.3%, and 96.03% respectively, with an average accuracy rate of 92.5%.

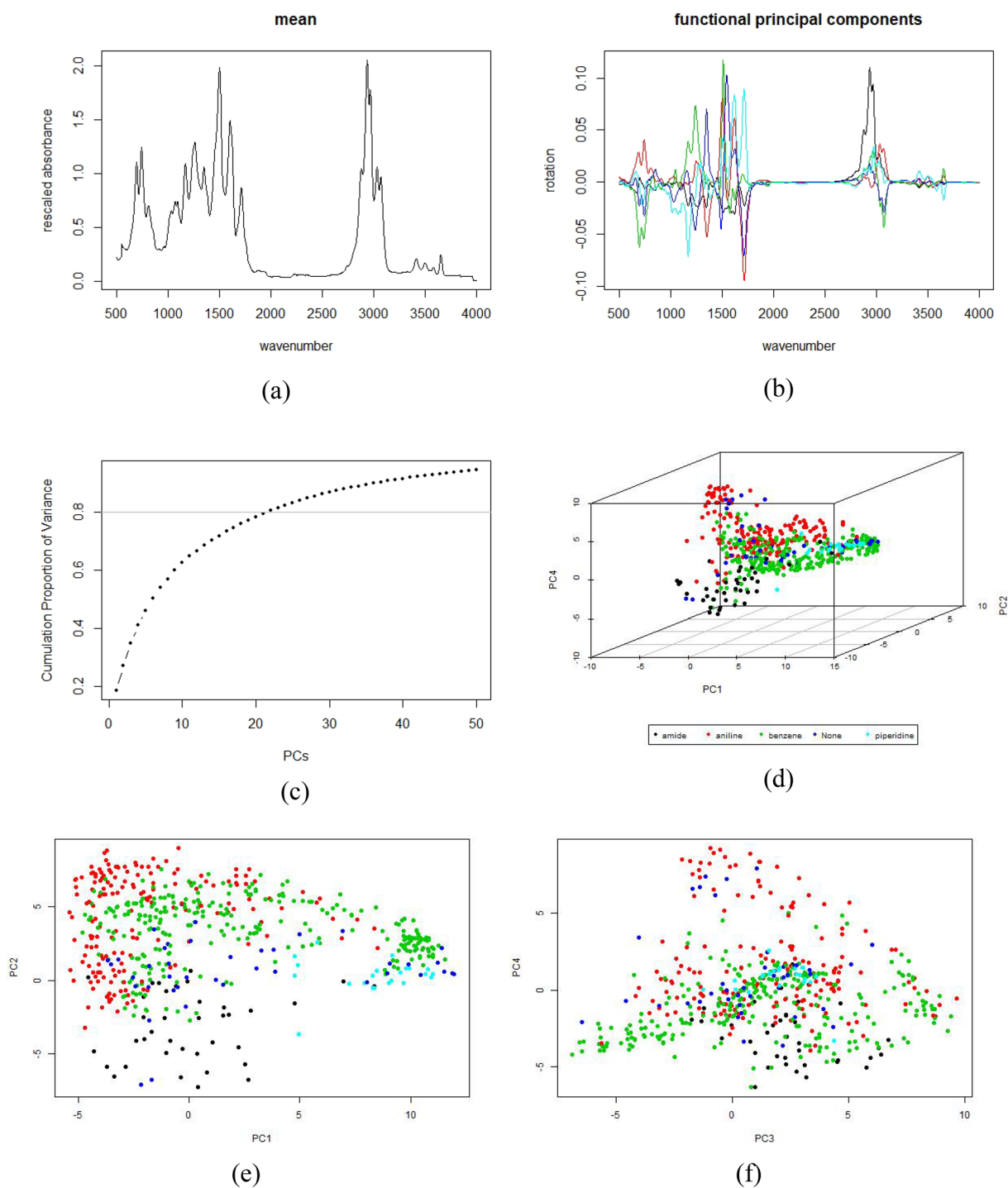


Figure 4. Functional PCA results: (a) the mean function of the rescaled absorbance; (b) the first five functional principal components; (c) the cumulative proportion of variance explained by the first 50 functional PCs; (d) the 3D plot of PC1, PC2 and PC4 scores of the compounds consisting of at most one functional groups of interest; (e) PC1 and PC2 scores of the compounds consisting of at most one functional groups of interest; (f) PC1 and PC4 scores of the compounds consisting of at most one functional groups of interest.

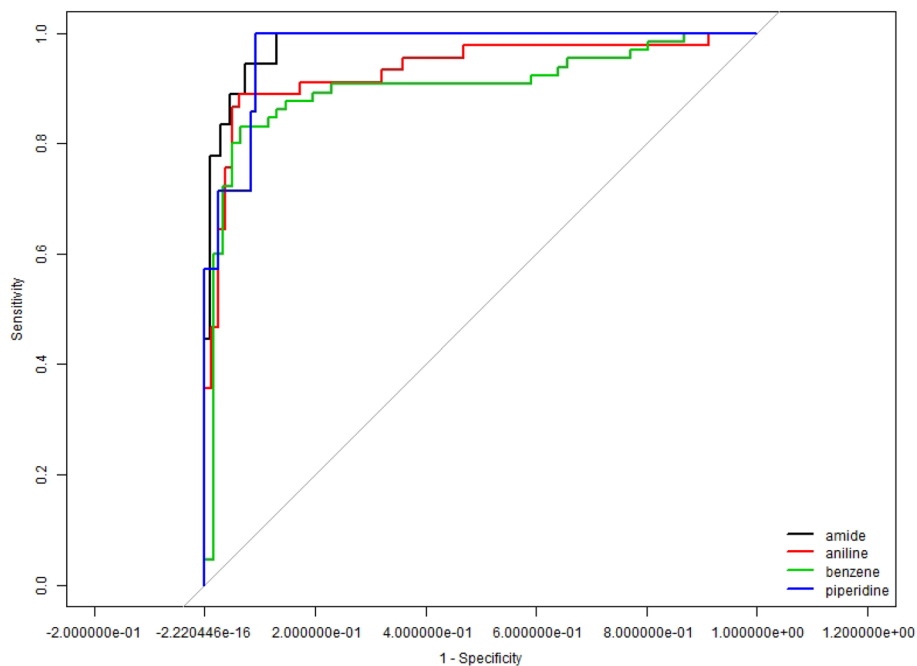


Figure 5. ROC curves for models 1.1 to 1.4.

	1—Amide	2—Aniline	3—Benzene	4—Piperidine
AUC	0.9815	0.9336	0.9019	0.9712

Table 3. AUC of models 1–4.

1—Amide		Prediction	
		False	True
Ground truth	False	107	1
	True	4	14

Table 4. Confusion matrix for Model 1—predicting appearance of amide.

2—Aniline		Prediction	
		False	True
Ground truth	False	74	7
	True	5	40

Table 5. Confusion matrix for Model 2—predicting appearance of aniline.

3—Benzene		Prediction	
		False	True
Ground truth	False	58	6
	True	13	50

Table 6. Confusion matrix for Model 3—predicting appearance of benzene.

4—Piperidine		Prediction	
		False	True
Ground truth	False	115	4
	True	0	8

Table 7. Confusion matrix for Model 4—predicting appearance of piperidine.

		Prediction								
		Amide only	Aniline only	Benzene only	Piperidine only	Amide and aniline	Amide and benzene	Aniline and benzene	Benzene and piperidine	None
Ground truth	Amide only	5	0	0	1	1	0	0	0	0
	Aniline only	1	32	1	0	0	0	1	0	0
	Benzene only	0	3	49	0	0	0	1	0	1
	Piperidine only	0	0	0	4	0	0	0	0	1
	Amide and aniline	1	0	0	0	3	1	0	0	1
	Amide and benzene	1	1	0	0	1	1	0	0	1
	Aniline and benzene	0	3	0	0	0	0	1	0	0
	Benzene and piperidine	0	0	1	1	0	0	0	0	0
	None	0	0	1	2	0	0	0	0	5

Table 8. Confusion matrix of the classification in the testing set.

Joint classification into non-overlapping groups.. We combine the four binary classification results described in Tables 4, 5, 6, 7 and group the data as shown in Table 1. The confusion matrix obtained from prediction in the 126 testing molecules is summarized in Table 8.

From Table 8 one can find that, out of the 126 predicted molecules, 100 are classified exactly correctly, with an accuracy rate of 79.37%. Among the misclassified 26 molecules, 12 are classified with partial mistakes by incorrectly including or excluding a functional group while correctly identify another, 4 are misclassified by missing the correct functional group, 3 is misclassified by identifying a wrong functional group and 7 are misclassified by both missing the truth and identify a false. The misclassified molecules are listed in Table 9.

Table 9 shows a relatively high error rate in distinguishing simultaneously appeared functional groups from the cases where only one of the functional groups of interest is present. The algorithm also has a relatively high error rate in distinguishing the aniline group and benzene, due to fact that these functional groups share common structure.

Summary

Today, synthetic opioid analogues such as, fentanyl, have been the cause of many accidental deaths across the world, and thus the detection of low concentrations of these harmful substances at a distance via spectroscopic techniques is crucial for law enforcement. Unfortunately, new fentanyl related compounds are being created and distributed frequently with slight modifications to the functional groups present in them. Therefore, machine learning based intelligent detection schemes must be employed for intelligent detection of such molecules. Here, we applied a functional generalized linear model with smoothed functional principal component basis to classify functional groups of molecules from their IR absorption data. The result serves as a basis of the identification of fentanyl and its derivatives, which could be accomplished in future. This effort demonstrated the efficacy of a functional data analysis to identify molecules containing one or more specific functional groups from their infrared absorption spectra. The accuracy rate of classification into the 9 distinct classes is 79.4%. The average rate of accurately identifying the four functional groups is 92.5%. Continued efforts on this model will seek to also expand the utility from gas phase compound analysis to solid phase, with spectral properties of fentanyl and its analogues included. Such expansions will require additional consideration.

Compound	Ground truth	Prediction
1-Adamantanecarboxamide, N,N-dimethyl-,	Amide only	Piperidine only
Acetamide, N-butyl-	Amide only	Amide and aniline
Azobenzene	Aniline only	Benzene
4-Amino-2,6-dichlorophenol	Aniline only	Aniline and Benzene
1,3-Diphenylguanidine	Aniline only	Amide only
Benzene, pentafluoromethyl-	Benzene only	Aniline only
Benzene, 1-fluoro-3-methyl-	Benzene only	Aniline and Benzene
Diethyl Phthalate	Benzene only	None
Phenol, 2,6-dimethoxy-	Benzene only	Aniline only
Benzene, 1,3-difluoro-	Benzene only	Aniline only
Cyclohexanamine	None	Piperidine only
Nitrous oxide	None	Benzene only
Triethanolamine	None	Piperidine only
4-Piperidinol, 2,2,6,6-tetramethyl-, 1-oxide	Piperidine only	None
4-Benzylpiperidine	Benzene and piperidine	Benzene only
4-(4-Chlorophenyl)-4-hydroxypiperidine	Benzene and piperidine	Piperidine only
Metazachlor	Amide and aniline	Amide and benzene
Acetamide, N-(2,6-dimethylphenyl)-	Amide and aniline	Amide only
4'-Bromoacetanilide	Amide and aniline	None
Benzamide	Amide and benzene	Amide only
Benzhydroxamic acid	Amide and Benzene	None
Dibucaine	amide and Benzene	Aniline only
Acetanilide, 4'-fluoro-	Amide and benzene	Amide and aniline
p-Butoxybenzylidene p-butylaniline	Aniline and benzene	Aniline only
p-Hexyloxybenzylidene p-butylaniline	Aniline and benzene	Aniline only
2(o-Aminophenyl)-benzimidazole	Aniline and benzene	Aniline only

Table 9. Misclassified compounds and the classification errors.

Received: 17 March 2020; Accepted: 28 July 2020

Published online: 11 August 2020

References

- Ogawa, N., Nagase, H., Endo, T., Loftsson, T. & Ueda, H. Crystal structure of fentanyl base. *X-Ray Struct. Anal. Online* **25**, 83–84 (2009).
- Uğur, B., Oğurlu, M., Yılmaz, S. & Kıvrak, V. Determining the optimal fentanyl dose for dilation and curettage procedures. *Clin. Exp. Obstet. Gynecol.* **39**, 509–511 (2012).
- Labroo, R. B., Paine, M. F., Thummel, K. E. & Kharasch, E. D. Fentanyl metabolism by human hepatic and intestinal cytochrome P450 3A4: implications for interindividual variability in disposition, efficacy, and drug interactions. *Drug Metab. Dispos.* **25**, 1072–1080 (1997).
- Tateishi, T. *et al.* Identification of human liver cytochrome P-450 3A4 as the enzyme responsible for fentanyl and sufentanil N-dealkylation. *Anesth. Analg.* **82**, 167–172 (1996).
- Poklis, A. & Backer, R. Urine concentrations of fentanyl and norfentanyl during application of Duragesic transdermal patches. *J. Anal. Toxicol.* **28**, 422–425 (2004).
- Marquardt, K. A. & Steven Tharratt, R. Inhalation abuse of fentanyl patch. *J. Toxicol. Clin. Toxicol.* **32**, 75–78. <https://doi.org/10.3109/15563659409000433> (1994).
- Bucerius, S. M. & Haggerty, K. D. Fentanyl behind bars: the implications of synthetic opiates on prisoners and correctional officers. *Int. J. Drug Policy* **71**, 133–138 (2019).
- Armenian, P., Vo, K. T., Barr-Walker, J. & Lynch, K. L. Fentanyl, fentanyl analogs and novel synthetic opioids: a comprehensive review. *Neuropharmacology* **134**, 121–132 (2018).
- Stanley, T. H., Egan, T. D. & Van Aken, H. A tribute to Dr. Paul A. J. Janssen: entrepreneur extraordinaire, innovative scientist, and significant contributor to anesthesiology. *Anesth. Analg.* **106**, 451–462 (2008).
- Peeters, O. M., Blaton, N. M. & De Ranter, C. J. N-{4-(Methoxymethyl)-1-[2-(2-thienyl)ethyl]-4-piperidinyl}-N-phenylpropanamide: sufentanil. *Acta Crystallogr. Sect. B* **35**, 999–1001 (1979).
- Durant, F., Michel, A., Lebrun, B. & Evrard, G. X-ray crystal structure of alfentanil (R39209), an analgesic of the fentanyl family. *Bull. Soc. Chim. Belg.* **96**, 331–332 (1987).
- Bürkle, H., Dunbar, S. & Van Aken, H. Remifentanyl: a novel, short-acting, μ -opioid. *Anesth. Analg.* **83**, 646–651 (1996).
- Johnson, H. E. *et al.* High-throughput metabolic fingerprinting of legume silage fermentations via Fourier transform infrared spectroscopy and chemometrics. *Appl. Environ. Microbiol.* **70**, 1583–1592 (2004).
- Socrates, G. *Infrared and Raman Characteristic Group Frequencies: Tables and Charts* (Wiley, New York, 2004).
- Howley, T., Madden, M. G., O'Connell, M.-L. & Ryder, A. G. *International Conference on Innovative Techniques and Applications of Artificial Intelligence* 209–222 (Springer, New York, 2005).
- Markey, M. K., Tourassi, G. D. & Floyd, C. E. Jr. Decision tree classification of proteins identified by mass spectrometry of blood serum samples from people with and without lung cancer. *Proteomics* **3**, 1678–1679 (2003).

17. Yang, H., Griffiths, P. R. & Tate, J. Comparison of partial least squares regression and multi-layer neural networks for quantification of nonlinear systems and application to gas phase Fourier transform infrared spectra. *Anal. Chim. Acta* **489**, 125–136 (2003).
18. Zou, T., Dou, Y., Mi, H., Zou, J. & Ren, Y. Support vector regression for determination of component of compound oxytetracycline powder on near-infrared spectroscopy. *Anal. Biochem.* **355**, 1 (2006).
19. Madden, M. G. & Ryder, A. G. Machine learning methods for quantitative analysis of Raman spectroscopy data. *Opto-Ireland 2002 Opt. Photon. Technol. Appl.* **4876**, 1130–1139 (2003).
20. 20Ratle, F., Terrettaz, A.-L., Kanevski, M. F., Esseiva, P. & Ribaux, O. Pattern analysis in illicit heroin seizures: a novel application of machine learning algorithms. *ESANN* 665–670 (2006).
21. Ryder, A. G. Classification of narcotics in solid mixtures using principal component analysis and Raman spectroscopy. *J. Forensic Sci.* **47**, 275–284 (2002).
22. Cadet, X. F., Lo-Thong, O., Bureau, S., Dehak, R. & Bessafi, M. Use of machine learning and infrared spectra for rheological characterization and application to the apricot. *Sci. Rep.* **9**, 1–12 (2019).
23. Chu, P. M., Guenther, F. R., Rhoderick, G. C. & Lafferty, W. J. The NIST quantitative infrared database. *J. Res. Nat. Inst. Stand. Technol.* **104**, 59 (1999).
24. Terracciano, A. C. *et al.* Hazardous gas detection sensor using broadband light-emitting diode-based absorption spectroscopy for space applications. *New Space* **6**, 28–36 (2018).
25. Ramsay, J. O. & Silverman, B. W. *Applied Functional Data Analysis: Methods and Case Studies* 147–186 (Springer, New York, 2007).
26. McCullagh, P. & Nelder, J. *Generalized Linear Models* 2nd edn, 98–148 (Chapman and Hall, London, 1989).
27. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* 119–127 (Springer Science and Business Media, New York, 2009).
28. Febrero Bande, M. & Oviedo de la Fuente, M. Statistical computing in functional data analysis: the R package fda. *usc. J. Stat. Softw.* **51**, 1–28 (2012).
29. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **12**, 77 (2011).

Acknowledgements

The project or effort depicted was or is sponsored by the Defense Advanced Research Projects Agency (DARPA) Agreement Number: D18AP00040. The readers should note that the content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

Author contributions

S.S.V., A.C.T., and M.X. designed the research. M.X., C.-H.W., A.C.T., A.E.M., and S.S.V. conducted research and wrote the paper. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-70471-7>.

Correspondence and requests for materials should be addressed to S.S.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020