Supplementary material

Fecal volatile organic compound–based machine learning model for noninvasive detection of colorectal cancer

Qi-Jun Li, Zu-Bai Li, Bo-Rong Yu, Xiao-Hong Wang, Xiao-Wen Li

Qi-Jun Li, Xiao-Hong Wang, Department of General Surgery, Huadong Hospital, Fudan University, Shanghai 200040, China

Zu-Bai Li, College of Chemistry and Chemical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

Xiao-Wen Li, Bo-Rong Yu Department of Gastroenterology, Huadong Hospital, Fudan University, Shanghai 200040, China

Xiao-Wen Li, Shanghai Key Laboratory of Clinical Geriatric Medicine. Shanghai Institute of Geriatrics and Gerontology, Shanghai, China

Co-first authors: Qi-Jun Li, Zu-Bai Li

Co-corresponding authors: Xiao-Hong Wang, Xiao-Wen Li

Corresponding author: Xiao-Wen Li, Department of Gastroenterology, Huadong Hospital, Fudan University, Shanghai 200040, China. xiaowenli12@fudan.edu.cn

Corresponding author: Xiao-Hong Wang, Department of General Surgery, Huadong Hospital, Fudan University, Shanghai 200040, China. xhwangwy@163.com

Supplementary Tables:

Table S1.Statistics of Subjects' Age, Gender, and Body Mass Index

Group	HC，n=57	CRC，n=78
Age	61.5 ±14.2（33-88）	65.4 18.3（33-94）
Males	21（36.8%）	54（69.2%）
Females	36（63.2%）	24（30.8%）
BMI	23.1 ± 2.8（17.9-31.9）	24.3 ± 3.5（17.8-31.1）

Table S2.Classification of 85 Volatile Compounds

VOCS	Classification
2-Butanone	Ketone
1-Penten-3-one	Ketone
4-Methyl-3-penten-2-one	Ketone
2-Heptanone	Ketone
2-Octanone	Ketone
3-Hydroxy-2-butanone	Ketone
1-hydroxy-2-propanone	Ketone
Cyclohexanone	Ketone
2-Nonanone	Ketone
Butanoic acid-M	Fatty acids
Butanoic acid-D	Fatty acids
2-Methylpropanoic acid-M	Fatty acids
2-Methylpropanoic acid-D	Fatty acids
Propanoic acid-M	Fatty acids
Propanoic acid-D	Fatty acids
Pentanoic acid-M	Fatty acids
Pentanoic acid-D	Fatty acids
2-Methylbutanoic acid-M	Fatty acids
2-Methylbutanoic acid-D	Fatty acids
3-Methylbutanoic acid-M	Fatty acids
3-Methylbutanoic acid-D	Fatty acids
Hexanoic acid-M	Fatty acids
Hexanoic acid-D	Fatty acids
Nonanoic acid	Fatty acids
2-Ethyl-1-hexanol	Alcohols
Phenol	Alcohols
2-Octanol	Alcohols
2-Methyl-1-butanol	Alcohols
1-Butanol	Alcohols
1-Propanol	Alcohols
1-Pentanol	Alcohols
3-Methyl-1-butanol	Alcohols
2-Hexanol	Alcohols
1-Octen-3-ol	Alcohols
Benzenemethanol	Alcohols
1-Hexanol-M	Alcohols
1-Hexanol-D	Alcohols
2-Butoxyethanol-M	Alcohols
Linalool	Alcohols

Table S2 (Continued)

VOCS	Classification
4-Methylphenol	Alcohols
2-Butoxyethanol-D	Alcohols
Nonanal-M	Aldehydes
2-Phenylacetaldehyde	Aldehydes
Benzaldehyde-M	Aldehydes
Benzaldehyde-D	Aldehydes
Hexanal	Aldehydes
2-Methylpentanal	Aldehydes
3-Methylbutanal	Aldehydes
Heptanal	Aldehydes
(E)-2-Pentenal	Aldehydes
(E)-2-Decenal	Aldehydes
Nonanal-D	Aldehydes
2-Methylbutanal	Aldehydes
Propyl acetate	Esters
Amyl acetate	Esters
Ethyl trans-2-butenoate	Esters
Ethyl acetate	Esters
Ethyl butanoate	Esters
Methyl salicylate	Esters
Benzyl acetate-M	Esters
Benzyl acetate-D	Esters
Propyl butanoate	Esters
Dimethyl trisufide	Sulfocompound
Propyl sulfide	Sulfocompound
Methional	Sulfocompound
Thiazole	Sulfocompound
2-Methyl-3-furanthiol	Sulfocompound
Indole	others
Octamethylcyclotetrasiloxane	others
2-Ethyl-6-methylpyrazine	others
4,5-Dihydro-3(2H)-thiophenone	others
3-Butenenitrile	others
Limonene-M	others
Limonene-D	others
alpha-Pinene	others
gamma-Terpinene	others
p-Cymene	others
Hexanenitrile	others
Dimethylamine	others
1.8-Cineole-M	others
1,8-Cineole-D	others
beta-Pinene	others
2-Pentylfuran	others
2-Isopropyl-3-methoxypyrazine	others
Hexamethylcyclotrisiloxane	others

Table S3. Feature VOC Selection Results of Three Algorithms

RFE-RF	RFE-SVM_Selected	RFE-LASSO	Common Selected
Indole	Indole	Indole	(E)-2-Pentenal
Nonanal-M	Nonanal-M	Nonanal-M	Nonanal-D
2-Phenylacetaldehyde	2-Phenylacetaldehyde	2-Phenylacetaldehyde	2-Phenylacetaldehyde
Phenol	2-Ethyl-1-hexanol	2-Ethyl-1-hexanol	Indole
Dimethyl trisufide	Octamethylcyclotetrasiloxane	Octamethylcyclotetrasiloxane	Dimethyl trisufide
Pentanoic acid-M	Dimethyl trisufide	Dimethyl trisufide	1-Butanol
3-Methylbutanoic acid-M	2-Octanol	4,5-Dihydro-3(2H)-thiophenone	4-Methylphenol
Propanoic acid-D	4,5-Dihydro-3(2H)-thiophenone	Pentanoic acid-M	Pentanoic acid-M
2-Butoxyethanol-D	Pentanoic acid-M	Pentanoic acid-D	Propanoic acid-D
1-Butanol	Pentanoic acid-D	3-Methylbutanoic acid-M	Nonanal-M
(E)-2-Pentenal	3-Methylbutanoic acid-M	3-Methylbutanoic acid-D	3-Methylbutanoic acid-M
p-Cymene	3-Methylbutanoic acid-D	Butanoic acid-M	
Nonanal-D	Butanoic acid-D	Butanoic acid-D	
4-Methylphenol	2-Methylpropanoic acid-M	2-Methylpropanoic acid-M	
	2-Methylpropanoic acid-D	2-Methylpropanoic acid-D	
	Propanoic acid-M	3-Butenenitrile	
	3-Butenenitrile	Propanoic acid-D	
	Propanoic acid-D	Hexanoic acid-M	
	Hexanoic acid-M	Benzaldehyde-D	
	2-Butoxyethanol-M	2-Butoxyethanol-M	
	Propyl acetate	2-Butoxyethanol-D	
	3-Methylbutanal	Hexanal	
	2-Butanone	Propyl acetate	
	1-Penten-3-one	3-Methylbutanal	
	1-Butanol	2-Butanone	
	2-Heptanone	1-Penten-3-one	
	2-Octanone	Amyl acetate	
	Hexanenitrile	4-Methyl-3-penten-2-one	
	1-Hexanol-M	1-Butanol	
	Ethyl trans-2-butenoate	1-Propanol	
	(E)-2-Pentenal	1-Pentanol	
	Propyl sulfide	alpha-Pinene	
	Benzenemethanol	2-Octanone	
	2-Nonanone	Hexanenitrile	
	Ethyl acetate	1-Hexanol-M	

Table S3 (Continued)

RFE-RF	RFE-SVM_Selected	RFE-LASSO	Common Selected
	3-Methyl-1-butanol	(E)-2-Pentenal	
	2-Methylbutanal	Propyl sulfide	
	p-Cymene	Benzenemethanol	
	Linalool	Ethyl acetate	
	Thiazole	Dimethylamine	
	Cyclohexanone	3-Methyl-1-butanol	
	beta-Pinene	Methional	
	3-Hydroxy-2-butanone	2-Hexanol	
	Nonanoic acid	1.8-Cineole-M	
	Nonanal-D	1-Octen-3-ol	
	4-Methylphenol	2-Methylbutanal	
	2-Isopropyl-3-methoxypyrazine	gamma-Terpinene	
		Linalool	
		Thiazole	
		Cyclohexanone	
		2-Methyl-3-furanthiol	
		3-Hydroxy-2-butanone	
		1-Hexanol-D	
		Limonene-D	
		Nonanoic acid	
		Methyl salicylate	
		Nonanal-D	
		(E)-2-Decenal	
		Benzyl acetate-D	
		4-Methylphenol	
		2-Isopropyl-3-methoxypyrazine	
		1-hydroxy-2-propanone	

Table S4.Subjects' CEA Data

HC	CEA (ng/mL)	CRC	CEA (ng/mL)
1	2.3	1	1.5
2	1.3	2	2.7
3	1.8	3	1.8
4	2	4	10.2
5	1	5	1.2
6	0.7	6	3
7	3.2	7	10.9
8	3.9	8	1.9
9	3.8	9	3.6
10	1.1	10	0.9
11	4.8	11	26.3
12	3.1	12	74.5
13	2.1	13	2.5
14	0.9	14	1.6
15	0.3	15	4.6
16	2	16	28
17	0.5	17	2.7
18	1.5	18	1.4
19	3.2	19	4.6
20	3.1	20	33.1
21	2	21	26.2
22	2.5	22	35.3
23	3.5	23	1.3
24	2.5	24	3
25	0.9	25	7.2
26	3.1	26	3
27	4	27	3.3
28	2	28	12.9
29	0.6	29	1.8
30	1.2	30	2.5
31	2.2	31	2.8
32	2.6	32	5.4
33	2.9	33	27.4
34	2.2	34	291.7
35	3.4	35	66.4
36	2.6	36	17
37	1.92	37	1.1
38	3.01	38	127.4
39	2.76	39	5.2
40	1.45	40	2.8
41	2.31	41	15
42	0.88	42	3.7
43	2.65	43	1.52
44	3.37	44	1.65
45	2.19	45	3.27
46	1.74	46	2.45
47	2.44	47	0.86
48	1.93	48	3.12

Table S4 (Continued)

HC	CEA (ng/mL)	CRC	CEA (ng/mL)
49	1.68	49	2.16
50	1.33	50	4.56
51	3.06	51	1.26
52	2.56	52	2.36
53	3.57	53	2.93
54	4.06	54	11.37
55	3.21	55	4.12
56	0.40	56	35.67
57	0.65	57	1.61
		58	202.98
		59	3.16
		60	8.93
		61	2.48
		62	26.87
		63	13.39
		64	1.91
		65	0.6
		66	100.17
		67	3.7
		68	12.48
		69	3.4
		70	2.3
		71	1.1
		72	3.2
		73	1.4
		74	5.8
		75	2.8
		76	0.56
		77	0.2
		78	2.3