Supporting Information for: Instrument-independent chemometric models for rapid, calibration-free NPS isomer differentiation from mass spectral GC-MS data

Jennifer L. Bonetti,[*,†,‡] Ruben F. Kranenburg,[†,¶] Esmee Schoonderwoerd,[†] Saer Samanipour,[†] and Arian C. van Asten[†,§]

†*Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, P.O. Box 94157, Amsterdam 1090 GD, The Netherlands*

‡*Virginia Department of Forensic Science, Norfolk, VA, 23606, USA*

¶*Dutch National Police, Unit Amsterdam, Forensic Laboratory, Kabelweg 25, Amsterdam, 1014 BA, the Netherlands*

§*Co van Ledden Hulsebosch Center (CLHC), Amsterdam Center for Forensic Science and Medicine, 1098 XH Amsterdam, The Netherlands*

E-mail: J.L.Bonetti@uva.nl

List of Figures

List of Tables

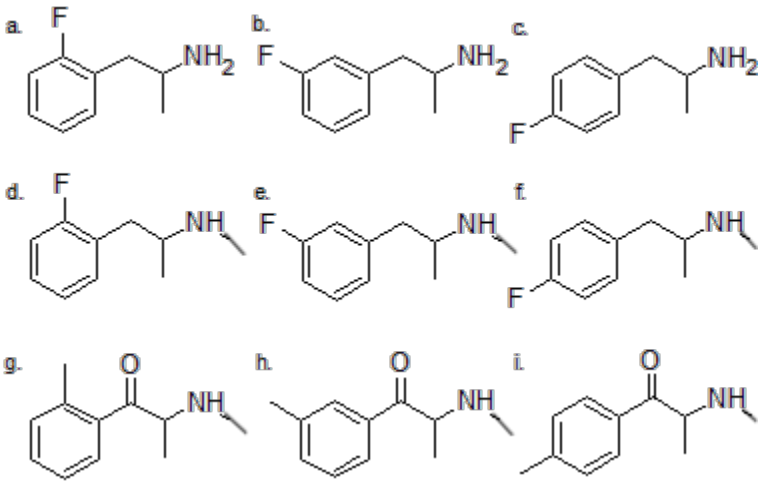

Figure S1: Structures of isomers studied. a-c: 2-, 3-, 4- fluoroamphetamine, respectively, d-f: 2-, 3-, 4- fluoromethamphetamine, respectively, and g-i: 2-, 3-, 4-, methylmethcathinone, respectively.

Table S2 shows all considered preprocessing methods. For tune standardization, two options for applying the step were "mean" or "one" which both reference how any data generated under standard tune conditions were to be adjusted. These data were adjusted by subtracting either by the mean abundance difference between data generated under both tune conditions for the same compound on the same day or by a representative single difference. The "multiply" method was applied by taking the ratio between the average autotuned data and standard tune data from the same day and multiplying each standard tune spectra by this ratio.

For MS selection, all combinations of front/back percentage removal were tested when filter = null. All other filter/threshold combinations were tested separately wherein a m/z was kept if the threshold (set at % of base peak abundance) was reached by the set filter percentage of samples. For normalization, alternative methods of spectrum normalization were investigated. "Ion current" refers to dividing the abundance value for each remaining m/z value by the sum of the abundance values of all m/z values remaining after the MS selection step. "Unit vector" refers to treating the remaining spectrum as a vector and dividing each abundance value by the length of the vector (square root of the sum of the

Table S1: Design matrix

Experiment	Tune Standardization	MS Selection	Normalization	Scaling
1	+	+	+	+
2	+	+	+	-
3	+	+	-	+
4	+	+	-	-
5	+	-	+	+
6	+	-	+	-
7	+	-	-	+
8	+	-	-	-
9	-	+	+	+
10	-	+	+	-
11	-	+	-	+
12	-	+	-	-
13	-	-	+	+
14	-	-	+	-
15	-	-	-	+
16	-	-	-	-

Table S2: All considered preprocessing methods. *9999 normalization performed before DoE assessment. Unit vector normalization refers to treating each spectrum as a vector and normalizing to a vector of length = 1. Unit vector scaling refers to treating each m/z peak as a vector and normalization to a vector of length = 1.

Tune Standardization	MS Selection	Normalization	Scaling
None	None	9999*	none
Mean	Low m/z (0, 5, 10, 15, 20, 25%)	Ion Current	Mean centering
Subtract	High m/z (0, 5, 10, 15, 20, 25%)	Unit vector	Standard
Multiply	Filter (null, 10%, 20%, 50%, 90%)	New base peak	Pareto
	Threshold (0, 1, 5%)	Remove base peak/ Ion current	Vast
		Remove base peak/ Unit vector	Level
		New base peak/ Ion current	MinMax
		New base peak/ Unit vector	Robust
			Quantile
			Unit vector
			Power

squares of the abundance values). In addition, removing the base peak and assigning a new base peak (with normalization to 9999 abundance) was investigated as well as following the base peak removal and/or new base peak assignment with ion current or unit vector normalization.

Multiple scaling options were attempted based on the available options from the Scikit-learn package in Python.[1] Pareto, vast, and level scaling were performed as discussed by van den Berg et al..[2]

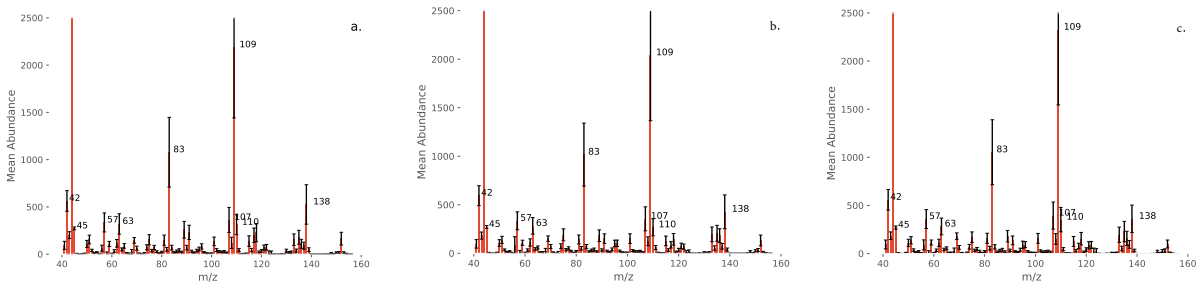

Figure S2: Average m/z spectra for Fluoroamphetamine isomers. a. 2-FA, b. 3-FA, c. 4-FA. Error bars represent one standard deviation. Spectra are zoomed in to show detail, so the base peak at an abundance of 9999 is not seen in full.

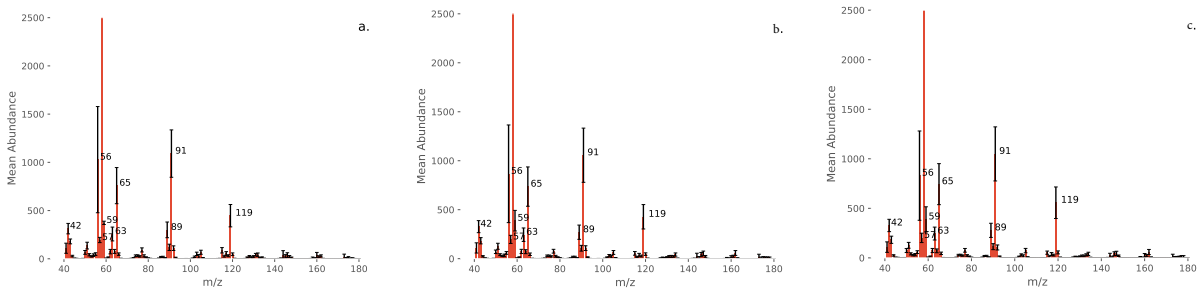

Figure S3: Average m/z spectra for Methylmethcathinone isomers. a. 2-MMC, b. 3-MMC, c. 4-MMC. Error bars represent one standard deviation. Spectra are zoomed in to show detail, so the base peak at an abundance of 9999 is not seen in full.

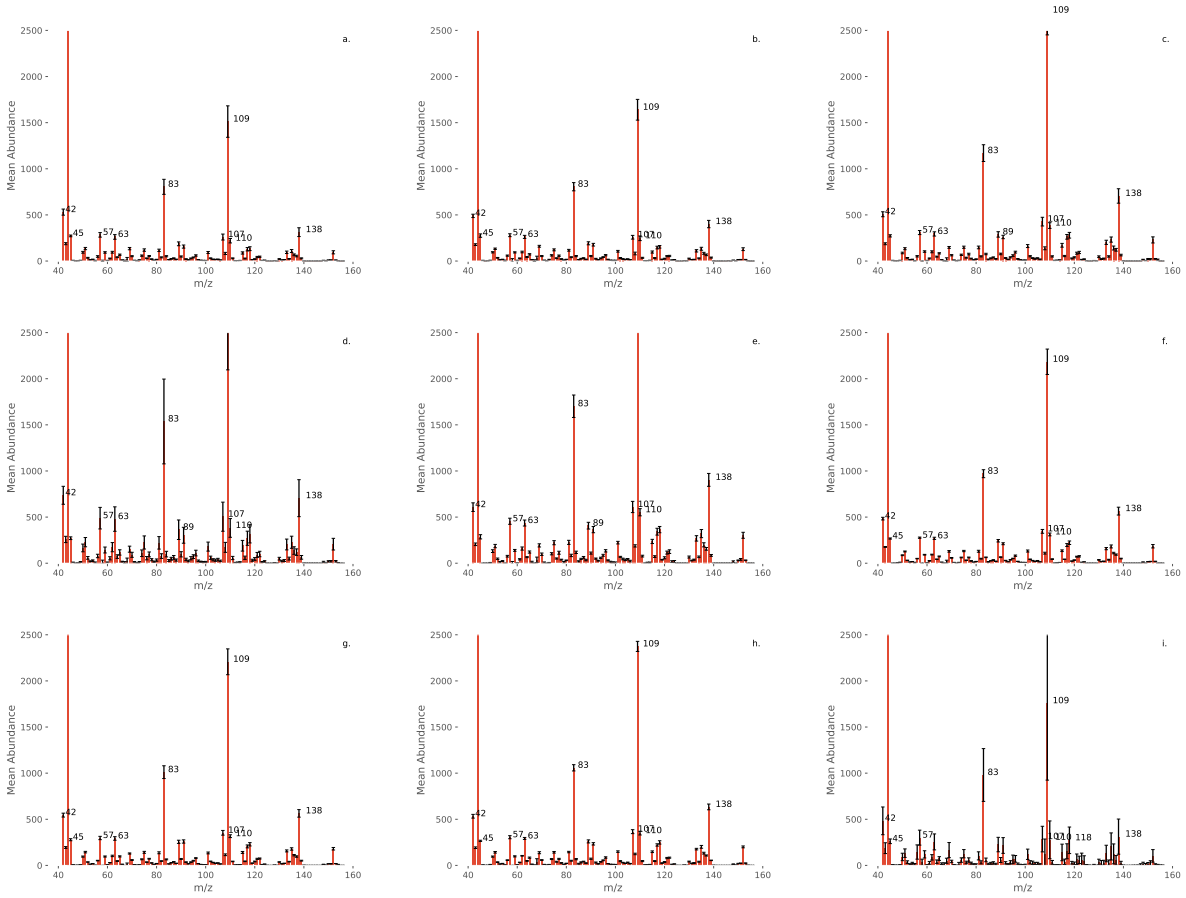

Figure S4: Average m/z spectra for Fluoroamphetamine isomers per instrument. a. Instrument 2, b. Instrument 3, c. Instrument B, d. Instrument C, e. Instrument D, f. Instrument F, g. Instrument G, h. Instrument H, i. Instrument U. Error bars represent one standard deviation. Spectra are zoomed in to show detail, so the base peak at an abundance of 9999 is not seen in full.

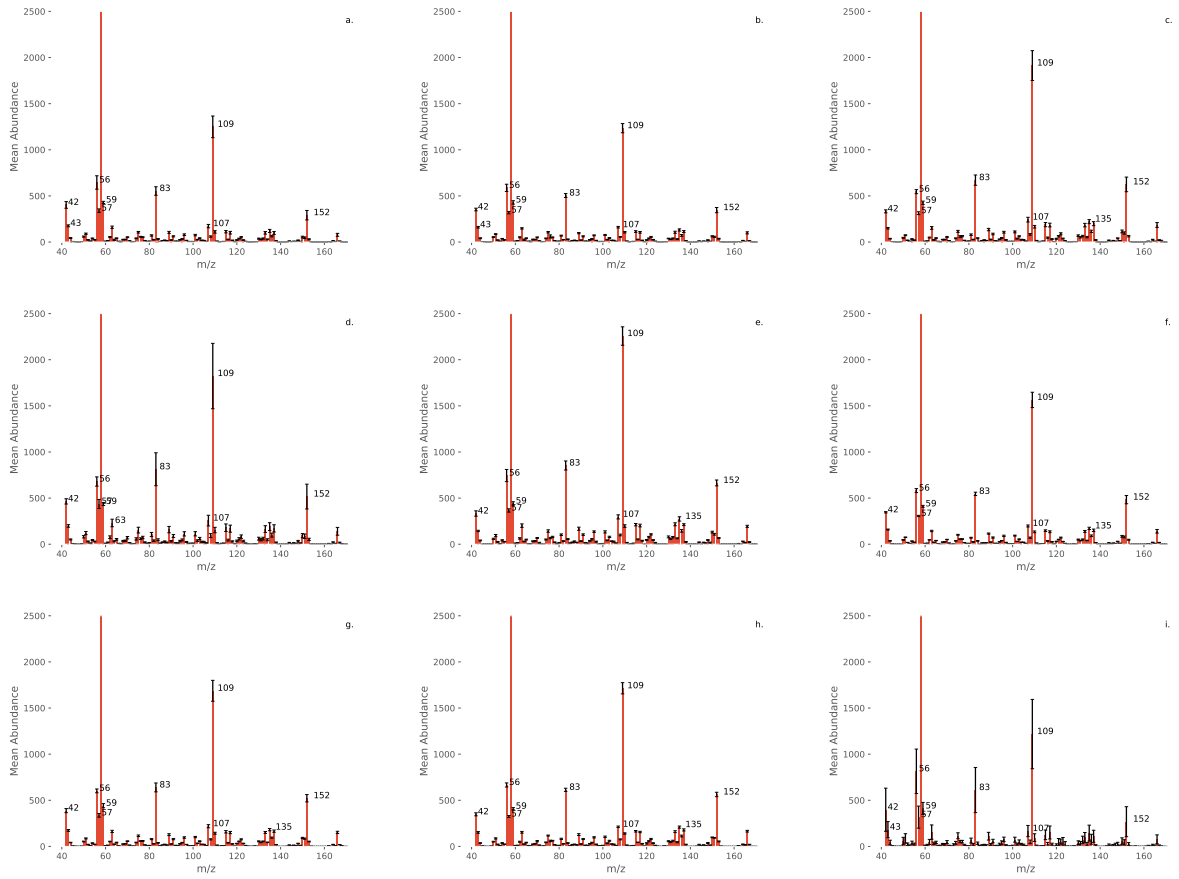

Figure S5: Average m/z spectra for 2-Fluoromethamphetamine isomers per instrument. a. Instrument 2, b. Instrument 3, c. Instrument B, d. Instrument C, e. Instrument D, f. Instrument F, g. Instrument G, h. Instrument H, i. Instrument U. Error bars represent one standard deviation. Spectra are zoomed in to show detail, so the base peak at an abundance of 9999 is not seen in full.

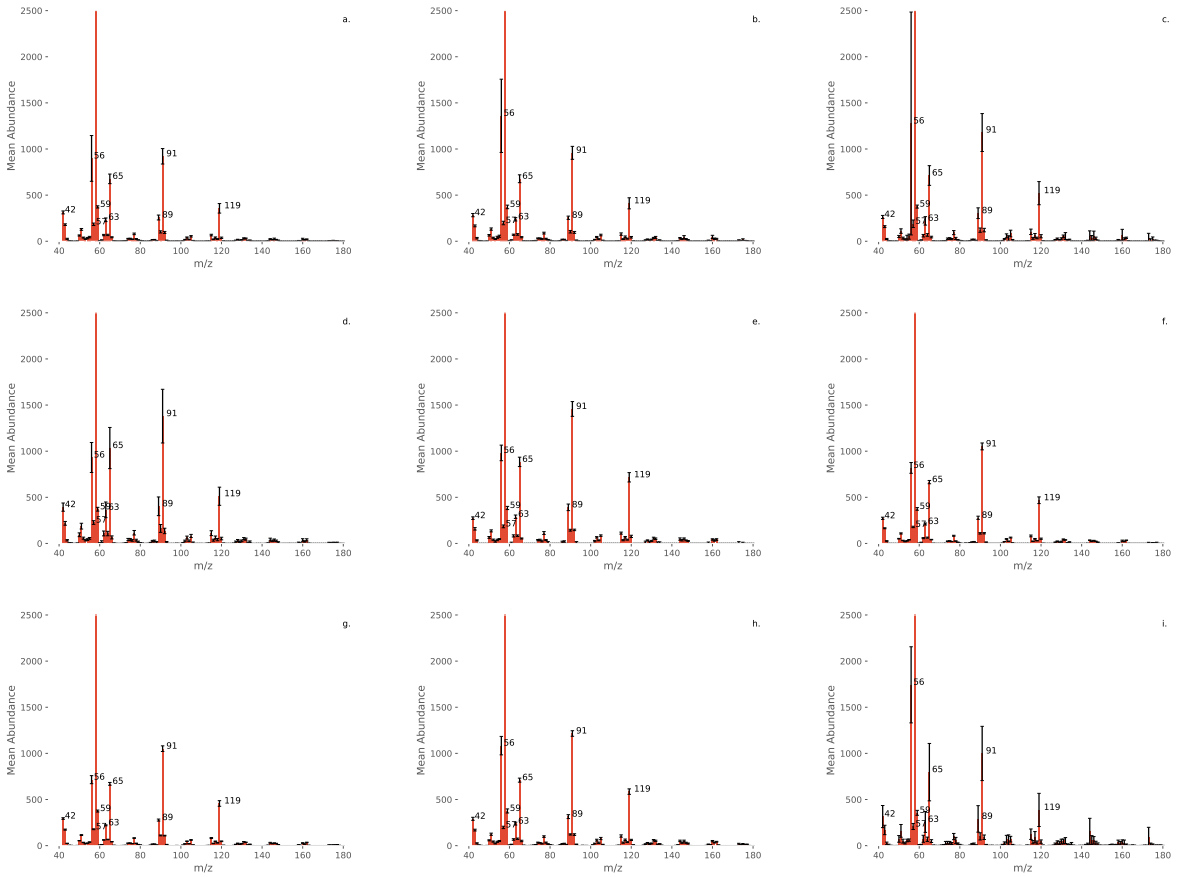

Figure S6: Average m/z spectra for 2-Methylmethcathinone isomers per instrument. a. Instrument 2, b. Instrument 3, c. Instrument B, d. Instrument C, e. Instrument D, f. Instrument F, g. Instrument G, h. Instrument H, i. Instrument U. Error bars represent one standard deviation. Spectra are zoomed in to show detail, so the base peak at an abundance of 9999 is not seen in full. The $m/z = 56$ peak shows abundant variation, likely due to the well-characterized potential for thermal oxidative degradation to the enamine or imine species which can often coelute with the primary compound. This variation is most pronounced in the older instruments (B and U).[3–5]

Table S3: Total Welch and Levene Test results. Percentage of comparisons found to be not significantly different (p>0.05) when comparing mean (Welch) and variance (Levene) of abundance values from one instrument to the total dataset. Comparisons were performed for each *m/z* fragment and each isomer identity (ortho/meta/para) was compared separately.

	FA		FMA		MMC	
Instrument	Welch	Levene	Welch	Levene	Welch	Levene
2	11.3%	13.9%	19.9%	13.7%	27.6%	35.3%
3	18.0%	24.1%	19.9%	24.5%	50.1%	61.9%
B	30.1%	27.5%	24.5%	35.9%	39.8%	75.1%
C	16.2%	52.5%	22.7%	58.9%	21.8%	61.4%
D	18.0%	61.7%	16.5%	57.6%	19.9%	75.5%
F	39.1%	49.0%	44.2%	37.2%	27.6%	69.8%
G	44.3%	14.2%	35.4%	14.0%	24.5%	35.7%
H	53.3%	68.7%	52.7%	63.3%	42.7%	85.6%
U	57.7%	49.0%	44.4%	48.8%	33.6%	32.9%
Total	40.1%	40.1%	31.2%	39.3%	31.9%	59.2%

Table S4: Total Welch and Levene Test results after preprocessing. Percentage of comparisons found to be not significantly different (p>0.05) when comparing mean (Welch) and variance (Levene) of abundance values from one instrument to the total dataset. Comparisons were performed for each *m/z* fragment and each isomer identity (ortho/meta/para) was compared separately.

	FA		FMA		MMC	
Instrument	Welch	Levene	Welch	Levene	Welch	Levene
2	25.4%	31.9%	23.8%	32.0%	33.1%	46.5%
3	43.9%	65.8%	41.6%	63.3%	59.2%	82.3%
B	21.1%	46.2%	22.5%	48.8%	34.3%	77.5%
C	40.1%	57.0%	48.8%	57.1%	31.4%	58.5%
D	32.5%	63.5%	31.3%	56.8%	23.0%	63.5%
F	40.1%	62.3%	34.1%	55.6%	29.5%	57.6%
G	25.7%	14.0%	28.4%	14.0%	30.0%	35.0%
H	50.9%	80.4%	51.4%	78.3%	39.1%	80.1%
U	59.6%	12.9%	51.4%	18.6%	31.7%	40.3%
Total	37.7%	48.2%	37.0%	47.2%	34.6%	60.1%

Table S5: 2-FMA Welch Test Results. p-values comparing the mean abundance values for 2-FMA from each instrument to the total dataset at selected m/z fragments. Totals represent the number of times the associated row or column was not found to have a statistically different mean from the total primary dataset.

	m/z fragments											
Instrument	42	43	56	57	59	63	83	107	109	135	152	Total:
2	4.81×10^{-02}	8.49×10^{-02}	3.11×10^{-01}	8.78×10^{-03}	7.55×10^{-01}	7.00×10^{-06}	6.73×10^{-18}	6.40×10^{-23}	6.29×10^{-34}	2.36×10^{-38}	1.04×10^{-41}	3
3	1.43×10^{-15}	5.06×10^{-09}	5.00×10^{-09}	3.49×10^{-16}	7.77×10^{-01}	2.70×10^{-20}	1.13×10^{-42}	3.28×10^{-43}	2.93×10^{-43}	1.42×10^{-25}	3.56×10^{-23}	1
B	4.00×10^{-29}	4.55×10^{-25}	6.48×10^{-37}	4.48×10^{-19}	1.23×10^{-01}	3.15×10^{-09}	7.83×10^{-03}	7.44×10^{-11}	4.73×10^{-21}	2.28×10^{-24}	9.46×10^{-23}	1
C	2.09×10^{-33}	1.11×10^{-23}	1.63×10^{-07}	1.10×10^{-23}	1.19×10^{-02}	2.61×10^{-18}	4.47×10^{-11}	9.18×10^{-09}	2.74×10^{-07}	1.80×10^{-04}	1.38×10^{-04}	0
D	1.16×10^{-04}	1.24×10^{-08}	7.21×10^{-04}	3.54×10^{-02}	1.08×10^{-01}	9.68×10^{-04}	1.30×10^{-08}	1.59×10^{-07}	3.44×10^{-12}	4.29×10^{-08}	2.84×10^{-14}	1
F	4.28×10^{-16}	5.10×10^{-06}	3.88×10^{-06}	1.01×10^{-17}	2.24×10^{-04}	2.69×10^{-09}	1.30×10^{-13}	2.56×10^{-02}	9.94×10^{-01}	5.28×10^{-01}	3.94×10^{-02}	2
G	5.75×10^{-01}	6.55×10^{-01}	5.76×10^{-08}	1.81×10^{-03}	2.18×10^{-03}	7.15×10^{-04}	9.35×10^{-01}	7.76×10^{-03}	3.29×10^{-06}	5.14×10^{-05}	3.87×10^{-13}	3
H	2.83×10^{-06}	2.53×10^{-05}	3.54×10^{-03}	9.16×10^{-06}	2.94×10^{-04}	7.05×10^{-05}	9.28×10^{-03}	3.77×10^{-01}	7.08×10^{-06}	2.08×10^{-08}	2.10×10^{-12}	1
U	8.98×10^{-01}	6.49×10^{-01}	4.57×10^{-03}	2.47×10^{-01}	2.36×10^{-01}	3.55×10^{-01}	6.14×10^{-01}	7.27×10^{-03}	7.92×10^{-04}	2.42×10^{-01}	1.49×10^{-04}	7
Total:	2	3	1	1	5	1	2	1	1	2	0	

Table S6: 2-FMA Levene Test Results. p-values comparing the variance in abundance values for 2-FMA from each instrument to the total dataset at selected m/z fragments. Totals represent the number of times the associated row or column was not found to have a statistically different variance from the total primary dataset.

	m/z fragments											
Instrument	42	43	56	57	59	63	83	107	109	135	152	Total:
2	1.43×10^{-04}	4.33×10^{-06}	3.35×10^{-01}	1.47×10^{-06}	1.04×10^{-03}	2.89×10^{-06}	1.66×10^{-08}	3.63×10^{-12}	1.58×10^{-23}	2.92×10^{-21}	5.99×10^{-26}	1
3	2.52×10^{-06}	2.93×10^{-04}	4.62×10^{-03}	1.43×10^{-05}	3.27×10^{-01}	1.55×10^{-05}	3.06×10^{-08}	1.02×10^{-12}	3.51×10^{-22}	2.22×10^{-17}	9.46×10^{-22}	1
B	1.04×10^{-05}	1.72×10^{-04}	2.13×10^{-05}	4.98×10^{-05}	3.05×10^{-01}	1.59×10^{-03}	9.55×10^{-05}	7.24×10^{-06}	9.90×10^{-12}	2.70×10^{-10}	4.57×10^{-10}	1
C	9.19×10^{-06}	5.13×10^{-04}	1.09×10^{-02}	8.75×10^{-01}	2.71×10^{-01}	1.42×10^{-01}	1.03×10^{-01}	3.89×10^{-01}	1.54×10^{-01}	7.91×10^{-03}	9.12×10^{-03}	6
D	1.79×10^{-01}	4.79×10^{-02}	5.63×10^{-01}	8.82×10^{-02}	7.51×10^{-01}	3.32×10^{-01}	5.73×10^{-02}	1.09×10^{-02}	9.18×10^{-05}	4.46×10^{-03}	1.12×10^{-05}	6
F	1.72×10^{-02}	3.26×10^{-02}	5.14×10^{-02}	2.25×10^{-02}	1.45×10^{-01}	4.85×10^{-02}	8.39×10^{-03}	1.72×10^{-03}	2.90×10^{-05}	1.67×10^{-04}	9.29×10^{-05}	2
G	1.15×10^{-04}	3.96×10^{-04}	1.10×10^{-05}	2.39×10^{-04}	6.18×10^{-01}	2.99×10^{-04}	1.44×10^{-05}	4.87×10^{-09}	9.03×10^{-16}	1.67×10^{-16}	3.96×10^{-18}	1
H	4.45×10^{-02}	8.74×10^{-02}	5.35×10^{-02}	2.90×10^{-02}	3.14×10^{-01}	5.88×10^{-02}	7.50×10^{-03}	8.25×10^{-04}	6.89×10^{-06}	4.48×10^{-05}	2.53×10^{-06}	4
U	1.12×10^{-09}	1.96×10^{-10}	6.24×10^{-13}	3.49×10^{-07}	3.66×10^{-12}	1.74×10^{-03}	1.03×10^{-02}	3.21×10^{-01}	4.53×10^{-01}	1.22×10^{-06}	7.06×10^{-01}	3
Total:	1	1	4	2	7	3	2	2	2	0	1	

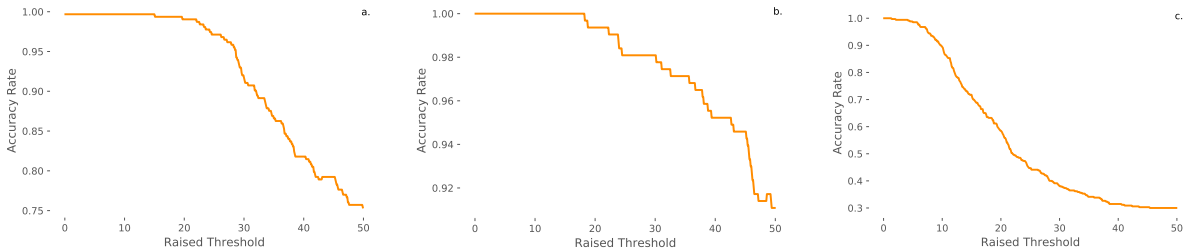

Figure S7: Decreasing validation accuracy as abundance threshold is incrementally raised for a) FA b) FMA and c) MMC required fragments. Abundance thresholds are represented as the value subtracted from all normalized abundance values across the spectrum. Negative values were reset to zero and the remaining spectrum re-normalized to a base peak abundance of 9999.

Table S7: Required m/z fragments to reach abundance threshold.

FA	FMA	MMC
42, 43, 44, 45, 50, 51, 52, 56, 57, 59, 62, 63, 64, 65, 69, 70, 74, 75, 77, 81, 83, 84, 89, 91, 95, 96, 101, 107, 108, 109, 110, 115, 117, 118, 133, 135, 138	42, 43, 44, 50, 51, 56, 57, 58, 59, 62, 63, 70, 75, 77, 81, 83, 89, 96, 101, 107, 108, 109, 110, 115, 133, 135, 152	42, 43, 44, 50, 51, 52, 53, 54, 55, 56, 58, 59, 62, 62, 64, 65, 66, 74, 75, 77, 78, 87, 89, 90, 91, 92, 103, 105, 115, 116, 117, 118, 119, 120, 130, 131, 132, 144, 146, 147, 160, 162

Table S8: Confusion matrices for Test and Validation Sets with spectra under Abundance threshold

		Actual										
Validation		2-FA	3-FA	4-FA		2-FMA	3-FMA	4-FMA		2-MMC	3-MMC	4-MMC
	2-FA	2	1	0	2-FMA	0	0	0	2-MMC	2	0	0
Predicted	3-FA	0	0	0	3-FMA	0	0	0	3-MMC	0	0	0
	4-FA	0	0	0	4-FMA	0	0	0	4-MMC	0	0	1
		Actual										
Test		2-FA	3-FA	4-FA		2-FMA	3-FMA	4-FMA		2-MMC	3-MMC	4-MMC
	2-FA	12	0	0	2-FMA	9	0	0	2-MMC	9	0	0
Predicted	3-FA	0	15	0	3-FMA	0	11	0	3-MMC	0	1	1
	4-FA	0	0	12	4-FMA	0	0	10	4-MMC	0	0	2

References

(1) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(2) van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. Centering, scaling, and transformation: improving the biological information content of metabolomics data. *BMC Genomics* **2006**, *7*.

(3) Recommended Methods for the Identification and Analysis of Synthetic Cathinones in Seized Materials (Revised and Updated). 2020; `https://www.unodc.org/documents/scientific/Recommended_methods_for_the_Identification_and_Analysis_of_Synthetic_Cathinones_in_Seized_Materials-Rev..pdf`, [Accessed: 2022-07-27].

(4) Kerrigan, S.; Savage, M.; Cavazos, C.; Bella, P. Thermal Degradation of Synthetic Cathinones: Implications for Forensic Toxicology. *J. Anal. Toxicol.* **2016**, *40*, 1–11.

(5) Tsujikawa, K.; Kuwayama, K.; Kanamori, T.; Iwata, Y. T.; Inoue, H. Thermal Degradation of -pyrrolidinopentiophenone During Injection in Gas Chromatography/Mass Spectrometry. *Forensic Sci. Int.* **2013**, *231*, 296–299.