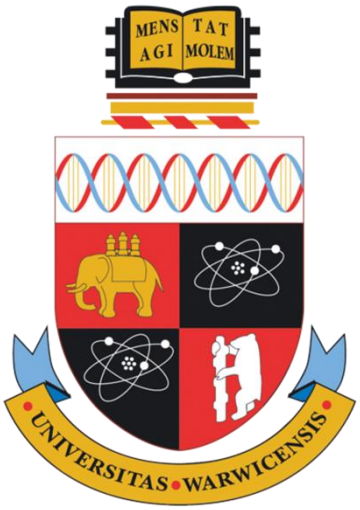
Diagnosis of Human Disease by Odour Analysis Employing Machine Learning



by

Alfian Nur Wicaksono

Dissertation Submitted for the Degree of

Doctor of Philosophy

School of Engineering

The University of Warwick

September 2021

Table of Contents

List of Figures vi

List of Tables x

Declaration xii

Acknowledgement xiii

List of Publications and Research Outputs xiv

Abstract xviii

Abbreviations xx

Chapter 1 INTRODUCTION 1

Introduction 2

Research Aims 4

Thesis Outline 4

Reference 6

Chapter 2 LITERATURE REVIEW 9

2.1. Introduction 10

2.2. The Fundamental of Olfactory Systems 10

2.3. The Origin of Human Waste Biomarker 12

2.3.1. Exhaled Breath 12

2.3.2. Urine 13

2.3.3. Faecal 14

2.4. Previous Research on Human Waste Odour for Disease Detection 15

2.5. Current Data Analysis Approaches 17

2.5.1. Pre-processing 17

2.5.2. Multivariate Analysis and Machine Learning 18

2.6. Conclusion 19

2.7. Reference 19

Chapter 3 SENSING INSTRUMENT 26

3.1. Introduction 27
3.2. Electronic Nose (E-nose) 27
3.2.1. Operating Principle 28
3.2.2. WOLF Electronic Nose 30
3.3. Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) 33
3.3.1. Operating Principle 34
3.3.2. Commercial Products 36
3.4. Gas Chromatography Ion Mobility Spectroscopy (GC-IMS) 37
3.4.1. Operating Principle 38
3.4.2. Commercial Products 39
3.5. Gas Chromatography Time-Of-Flight Mass Spectrometry (GC-TOF-MS) 41
3.5.1. Operating Principle 42
3.5.2. Commercial Products 43
3.5.3. Thermal Desorption 43
3.6. Quality Control and Calibration 44
3.7. Comparison 45
3.8. Conclusion 48
3.9. Reference 48
Chapter 4 DATA ANALYSIS 53
4.1. Introduction 54
4.2. Data Structure 54
4.2.1. E-nose 54
4.2.2. FAIMS 56
4.2.3. GC-IMS 57
4.2.4. GC-TOF-MS 59
4.3. Data Analysis Pipeline 60
4.3.1. Pre-processing 61
4.3.2. Data Splitting 65

4.3.3. Cross-Validation 66
4.3.4. Imbalance Dataset 68
4.3.5. Feature Selection 71
4.3.6. Classification Models 73
4.3.7. Performance Metrics and ROC Curve 76
4.3.8. Significant Feature 78
4.4. Conclusion 79
4.5. Reference 80
Chapter 5 APPLICATION ON CANCER DETECTION 85
5.1. Introduction 86
5.2. Non-invasive Diagnosis of pancreatic cancer through detection of volatile organic compounds in urine 87
5.2.1. Methods 87
5.2.2. Result 89
5.3. Volatile Organic Compounds (VOCs) for the noninvasive detection of pancreatic cancer from urine 90
5.3.1. Methods 91
5.3.2. Results 93
5.4. Discussion of Pancreatic Cancer Studies 96
5.5. Colorectal cancer and adenoma screening using urinary volatile organic compounds (VOC) detection: early result from a singles-centre bowel screening population (UK BCSP) 98
5.5.1. Methods 98
5.5.2. Results 99
5.6. Early detection and follow up of colorectal neoplasia based on faecal volatile organic compounds 102
5.6.1. Method 103
5.6.2. Results 105
5.7. Discussion of Colorectal Cancer Studies 111

5.8. Conclusion 117
5.9. References 117
Chapter 6 APPLICATION ON INFLAMMATORY BOWEL DISEASE DETECTION .. 124
6.1. Introduction 125
6.2. Differentiation Between Paediatric Irritable Bowel Syndrome and Inflammatory Bowel Disease Based on Faecal Scent (Study 1) 127
6.2.1. Methods 127
6.2.2. Results 129
6.3. Breath Analysis Using E-nose and Ion Mobility Technology to Diagnose Inflammatory Bowel Disease – a Pilot Study (Study 2) 134
6.3.1. Methods 134
6.3.2. Results 140
6.4. Simultaneous Assessment of Urinary and Faecal Volatile Organic Compound Analysis in De Novo Paediatric IBD (Study 3) 144
6.4.1. Methods 144
6.4.2. Result 146
6.5. The faecal scent of inflammatory bowel disease detection and monitoring based on volatile organic compound analysis (Study 4) 149
6.5.1. Methods 149
6.5.2. Result 152
6.6. Discussion 156
6.7. Conclusion 165
6.8. Reference 165
Chapter 7 CONCLUSION AND FURTHER WORK 176
7.1. Introduction 177
7.2. Data Analysis Pipeline 177
7.3. Pancreatic Cancer Studies 178
7.4. Colorectal Cancer Studies 179
7.5. IBD Studies 180

7.6. Further Work and Future Prospect 182

7.7. Reference 183

Appendix 184

List of Figures

Figure 2.1. An illustration of Human Olfactory System 11

Figure 2.2. An Illustration of Urine Production in Nephron Inside Kidney [24]. 13

Figure 3.1. Electronic Nose Concept[1] 27

Figure 3.2. The Basic Stage of How E-Nose System Works [8] 29

Figure 3.3. An Example of Electronic Nose Sensor Array Setup; (a) Gas Sensor Array Exposed to VOCs (b) Gas Sensor Array Response to a VOCs 29

Figure 3.4. WOLF E-Nose in a Laboratory Setting 30

Figure 3.5. Schematic of an Electrochemical Sensor [12]. 31

Figure 3.6. Schematic Diagram of an NDIR Gas Sensor [16]. 32

Figure 3.7. The Schematic of PID's Basic Working Principle [19] 33

Figure 3.8. Schematic of Working Principle of FAIMS. 34

Figure 3.9. The FAIMS Systems with ATLAS Sampling System of the Left Side 37

Figure 3.10. The Schematic of Working Principle of GC-IMS 39

Figure 3.11. G.A.S. Breathspec GC-IMS Instrument [30] 40

Figure 3.12. G.A.S. GC-IMS-SILOX Instrument [31] 40

Figure 3.13. G.A.S. FlavourSpec GC-IMS Instrument [32] 41

Figure 3.14. Schematic Diagram of GC-TOF-MS. Sample is First Pre-Separated by GC Before Being Ionised. Then, the Time it Takes for the Ion to Travel from Ionisation Chamber to the Detector is Measured. In the Flight Tube, Ions Separate into Groups According to Velocity. 42

Figure 3.15. GC-TOF-MS Instrument Used in This Thesis: (a) TRACE 1300 GC; (b) BenchTOF-HD TOF-MS; (c) ULTRA-xr and UNITY-xr 43

Figure 3.16. Un-Capped and Capped Thermal Desorption (TD) Tubes 44

Figure 3.17. Comparison and Schematic of VOC Detection Technologies Used in This Thesis. (a) Electronic Nose (E-nose): Generates a Single from a Cross-Reactive Gas Sensor Array, (b) Field Asymmetric Ion Mobility Spectrometry (FAIMS): the Separation of Compounds Based on Different Ion Mobility in a High and Low Electric Field; (c) Gas Chromatography Ion Mobility Spectrometry (GC-IMS): GC Which Perform the Pre-

Separation is Represented in Blue and the Separation of Compounds Based on Ion Mobility Represent as a Group of VOCs with a Different Colour. (d) Gas Chromatography Time of Flight Mass Spectrometry (GC-TOF-MS): Pre-Separation is Performed by GC and Compounds Separation Based on the Time it Takes to Reach the Detector. A Typical Response is Shown for Each Method. VOCs are Illustrated as Red, Green, and Blue Circles. 46

Figure 4.1. Schematic of Single E-Nose Gas Sensor Response 55

Figure 4.2. Typical Sensor Output Responses of WOLF E-Nose 56

Figure 4.3. Typical FAIMS Response to Urinary VOC Sample from a Patient with Pancreatic Cancer 57

Figure 4.4. 3D Representation of GC Data Output with Corresponding IMS Chromatogram. (a) Single IMS Spectra Data is Combined with GC Run Time Peaks (b) Heatmap Corresponding to GC-IMS Peaks (Yellow and Blue Lines) (Image Adapted with Permission from IMSPEX, UK) 58

Figure 4.5. An Example Output of GC-IMS Instrument to Faecal VOC Sample from a Patient with Refractory Coeliac Disease. 59

Figure 4.6. A Typical Raw Response of GC-TOF-MS when Analysing the Urinary VOC Patient with PDAC. 59

Figure 4.7. An Example of Chemical Interpretation of GC-TOF-MS Peak for Patient with PDAC 60

Figure 4.8. Diagram of VOC Data Analysis Pipeline 61

Figure 4.9. FAIMS Pre-Processing Step Involving: (a) Raw Positive and Negative Ions Data; (b) Combining Positive and Negative Ion Data Into 102 x 512 Matrix; (c) Padding the Data with Zero to Make 128 x 512 Matrix; (d) Cropping the Data Into 128 x 128 Matrix for 2D DWT. 63

Figure 4.10. The Pre-Processing Steps of GC-IMS Dataset Involve: (a) Raw Data of GC-IMS; (b) RIP Alignment; (c) Cropping Area of Interest; (d) Thresholding to Remove the Noise. 65

Figure 4.11. An Illustration of Leave One Out Cross Validation (LOOCV) 67

Figure 4.12. An Illustration of K-fold Cross-Validation 68

Figure 4.13. Illustration of (a) Undersampling in Which the Number or Majority Sample is Reduced to Match the Minority Sample; (b) Oversampling Where the Minority Sample is Duplicated Until it Matches with the Majority Sample. 70

Figure 4.14. An illustration of SMOTE in creating synthetic samples from minority class.. 71

Figure 4.15. Illustrate the Feature Selection Method Using Wilcoxon Rank Sum Test by Comparing Each Feature Between Groups to Get the p-value. 72

Figure 4.16. The Overall Performance of Random Forest Classifier is Represented as the Sum of AUC, Sensitivity, and Specificity for Variation of Features from 2-100 when Detecting Pancreatic Cancer from Healthy Control Using FAIMS. 73

Figure 4.17. Confusion Matrix 76

Figure 4.18. Example of ROC Curve with no, Good, and Perfect Separation Represented by Red, Black, and Green Line Respectively 78

Figure 4.19. Selecting Features are Plotted Back onto the GC-IMS Output: (a) Excellent Separation AUC = 0.91; (b) Poor Separation AUC = 0.52. 79

Figure 5.1. (A) Receiver operating characteristic curve showing performance of volatile organic compounds in differentiating healthy individuals from PDAC with an AUC of 0.92 (95% CI, 0.88-0.96), sensitivity of 0.91 (95% CI, 0.83-0.96), and specificity of 0.83 (95% CI, 0.73-0.90). (B) Healthy individuals from early stage I/II with an AUC of 0.89 (95% CI, 0.83-0.94), a sensitivity of 0.91 (95% CI, 0.78-0.97), and a specificity of 0.78 (95% CI, 0.69-0.86). (C) Early stage PDAC (I/II) could also be successfully separated from advanced stage PDAC (III/IV) with an AUC of 0.92 (95% CI, 0.86-0.97), sensitivity of 0.82 (95% CI, 0.67-0.92), and specificity of 0.89 (95% CI, 0.75-0.97). 90

Figure 5.2. ROC for (a) PDAC vs Healthy and (b) PDAC vs CP using GC-IMS 94

Figure 5.3. ROC for (a) PDAC vs Healthy and (b) PDAC vs CP using GC-TOF-MS 95

Figure 5.4. ROC curve for (a) classification of CRC vs healthy control in bowel cancer screening programme (BCSP) patients (balanced) using the sparse logistic regression classifier using FAIMS and (b) classification of CRC vs healthy control using GC-IMS with Gaussian process classifier 101

Figure 5.5. An example output of the GC-IMS instrument of the colorectal cancer group (CRC), advanced adenoma group (AA) and control group. The y-axis represents retention time from the gas chromatography column, and the x-axis represents drift time through the ion mobility spectrometer. Volatile organic compound levels in the sample are represented by colour intensity 109

Figure 5.6. ROC curve for the comparison between colorectal cancer (CRC), advanced adenoma (AA), large adenoma (LA), small adenoma (SA) and healthy control (HC) (AUC, area under the curve) 110

Figure 5.7. ROC curves for the polypectomy follow-up (AUC, area under the curve; HC, healthy control) 111

Figure 6.1. ROC for IBS/FAP-NOS vs IBD, UC, and CD and IBD vs HC. AUC are reported for the SLR analysis. 133

Figure 6.2. CD and UC boxplots for FCP scores 136

Figure 6.3. WOLF E-nose and custom Bio-VOC injection system 137

Figure 6.4. G.A.S Breathspec GC-IMS: (a) Plastic disposal mouthpiece; (b) mouthpiece holder; (c) GC-IMS instrument; (d) Nitrogen generator 138

Figure 6.5. IBD analysis: G.A.S. Breathspec GC-IMS ROC curve; (a) IBD vs HC; (b) CD vs UC 140

Figure 6.6. IBD analysis: WOLF E-nose ROC curves; (a) IBD vs HC; (b) CD vs UC 141

Figure 6.7. WOLF E-nose sensor response radar plot 142

Figure 6.8. (A) Display the ROC curve for the faecal VOC profile, and (B) display the ROC curve for urinary VOC profile for the discrimination of IBD vs HC. 148

Figure 6.9. Examples of GC-IMS output. Here, the GC-IMS of feacal (A) and urinary (B) VOC profiles of one IBD patient are displayed 149

Figure 6.10. ROC curves for the differentiation between study groups based on SLR classification using the 100 most discriminative features. 155

List of Tables

Table 2.1. A list of descriptive aromas with its associate disease and sample type 15

Table 3.1. Sensors Deployed in WOLF E-Nose 30

Table 3.2. Summarise the Pros and Cons for Each Device Used in This Thesis. 46

Table 5.1. Demographic Information of All Study Participants 87

Table 5.2. Statistical Analysis of FAIMS Data Using SVM 89

Table 5.3. Demographics of Sample Groups and Cancer Stage 91

Table 5.4. Statistical Analysis of GC-IMS Data (Best Results) 94

Table 5.5. Statistical Analysis of GC-TOF-MS Data (Best Results) 95

Table 5.6. Chemicals Used to Separate Sample Groups 95

Table 5.7. Diagnostic Outcomes for Study Participants and Distribution of CRC by Site (Total of 13 Cancer Sites as One Patient Had a Synchronous Tumour) 99

Table 5.8. Classification of BCSP Study Participants by Outcome Using FAIMS 100

Table 5.9. Classification of BCSP Study Participants Using GC-IMS using Gaussian Process or Support Vector Machine 102

Table 5.10. Demographics of All Study Participants 106

Table 5.11. Demographics of Participants Included in Follow-Up Study 107

Table 5.12. Differences Between All Subgroups of Colorectal Neoplasia, Polyps and Healthy Controls Based on Faecal Volatile Organic Compounds 108

Table 5.13. Summary of Two CRC Studies 112

Table 5.14. Summary of Urinary and Faecal VOC Analysis Performance in Detecting CRC from Healthy Controls 112

Table 5.15. Summary of CRC vs Adenoma 113

Table 5.16. Summary of Adenoma vs Healthy Control 115

Table 6.1. Baseline Characteristics 130

Table 6.2. Performance Characteristics for the Discrimination of IBS/FAP-NOS, IBD, and HC by Faecal VOC Analysis 132

Table 6.3. Demographic data of IBD patients and healthy controls 135

Table 6.4. Summary of confounding factor group 139
Table 6.5. IBD analysis G.A.S. Breathspec GC-IMS result 141
Table 6.6. IBD analysis WOLF E-nose result 141
Table 6.7. Confounding factors: G.A.S. BreathSpec GC-IMS results 142
Table 6.8. Confounding factors: WOLF E-nose results 143
Table 6.9. CD vs UC (without and with FCP): WOLF E-nose results 143
Table 6.10. CD vs HC & UC vs HC (without and with FCP): WOLF E-nose results 143
Table 6.11. Baseline characteristics 146
Table 6.12. Diagnoses Controls 147
Table 6.13. Medication usage 147
Table 6.14. Baseline characteristics 152
Table 6.15. Inflammatory Bowel Disease Activity Scores 154
Table 6.16. Differences in Faecal Volatile Organic Compound Patterns 154
Table 6.17. Summary of Four Studies 157
Table 6.18. Summarized of GP Classifier Performance When Detecting IBD from HC in Faecal Samples 157
Table 6.19. The Summary of FAIMS and GC-IMS Performance in Differentiating UC vs HC, CD vs HC, and UC vs CD from Patient's Faecal VOC Profile Using Gaussian Process Classifier 158
Table 6.20. The Result of Urinary VOC Analysis Using E-nose, FAIMS, and GC-IMS to Detect IBD from HC 161
Table 6.21. Discrimination Performance of IBD vs HC Based on Sample Type 164

Declaration

This thesis is submitted to the University of Warwick in support of the application for the degree of Doctor of Philosophy. It has not been submitted in part, or in whole, for a degree or other qualification at any other University. Parts of this thesis are published by the author in peer-reviewed journals. Apart from commonly understood and accepted ideas, or where reference is made to the work of others, the work described in this thesis is carried out by the author in the School of Engineering at the University of Warwick.

Alfian Nur Wicaksono

2016 – 2021

Acknowledgement

The path toward completing this thesis has been a rollercoaster for me. There were countless moments that I felt like giving up and the pandemic surely had pushed me to hit rock bottom in my life. However, due to the tremendous supports from many people, I finally finish my thesis. Here, I would like to acknowledge everyone who played a role in my academic accomplishments.

الحمد لله ربّ العالمين

Firstly, all praises and thanks to all mighty God, Allah S.W.T., for giving me the blessing and strength to complete this thesis.

I would like to thank my PhD supervisor, Prof. James Covington, without whom this PhD would not have been possible. Undertaking this PhD has provided me with loth of experiences and knowledges in disease diagnosis and data analysis that I am hugely grateful for, and without doubt James is the one the make all of this possible. I am grateful for his constant guidance and support during my entire academic career at Warwick. I am very fortunate to have him as my supervisor and I sincerely hope that our professional relationship can continue in the future.

Furthermore, I want to acknowledge and thank Prof. Ramesh Arasaradnam and his team at UHCW, and Dr Sofie Bosch and her team at Vrije Universiteit Amsterdam, with whom I worked in close collaboration for this research.

I also would like to thank Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan (LPDP)) for financial support during my PhD.

I would like to thank to my dearest friends, Dr Akira Tiele, Dr Emma Daulton, Dr Siavash Esfahani, Dr Wangi Pandan Sari, Dr Sam Agbroko and Sai Kiran Ayyala for all the support and happiness that you brought in and beyond the lab. I would also like to thank Norma Dani Risdiandita, Khoirul Faiq Muzakka, Azizon, Nelvia Sari, Dini Rachmadhani, Risa Umari Yuli Aliviyanti, and Dwi Ardhanariswari Sundrijo for all your encouragements and support.

Last but not least, unreserved thanks and gratitude must go to my lovely family. Word can’t describe how grateful I am to have Kartini as my Mum and Sugiarto as my Dad for their unconditional love and support throughout my life. To my lovely brothers Irfan Darmawan and Despiyadi, thank you for your constant support and caring to me. Finally, I want to thank my partner, Juliyani Purnama Ramli, for supporting, encouraging, motivating and always being there for me.

List of Publications and Research Outputs

Journal Articles

[1] S. Bosch *et al*., "Prediction of Inflammatory Bowel Disease Course Based on Fecal Scent," Sensors, vol. 22, no. 6, p. 2316, Mar. 2022, doi: 10.3390/s22062316.

[2] M. McFarlane *et al.*, "Minimal Gluten Exposure Alters Urinary Volatile Organic Compounds in Stable Coeliac Disease," Sensors, vol. 22, no. 3, p. 1290, Feb. 2022, doi: 10.3390/s22031290.

[3] J. Nazareth, D. Pan, J. W. Kim, J. Leach, and J. G. Brosnan, "Discriminatory ability of gas chromatography-ion mobility spectrometry to identify patients hospitalised with COVID-19 and predict prognosis," 2022, doi: https://doi.org/10.1101/2022.02.28.22271571.

[4] K. A. Goggin, E. Brodrick, A. Wicaksono, J. A. Covington, A. N. Davies, and D. J. Murphy, "A PROOF-OF-CONCEPT STUDY: DETERMINING THE GEOGRAPHICAL ORIGIN OF CRUDE PALM OIL WITH THE COMBINED USE OF GC-IMS FINGERPRINTING AND CHEMOMETRICS," J. Oil Palm Res., vol. 33, no. June, pp. 227–234, Apr. 2021, doi: 10.21894/jopr.2021.0013.

[5] D. A. van den Brink *et al.*, "Prediction of mortality in severe acute malnutrition in hospitalized children by faecal volatile organic compound analysis: proof of concept," Sci. Rep., vol. 10, no. 1, pp. 1–9, 2020, doi: 10.1038/s41598-020-75515-6.

[6] N. M. Sagar *et al.*, "The pathophysiology of bile acid diarrhoea: differences in the colonic microbiome, metabolome and bile acids," Sci. Rep., vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-77374-7.

[7] E. Daulton *et al.*, "Volatile organic compounds (VOCs) for the non-invasive detection of pancreatic cancer from urine," Talanta, vol. 221, no. August 2020, p. 121604, 2021, doi: 10.1016/j.talanta.2020.121604.

[8] E. Daulton, A. Wicaksono, J. Bechar, J. A. Covington, and J. Hardwicke, "The detection of wound infection by ion mobility chemical analysis," Biosensors, vol. 10, no. 3, pp. 1–9, 2020, doi: 10.3390/bios10030019.

[9] S. Bosch *et al.*, "Early detection and follow-up of colorectal neoplasia based on faecal volatile organic compounds," Color. Dis., pp. 0–1, 2020, doi: 10.1111/codi.15009.

[10] D. J. C. Berkhout *et al.*, "Preclinical Detection of Non-catheter Related Late-onset Sepsis in Preterm Infants by Fecal Volatile Compounds Analysis: A Prospective, Multi-center Cohort Study," Pediatr. Infect. Dis. J., no. February, pp. 330–335, 2020, doi: 10.1097/INF.0000000000002589.

[11] A. Tiele, A. Wicaksono, S. K. Ayyala, and J. A. Covington, “Development of a compact, iot-enabled electronic nose for breath analysis,” Electron., vol. 9, no. 1, 2020, doi: 10.3390/electronics9010084.

[12] A. Tiele *et al.*, “Breath-based non-invasive diagnosis of Alzheimer’s disease: A pilot study,” J. Breath Res., vol. 14, no. 2, p. 026003, Feb. 2020, doi: 10.1088/1752-7163/ab6016.

[13] S. el Manouni el Hassani *et al.*, “Simultaneous Assessment of Urinary and Fecal Volatile Organic Compound Analysis in De Novo Pediatric IBD,” Sensors, vol. 19, no. 20, p. 4496, Oct. 2019, doi: 10.3390/s19204496.

[14] L. Lacey, E. Daulton, A. Wicaksono, J. A. Covington, and S. Quenby, “Detection of Group B Streptococcus in pregnancy by vaginal volatile organic compound analysis: a prospective exploratory study,” Transl. Res., vol. 216, pp. 23–29, 2020, doi: 10.1016/j.trsl.2019.09.002.

[15] E. Pérez-Calvo *et al.*, “The measurement of volatile organic compounds in faeces of piglets as a tool to assess gastrointestinal functionality,” Biosyst. Eng., vol. 184, pp. 122–129, 2019, doi: 10.1016/j.biosystemseng.2019.06.005.

[16] M. D. Rouvroye *et al.*, “Faecal scent as a novel non-invasive biomarker to discriminate between coeliac disease and refractory coeliac disease: A proof of principle study,” Biosensors, vol. 9, no. 2, 2019, doi: 10.3390/bios9020069.

[17] E. Mozdiak, A. N. Wicaksono, J. A. Covington, and R. P. Arasaradnam, “Colorectal cancer and adenoma screening using urinary volatile organic compound (VOC) detection: early results from a single-centre bowel screening population (UK BCSP),” Tech. Coloproctol., vol. 23, no. 4, pp. 343–351, 2019, doi: 10.1007/s10151-019-01963-6.

[18] A. Tiele, A. Wicaksono, J. Kansara, R. P. Arasaradnam, and J. A. Covington, “Breath analysis using enose and ion mobility technology to diagnose inflammatory bowel disease — A pilot study,” Biosensors, vol. 9, no. 2, 2019, doi: 10.3390/bios9020055.

[19] E. Stark, J. Pitt, A. Nur Wicaksono, K. Milanovic, V. Lush, and S. Hoover, “Odorveillance and the Ethics of Robotic Olfaction [Opinion],” IEEE Technol. Soc. Mag., vol. 37, no. 4, pp. 16–19, 2018, doi: 10.1109/MTS.2018.2876103.

[20] S. Bosch *et al.*, “P132 Detection and monitoring of IBD based on faecal volatile organic compounds,” J. Crohn’s Colitis, vol. 13, no. Supplement_1, pp. S154–S154, 2019, doi: 10.1093/ecco-jcc/jjy222.256.

[21] S. Esfahani, A. Wicaksono, E. Mozdiak, R. P. Arasaradnam, and J. A. Covington, “Non-invasive diagnosis of diabetes by volatile organic compounds in urine using FAIMs and FOX4000 electronic nose,” Biosensors, vol. 8, no. 4, 2018, doi: 10.3390/bios8040121.

[22] M. M. Widlak *et al.*, "Risk stratification of symptomatic patients suspected of colorectal cancer using faecal and urinary markers," Color. Dis., vol. 20, no. 12, pp. O335–O342, 2018, doi: 10.1111/codi.14431.

[23] D. J. C. Berkhout *et al.*, "Late-onset sepsis in preterm infants can be detected preclinically by fecal volatile organic compound analysis: A prospective, multicenter cohort study," Clin. Infect. Dis., vol. 68, no. 1, pp. 70–77, 2019, doi: 10.1093/cid/ciy383.

[24] S. Bosch *et al*., "Optimized Sampling Conditions for Fecal Volatile Organic Compound Analysis by Means of Field Asymmetric Ion Mobility Spectrometry," Anal. Chem., vol. 90, no. 13, pp. 7972–7981, 2018, doi: 10.1021/acs.analchem.8b00688.

[25] S. Bosch *et al*., "Differentiation between pediatric irritable bowel syndrome and inflammatory bowel disease based on fecal scent: Proof of principle study," Inflamm. Bowel Dis., vol. 24, no. 11, pp. 2468–2475, 2018, doi: 10.1093/IBD/IZY151.

[26] R. P. Arasaradnam, A. Wicaksono, H. O'Brien, H. M. Kocher, J. A. Covington, and T. Crnogorac-Jurcevic, "Noninvasive Diagnosis of Pancreatic Cancer Through Detection of Volatile Organic Compounds in Urine," Gastroenterology, vol. 154, no. 3, pp. 485-487.e1, 2018, doi: 10.1053/j.gastro.2017.09.054.

[27] L. Lacey, E. Daulton, A. Wicaksono, J. A. Covington, and S. Quenby, "Volatile organic compound analysis, a new tool in the quest for preterm birth prediction—an observational cohort study," Sci. Rep., vol. 10, no. 1, pp. 1–9, 2020, doi: 10.1038/s41598-020-69142-4.

[28] S. Bosch *et al.*, "The faecal scent of inflammatory bowel disease: Detection and monitoring based on volatile organic compound analysis," Dig. Liver Dis., no. xxxx, 2020, doi: 10.1016/j.dld.2020.03.007.

Conference Oral and Poster Presentations

[1] Tiele, A. Wicaksono, K. A. Ayyala, and J. A. Covington, "Development of a Compact, IoT-Enabled Electronic Nose for Breath Analysis," Poster Presentation at Owlstone Breath Biopsy Conference 2019, (Cambridge, UK), 13– 14 November 2019.

[2] J. Covington, A. Tiele, A. Wicaksono, E. Daulton, X. Li, V. Eyre, S. Clarke and S. Pearson, "Breath-based non-invasive diagnosis of Alzheimer's disease: A pilot study," Poster Presentation at Owlstone Breath Biopsy Conference 2018, (Cambridge, UK), 8 November 2018.

[3] A. Tiele, A. Wicaksono, J. Kansara, K. James, R.P. Arasaradnam and J.A. Covington, "Real-time biologic drug monitoring using Breath in Inflammatory Bowel Disease (IBD) – Early results," Poster Presentation at UEG Week 2018 – United European Gastroenterology, (Vienna, Austria), 20–24 October 2018.

[4] A. Tiele, A. Wicaksono, J. Kansara, R. Arasaradnam and J. Covington, “Breath analysis using eNose technology to diagnose inflammatory bowel disease – Early results,” Poster Presentation at 28th Anniversary World Congress on Biosensors (Miami, Florida, USA), 12–15 June 2018.

[5] S. Esfahani, A. Wicaksono, E. Mozdiak, R. Arasaradnam and J. Covington, “Non-invasive Diagnosis of Diabetes by Volatile Organic Compounds in Urine using FAIMS and FOX4000 Electronic Nose,” Poster Presentation at 28th Anniversary World Congress on Biosensors (Miami, Florida, USA), 12-15 June 2018.

Abstract

The scientific community has long been intrigued by the potential use of human waste odour in medical diagnosis and disease monitoring. Human waste odour analysis offers a fast and non-invasive way of testing and has the potential as an early disease diagnostic method. Cancer (particularly pancreatic and colorectal cancer), which has been known to be one of the deadly diseases of all time due to the late prognosis, may benefit from odour analysis as the current diagnostic method using highly invasive procedures such as colonoscopy. Other diseases, such as inflammatory bowel disease (IBD), which also employed colonoscopy in its diagnostic procedure, could also benefit from the application of human waste-based investigation.

In this work, data analysis pipelines employing five machine learning algorithms were developed for analytical instrument electronic nose (E-nose), field asymmetric ion mobility spectrometry (FAIMS), gas chromatography ion mobility spectrometry (GC-IMS), and gas chromatography time of flight mass spectrometry (GC-TOF-MS). These methods were used in 8 studies to investigate the potential of non-invasive diagnosis and monitoring of pancreatic cancer (PDAC), colorectal cancer (CRC), and IBD through urine, breath, and faecal odour analysis.

The PDAC studies included 285 subjects: 126 with PDAC, 45 with Chronic Pancreatitis (CP), and 114 controls, and utilised FAIMS, GC-IMS, and GC-TOF-MS to distinguish between groups based on the urinary VOC. All of the technologies were consistently able to distinguish between CRC and healthy control with an area under the curve (AUC) greater than 86%). FAIMS was also able to differentiate between early-stage PDAC from healthy and from late-stage PDAC (AUC 89% and 92%, respectively). However, urinary VOC analysis couldn't differentiate PDAC from CP (AUC 0.58). Further studies investigating urinary and faecal VOC of patient for detecting CRC using FAIMS and GC-IMS included 728 subjects which consisted of 26 patients with CRC, 340 with adenoma, 32 polypectomy patients, 296 healthy control and 33 with other gastrointestinal diseases. Both FAIMS and GC-IMS methods were able to separate between CRC and healthy control with AUC above 82%. Patients with adenoma could only be separated from healthy control in faecal odour analysis. Interestingly, odour analysis showed that the faecal VOC profile of patients whom underwent polypectomy returned to its physiological state after three months of polypectomy. This demonstrated the potential of faecal VOC analysis for disease monitoring, especially in adenoma and CRC. Final studies investigated the urinary, breath, and faecal VOC of patients with IBD. Six hundred forty-one subjects in total were recruited: 350 IBD, 276 healthy control, and 15 irritable bowel syndromes (IBS)/Functional Abdominal Pain-Not Otherwise Specified (FAP-

NOS). All three types of human waste were consistently able to detect IBD, both Crohn's Disease (CD) and Ulcerative Colitis (UC), from healthy control. Faecal odour analysis was also able to separate between IBD patients with IBS/FAP-NOS, which share many common symptoms.

In this work, the potential of utilising urine, breath, and faecal odour-based methods was successfully demonstrated for the non-invasive diagnosis of PDAC, CRC and IBD. The potential use of odour analysis in healthy monitoring was also shown in PDAC patients and patients undertaking polypectomy. Further work in this area will need to focus on sampling and storing samples as well as machine learning approach in order to speed up the advances in the application of modern E-nose technology in a clinical setting. Furthermore, a large-scale multi-centre study involving patients with similar symptoms will push further the knowledge of human waste odour for diagnosis.

Abbreviations

Before Century	BC
Tuberculosis	TB
Volatile Organic Compounds	VOC
Gas Chromatography Mass Spectrometry	GC-MS
Ion Mobility Spectrometry	IMS
Electronic Nose	E-nose
Inflammatory Bowel Disease	IBD
Crohn's Disease	CD
Ulcerative Colitis	UC
Field Asymmetric Ion Mobility Spectrometry	FAIMS
Gas Chromatography Ion Mobility Spectrometry	GC-IMS
Gas Chromatography Time Of Flight Mass Spectrometry	GC-TOF-MS
Pancreatic Ductal Adenocarcinoma	PDAC
Colorectal Cancer	CRC
Irritable Bowel Syndrome	IBS
Warwick OLFaction	WOLF
Non Dispersive Infra Red	NDIR
Photoionization Detector	PID
Ultraviolet	UV
Electron Volt	eV
Compensation Voltage	CV
Dispersion Field	DF
Direct Current	DC
Gas Chromatography	GC
Circular Gas Flow Unit	CGFU
Time of Flight	TOF
Time of Flight Mass Spectrometry	TOF-MS
Thermal Desorption	TD
Arbitrary Unit	A.U.
Reactant Ion Peak	RIP
Laboratory Analytical Viewer	LAV
Three Dimensions	3D
Two Dimensions	2D
Discrete Wavelet Transform	DWT
One Dimension	1D
Leave One Out Cross Validation	LOOCV
Condensed Nearest Neighbour Rule	CNN
Edited Nearest Neighbour Rule	ENN
Neighbourhood Cleaning Rule	NCL
Synthetic Minority Oversampling Technique	SMOTE
Sparse Logistic Regression	SLR
Random Forest	RF
Support Vector Machine	SVM
Neural Network	NN
Gaussian Process	GP
Malignant Pleural Mesothelioma	MPM
Least Absolute Shrinkage and Selection Operator	LASSO
Positive Predictive Value	PPV
Negative Predictive Value	NPV

Receiver Operating Characteristic	ROC
Area Under the ROC Curve	AUC
True Positive	TP
True Negative	TN
False Positive	FP
False Negative	FN
Confidence Interval	CI
Computed Tomography	CT
Magnetic Resonance Imaging	MRI
Advanced Adenoma	AA
Immunochemical Faecal Occult Blood Test	iFOBT
Bowel Cancer Screening Program	BCSP
Non-Applicable	NA
Standard Deviation	SD
American Joint Committee on Cancer	AJCC
Tumor Nodes Metastasized	TNM
Chronic Pancreatitis	CP
Pound per Square Inch	PSI
Atomic mass units	atu
Nuclear Magnetic Resonance	NMR
Liquid chromatography-tandem mass spectrometry	GeLC-MS/MS
Faecal Occult Blood Test	FOBT
The British Society of Gastroenterology	BSG
High risk	hr
Intermediate risk	ir
Low risk	lr
Medical Ethical Review Committee	METc
European Society of Gastrointestinal Endoscopy	ESGE
Large Adenoma	LA
Small Adenoma	SA
Body Mass Index	BMI
Antibiotics	ABx
High-grade dysplasia	HGD
Low-grade dysplasia	LGD
Faecal Immunochemical Test	FIT
Time before polypectomy	T0
Three months after polypectomy	T1
Faecal calprotectin	FCP
The National Health Service	NHS
Academic Medical Centre	AMC
Functional Abdominal Pain-Not Otherwise Specified	FAP-NOS
Magnetic Resonance Enteroclysis	MRE
Physician Global Assessment	PGA
C reactive protein	CRP
Healthy Control	HC
Interquartile range	IQR
Nonstricturing nonpenetrating	NSNP
Penetrating	P
Stricturing	S
perianal	p
University Hospitals Coventry and Warwickshire	UHCW
Simple colitis activity index	SCAI
Harvey Bradshaw Index	HBI

Part per billion	ppb
VU medical centre	VUmc
Crohn's Disease active	CDa
Crohn's Disease remission	CDr
Ulcerative Colitis active	UCa
Ulcerative Colitis remission	UCr
IBD active	IBDa
IBD remission	IBDr
Diarrhoea-predominant irritable bowel syndrome	IBS-D
Clostridium difficile infection	CDI
Linear Retention Index	LRI
Principal Component Analysis	PCA
Independent Component Analysis	ICA
Partial Least Squares Discriminant Analysis	PLS-DA
Multicapillary Column Ion Mobility Spectrometry	MCC-IMS
Principal Component	PC
Linear Discrimination Analysis	LDA

Chemicals

Oxygen	O_2
Ammonia	NH_3
Ethylene oxide	ETO
Sulphur dioxide	SO_2
Ozone	O_3
Nitric oxide	NO
Nitrogen dioxide	NO_2
Hydrogen sulphide	H_2S
Carbon Monoxide	CO
Hydrogen	H_2
Carbon dioxide	CO_2
ethane	CH_4

Chapter 1
INTRODUCTION

Introduction

Until the development of modern analytical methods, doctors mainly depended on their senses to diagnose and treat patients [1]. Doctors often touched, tasted, or smelled their patients' excrement to diagnose their illnesses [2]. This practice has been applied since 400 BC by the famous Greek physician, Hippocrates, who is known as the "Father of Medicine" [3]. He would set human sputum over hot coal to produce a unique smell in order to identify Tuberculosis (TB) in patients [4]. Smell has also been used by doctors to diagnose other illnesses. For example, by smelling the faecal Volatile Organic Compounds (VOC), Poulton *et al*. found that it is possible to identify rotavirus gastroenteritis before culture results were available [5]. It is not just humans who smell and detect different illnesses; animals such as dogs, rats, and even bees are known to sniff out and identify various diseases such as cancer, TB, and diabetes [6,7,8].

The detection of VOC from patients' waste allows for the identification of disease in a quick and non-invasive manner. However, employing a human or animal nose to perform disease diagnosis has several drawbacks. When dealing with infectious diseases, the use of humans or animal noses may not be possible due to the increased risk of transmission into the environment or among humans [9]. Furthermore, trained animals like canines are unable to perform for extended periods of time due to body and olfactory fatigue, which is the inability to identify specific smells after being exposed to the same smell for too long. These pros and cons of smell detection for disease diagnosis have inspired many researchers to investigate more into analytical tools as an alternative method to using human and animal olfactory systems.

Gas Chromatography-Mass Spectrometry (GC-MS) is frequently referred to as the "gold standard" of VOC analysis due to its excellent sensitivity, accuracy, and repeatability. However, this method has several constraints, including the high price tag, complexity, and long duration of analysis, which may restrict its popular use as a disease diagnostic tool [10]. Electronic nose (E-nose) instruments are a bioinspired analytical tool potentially offering a faster, smaller, more portable, and cheaper alternative. This instrument imitates mammalian olfactory systems by using several gas sensors to function as olfactory receptors and pattern recognition serving as the brain. The data produced by an E-nose's multiple non-specific gas sensor array, combined with pattern recognition, enable it to differentiate patients with specific diseases from healthy control. However, E-nose cannot provide the chemical information from

the VOCs detected, often suffer from sensor drift, and have poor selectivity due to the nature of the non-specific gas sensor array [11].

With the advances of technology, one of the analytical tools gaining traction for analysing VOCs and having high potential as an early diagnostic tool is Ion Mobility Spectrometry (IMS). Analytical tools based on IMS have better sensitivity, specificity, and do not experience as much sensor drift when compared to traditional E-noses. IMS also offers the possibility of chemical identification, which can help the understanding of the correlation between human waste VOCs and targeted disease. The use of IMS in medical diagnosis can be treated as an E-nose by combining the instrument with a pattern recognition/ machine learning algorithm. However, the IMS is generally more expensive than traditional E-nose and produces a high dimensional megavariate dataset that becomes challenging to process [12].

Cancer has significantly impacted human health over the course of human history. According to the most recent data, an estimated 19 million individuals in the world were diagnosed with cancer in 2020, and another 10 million people died of the disease in the same year [13]. Late diagnosis and limited access to appropriate treatments are probably the main factors for this high cancer death rate [14]. Thus, diagnosing cancer at an early stage, with high accuracy, is vital to improve the effectiveness of cancer therapy. A variety of cancer diagnostic techniques are now available, including biopsy and cancer imaging as well as mass spectrometry-based proteomics and fluorescence spectrometry-based tests. However, these techniques involve costly equipment and complex procedures, discouraging most people from seeking regular medical tests for early cancer signs, allowing it to develop to an irreversible stage later. Thus, the development of simple and affordable cancer diagnostic technologies based on human waste VOC analysis would have high clinical value in fighting cancer.

Furthermore, one of the diseases that often lead to cancer is inflammatory bowel disease (IBD) [15]. IBD patients have an up to 18% increased risk of getting colorectal cancer 30 years after their IBD diagnosis [16]. In addition, Over 6.8 million people worldwide, with 2 million Europeans, have inflammatory bowel disease (IBD) [17]. The two forms of IBD, Crohn's disease (CD) and ulcerative colitis (UC), are both characterised by inflammation of the gut. To diagnose IBD, colonoscopy with histology is regarded to be the "gold standard" [18]. Many people experience dread while facing this invasive procedure. Therefore, the development of VOC analysis techniques for IBD detection would help patients to get diagnosed without going through such an uncomfortable and embarrassing process. Thus, the studies reported in this thesis were primarily concerned with investigating cancer (particularly pancreatic and colorectal cancer) and IBD using human waste VOC analysis.

Research Aims

As previously stated, the use of VOC analysis of human waste for disease detection is gaining popularity in the scientific community and has received attention because of its potential for future medical applications. Despite this, the advances of VOC analytical tools create challenging problems in analysing the increasingly complex and high dimensional dataset. Hence, the aims of this research were to utilise modern E-nose instruments for disease diagnosis and monitoring applications by analysing human waste's VOCs from different diseases and analyse the data generated by them. To do that, this thesis reported the investigation of the possible use of human waste's VOC for disease diagnosis by combining electronic nose technology with chemical identification. The type of human waste investigated in this thesis are urine, stool, and breath. Additionally, the studies in this thesis concentrated on applications of cancer and IBD detection (sample media defined by availability).

To make it clear, below is the list of aims in cancer and IBD studies related to this thesis

1. To investigate the potential use of urinary VOC analysis to detect pancreatic cancer.
2. To look at possible biomarkers that may be associated with pancreatic cancer.
3. To investigate the potential use of urinary and faecal VOC analysis to detect colorectal cancer.
4. To investigate the use of faecal VOC analysis for secondary non-invasive follow up after polypectomy in patients with adenoma.
5. To investigate the potential use of urine, stool, and breath VOC analysis in IBD detection.
6. To look at possible breath biomarkers that may be associated with IBD.
7. To investigate the potential use of faecal VOC in monitoring the IBD activity in patients.

Thesis Outline

This thesis starts with a comprehensive literature review followed by the description of sensing instrumentation and data analysis pipeline utilised in the studies relating to this thesis. Following that, two major body chapters focus on discussing the application of the developed method for cancer and inflammatory bowel disease (IBD) detection. The outline of this thesis is provided below.

Chapter 2: Literature Review

Chapter 2 provides a literature review, which first describes the fundamental of the olfactory system. The Origin of human waste biomarkers, including exhaled breath, urine, and faecal, are summarised. Then, previous research on human waste odour for disease detection is also summarised. Finally, the current data analysis approaches, including feature extraction methods, multivariate analysis, and the use of machine learning in artificial olfaction, are then reviewed.

Chapter 3: Sensing Instrumentation

Chapter 3 present an overview of the analytical instrumentations utilised in this thesis. They are E-nose, FAIMS, GC-IMS, and GC-TOF-MS. Additionally, each instrument's working principle, medical use, and commercial product are explained. Finally, the advantages and disadvantages of each analytical instrument are reviewed.

Chapter 4: Data Analysis

Chapter 4 covers the newly developed data analysis pipeline for studies in chapters 5 and 6. This chapter begins with an explanation of the structure and dimension of each raw data produced by each instrument. Then, for each step in the data analysis pipeline, a thorough explanation is provided. Lastly, the data analysis pipeline for each instrument is summarised.

Chapter 5: Application on Cancer Detection

Chapter 5 describes new clinical studies examining the possibility of urine and faecal Volatile Organic Compounds (VOCs) as a non-invasive biomarker for the detection of Pancreatic Ductal Adenocarcinoma (PDAC) and Colorectal cancer (CRC). The instruments from Chapter 3 and the data analysis method from chapter 4 are applied in these four studies. The first and second studies focus on analysing the urinary VOC for PDAC detection using FAIMS (study 1), GC-IMS (study 2), and GC-TOF-MS (study 2). The third and fourth studies focus on detecting CRC by analysing urinary VOC using FAIMS and GC-IMS (study 3) and by analysing faecal VOC using GC-IMS (study 4). Additionally, the use of faecal VOC analysis for secondary non-invasive follow-up following polypectomy is also assessed. Finally, the result and comparisons of these studies are addressed, focusing on the various tools and biological types utilised.

Chapter 6: Application on Inflammatory Bowel Disease Detection

Chapter 6 discuss new clinical studies investigating how non-invasive VOC analysis from faecal, urine, and breath samples may be used to identify Inflammatory Bowel Disease (IBD). There are four studies that make use of the instruments described in Chapter 3 as well as the data analysis technique described in Chapter 4. The first study showed how FAIMS was used to investigate the faecal VOC to differentiate IBD from irritable bowel syndrome (IBS) in paediatrics. The second study describes the possibility of exhaled breath sample testing to diagnose IBD using E-nose and GC-IMS. The third study compares the performance of urine and faecal VOC analysis using GC-IMS to detect IBD. The last study discusses the possibility of faecal VOC analysis to be used to monitor IBD activity in patients. Finally, the results from all of these studies are compared and discussed, focusing on the various tools and biological types utilised.

Chapter 7: Conclusion and Further Work

Chapter 7 summarises the thesis covering the data analysis pipeline in Chapter 4 and the findings of the studies provided in Chapters 5 and 6, with particular emphasis on their originality and importance in the context of the broader research area. Afterwards, new study opportunities and suggestions for future work are discussed.

Reference

[1] L. D. J. Bos, P. J. Sterk, and M. J. Schultz, "Volatile Metabolites of Pathogens: A Systematic Review," *PLoS Pathog.*, vol. 9, no. 5, pp. 1–8, 2013, doi: 10.1371/journal.ppat.1003311.

[2] J. K. Nicholson and J. C. Lindon, "Metabonomics," *Nature*, vol. 455, no. 7216, pp. 1054–1056, Oct. 2008, doi: 10.1038/4551054a.

[3] F. Adams, "Aphorisms By Hippocrates." .

[4] E. Hong-Geller and S. Adikari, "Volatile Organic Compound and Metabolite Signatures as Pathogen Identifiers and Biomarkers of Infectious Disease," in *Biosensing Technologies for the Detection of Pathogens - A Prospective Way for Rapid Analysis*, vol. i, no. tourism, InTech, 2018, p. 13.

[5] J. Poulton and M. J. Tarlow, "Diagnosis of rotavirus gastroenteritis by smell," *Arch. Dis. Child.*, vol. 62, no. 8, pp. 851–852, 1987, doi: 10.1136/adc.62.8.851.

[6] M. McCulloch, T. Jezierski, M. Broffman, A. Hubbard, K. Turner, and T. Janecki, "Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers," *Integr. Cancer Ther.*, vol. 5, no. 1, pp. 30–39, 2006, doi: 10.1177/1534735405285096.

[7] Y. Oh, O. Kwon, S. S. Min, Y. B. Shin, M. K. Oh, and M. Kim, "Multi-odor discrimination by rat sniffing for potential monitoring of lung cancer and diabetes," *Sensors*, vol. 21, no. 11, pp. 1–14, 2021, doi: 10.3390/s21113696.

[8] D. M. Suckling and R. L. Sagar, "Honeybees Apis mellifera can detect the scent of *Mycobacterium tuberculosis*," *Tuberculosis*, vol. 91, no. 4, pp. 327–328, 2011, doi: 10.1016/j.tube.2011.04.008.

[9] Y. Salgirli Demirbaş *et al.*, "The role of bio-detection dogs in the prevention and diagnosis of infectious diseases: A systematic review," *Ankara Univ. Vet. Fak. Derg.*, vol. 68, no. 2, pp. 185–192, 2021, doi: 10.33988/auvfd.834133.

[10] S. O. Agbroko and J. Covington, "A novel, low-cost, portable PID sensor for the detection of volatile organic compounds," *Sensors Actuators, B Chem.*, vol. 275, no. July, pp. 10–15, 2018, doi: 10.1016/j.snb.2018.07.173.

[11] L. Capelli *et al.*, "Application and uses of electronic noses for clinical diagnosis on urine samples: A review," *Sensors (Switzerland)*, vol. 16, no. 10, pp. 1–23, 2016, doi: 10.3390/s16101708.

[12] E. Szymańska, A. N. Davies, and L. M. C. Buydens, "Chemometrics for ion mobility spectrometry data: Recent advances and future prospects," *Analyst*, vol. 141, no. 20, pp. 5689–5708, 2016, doi: 10.1039/c6an01008c.

[13] J. Ferlay *et al.*, "Global Cancer Observatory: Cancer Today.," *Lyon, France: International Agency for Research on Cancer.*, 2020. https://gco.iarc.fr/today (accessed Jan. 20, 2021).

[14] R. L. Siegel, A. Jemal, R. C. Wender, T. Gansler, J. Ma, and O. W. Brawley, "An assessment of progress in cancer control," *CA. Cancer J. Clin.*, vol. 68, no. 5, pp. 329–339, 2018, doi: 10.3322/caac.21460.

[15] M. Yalchin, A. M. Baker, T. A. Graham, and A. Hart, "Predicting colorectal cancer occurrence in ibd," *Cancers (Basel).*, vol. 13, no. 12, pp. 1–28, 2021, doi: 10.3390/cancers13122908.

[16] J. A. Eaden, "The risk of colorectal cancer in ulcerative colitis: a meta-analysis," *Gut*,

vol. 48, no. 4, pp. 526–535, Apr. 2001, doi: 10.1136/gut.48.4.526.

[17] V. Jairath and B. G. Feagan, “Global burden of inflammatory bowel disease,” *Lancet Gastroenterol. Hepatol.*, vol. 5, no. 1, pp. 2–3, 2020, doi: 10.1016/S2468-1253(19)30358-9.

[18] Y. G. Kim and B. I. Jang, “The Role of Colonoscopy in Inflammatory Bowel Disease,” *Clin. Endosc.*, vol. 46, no. 4, p. 317, 2013, doi: 10.5946/ce.2013.46.4.317.

Chapter 2
LITERATURE REVIEW

2.1. Introduction

This chapter provides an overview of the fundamental of artificial olfaction as well as a summary of previous odour analyses for disease detection. The fundamental of artificial olfaction covers the mechanism of the human olfactory system, the origin of breath, urine, and stool VOC. Then, the unique aroma generated from human waste, which is associated with certain diseases or disorders, are summarised. Lastly, the summary of pre-processing techniques, multivariate analysis, and machine learning are presented under the current VOC data analysis approaches section.

2.2. The Fundamental of Olfactory Systems

Since the basic concept of the method used in this thesis is heavily inspired by the mammalian/human olfactory system, it is important to first review on the fundamental of how olfactory systems works. The peripheral olfactory system was first described in detail by Santiago Cajal in 1890, which consist of the nostrils, ethmoid bone (cribriform plate), nasal cavity, and the olfactory epithelium, as illustrated in Figure 2.1 [1]. The olfactory epithelium is a thin layer of tissue which is covered by mucus that is located at the top of the nasal cavity. The main components of the olfactory epithelium are mucous membrane, olfactory glands (bowman's glands), and olfactory neurons. The mucous membrane is responsible for mucus storage while olfactory glands produce mucus and release metabolic enzymes present in mucus [2].

There are two ways for odour molecules to get into the nasal cavity: via nostril during inhalation or throat during chewing or swallowing food. Once these odour molecules enter the nasal cavity, odour are dissolved into mucus layer and detected by olfactory sensory neurons, which then send the signal into the brain in a process called sensory transduction [3]. Olfactory neurons are bipolar neurons which each neuron has one axon and one dendrite. The dendrites of these neurons have cilia (tiny hairs) that contain olfactory receptors. When odour molecules interact with these receptors, neuron fires an electrical signal that travels to olfactory bulb at the back of nasal cavity to be processed [4].

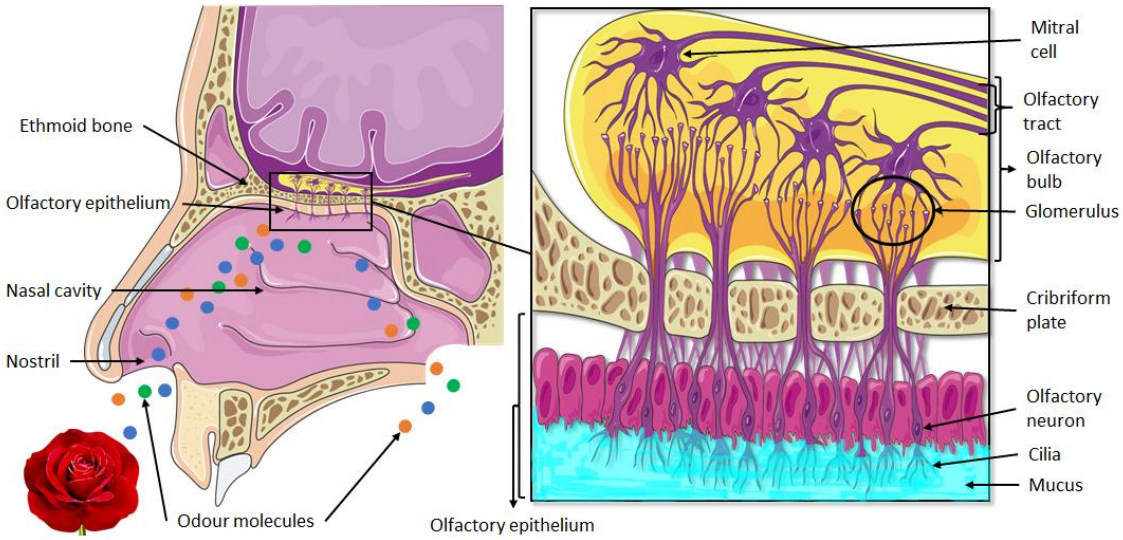


Figure 2.1. An illustration of Human Olfactory System

Until the beginning of the 1990s, researchers were not understood how humans had the capability of recognising and remembering thousands of different odours. However, in 1991 Linda Buck and Richard Axel published a paper titled "A novel multigene family may encode odorant receptors: a molecular basis for odour recognition" that answered this question [5]. They discovered that around 3% of our genes are responsible for the fine-tuning of olfactory receptor types. According to them, olfactory neurons have many types depending on their olfactory receptor. Each olfactory neuron has a specific odorant receptor protein in its cilia that is able to detect a limited number of odorant substances making it highly specialised for a few odours. When some odour molecules interact with the receptor protein, olfactory neurons exhibit an electric signal with various intensities depending on the type of odorant molecule. Olfactory neurons which have the same type of olfactory receptors combine and cluster their axons into the glomerulus inside the olfactory bulb. Inside the glomerulus, the electric signals from multiple olfactory neuron axons interact with mitral and tufted cell. These cells then activate and pass the signal to the brain through olfactory tract. Both of these types of cells help our brain later to determine how strong an odour is depending on the time-specific neuron clusters activated (referred to as the 'timing code'). Odours are comprised of a large number of different chemical odorant molecules, each of which stimulates numerous odorant receptors in the epithelium. All of these processes allow the olfactory system to create a combinatorial code of the odorant response, often called "odorant pattern". Because of this, we are capable of recognising and remembering thousand distinct odours. These contributions had brought Linda Buck and Richard Axel to win Nobel Prize in Physiology or Medicine in 2004 [6]. This understanding of biological olfactory system has inspired the development of the advanced electronic nose.

2.3. The Origin of Human Waste Biomarker

It is believed that body odours are caused by a mixture of thousands of odorous volatile organic compounds (VOCs) that are produced from different cells inside the body via metabolic pathways. Breath, urine, and stool are the main sources of VOCs. Knowing the biological origin and function of VOCs is important when it comes to something as vital as medical diagnosis. Here, the origin of biomarkers in breath, urine, and stool are explained.

2.3.1. Exhaled Breath

Respiratory gas exchange occurs when oxygen (O_2) enriches the blood while excess carbon dioxide (CO_2) is eliminated from the body through the lungs. This activity takes place in the lung's alveoli, and it is controlled by molecules' partial pressure and concentration gradients [7]. The diffusion of O_2 and CO_2 between alveolar gas and mixed venous blood enables the transfer of thousands of VOCs from blood to breath and vice versa. Because of this, the fingerprint of metabolic activity that is taking place endogenously must be present in the exhaled breath as the composition of exhaled breath is mostly a reflection of the volatile content of bloodstream. This gas exchange process is the fundamental concept behind biomarker presence in exhaled breath. This has been verified by comparing the amounts of certain VOCs in exhaled breath with the concentration of the same chemicals in the blood [8].

Exhaled breath biomarker research is primarily concerned with the identification and monitoring of gas-phase compounds. Nitrogen, oxygen, carbon dioxide, hydrogen [9], inert gases[10], and water vapour make up the majority of human exhaled breath. In addition, inorganic VOCs such as nitric oxide [11], nitrous oxide [11], ammonia [12], carbon monoxide [10], hydrogen sulphide [13], etc., and organic VOC such as acetone[14], ethanol, isoprene [15], etc., also presented in the exhaled breath. The latest study has reported a total of 1448 VOCs found in exhaled breath [16]. VOCs are produced as by-products of normal metabolic activity that happen inside the human body. However, illnesses may cause organs or surrounding tissue to produce VOCs in a different manner. In turn, this leads to the production of novel VOCs that the body does not naturally generate during regular physiological processes, as well as the variation of VOC concentrations that are greater or lower than normal levels [17]. In addition, many of the VOC in exhaled breath may have exogenous origins [18],

be generated by medicine [19], be released by bacteria in the airways [20], the oral cavity [21], or originated from the gut [22].

2.3.2. Urine

As previously explained in section 2.3.1., diseases can cause organs or surrounding tissue to generate slightly different metabolites waste that can be used as VOC biomarkers for disease diagnostic and monitoring. VOCs released by tissues are circulated around the body together with blood to be able to remove them from the body. One of the end results of metabolic pathways is urine. The production of urine happens in kidneys, in which blood experiences three stages of the process, namely filtration, reabsorption, and secretion, as shown in Figure 2.2 [23]. These stages assure that the body is cleansed of just waste, toxins, and excess water.

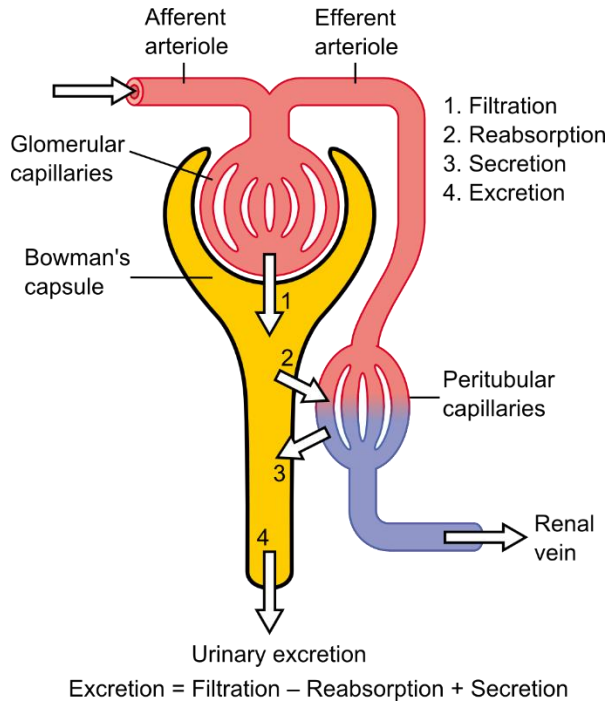


Figure 2.2. An Illustration of Urine Production in Nephron Inside Kidney [24].

The first stage is filtration. The filtering units inside the kidneys are known as nephrons which are made up of two parts: glomerulus and renal tubule. In each kidney of a healthy adult, there are 1-1.5 million nephrons [23]. Filtration happens when blood passes through glomerulus, which is a tiny capillary blood vessel. Due to the pressure of the blood, water and

small solutes are pushed through the filtration membrane, leaving only blood cells and large proteins in the bloodstream.

The result of filtration includes not only waste but also other important things that the body needs, such as ions, small protein, amino acids, and glucose. Hence, this filtrate fluid flows into the renal tubule for the reabsorption process, where the necessary chemicals are reabsorbed through the wall of the renal tube. Simultaneously, waste ions exit the capillaries and enter the renal tubule. This process is referred to as secretion. Finally, these remaining filtrate and waste ions exit the kidney through the renal pelvis, the ureter, and the bladder. Urine is about 91-96% water, and the remaining composition is waste products [25]. Urine output in adults ranges from 1-2 litres per day, depending on hydration, activity level, environmental variables, body weight, and overall health. The latest study has reported a total of 444 VOCs found in urine with the largest compounds identified belong to ketone group [16]. Because ketones were detected in much lower quantities in the urine of 'germ-free' rats, it is probable that ketones in urine are produced at least in part by bacterial activity in the gut, perhaps via decarboxylation of the corresponding oxo-acids [26].

2.3.3. Faecal

There are many species of bacteria in the human gut microbiota, and it is believed that these bacteria contribute to the integrity of mucosal as well as the maintenance and protection of host health as a result of defending against invading organisms [27]. Because of this, it is possible that the interaction between food and these gut microbiotas will result in the formation of faecal VOC that are shared by various people. For example, *Bacteroides*, in particular, are responsible for the fermentation of carbohydrates in the gut, which results in the production of ethanoic, propionic, butanoic, pentanoic, and hexanoic acids [28]. On the other hand, it is possible that gastrointestinal illnesses will result in a distinct pattern of VOC production [29]. A recent study by Petal *et al*. showed a change in faecal VOC profile of patients with *Clostridium difficile* infection [30].

A recent study reported a total of 443 VOC found in faecal of healthy people [16]. Compared to the number of VOC found in the breath, there is still more qualitative analysis study that needs to be done on faecal samples. A recent study suggests that the mouth microbiome may have an effect on the gut microbiota [31]. Olsen and Yamazaki investigated patients with chronic periodontitis and they discovered that the *Porphyromonas gingivalis*

bacteria from mouth produces a dysbiosis in the stomach, which results in dysregulation of the gut microbiota [31].

2.4. Previous Research on Human Waste Odour for Disease Detection

For many years, in addition to physical exams, diagnosing human illnesses using the sense of smell remained one of the most reliable techniques used in clinical medicine. The smells released by the patient's body were previously regarded as one of the most important early diagnostic indicators. As a result of the important emphasis on the assessment of patient odour, back in 1982, Fitgerald and Tierney presented a variety of signs and additional diagnostic methods that offered valuable relevant information for clinical diagnostic examinations [32]. Their publication was inspired by the worry that smell prognosis might be forgotten to the future generation physicians. They also provided several examples of smell bioindicators for various disease to highlight the need of maintaining the smell diagnostic skill in the medical community. Numerous studies have linked unique individual human smells to certain human diseases and disorder. Table 2.1 shows the list of some diseases with sample type and their associate aromas.

Table 2.1. A list of descriptive aromas with its associate disease and sample type

Sources	VOC sample type	Disease	Descriptive aroma
Liddell *et al*. [33]	Breath	Diabetic ketoacidosis	Rotting apples, acetone
Pavlou and Turner [34]	Breath	Diabetes mellitus	Acetone-like
Hayden [35]	Breath	Empyema (anaerobic)	Foul, putrid
Smith *et al*. [36]	Breath	Liver failure	Dimethyl sulphide
Hayden [35]	Breath	Uremia	Fishy, ammonia, urine-like
Pavlou and Turner [34]	Urine	Trimethylaminuria	Fishy
Pavlou and Turner [34]	Urine	Bladder infection	Ammonia
Najarian [37]	Urine	Azotemia (prerenal)	Concentrated urine odour
Pavlou and Turner [34]	Stool	Rotavirus gastroenteritis	Full

As shown in Table 2.1, a distinctive odour is linked to a variety of diseases. Liddel *et al*. for example, described in their report that fruity aroma rotting apples could be an early sign of diabetic ketosis. However, smelling patient waste directly through the nose imposed many drawbacks from the increasing risks of disease transmission to the high possibility of smelling

and decision fatigue experienced by physicians which may lead to misdiagnosis. Replacing human olfactory systems with machine is more desirable as to avoid potential problems mentioned above. This machine is often called an electronic nose (E-nose). A deep explanation of the working principle of E-nose can be found in Chapter 3. In 1997, Ping *et al*. used E-nose technology to examine exhaled breath odour of 32 subjects which consisted of 18 diabetic patients and 14 healthy normal persons to detect diabetes [38]. Their study shows that analysing the exhaled breath after the meal might be a reliable monitoring method for diabetes diagnosis. More modern analytical tool that implements the principle of E-nose has been emerging due to the simple and fast analysis time nature of E-nose. Some of them are field asymmetric ion mobility spectrometry (FAIMS), gas chromatography ion mobility spectrometry (GC-IMS), and gas chromatography time of flight mass spectrometry (GC-TOF-MS). The detailed working principle of these methods can be found in Chapter 3.

In 2014, Ramesh *et al*. investigated the urinary VOC profile of patients in order to differentiate those with coeliac disease from irritable bowel syndrome using FAIMS [39]. They found out that FAIMS was able to achieve 0.91 (0.83 - 0.99) ROC Curve AUC with 85% of sensitivity and specificity when using sparse logistic regression as the classification model. On the other hand, Marije *et al*. used FAIMS to diagnose *Clostridium difficile* Infection (CDI) on hospital wards by analysing the stool VOC of patients [40]. Using a random forest algorithm, FAIMS can detect CDI with an AUC of 91% (0.86 - 0.97). When deploying FAIMS to analyse breath VOC samples, Sahota *et al*. found that FAIMS is capable of detecting patients with tuberculosis from healthy with 81% sensitivity and 79% specificity [41]. From these, FAIMS shows promising potential as one of the tools for noninvasive diagnostic.

On the other hand, Baumbach *et al*. used GC-IMS to investigate volatile metabolites in human exhaled air [42]. In their research, they found out that GC-IMS was capable of differentiating patients suffering from pneumonia from healthy persons. As pneumonia is one of the symptoms of lung cancer, this finding gave hope that metabolic VOC analysis has the potential for an early diagnosis of lung cancer. This research has inspired other researchers to explore more into the capability of GC-IMS in terms of medical diagnosis, from cancer detection [16,17] bacterial infection monitoring [18,19] to detecting gastrointestinal diseases [20,21].

When it comes to GC-TOF-MS, one of the studies that used this method is a study by Haiyan *et al*. in 2019 [49]. They conducted research that investigated the VOC of vitreous and aqueous humour in patients as potential biomarkers to detect diabetic retinopathy. It turned

out that GC-TOF-MS coupled with logistic regression was able to differentiate the diabetic retinopathy patient from healthy control with the AUC, sensitivity, and specificity of 0.965, 88%, and 95.7%, respectively, when using aqueous humour. Similar performance of GC-TOF-MS was also achieved when using vitreous, with 0.951 AUC, 95.5% sensitivity, and 85.7% specificity [49].

GC-TOF-MS was also used to detect inflammation biomarkers in the lung by in vitro stimulation using A549 epithelial cells [50]. Delphine *et al*. introduced chemical and biological stress into the cell by exposing it to H_2O_2 and an inflammatory pool of sputum supernatants, making the cell inflamed. Based on the type of induced inflammation, the VOC profile detected by GC-TOF-MS differs significantly when compared to healthy control cells. Using NIST 2014 mass spectra library combined with chromatographic data such as the linear retention index (LRI), an increase of carbonyl compounds and hydrocarbons were observed. This shows the potential of GC-TOF-MS as a biomarker discovery tool and a highly effective tool for VOC analysis.

2.5. Current Data Analysis Approaches

The fundamental theory of metabolomics waste of humans for disease diagnosis has been explained in the previous section and shown promising potential. Nevertheless, the advances of technology in gas analytical instrument creates issues in terms of interpreting VOC metabolomics data. Even though the sensitivity and selectivity of instruments increase, the resolution and complexity of raw data generated by these analytical methods are also increased, posing a new challenge for researcher to solve. Below, a brief overview of pre-processing techniques, multivariate analysis, and machine learning are presented.

2.5.1. Pre-processing

Pre-processing process becomes a crucial stage in the process of analysing odour metabolomics as the data is increasingly complex. In 2002, Marthinelli *et al*. proposed a pre-processing technique called Independent Component Analysis (ICA) [51]. This pre-processing technique had improved the separation of apple based on their ripeness stages using Partial Least Squares Discriminant Analysis (PLS-DA) and Principal Component Analysis (PCA). More traditional method of pre-processing in gas sensor array E-nose involves filtering, noise reduction, and standardisation [52].

More modern analytical tools like IMS requires more complex pre-processing step due to the nature of high dimensionality raw dataset that it produced. Ewa *et al*. proposed data size reduction strategy for the classification of exhaled breath-based biomarkers [53]. The step she suggested involves data alignment, denoising and compression in the wavelet domain, background correction, and mask construction. Using these pre-processing steps, they claimed to be able to reduce the 500.000 data points in Multicapillary Column Ion mobility Spectrometry (MCC-IMS) to just around 200 data points. Employing this technique may speed up the process of analysis.

2.5.2. Multivariate Analysis and Machine Learning

In the process of traditional E-nose data, multivariate technique is typically better suited since hundreds of variables may be handled concurrently. Multivariate method examines the correlation and covariance of the whole dataset. One of the most famous multivariate techniques used in dealing with E-nose data is Principal Component Analysis (PCA). It is an unsupervised linear technique that reduce the number of dimensions in the data by focusing on a limited number of principal components (PCs) that are linearly uncorrelated and account for the most of the variance [54]. The first PC is a linear collection of characteristics that represents that data's greatest variance. The next PC is orthogonal to the preceding one and reflect the greatest variance left after the preceding PC, and so on. The PCA result is shown using a score plot, in which each sample is represented by a single point indicating the relationships between all samples. The data points that clustered have similar characteristics, whereas points farther apart have dissimilar properties. A successful PCA implementation will create a distinct cluster of points in which little or no overlap is observed. In this case, the first two PCs usually will be accounted for >90% of the variance [55].

PCA and other unsupervised techniques reduce dimensionality and also show group clusters when within-group variance is less than the variation between groups. Meanwhile, supervised learning methods may be used to investigate pattern within samples and associate it with certain health conditions. Linear Discriminant Analysis (LDA) is a supervised linear technique that attempts to discover a linear discriminant function that may be used to distinguish certain groups [56]. Although this is a quick and effective method, it should be used only when the number of samples is bigger than the features [57]. Partial least-square discriminant analysis (PLS-DA) is another linear supervised method that works by creating new dimensions based on the linear combinations of input features. Although these techniques are powerful, it usually not capable when dealing with complex non-linear data. Classification

algorithms used in non-linear statistical learning methods are classified into 3, which are tree-based technique, neural networks, and kernel-based method.

2.6. Conclusion

This section has reviewed the fundamental theory behind human waste odours and olfactory systems. The human waste aroma associated with specific diseases and disorders has also been summarised. The advances of analytical tool technology have also pushed the advances of data analysis field. The potential of odour analysis as a diagnostic and monitoring method looks promising. The next chapter will explain more about each modern analytical technology used throughout this thesis.

2.7. Reference

[1] C. Levine and A. Marcillo, "Origin and endpoint of the Olfactory Nerve fibers: As described by Santiago Ramón y Cajal," *Anat. Rec.*, vol. 291, no. 7, pp. 741–750, 2008, doi: 10.1002/ar.20660.

[2] D. Purves *et al.*, *Neuroscience 6th Edition*, vol. 52, no. 10. 2018.

[3] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, *Molecular Cell Biology 4th edition*. New York: W.H. Freeman, 2000.

[4] A. Rinaldi, "The scent of life. The exquisite complexity of the sense of smell in animals and humans," *EMBO Rep.*, vol. 8, no. 7, pp. 629–633, 2007, doi: 10.1038/sj.embor.7401029.

[5] L. B. Buck, "A novel multigene family may encode odorant receptors.," *Soc. Gen. Physiol. Ser.*, vol. 47, pp. 39–51, 1992.

[6] NobelPrize.org, "Press release," *Nobel Prize Outreach AB*, 2021. https://www.nobelprize.org/prizes/medicine/2004/press-release/ (accessed Sep. 25, 2021).

[7] J. A. Collins, A. Rudenski, J. Gibson, L. Howard, and R. O'Driscoll, "Relating oxygen partial pressure, saturation and content: The haemoglobin–oxygen dissociation curve," *Breathe*, vol. 11, no. 3, pp. 194–201, 2015, doi: 10.1183/20734735.001415.

[8] P. Mochalski *et al.*, “Blood and breath levels of selected volatile organic compounds in healthy volunteers,” *Analyst*, vol. 138, no. 7, pp. 2134–2145, 2013, doi: 10.1039/c3an36756h.

[9] T. Hibbard and A. J. Killard, “Breath ammonia analysis: Clinical application and measurement,” *Crit. Rev. Anal. Chem.*, vol. 41, no. 1, pp. 21–35, 2011, doi: 10.1080/10408347.2011.521729.

[10] W. Lindinger and A. Hansel, “Analysis of trace gases at ppb levels by proton transfer reaction mass spectrometry (PTR-MS),” *Plasma Sources Sci. Technol.*, vol. 6, no. 2, pp. 111–117, 1997, doi: 10.1088/0963-0252/6/2/004.

[11] M. Mürtz, “Breath Diagnostics Using Laser Spectroscopy,” *Opt. Photonics News*, vol. 16, no. 1, p. 30, Jan. 2005, doi: 10.1364/OPN.16.1.000030.

[12] W. Lindinger, A. Hansel, and A. Jordan, “On-line monitoring of volatile organic compounds at pptv levels by means of Proton-Transfer-Reaction Mass Spectrometry (PTR-MS) Medical applications, food control and environmental research,” *Int. J. Mass Spectrom. Ion Process.*, vol. 173, no. 3, pp. 191–241, 1998, doi: 10.1016/s0168-1176(97)00281-4.

[13] C. Warneke, J. Kuczynski, A. Hansel, A. Jordan, W. Vogel, and W. Lindinger, “Proton transfer reaction mass spectrometry (PTR-MS): Propanol in human breath,” *Int. J. Mass Spectrom. Ion Process.*, vol. 154, no. 1–2, pp. 61–70, 1996, doi: 10.1016/0168-1176(96)04369-8.

[14] Z. Wang and C. Wang, “Is breath acetone a biomarker of diabetes? A historical review on breath acetone measurements,” *J. Breath Res.*, vol. 7, no. 3, 2013, doi: 10.1088/1752-7155/7/3/037109.

[15] A. Bajtarevic *et al.*, “Noninvasive detection of lung cancer by analysis of exhaled breath,” *BMC Cancer*, vol. 9, p. 348, 2009, doi: 10.1186/1471-2407-9-348.

[16] N. Drabińska *et al.*, “A literature survey of all volatiles from healthy human breath and bodily fluids: The human volatilome,” *J. Breath Res.*, vol. 15, no. 3, 2021, doi: 10.1088/1752-7163/abf1d0.

[17] K. Schmidt and I. Podmore, “Current Challenges in Volatile Organic Compounds Analysis as Potential Biomarkers of Cancer,” *J. Biomarkers*, vol. 2015, pp. 1–16, 2015, doi: 10.1155/2015/981458.

[18] J. D. Pleil, M. A. Stiegel, and T. H. Risby, “Clinical breath analysis: Discriminating between human endogenous compounds and exogenous (environmental) chemical confounders,” *J. Breath Res.*, vol. 7, no. 1, 2013, doi: 10.1088/1752-7155/7/1/017107.

[19] J. Beauchamp, “Inhaled today, not gone tomorrow: Pharmacokinetics and environmental exposure of volatiles in exhaled breath,” *J. Breath Res.*, vol. 5, no. 3, 2011, doi: 10.1088/1752-7155/5/3/037103.

[20] W. Filipiak *et al.*, “Characterisation of volatile metabolites taken up by or released from Streptococcus pneumoniae and Haemophilus influenzae by using GC-MS,” *Microbiol. (United Kingdom)*, vol. 158, no. 12, pp. 3044–3053, 2012, doi: 10.1099/mic.0.062687-0.

[21] T. Wang, A. Pysanenko, K. Dryahina, P. Španěl, and D. Smith, “Analysis of breath, exhaled via the mouth and nose, and the air in the oral cavity,” *J. Breath Res.*, vol. 2, no. 3, 2008, doi: 10.1088/1752-7155/2/3/037013.

[22] B. P. J. De Lacy Costello, M. Ledochowski, and N. M. Ratcliffe, “The importance of methane breath testing: A review,” *J. Breath Res.*, vol. 7, no. 2, 2013, doi: 10.1088/1752-7155/7/2/024001.

[23] C. Lote, *Principles of Renal Physiology Fifth Edition*. Springer-Verlag New York, 2012.

[24] Madhero, “Physiology of Nephron.” https://en.wikipedia.org/wiki/File:Physiology_of_Nephron.png (accessed Sep. 29, 2021).

[25] C. Rose, A. Parker, B. Jefferson, and E. Cartmell, “The characterisation of feces and urine: A review of the literature to inform advanced treatment technology,” *Crit. Rev. Environ. Sci. Technol.*, vol. 45, no. 17, pp. 1827–1879, 2015, doi: 10.1080/10643389.2014.1000761.

[26] G. A. Mills and V. Walker, “Headspace solid-phase microextraction profiling of volatile compounds in urine: Application to metabolic investigations,” *J. Chromatogr. B Biomed. Sci. Appl.*, vol. 753, no. 2, pp. 259–268, 2001, doi: 10.1016/S0378-4347(00)00554-5.

[27] C. S. J. Probert, F. Ahmed, T. Khalid, E. Johnson, S. Smith, and N. Ratcliffe,

"Volatile organic compounds as diagnostic biomarkers in gastrointestinal and liver diseases," *J. Gastrointest. Liver Dis.*, vol. 18, no. 3, pp. 337–343, 2009.

[28] F. Bäckhed, R. E. Ley, J. L. Sonnenburg, D. A. Peterson, and J. I. Gordon, "Host-bacterial mutualism in the human intestine," *Science (80-.).*, vol. 307, no. 5717, pp. 1915–1920, 2005, doi: 10.1126/science.1104816.

[29] C. E. Garner *et al.*, "Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease," *FASEB J.*, vol. 21, no. 8, pp. 1675–1688, 2007, doi: 10.1096/fj.06-6927com.

[30] M. Patel, D. Fowler, J. Sizer, and C. Walton, "Faecal volatile biomarkers of *Clostridium difficile* infection," *PLoS One*, vol. 14, no. 4, pp. 1–15, 2019, doi: 10.1371/journal.pone.0215256.

[31] I. Olsen and K. Yamazaki, "Can oral bacteria affect the microbiome of the gut?," *J. Oral Microbiol.*, vol. 11, no. 1, 2019, doi: 10.1080/20002297.2019.1586422.

[32] F. T. Fitzgerald and L. M. Tierney, "The bedside Sherlock Holmes.," *West. J. Med.*, vol. 137, no. 2, pp. 169–175, 1982, doi: 10.1016/s0196-0644(83)80631-3.

[33] K. Liddell, "Smell as a diagnostic marker," *Postgrad. Med. J.*, vol. 52, no. 605, pp. 136–138, Mar. 1976, doi: 10.1136/pgmj.52.605.136.

[34] A. K. Pavlou and A. P. F. Turner, "Sniffing out the Truth: Clinical Diagnosis Using the Electronic Nose," *Clin. Chem. Lab. Med.*, vol. 38, no. 2, pp. 99–112, Jan. 2000, doi: 10.1515/CCLM.2000.016.

[35] G. F. Hayden, "Olfactory diagnosis in medicine," *Postgrad. Med.*, vol. 67, no. 4, pp. 110–118, 1980, doi: 10.1080/00325481.1980.11715427.

[36] M. Smith, L. Smith, and B. Levinson, "THE USE OF SMELL IN DIFFERENTIAL DIAGNOSIS," *Lancet*, vol. 320, no. 8313, pp. 1452–1453, Dec. 1982, doi: 10.1016/S0140-6736(82)91343-5.

[37] M. Major Robert H; Moser, "The Diagnostic Importance of the Odor of Urine," *N. Engl. J. Med.*, vol. 303, no. 19, pp. 1128–1128, Nov. 1980, doi: 10.1056/NEJM198011063031925.

[38] W. Ping, T. Yi, X. Haibao, and S. Farong, "A novel method for diabetes diagnosis based on electronic nose," *Biosens. Bioelectron.*, vol. 12, no. 9–10, pp. 1031–1036,

1997, doi: 10.1016/S0956-5663(97)00059-6.

[39] R. P. Arasaradnam *et al.*, "Differentiating coeliac disease from irritable bowel syndrome by urinary volatile organic compound analysis - A pilot study," *PLoS One*, vol. 9, no. 10, pp. 1–9, 2014, doi: 10.1371/journal.pone.0107312.

[40] M. K. Bomers *et al.*, "Rapid, Accurate, and on-site detection of *C. difficile* in stool samples," *Am. J. Gastroenterol.*, vol. 110, no. 4, pp. 588–594, 2015, doi: 10.1038/ajg.2015.90.

[41] A. S. Sahota *et al.*, "A simple breath test for tuberculosis using ion mobility: A pilot study," *Tuberculosis*, vol. 99, pp. 143–146, 2016, doi: 10.1016/j.tube.2016.05.005.

[42] J. I. Baumbach, W. Vautz, V. Ruzsanyi, and L. Freitag, "Early Detection of Lung Cancer: Metabolic Profiling of Human Breath with Ion Mobility Spectrometers," *Mod. Biopharm. Des. Dev. Optim.*, vol. 3, pp. 1343–1358, 2008, doi: 10.1002/9783527620982.ch55.

[43] H. Ha, A. Usuba, S. Maddula, J. I. Baumbach, M. Mineshita, and T. Miyazawa, "Exhaled breath analysis for lung cancer detection using ion mobility spectrometry," *PLoS One*, vol. 9, no. 12, pp. 1–13, 2014, doi: 10.1371/journal.pone.0114555.

[44] A. S. Bannaga, H. Tyagi, E. Daulton, J. A. Covington, and R. P. Arasaradnam, "Exploratory Study Using Urinary Volatile Organic Compounds for the Detection of Hepatocellular Carcinoma," *Molecules*, vol. 26, pp. 1–11, 2021, [Online]. Available: https://doi.org/10.3390/molecules26092447.

[45] J. M. Lewis, R. S. Savage, N. J. Beeching, M. B. J. Beadsworth, N. Feasey, and J. A. Covington, "Identifying volatile metabolite signatures for the diagnosis of bacterial respiratory tract infection using electronic nose technology: A pilot study," *PLoS One*, vol. 12, no. 12, pp. 1–10, 2017, doi: 10.1371/journal.pone.0188879.

[46] E. Daulton, A. Wicaksono, J. Bechar, J. A. Covington, and J. Hardwicke, "The detection of wound infection by ion mobility chemical analysis," *Biosensors*, vol. 10, no. 3, pp. 1–9, 2020, doi: 10.3390/bios10030019.

[47] M. D. Rouvroye *et al.*, "Faecal scent as a novel noninvasive biomarker to discriminate between coeliac disease and refractory coeliac disease: A proof of principle study," *Biosensors*, vol. 9, no. 2, 2019, doi: 10.3390/bios9020069.

[48] A. B. Ballinger and C. Anggiansah, “Colorectal cancer,” *Br. Med. J.*, vol. 335, no. 7622, pp. 715–718, 2007, doi: 10.1136/bmj.39321.527384.BE.

[49] H. Wang *et al.*, “Metabolomic profile of diabetic retinopathy: a GC-TOFMS-based approach using vitreous and aqueous humor,” *Acta Diabetol.*, vol. 57, no. 1, pp. 41–51, 2020, doi: 10.1007/s00592-019-01363-0.

[50] D. Zanella *et al.*, “Comparison of the effect of chemically and biologically induced inflammation on the volatile metabolite production of lung epithelial cells by GC×GC-TOFMS,” *Analyst*, vol. 145, no. 15, pp. 5148–5157, 2020, doi: 10.1039/d0an00720j.

[51] E. Martinelli, C. Falconi, A. D’Amico, and C. Di Natale, “Pre-processing of Electronic Nose Data by Independent Component Analysis,” *Proc. IEEE Sensors*, vol. 1, no. 2, pp. 1339–1342, 2002, doi: 10.1109/icsens.2002.1037313.

[52] L. Cheng, Q. H. Meng, A. J. Lilienthal, and P. F. Qi, “Development of compact electronic noses: A review,” *Meas. Sci. Technol.*, vol. 32, no. 6, 2021, doi: 10.1088/1361-6501/abef3b.

[53] E. Szymańska, E. Brodrick, M. Williams, A. N. Davies, H. J. Van Manen, and L. M. C. Buydens, “Data size reduction strategy for the classification of breath and air samples using multicapillary column-ion mobility spectrometry,” *Anal. Chem.*, vol. 87, no. 2, pp. 869–875, 2015, doi: 10.1021/ac503857y.

[54] K. A. Wlodzimirow *et al.*, “Exhaled breath analysis with electronic nose technology for detection of acute liver failure in rats,” *Biosens. Bioelectron.*, vol. 53, pp. 129–134, 2014, doi: 10.1016/j.bios.2013.09.047.

[55] G. Peng *et al.*, “Diagnosing lung cancer in exhaled breath using gold nanoparticles,” *Nat. Nanotechnol.*, vol. 4, no. 10, pp. 669–673, 2009, doi: 10.1038/nnano.2009.235.

[56] A. Smolinska, A. C. Hauschild, R. R. R. Fijten, J. W. Dallinga, J. Baumbach, and F. J. Van Schooten, “Current breathomics - A review on data pre-processing techniques and machine learning in metabolomics breath analysis,” *J. Breath Res.*, vol. 8, no. 2, 2014, doi: 10.1088/1752-7155/8/2/027105.

[57] A. W. Boots, J. J. B. N. Van Berkel, J. W. Dallinga, A. Smolinska, E. F. Wouters, and F. J. Van Schooten, “The versatile use of exhaled volatile organic compounds in human health and disease,” *J. Breath Res.*, vol. 6, no. 2, 2012, doi: 10.1088/1752-

7155/6/2/027108.

Chapter 3
SENSING INSTRUMENT

3.1. Introduction

This chapter will provide an overview of the analytical methods used for the work conducted throughout this thesis. This will include the Electronic Nose (E-nose), Field Asymmetric Ion Mobility Spectrometry (FAIMS), Gas Chromatography Ion Mobility Spectrometry (GC-IMS), and Gas Chromatography Time-Of-Flight Mass Spectrometry (GC-TOF-MS). The operating principle, medical application, and description of the commercial products for each method are also described. Lastly, the pros and cons of each method are compared.

3.2. Electronic Nose (E-nose)

An electronic nose (E-nose) is an instrument that attempts to mimic the mammalian olfactory system. It consists of an array of gas sensors with partial specificity that performs the same task as olfactory receptor cells in the roof of the nasal cavity, a data acquisition system that acts as an olfactory bulb, and a pattern recognition system that acts as the brain, capable of recognising simple and complex odours. Figure 3.1 shows the similarity between the biological olfactory system and the electronic nose technology [1].

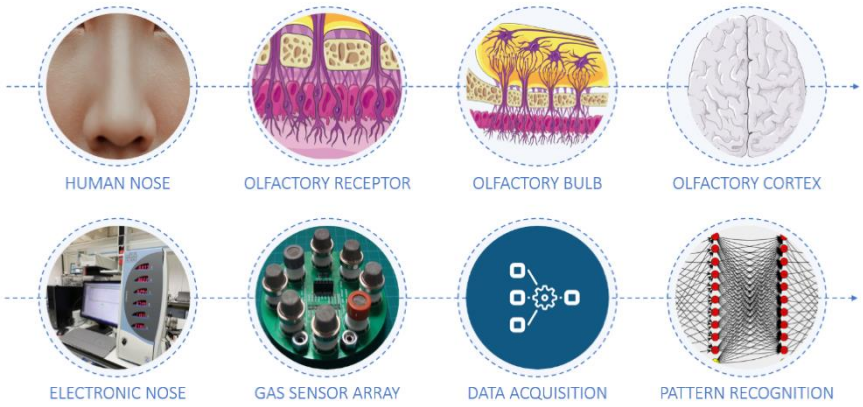


Figure 3.1. Electronic Nose Concept[1]

Moncrieff's work in 1961 on the development of a mechanical nose for detecting odours is arguably to be the earliest known work on a mammalian inspired instrument [2]. Then, research on redox reaction of odorants at an electrode and modulation of contact potential odorants, in 1964 and 1965 progressed the maturity of this instrument [3,4]. However, it was not until almost 20 years later that Persaud and Dodd (1982) at Warwick University, UK [5] and Kaneyasu *et al.* (1987) at Hitachi, Japan [6] published papers on the idea of using an

electronic nose with a chemical array sensor system for odour classification. Then, in 1987, the ‘electronic nose’ term was introduced by Gardner through his work in an international conference [7]. Since then, the food and beverage industry has been the first to apply E-nose technology and it remains the largest sector ever since. For example, it has been used to detect adulteration in meat, assessing the quality of tea and coffee, determining the roasting level in cocoa beans and many more [8]. The key point of E-nose technology is that it can be taught to recognise new smells. It is also highly customisable, can be made portable, easy to use and at a relatively low-cost compared to other gas analysis technologies. Recently, the number of papers on using the E-nose for medical applications has considerably increased [9]. This trend is likely to continue due to the expanding need for rapid, low-cost, and non-invasive techniques of diagnosis.

3.2.1. Operating Principle

The basic stages of how an E-nose system works is shown in Figure 3.2. First, the sample release a multitude of VOCs. Then, the VOCs will be drawn into the instrument and interact with the sensor array in some way. The most common response will cause an electrical change (e.g., change in voltage, current, frequency or resistance parameter) to occur during the interaction. This will create a signal that is then measured to form a raw dataset. Each sensor used in an E-nose is cross-reactive, meaning that it will respond to a wide range of chemical compounds. It is also important to have sensor diversity in the array, which means each sensor has a different response magnitude to other sensors when being exposed to certain VOC. This will enable the E-nose to identify complex VOCs that may comprise hundreds of individual chemicals. Instead of only analysing an individual sensor signal, the combination of signals from all of the sensors will create a unique response pattern, often called a "smell-print". Through recognising this smell-print pattern, using an appropriate pattern recognition system, the E-nose has the capability to classify and identify complex VOC without having to separate it into its individual chemical components. This allows for simpler and faster VOC analysis. An illustration of a typical gas sensor array response to VOC can be seen in Figure 3.3.

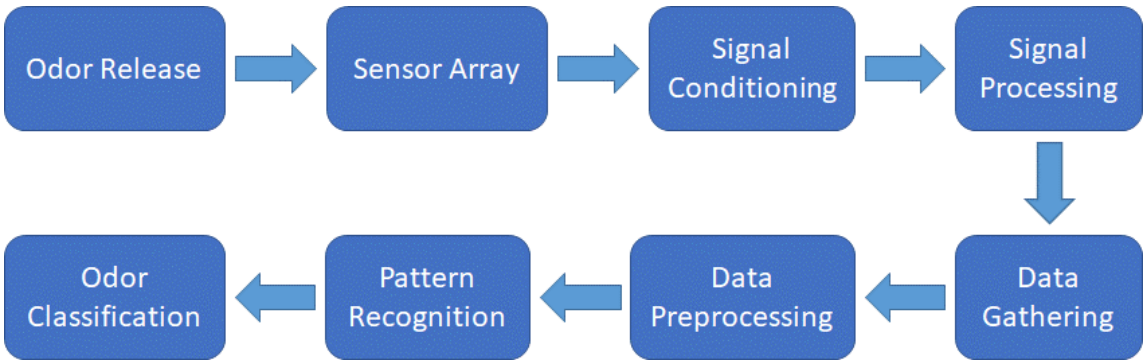


Figure 3.2. The Basic Stage of How E-Nose System Works [8]

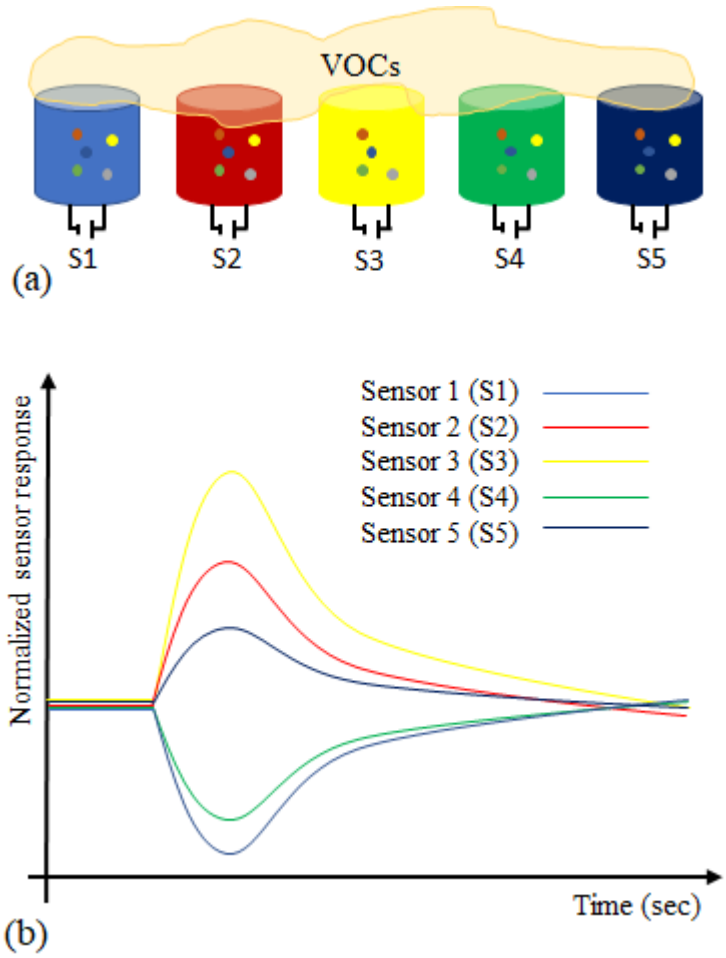


Figure 3.3. An Example of Electronic Nose Sensor Array Setup; (a) Gas Sensor Array Exposed to VOCs (b) Gas Sensor Array Response to a VOCs

3.2.2. WOLF Electronic Nose

The WOLF (Warwick OLFaction) E-nose system was built in house at the school of engineering, the University of Warwick in 2014. The WOLF system is shown in Figure 3.4.

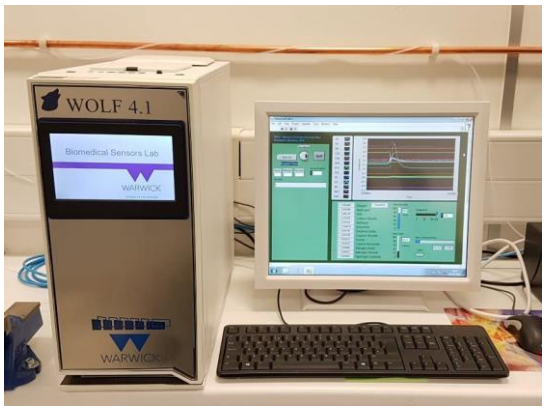


Figure 3.4. WOLF E-Nose in a Laboratory Setting

In brief, the instrument consists of 13 gas sensors, employing a range of different sensor technologies, including 8 amperometric electrochemical sensors (AlphaSense, Essex, UK), 2 non-dispersive infra-red (NDIR) optical devices (Clairair, Essex, UK) and 1 photoionisation detector (Mocon, Minneapolis, MN, USA). The sensors deployed in the WOLF are summarised in Table 3.1. The working principles for each sensing method are explained below.

Table 3.1. Sensors Deployed in WOLF E-Nose

Manufacturer	**Sensing Method**	**Target Gases**
Alphasense Ltd.	Electrochemical	$O_2, NH_3, ETO, SO_2, O_3, NO, NO_2, H_2S, CO, H_2$
Clairair Ltd.	Infrared optical	CO_2, CH_4
MOCON	Photoionisation	All

Electrochemical Gas Sensor

The amperometric electrochemical sensor works by measuring current response of the sensor based on electrochemical oxidation or reduction of a gas molecule when interacting with the sensor's catalytic electrode surface [10]. The amount of current generated is linearly correlated with the concentration of the gas measured when the sensor operated under constant potential conditions [11]. An electrochemical gas sensor typically consists of anti-condensation membrane, capillary, hydrophobic membrane, a sensing electrode (known as a

working electrode), reference electrode, counter electrode and electrolyte, as shown in Figure 3.5.

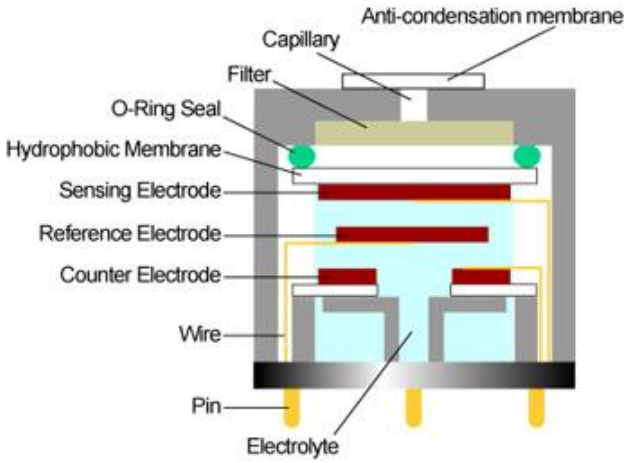


Figure 3.5. Schematic of an Electrochemical Sensor [12].

The VOC is firstly flowing through anti-condensation membranes, which also protect the sensor from dust. Then, the VOC diffuses through the capillary gap and hydrophobic membrane to interact with the sensing electrode. The hydrophobic membrane prevents the leak of electrolytes out of the sensor and allows a sufficient amount of gas to enter. The sensing electrode, which is made of catalytic material, reacts immediately when interacting with the VOC in the form of oxidation or reduction [10]. This process creates a current due to the loss or gain of electrons. The counter electrode balances the reaction by providing an equal opposite current to the sensing electrode. These transfer ions flow through the electrolyte between these electrodes, whereas the current moves through wiring with a pin connector [12]. The reference electrode keeps the electric potential between sensing electrode and reference electrode at a fixed rate as this is critical for the sensitivity and selectivity of the sensor[11]. The electrochemical gas sensors used in WOLF are produced by Alphasense Ltd and are sensitive to $O_2, NH_3, ETO, SO_2, O_3, NO, NO_2, H_2S, CO, H_2$ as previously shown in Table 3.1 [13].

Infrared Optical Gas Sensor

The basic working principle of infrared optical gas sensors is based on optical spectroscopy, without chemical reactions involved. The response generated by an infrared gas sensor is often less likely to be affected by environmental changes, with the exception of pressure [14]. This type of sensor typically has a longer lifetime compared to other types of gas sensor and can be used for real-time detection due to relatively short response time. The

infrared optical gas sensor works based on absorption and emission spectrometry. The details of absorption frequency for specific gases can be found in the HITRAN database [15]. The type of optical gas sensor that is used in WOLF is Non-Dispersive InfraRed (NDIR) sensor. The NDIR schematic diagram is shown in Figure 3.6.

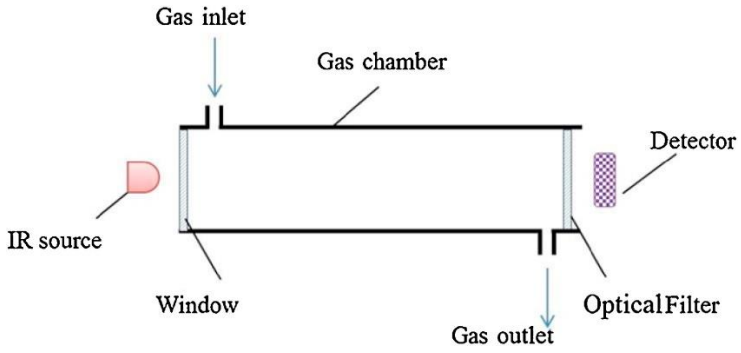


Figure 3.6. Schematic Diagram of an NDIR Gas Sensor [16].

The NDIR systems usually consist of a light source, gas chamber and detector. Mathematically, the interaction between infrared radiation with gas sample inside NDIR is formulated in Beer-Lambert law in Equation 3.1

$$I = I_0 \, x \, exp(-kCL) \qquad \text{Eq. 3.1}$$

Where I_0 is the initial intensity of light source, I is the final intensity captured by the detector after interacting with gas, k is absorption coefficient, C is a gas concentration and L is the length of the gas chamber. Using beer-lambert law, it is possible for NDIR gas sensors to identify particular gases based on their wavelength absorption [16]. The NDIR sensors used in WOLF were manufactured by Clairair Ltd. with commercial product name as Cirius 1 (CH_4) and Cirius 3 (CO_2).

Photoionization

The basic working principle of Photoionization detector (PID) is the use of light to ionise gaseous compounds to quantify the chemical in the gas samples by measuring the charge given off when these chemical molecules return to neutral charge. The light source that is used in PID is a high-energy proton ultraviolet (UV) light [17]. The choice of UV bulb is crucial since the PID is only able to detect the molecules that can be ionised with the ionisation energy lower than what the bulb can generate [18]. This ionisation energy is measured in

electron volts (eV). Each VOC molecule has its own ionisation potential. The schematic of PID's basic working principle is shown Figure 3.7.

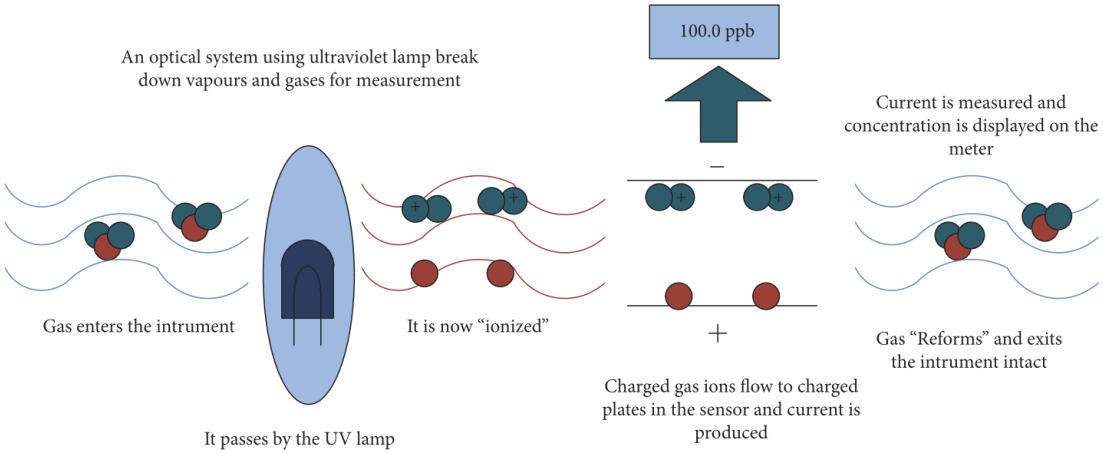


Figure 3.7. The Schematic of PID's Basic Working Principle [19]

When the VOC enters the sensor, it passes the UV lamp, absorbs the energy, and chemical molecules inside the VOC are ionised. These ionised VOCs then flow to electrically charged plates which act as a detector. When this ionised molecule gas hits the sensor, electrical current is produced. The measurement of this current provides information about concentration of the molecules [19]. This technical principle makes the PID gas sensor unique as we can specifically pick the bulb that the UV energy light slightly above the ionisation potential of targeted compounds. Even though this increases the specificity of PID, other compounds in the sample with ionisation potential below the targeted compound will also be measured, lowering the selectivity of the device. The PID sensor used in WOLF was manufactured by MOCON with commercial label as 10.6 eV piD-TECH plus.

3.3. Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS)

Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) is an analytic method to separate gas-phase ions based on their differing mobility in high and low electric fields [20]. Since its first conception in the 1980s, FAIMS has shown its potential as a gas and vapour separation tool that works in atmospheric pressure [21]. This technology has been used in many fields ranging from counterterrorism and law enforcement for explosives detection, environmental hazards and pollutants and to detecting disease in the biomedical field [22].

FAIMS has similar advantages to the electronic nose, which both undertake headspace analysis, use air as the carrier gas, portable, relatively cheap, and simple to operate. However, since FAIMS relies on the physical measurement of a chemical, it suffers less from drift compared to most electronic nose instruments. Due to this feature, FAIMS also holds the potential to be point-of-care tools in the medical field. Mostly, studies on FAIMS focused on analysing VOC from urine, stool, and breath samples.

3.3.1. Operating Principle

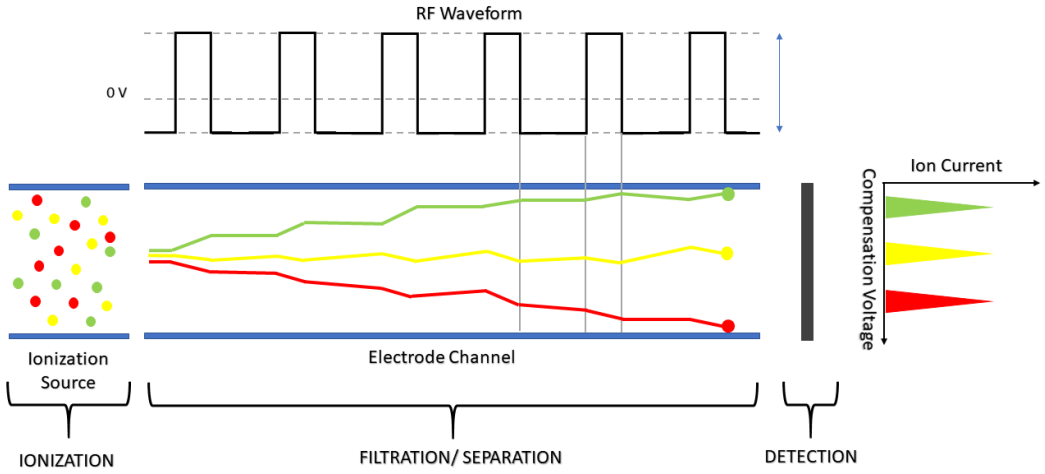


Figure 3.8. Schematic of Working Principle of FAIMS.

The operating principle of FAIMS can be divided into three stages, ionisation, filtration/separation, and detection as shown in Figure 3.8. The sample headspace is first drawn to the ionisation chamber using a clean/dry carrier gas and then passed between two parallel plates. It is known that an ion will experience constant force whenever an electrical field is applied. The force that acts on the ion can be formalised as

$$F = qE \qquad \text{Eq. 3.2}$$

Where q is the charge of the ion and E is the electrical field intensity. In a vacuum, this force will accelerate the ion with an acceleration of

$$a = \frac{qE}{m} \qquad \text{Eq. 3.3}$$

Hence, the motion of the ion will only depend upon its mass and charge. When an electric field is given to an ion in a buffer gas, the ion moves in the direction of the field with velocity

$$V = KE \qquad \text{Eq. 3.4}$$

Where k is a coefficient of ion mobility. The K value of an ion depends on ion structure. At low electric fields, K value is independent of electric field strength. However, at high electric fields, the K value shows a nonlinear dependence on field strength [23] which can be mathematically described as

$$K_h(E) = K_0[1 + f(E)] \qquad \text{Eq. 3.5}$$

Where K_h is the ion mobility at high electric field, K_0 is the coefficient of ion mobility at zero electric fields, and $f(E)$ is the ion mobility as a function of electric field strength.

In filtration/ separation stage, these ions will travel between two plates where the lower plate is usually held at ground potential while an alternating voltage (asymmetric waveform) is applied to the upper plate, creating an alternating voltage electric field in the space between them. The direction of this electric field is perpendicular to the ion's initial direction. This alternating voltage consists of high and low electrical fields such that

$$(E)_h t + (E)_l t_l = 0 \qquad \text{Eq. 3.6}$$

Where $(E)_h$ is the high-field voltage, t is the duration of the $(E)_h$, $(E)_l$ is the low-field voltage and t_l is the duration of $(E)_l$. Due to the perpendicular direction of electric field to the initial ion direction, the ion will be displaced at a distance of

$$d_h = K_h(E)_h t \quad \text{during the high-field} \qquad \text{Eq. 3.7}$$

$$d_l = K_l(E)_l t_l \quad \text{during the low-field} \qquad \text{Eq. 3.8}$$

From these equations, we can see that the total movement of the ion in one wave cycle will be determined by its differential mobility (K_h- K_l). When the K_h is greater than K_l, ions will experience a net movement toward the upper plate, and vice versa. If there is no other voltage is applied, the ions will eventually hit the upper or lower plate and be lost. The greater the difference between K_h and K_l, the faster for the ion to hit the side of the plate. However, when the K_h is the same with K_l, the total ion movement in vertical direction will be zero, allowing the ion to pass through the channel and be detected by the detector.

With this principle, it is possible to control which ion that can pass through the system and transmit to the detection device. Additional DC voltage can be applied between the plate, which act as compensation voltage (CV). The value of CV required to compensate for the vertical movement on a specific ion depends on the ion's ratio of K_h: K_l, $(E)_h$(also known as dispersion field, DF), the temperature, the pressure, the gas flow rates and the analyte concentration [24].

3.3.2. Commercial Products

A commercial FAIMS unit was utilised in projects relating to this work, specifically a Lonestar VOC Analyzer (Owlstone, Cambridge, UK). This device was set up to capture the headspace of sample by integrating it with an ATLAS sampling system (Owlstone, Cambridge, UK) that controls the flow rate and the temperature of the sample. Inside the ATLAS sampling system, clean/dry air is pushed over the surface of the sample into the Lonestar before it is ionised by radioactive sources (Ni-63), which emits beta particles. The term 'FAIMS' is used throughout this thesis when referring to the Lonestar VOC Analyzer. The Lonestar VOC and the ATLAS sampling system are shown in Figure 3.9 and have a dimension of 38.3 x 26.2 x 19.5 cm.

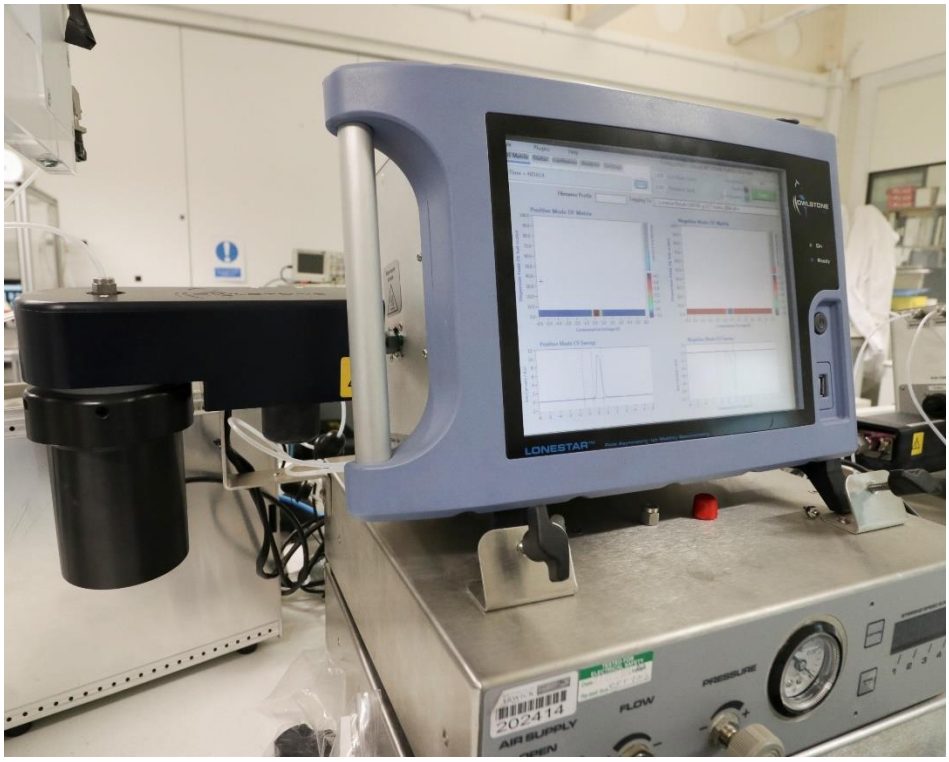


Figure 3.9. The FAIMS Systems with ATLAS Sampling System of the Left Side

3.4. Gas Chromatography Ion Mobility Spectroscopy (GC-IMS)

Gas Chromatography - Ion Mobility Spectrometry (GC-IMS) is a cutting-edge analytical technique that combines the high separation capacity of GC and the fast response of Ion Mobility Spectrometry (IMS). At the beginning of its invention, this technique was mainly used for environmental control and prohibited substance detection [25]. However, for the past decade, GC-IMS technology has grown significantly, attracting many people from other fields to employ it.

The GC-IMS was originally used for detecting food and environmental monitoring. However, in the early 2000s, Baumbach *et al*. used GC-IMS for the first time to investigate volatile metabolites in human exhaled air [26]. Since then, the number of studies using GC-IMS in the medical field has increased. The recent advances of the GC-IMS method is the addition of the NIST library to the system that makes it possible to detect compounds related to separation. NIST library is a standard reference database consisting of a comprehensive list of retention times and mass spectral information that is provided by National Institute of standard and technology, Gaithersburg, MD, USA. When certain unique unknown compounds, that are represented as peaks in GC-IMS dataset, want to be investigated, the

software will use a match factor to see the similarity of the peaks with the known library. This allows the GC-IMS to act like an electronic nose using a pattern recognition algorithm to detect "smell print" from the VOC with the additional features of compound detection, which were previously only possible on GC-MS and similar approaches. Thus, GC-IMS demonstrates the potential as a prominent and highly effective tool for disease detection based on VOC analysis capable of biomarker discovery.

3.4.1. Operating Principle

GC-IMS involves a two-stage analytical process: pre-separation and detection. In the first stage, the complex chemical mixture deriving from the sample headspace are injected into a gas chromatograph, where a carrier gas (mobile phase) pushes the sample through a capillary column. The column is either packed with finely divided solid or coated with a film or a liquid that act as stationary phase [27]. As the chemical mixture pass through the column, the gas mixture is separated into its component based on chemical interactions between each individual component with the stationary phase. As a result of this, compounds are eluded from the column at different times, which is known as retention time. The key factors that affect the pre-separation in GC are the stationary phase, column temperature, internal diameter of column, film thickness and column length [28].

The detection of the analyte happens inside the IMS. Within the IMS, VOC ions are created by soft chemical-ionisation (low-radiation tritium (^{3}H) source). These ions are fed into a drift tube, which are propelled along it by a uniform electric field. Against the flow of ions, a buffer gas is added (i.e. pure nitrogen). Under this circumstance, the ions which initially move with acceleration will soon achieve terminal velocity due to ion collision with buffer gas, which can be formulated as equation 3.4. The mobility of ion in buffer gas depends on the ion's collision cross-section [29]. In general, larger molecules are struck more times than smaller molecules, losing momentum and, thus, taking longer to travel along the tube. The time for the ion to hit the detector is known as drift time. As a result, each substance's drift time is dependent on the interaction of the ion with the electric field and the buffer gas. A Faraday plate is used to measure the resulting ion current, as a function of time [25]. The intensity of ion current is related to the concentration of the molecules. By knowing the drift time, the terminal velocity of each molecule can be calculated. Hence, the ion's mobility K can also be calculated. Using this principle, GC-IMS can detect substances down to parts-per-billion (ppb) range. The working principle of GC-IMS can be seen in Figure 3.10.

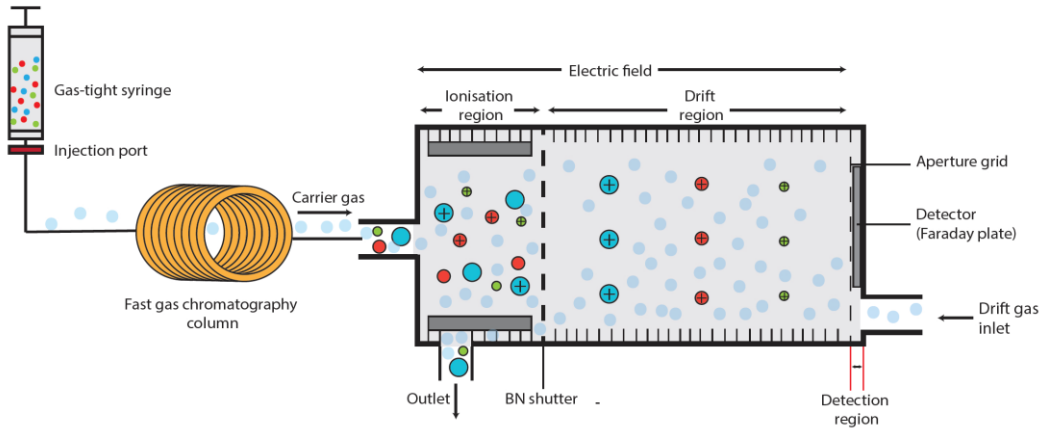


Figure 3.10. The Schematic of Working Principle of GC-IMS

3.4.2. Commercial Products

In this study, we used a commercially available GC-IMS instrument from G.A.S (Dortmund, Germany). Three different GC-IMS variants were employed including BreathSpec, GC-IMS-SILOX, and FlavourSpec.

The BreathSpec is a GC-IMS instrument that has been developed and optimised for the analysis of VOC of human exhaled breath. It has been designed for a direct handheld sampling method in which patients breathe directly into the mouthpiece, and the exhaled breath is sucked into the machine by an integrated pump. The BreathSpec also provides the leur-port for remote sampling method by injecting the breath sample using a disposable syringe for ultimate hygienic handling. A Circular Gas Flow Unit (CGFU) can be fitted onto the Breathspec system. CGFU filters and recirculates ambient air which provides a simpler alternative over traditional external gas supply. Figure 3.11 shows the BreathSpec (including CGFU) unit which has a dimension of 45 x 50 x 30 cm.

Figure 3.11. G.A.S. Breathspec GC-IMS Instrument [30]

An alternative form is the GC-IMS-SILOX, which is a stand-alone GC-IMS instrument that has been developed to measure siloxanes in biogas, landfill or sewage gas. The one-button design provides a user-friendly experience. The device can be used as an online monitoring system testing which automatically sampled according to user desirable intervals. As this device needs an external gas supply to run the system, CGFU or N_2 generator is usually equipped into the system. The dimension of this device is 44.9 x 43.5 x 17.7 cm. Figure 3.12 shows the GC-IMS-SILOX instrument that we used.

Figure 3.12. G.A.S. GC-IMS-SILOX Instrument [31]

Finally, the FlavourSpec is an ideal GC-IMS system to measure the VOC from a headspace of liquid and solid samples, like urine and stool. It comes with an automatic headspace injector that allows for fully automatic handling of multiple samples including sample preparation like heating up and mixing. The whole FlavourSpec system (including the automatic headspace injector) has a dimension of 101 x 80 x 79 cm. Figure 3.13 shows the

FlavourSpec system that was used in this work. The radioactive source used for ionisation in all these GC-IMS instruments is ^{3}H- Tritium, which emits beta radiation.

Figure 3.13. G.A.S. FlavourSpec GC-IMS Instrument [32]

3.5. Gas Chromatography Time-Of-Flight Mass Spectrometry (GC-TOF-MS)

GC-TOF-MS works in a similar way to traditional GC-MS methods, but instead of filtering ions by mass, the TOF utilises 'time of flight' and analyses all ions present. The popularisation of GC-TOF-MS began since 1995 when LECO Corp., Micromass UK Ltd., Jeol Ltd. and Themo Fisher Scientific introduced the first commercial GC-TOF-MS to the market [33]. Since then, GC-TOF-MS has been used in various fields ranging from food, environmental, to medical areas [34,35,36].

GC-TOF-MS provides an efficient separation technique for complex VOC sample analysis. This method has shown tremendous potential for biomarker discovery with excellent sensitivity, specificity, separation capacity, accuracy, and reliability in identifying small molecules in various biological samples. Another advantage of GC-TOF-MS is it can provide full-spectrum accurate-mass data, which enables it to investigate the presence of non-targeted compounds [37]. The use of GC-TOF-MS as a biomarker discovery tool is possible by combining the data from this instrument with NIST 2014 mass spectra library. Because of that, many researchers employed GC-TOF-MS in their projects.

3.5.1. Operating Principle

There are two separation steps that occur inside the GC-TOF-MS, pre-separation based on chemical interaction between VOC with stationary phase inside the GC/GC column and detection of each chemical component based on 'time of flight' of ions inside TOF-MS. The first part has been described in section 3.1.3.

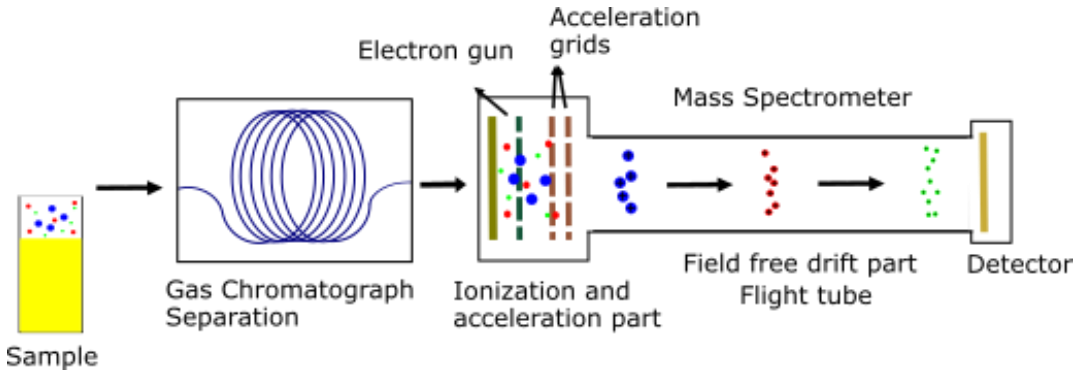


Figure 3.14. Schematic Diagram of GC-TOF-MS. Sample is First Pre-Separated by GC Before Being Ionised. Then, the Time it Takes for the Ion to Travel from Ionisation Chamber to the Detector is Measured. In the Flight Tube, Ions Separate into Groups According to Velocity.

VOC that has been pre-separated by GC, will be Ionised inside the ionisation chamber before it enters the TOF chamber. TOF-MS works based on the concept that lighter ions will travel faster than heavier ions in a vacuum when they have the same kinetic energy. Figure 3.14 shows a simplified conceptual diagram of a TOF-MS. Several kilovolt potentials are applied to accelerate the ions to travel along the flight tube and strike the detector at the end of the tube. The time it takes to traverse inside the tube from the ionisation chamber to detector is known as time of flight, which can be formulated as

$$t = L\sqrt{\frac{m}{2E}} \qquad \text{Eq. 3.9}$$

Where t is time of flight, L is the length of flight tube, m is the mass of the ion and E is kinetic energy of the ions. To accurately capture the signal from detector, the data acquisition system must operate at a nanosecond or shorter time scale [38]. The advances in signal processing technologies have made it possible to achieve such a level of accuracy at a modest

cost, making TOF-MS more popular for analytical purposes. The addition of the TOF-MS into the GC method allows for further separation of the ions present in a sample, increasing the sensitivity of the method.

3.5.2. Commercial Products

In the project relating to this thesis, we utilised a commercial GC-TOF-MS system which consists of a TRACE 1300 GC (Thermo Fisher Scientific, Loughborough, UK), combined with a BenchTOF-HD TOF-MS (Markes Intl., Llantrisant, UK). The GC column used was a 20 m, 0.18 mm ID, Rxi-624Sil MS column (Thames Restek, Saunderton, UK). This system also includes a high-throughput auto-sampler and a thermal desorption unit, ULTRA-xr and UNITY-xr, respectively (both from Markes Intl.). Figure 3.15 shows the picture of the GC-TOF-MS system.

Figure 3.15. GC-TOF-MS Instrument Used in This Thesis: (a) TRACE 1300 GC; (b) BenchTOF-HD TOF-MS; (c) ULTRA-xr and UNITY-xr

3.5.3. Thermal Desorption

Thermal Desorption (TD) is the most common pre-concentration tool for VOC analysis. TD works by collecting VOC into sorbent material, which then can be stored or used directly by heating up (around 200-300 C) the sorbent in the flow of carrier gas (typically helium) to release the high concentrated VOC. It was first used in the mid-1970s to concentrate sulphur dioxide [39] and nitrogen dioxide [40] before injecting them into GC. It helped increase the sensitivity of GC by reducing peak widths. Figure 3.16 shows the capped and uncapped TD

tube. TD is usually made of stainless-steel pack with sorbent material. With the advances of modern technology, there is a wide variety of sorbent material available. A few things to consider when choosing the best sorbent material for VOC analysis are: it has to be able to withstand higher temperature than required for desorption while generating minimal artefact of heating, the affinity of the sorbent should be low for water, and it has to have a low metal concentration [41]. There are three characteristics of sorbent material: volatile range, memory effects and breakthrough volume. The Volatile range of a sorbent tells us about the range of carbon numbers that suit the most. Memory effect happens when there is still remaining VOC after thermal desorption. Breakthrough volume is defined as the volume of carrier gas per gram of sorbent material, which causes the analyte molecules to travel through the TD. Knowing this value is important to make sure that the analyte of interest does not accidentally purge out of the system during data collection. Once the VOC is captured inside the TD, the sample can be stored for hours or days before it slowly degrades [42]. The main drawback of using this technique is that TD can only release the VOC once. It means if a problem occurs during the process, there is no second measurement using the same concentrated sample. However, modern TD systems allow partial recapture of the desorbed VOC for repeating analysis with slightly diluted sample.

Figure 3.16. Un-Capped and Capped Thermal Desorption (TD) Tubes

3.6. Quality Control and Calibration

All instruments were checked before any experiments were undertaken. These were either part of our standard QC checks or part of test to ensure that the instrument was fully functional. For the GC-TOF-MS, before any tests, the method was first checked, followed by

the vacuum level, voltages and flow rates to ensure there are no obvious issues with the instrument. Then a water/air check was undertaken to ensure there were no leaks in the system. This was followed by an autotune, where a known calibrant is injected into the mass spectrometer, and the instrument calibrates the high voltage components to optimise the signal to noise ratio. If the system passes these checks, then the instrument is ready to be used. Less regularly, a calibration standard formed of a "mega-mix" of chemicals is loaded onto a tube and run through the systems. This ensures that the instrument is fully functional, peaks elude at a similar time to previous experiments and checks the accuracy of the NIST list identification.

For the FAIMS unit, QC checks are undertaken on the internal temperature of the unit (sample and measurement) and that the pressure and flow rates are within an acceptable working range. Then, the output of the instrument is checked for contamination (which can be easily visually observed) and that the peak at zero dispersion filed is within the manufacturer's tolerance. The instrument does not need regular calibration, but a test standard is used, in line with the manufacturer's recommendation, periodically (once a month). In addition, air blank vials were used before and after a batch of samples to take a background reading.

The GC-IMS checking was similar to those earlier. First, temperature zones and flow rates were checked before starting a batch of samples. Then the RIP time and its magnitude were checked and compared with previous readings and the manufacturer's recommended value range. Though there were no recommended test chemicals, basic chemicals at low concentrations (specifically ethanol and acetone) were used to check the functionality of the unit. In addition, air blank vials were added at the beginning and end of a batch of samples to take both a room blank and to check the functionality of the instrument.

3.7. Comparison

A comparison and summary of instrumentations technology that is used in this thesis, with a visual explanation of their working principles, is shown in Figure 3.17. Each of these instruments has its own strengths and weaknesses. Table 3.2 below shows the pros and cons of each instrument.

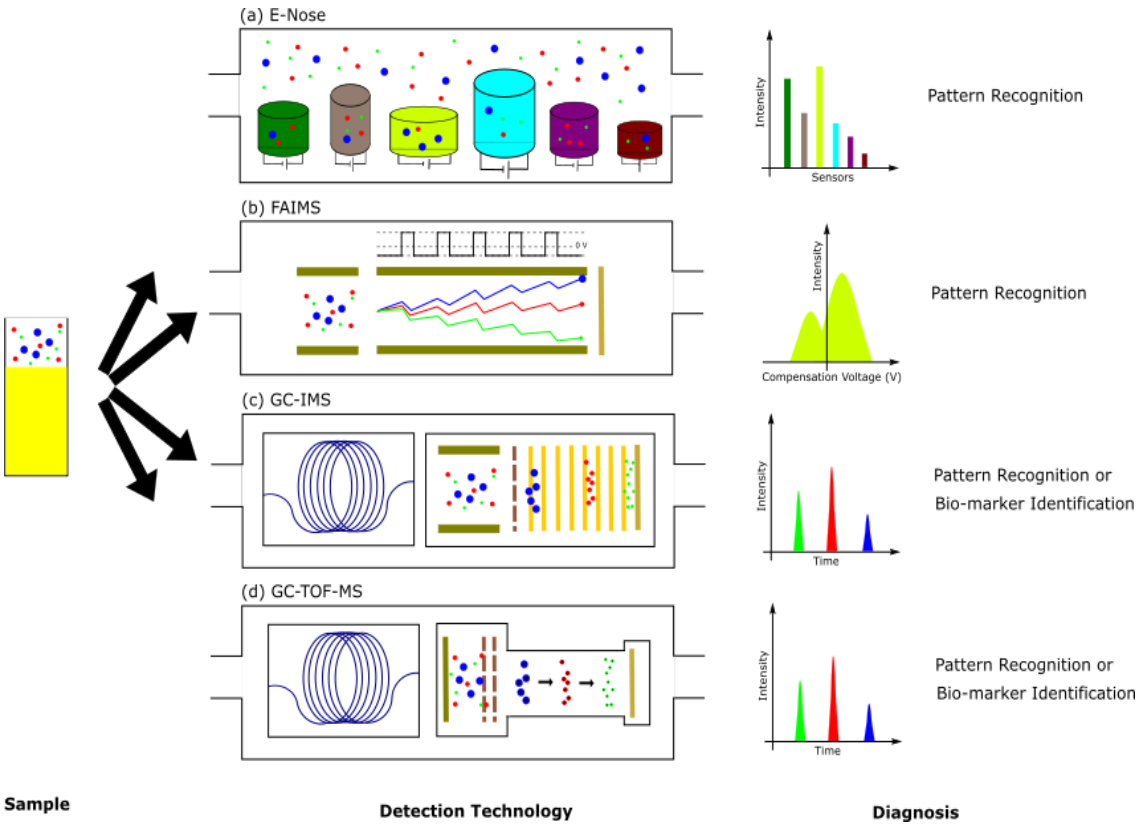


Figure 3.17. Comparison and Schematic of VOC Detection Technologies Used in This Thesis. (a) Electronic Nose (E-nose): Generates a Single from a Cross-Reactive Gas Sensor Array, (b) Field Asymmetric Ion Mobility Spectrometry (FAIMS): the Separation of Compounds Based on Different Ion Mobility in a High and Low Electric Field; (c) Gas Chromatography Ion Mobility Spectrometry (GC-IMS): GC Which Perform the Pre-Separation is Represented in Blue and the Separation of Compounds Based on Ion Mobility Represent as a Group of VOCs with a Different Colour. (d) Gas Chromatography Time of Flight Mass Spectrometry (GC-TOF-MS): Pre-Separation is Performed by GC and Compounds Separation Based on the Time it Takes to Reach the Detector. A Typical Response is Shown for Each Method. VOCs are Illustrated as Red, Green, and Blue Circles.

Table 3.2. Summarise the Pros and Cons for Each Device Used in This Thesis.

Technology	Advantages	Disadvantages
E-nose	• Relatively low-cost • Use ambient air as the carrier. • Quicker analysis time compared with traditional gas analytical equipment (such as GC-MS)	• Poor intra-device repeatability • Limited temporal stability • Poor selectivity due to the promiscuous nature of the chemical interaction between sample and the sensors

	• Provide a real-time method that requires little to no pre-treatment.	• Because of the poor repeatability, selectivity and stability, it's currently still hard to implement E-nose into clinical practice • Data processing is still a challenge that needs to address properly.
FAIMS	• Good intra device stability allowing a plug-and-play application in clinical practice after the initial validation • Produce high dimensionality data, rich with chemical information • High sensitivity with a reproducible detector in a point of care tool • Application can be easily used without requiring extensive training	• Chemical identification is not possible. Hence, FAIMS cannot be used as biomarker discovery tool (without the additional of a separate GC) • Cannot provide Collision Cross Section values • Requires high flow rate of clean air • Compound with low proton affinity is hard to detect • There is a risk of radioactive contamination.
GC-IMS	• Provide high information content • Physical measurement of complex chemical headspaces but still fulfil many of the requirements needed as a point of care tool • Combined the high separation of GC and fast response of IMS • The interface is simple and low cost, making the miniaturisation of its equipment possible. • Provides rapid analysis, high sensitivity and variable volume injection, with no pre-treatment • It can separate the compound into two dimensions • High separation efficiency • Atmospheric pressure operation	• Complex spectra provide a challenge in analysing the data • Compound with low proton affinity is hard to detect • There is a risk of radioactive contamination.
GC-TOF-MS	• High mass accuracy • Can provide full-spectrum accurate-mass data which enables it to investigate the presence of non-targeted compounds. • Fast acquisition rates	• The price is still considerably expensive. • Bulky/lab based • Requires specialised trained staff to operate

	• Spectral continuity and exceptional dynamic range • Moderate to high resolution	• Requires helium

3.8. Conclusion

Recent advances in the field of portable and low-cost gas analysers, including E-nose, FAIMS, GC-IMS and GC-TOF-MS, have improved the potential of utilising these instruments for diagnostic and monitoring purposes through analysing human metabolic waste VOC profile such as stool, urine and breath. This chapter has provided a detailed overview of the operating principle, description of commercial products and their application in the medical field for each analytical instrument used in this thesis. This chapter will be referred to in the respective method section in chapters 5 and 6. In the next chapter, the raw data and the method used to analyse the data for each analytical instrument will be discussed in detail.

3.9. Reference

[1] E. Stark, J. Pitt, A. Nur Wicaksono, K. Milanovic, V. Lush, and S. Hoover, "Odorveillance and the Ethics of Robotic Olfaction [Opinion]," *IEEE Technol. Soc. Mag.*, vol. 37, no. 4, pp. 16–19, 2018, doi: 10.1109/MTS.2018.2876103.

[2] R. W. Moncrieff, "An instrument for measuring and classifying odors," *J. Appl. Physiol.*, vol. 16, no. 4, pp. 742–749, Jul. 1961, doi: 10.1152/jappl.1961.16.4.742.

[3] F. Wilkens and D. Hartman, "An Electronic Analog for the Olfactory Processes," *Ann. N. Y. Acad. Sci.*, pp. 372–378, 1964.

[4] A. Dravnieks and P. J. Trotter, "Polar vapour detector based on thermal modulation of contact potential," *J. Sci. Instrum.*, vol. 42, no. 8, pp. 624–627, 1965, doi: 10.1088/0950-7671/42/8/335.

[5] K. Persaud and G. Dodd, "Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose," *Nature*, vol. 299, no. 5881, pp. 352–355, Sep. 1982, doi: 10.1038/299352a0.

[6] M. Kaneyasu, A. Ikegami, H. Arima, and S. Iwanaga, "Smell Identification Using a Thick-Film Hybrid Gas Sensor," *IEEE Trans. Components, Hybrids, Manuf. Technol.*, vol. 10, no. 2, pp. 267–273, 1987, doi: 10.1109/TCHMT.1987.1134730.

[7] J. W. Gardner and P. N. Bartlett, “A brief history of electronic noses,” *Sensors Actuators B Chem.*, vol. 18, no. 1–3, pp. 210–211, Mar. 1994, doi: 10.1016/0925-4005(94)87085-3.

[8] D. Karakaya, O. Ulucan, and M. Turkan, “Electronic Nose and Its Applications: A Survey,” *Int. J. Autom. Comput.*, vol. 17, no. 2, pp. 179–209, 2020, doi: 10.1007/s11633-019-1212-9.

[9] A. D. Wilson and M. Baietto, “Advances in electronic-nose technologies developed for biomedical applications,” *Sensors*, vol. 11, no. 1, pp. 1105–1176, 2011, doi: 10.3390/s110101105.

[10] F. R. Simões and M. G. Xavier, *Electrochemical Sensors*. Elsevier Inc., 2017.

[11] J. R. Stetter and J. Li, “Amperometric gas sensors - A review,” *Chem. Rev.*, vol. 108, no. 2, pp. 352–366, 2008, doi: 10.1021/cr0681039.

[12] “Electrochemical Gas Sensors.” https://www.membrapor.ch/electrochemical-gas-sensors/ (accessed Jun. 05, 2021).

[13] Alphasense Ltd., “Alphasense Application Note 109-02 - Interfering Gases,” 2010. http://www.alphasense.com/environmental-sensors/pdf/AAN_109-02.pdf.

[14] X. Liu, S. Cheng, H. Liu, S. Hu, D. Zhang, and H. Ning, “A survey on gas sensing technology,” *Sensors (Switzerland)*, vol. 12, no. 7, pp. 9635–9665, 2012, doi: 10.3390/s120709635.

[15] I. E. Gordon, “The HITRAN Database.” https://lweb.cfa.harvard.edu/hitran/ (accessed Jun. 15, 2021).

[16] T. V. Dinh, I. Y. Choi, Y. S. Son, and J. C. Kim, “A review on non-dispersive infrared gas sensors: Improvement of sensor detection limit and interference correction,” *Sensors Actuators, B Chem.*, vol. 231, pp. 529–538, 2016, doi: 10.1016/j.snb.2016.03.040.

[17] L. Spinelle, M. Gerboles, G. Kok, S. Persijn, and T. Sauerwald, “Review of portable and low-cost sensors for the ambient air monitoring of benzene and other volatile organic compounds,” *Sensors (Switzerland)*, vol. 17, no. 7, 2017, doi: 10.3390/s17071520.

[18] G. Coelho Rezende, S. Le Calvé, J. J. Brandner, and D. Newport, “Micro

photoionisation detectors," *Sensors Actuators, B Chem.*, vol. 287, no. February, pp. 86–94, 2019, doi: 10.1016/j.snb.2019.01.072.

[19] Q. Zhou, S. Zhang, X. Zhang, X. Ma, and W. Zhou, "Development of a novel micro photoionisation detector for rapid volatile organic compounds measurement," *Appl. Bionics Biomech.*, vol. 2018, 2018, doi: 10.1155/2018/5651315.

[20] B. M. Kolakowski and Z. Mester, "Review of applications of high-field asymmetric waveform ion mobility spectrometry (FAIMS) and differential mobility spectrometry (DMS)," *Analyst*, vol. 132, no. 9, pp. 842–864, 2007, doi: 10.1039/B706039D.

[21] I. A. Buryakov, E. V Krylov, A. L. Makas, V. V Pervukhin, and U. K. Rasulev, "ION DIVISION BY THEIR MOBILITY IN HIGH TENSION ALTERNATING ELECTRIC-FIELD," *Pisma v Zhurnal Tekhnicheskoi Fiz.*, vol. 17, no. 12, pp. 60–65, 1991.

[22] M. T. Costanzo, J. J. Boock, R. H. J. Kemperman, M. S. Wei, C. R. Beekman, and R. A. Yost, "Portable FAIMS: Applications and future perspectives," *Int. J. Mass Spectrom.*, vol. 422, pp. 188–196, 2017, doi: 10.1016/j.ijms.2016.12.007.

[23] E. A. Mason and E. W. McDaniel, *Transport Properties of Ions in Gases*. Wiley, 1988.

[24] R. W. Purves, R. Guevremont, S. Day, C. W. Pipich, and M. S. Matyjaszczyk, "Mass spectrometric characterisation of a high-field asymmetric waveform ion mobility spectrometer," *Rev. Sci. Instrum.*, vol. 69, no. 12, pp. 4094–4105, 1998, doi: 10.1063/1.1149255.

[25] J. I. Baumbach, "Process analysis using ion mobility spectrometry," *Anal. Bioanal. Chem.*, vol. 384, no. 5, pp. 1059–1070, 2006, doi: 10.1007/s00216-005-3397-8.

[26] J. I. Baumbach, W. Vautz, V. Ruzsanyi, and L. Freitag, "Early Detection of Lung Cancer: Metabolic Profiling of Human Breath with Ion Mobility Spectrometers," *Mod. Biopharm. Des. Dev. Optim.*, vol. 3, pp. 1343–1358, 2008, doi: 10.1002/9783527620982.ch55.

[27] M. M. Rahman, A. M. Abd El-Aty, J.-H. Choi, H.-C. Shin, S. C. Shin, and J.-H. Shim, "Basic Overview on Gas Chromatography Columns," *Anal. Sep. Sci.*, pp. 823–834, 2015, doi: 10.1002/9783527678129.assep024.

[28] M. Lorenzo and Y. Pico, *Gas Chromatography and Mass Spectroscopy Techniques for the Detection of Chemical Contaminants and Residues in Foods*. Elsevier Ltd, 2017.

[29] J. W. Lee, K. L. Davidson, M. F. Bush, and H. I. Kim, "Collision cross sections and ion structures: Development of a general calculation method via high-quality ion mobility measurements and theoretical modeling," *Analyst*, vol. 142, no. 22, pp. 4289–4298, 2017, doi: 10.1039/c7an01276d.

[30] G.A.S. Gesellschaft für analytische Sensorsysteme, "Breath Analysis using GC-IMS Technology Gas Chromatograph-Ion Mobility Spectrometer for VOC Trace Detection," 2021. https://www.gas-dortmund.de/data-live-gas/docs/pdf/Anwendungen/Atemluft/Flyer_BreathSpec_EU_1711.pdf (accessed Apr. 01, 2021).

[31] G.A.S. Gesellschaft für analytische Sensorsysteme, "GC-IMS-SILOX," 2021. https://www.gas-dortmund.de/data-live-gas/docs/pdf/Produkte/GC-IMS_SILOX_1803.pdf (accessed Apr. 01, 2021).

[32] G.A.S. Gesellschaft für analytische Sensorsysteme, "FlavourSpec® Gas Chromatograph-Ion Mobility Spectrometer (GC-IMS) Detection of Volatile Compounds in Food and Beverages," 2021. https://www.gas-dortmund.de/data-live-gas/docs/pdf/Produkte/Flavour/FlavourSpec_1803.pdf (accessed Apr. 01, 2021).

[33] T. Cajka, *Gas chromatography-time-of-flight mass spectrometry in food and environmental analysis*, 1st ed., vol. 61. Elsevier B.V., 2013.

[34] F. Lin, F. Cai, B. Luo, R. Gu, S. Ahmed, and C. Long, "Variation of Microbiological and Biochemical Profiles of Laowo Dry-Cured Ham, an Indigenous Fermented Food, during Ripening by GC-TOF-MS and UPLC-QTOF-MS," *J. Agric. Food Chem.*, vol. 68, no. 33, pp. 8925–8935, Aug. 2020, doi: 10.1021/acs.jafc.0c03254.

[35] L. Yang *et al.*, "Non-target screening of organic pollutants and target analysis of halogenated polycyclic aromatic hydrocarbons in the atmosphere around metallurgical plants by high-resolution GC/Q-TOF-MS," *Environ. Sci. Eur.*, vol. 32, no. 1, p. 96, Dec. 2020, doi: 10.1186/s12302-020-00376-9.

[36] M. Qiao *et al.*, "Antiliver Fibrosis Screening of Active Ingredients from Apium graveolens L. Seeds via GC-TOF-MS and UHPLC-MS/MS," *Evidence-based Complement. Altern. Med.*, vol. 2020, 2020, doi: 10.1155/2020/8321732.

[37] H. Wang *et al.*, “Metabolomic profile of diabetic retinopathy: a GC-TOFMS-based approach using vitreous and aqueous humor,” *Acta Diabetol.*, vol. 57, no. 1, pp. 41–51, 2020, doi: 10.1007/s00592-019-01363-0.

[38] A. L. Rockwood, M. M. Kushnir, and N. J. Clarke, “Mass spectrometry,” in *Principles and Applications of Clinical Mass Spectrometry: Small Molecules, Peptides, and Pathogens*, Elsevier, 2018, pp. 33–65.

[39] E. D. Palmes and A. F. Gunnison, “Personal Monitoring Device for Gaseous Contaminants,” *Am. Ind. Hyg. Assoc. J.*, vol. 34, no. 2, pp. 78–81, 1973, doi: 10.1080/0002889738506810.

[40] E. D. Palmes, A. F. Gunnison, J. Dimattio, and C. Tomczyk, “Personal sampler for nitrogen dioxide,” *Am. Ind. Hyg. Assoc. J.*, vol. 37, no. 10, pp. 570–577, 1976, doi: 10.1080/0002889768507522.

[41] A. Methods and C. Amctb, “Thermal desorption part 1: Introduction and instrumentation,” *Anal. Methods*, vol. 12, no. 26, pp. 3425–3428, 2020, doi: 10.1039/d0ay90082f.

[42] S. W. Harshman *ss.*, “Storage stability of exhaled breath on Tenax TA,” *J. Breath Res.*, vol. 10, no. 4, 2016, doi: 10.1088/1752-7155/10/4/046008.

Chapter 4
DATA ANALYSIS

4.1. Introduction

This chapter describes the data analysis approaches used to analyse data from the E-nose, FAIMS, GC-IMS, and GC-TOF-MS methods in studies undertaken related to this thesis. Firstly, the structure and dimension of the raw data generated by each method is explained. Then, each step of data analysis pipeline is discussed in detail, including the improvements for each step. Finally, the data analysis for each method is concluded.

4.2. Data Structure

Every instrument creates a different output response when gas or VOC is introduced. The dimension, size, amount of information, and characteristics are also different. While E-noses generally provide the smallest datasets, GC-IMS produces the largest data from a single sample. In this section, the structure, dimension, and characteristics of raw data produced by each instrument are described.

4.2.1. E-nose

As explained in chapter 3, E-nose systems operate in a manner similar to the mammalian olfactory system where an automated set of internal operations like sampling, purging, and cleaning of the sensor chamber between tests are undertaken during a test cycle. A schematic of single sensor response within a sensor array in an E-nose is shown in Figure 4.1. Typically, a reference gas is passed over the sensor array to establish a stable baseline response. The sampling procedure is then carried out by exposing the sensor array to a sample (usually continuously, but sometimes static), which causes changes in its output signal until the E-nose responses reach a steady state. The E-Nose output signal then returns to its baseline condition once the sample has been flushed out/removed with a reference gas, as shown in Figure 4.1. The time that the sensor is exposed to the sample is called the sampling period.

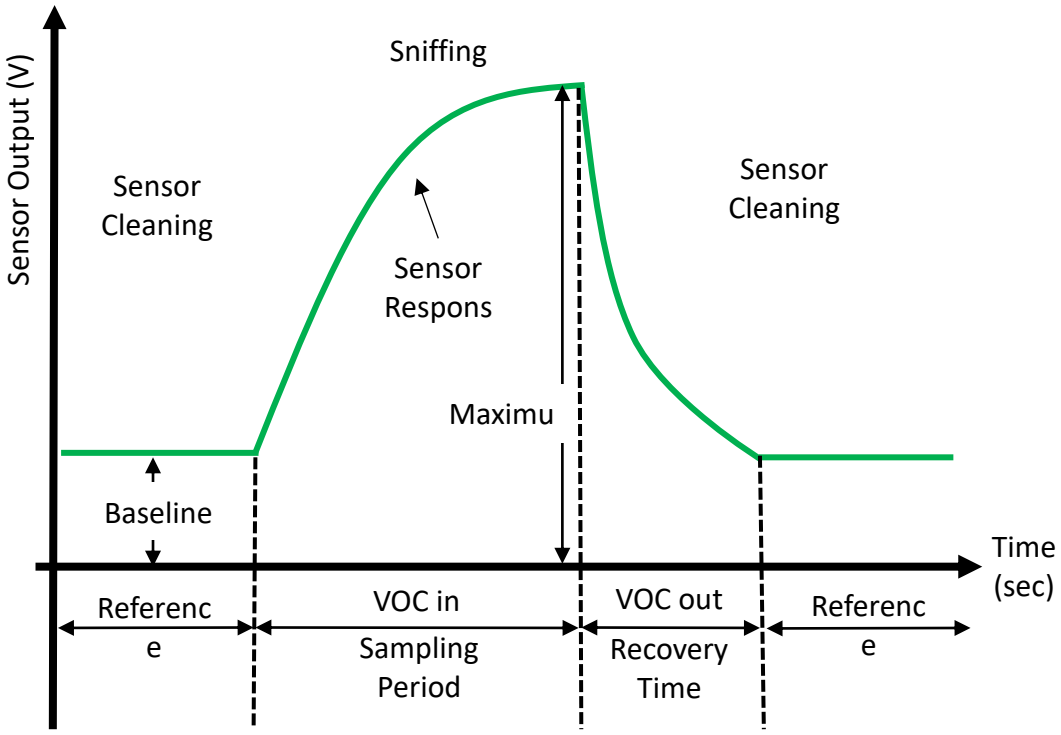


Figure 4.1. Schematic of Single E-Nose Gas Sensor Response

WOLF E-nose system uses 13 gas sensors in its sensor array. Figure 4.2 shows a typical sensor response of the WOLF E-nose system when analysing a breath sample from a patient with Crohn's disease. The chemicals list on the right side refers to the general gas/VOC that each sensor is designed to detect. As each sensor in the array has a unique reaction when a sample is introduced, the WOLF output signal can provide a fingerprint for a specific sample group. Data from the sensors are utilised to build a database for classification model training. A single WOLF E-nose output is a matrix in which its rows are the sensors' responses, and the columns are the E-nose sensor readings.

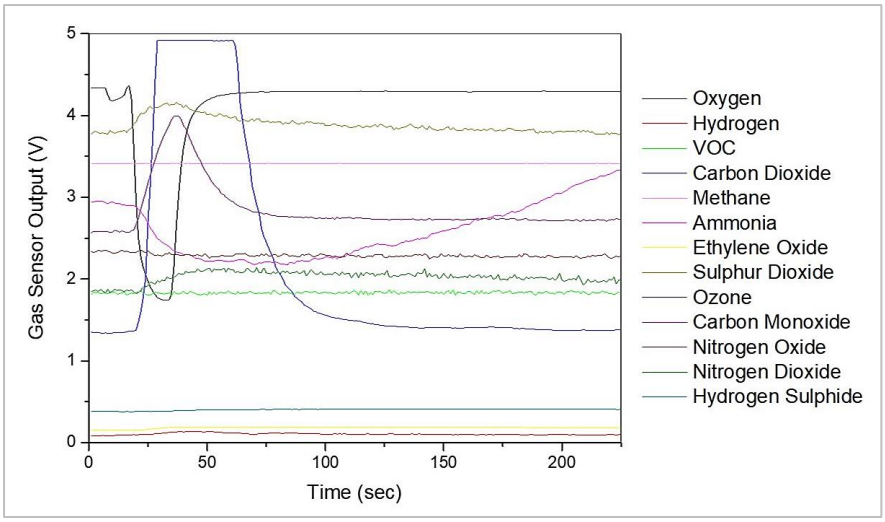


Figure 4.2. Typical Sensor Output Responses of WOLF E-Nose

4.2.2. FAIMS

In relation to the number of features/covariates measured per sample, FAIMS generates higher dimensional data that are rich in chemical information compared to an E-nose, as previously mentioned in chapter 3. For all studies in this work, the FAIMS was set with a dispersion field of 0% to 100% in 51 steps and -6 to +6 V compensation voltage in 512 steps, with both positive and negative fields applied to the samples. This created a dataset of 52,224 points per sample, with each sample taking less than 60 seconds to capture. Figure 4.3 shows a typical FAIMS response to urinary VOC sample from patient with pancreatic cancer. This instrument produces a three-dimensional matrix as a result of the molecules passing through an electric field between two plates and being detected. The x axis represents compensation voltage in volts, the y axis shows dispersion field in %, and the z axis represent the total ion current in arbitrary unit (A.U.). The blue area shown in Figure 4.3 represents the area with no chemical information.

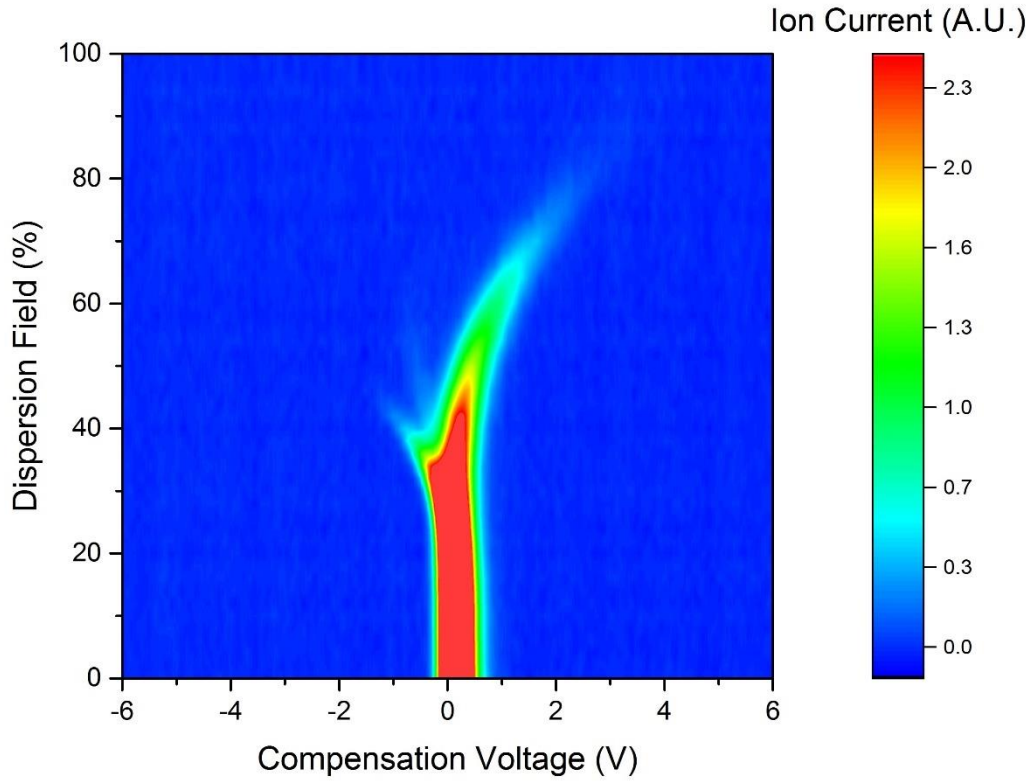


Figure 4.3. Typical FAIMS Response to Urinary VOC Sample from a Patient with Pancreatic Cancer

4.2.3. GC-IMS

As described in Chapter 3, the data output produced by GC-IMS are arranged in a three-dimensional array. Here each point is represented by gas chromatography (GC) data output with corresponding ion mobility spectrometry (IMS) chromatogram as shown in Figure 4.4. The drift time spectra from the IMS is measured in milliseconds (x axis), the retention time for the chromatographic column is measured in seconds (y axis), and the intensity of ion current signal in volts (z axis). The signal's intensity is shown by colour, with each high intensity region representing a single or a mixture of molecules (with the same properties). The extended red line represents the reactant ion peak (RIP). This is a constant background signal in the spectrum that occurs as a consequence of the carrier gas being present throughout the measurement procedure. The GC-IMS generates an output signal file in .mea extension, which can be viewed using Laboratory Analytical Viewer (LAV) software (v2.2.1, G.A.S., Dortmund, Germany). This .mea file is exported to .txt file using the same software, before

being processed. Figure 4.5 depicts an example GC-IMS response to a patient faecal sample with refractory coeliac disease.

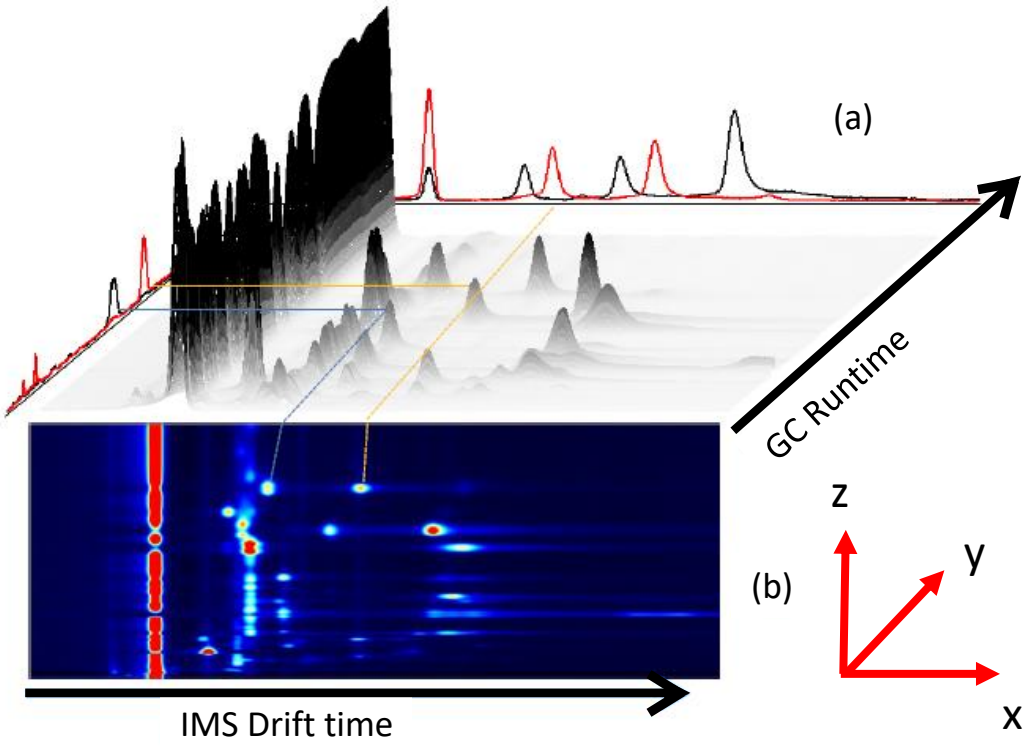


Figure 4.4. 3D Representation of GC Data Output with Corresponding IMS Chromatogram. (a) Single IMS Spectra Data is Combined with GC Run Time Peaks (b) Heatmap Corresponding to GC-IMS Peaks (Yellow and Blue Lines) (Image Adapted with Permission from IMSPEX, UK)

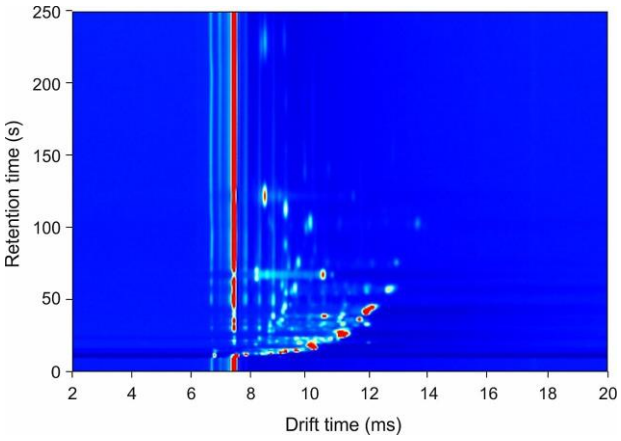


Figure 4.5. An Example Output of GC-IMS Instrument to Faecal VOC Sample from a Patient with Refractory Coeliac Disease.

4.2.4. GC-TOF-MS

In the case of GC-TOF-MS, a 2D spreadsheet dataset is generated in which the x-axis represents the retention time produced by the GC column (in minutes) and y-axis indicates the amount of chemical detected (in abundance or total ion count). Figure 4.6 depicts the typical raw response of the GC-TOF-MS to a urinary sample of a patient with Pancreatic Ductal Adenocarcinoma (PDAC). The number at the top of the peak corresponds to a specific retention time value that is associated with a particular chemical compound (as an example). The chemical identification is accomplished by comparing the data to a NIST library (2016). Figure 4.7 shows the chemical interpretation of peaks in a GC-TOF-MS dataset obtained from a PDAC patient sample.

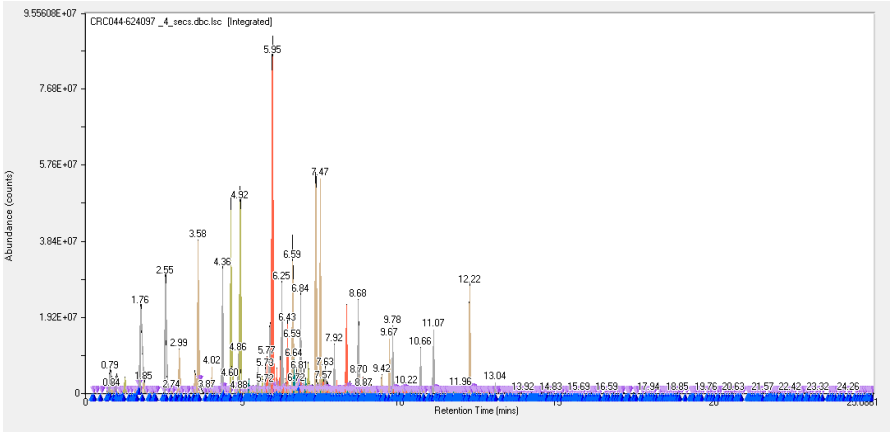


Figure 4.6. A Typical Raw Response of GC-TOF-MS when Analysing the Urinary VOC Patient with PDAC.

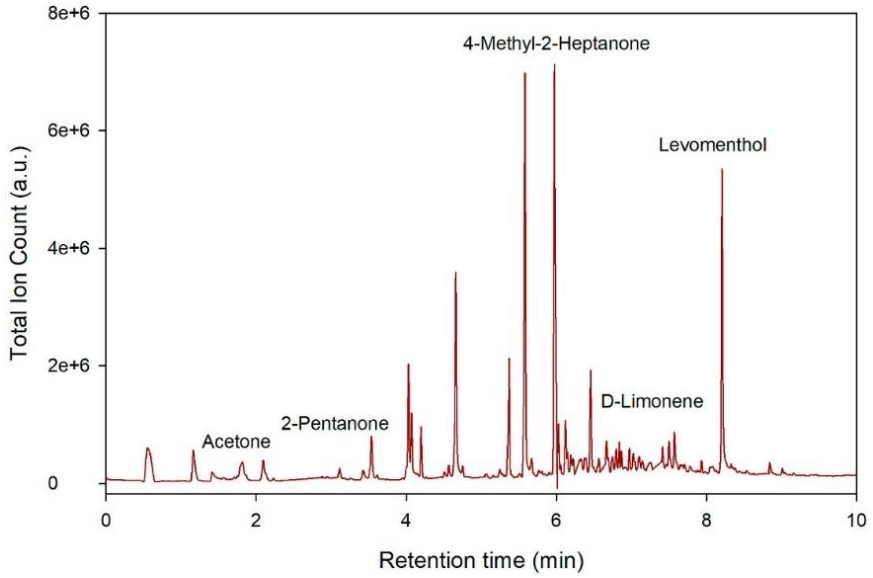


Figure 4.7. An Example of Chemical Interpretation of GC-TOF-MS Peak for Patient with PDAC

4.3. Data Analysis Pipeline

The majority of VOC analysis application in medical field seek to distinguish between disease states, such as a basic case-control design (disease group versus healthy group), specific disease subdivisions (cancer stages) or different disease groups. For these applications, classification analysis is required in order to discriminate between groups, and eventually try to develop models that may be utilised as a diagnostic tool.

The data analysis pipeline used to analyse raw data from the E-nose, FAIMS, GC-IMS, and GC-TOF-MS in studies relating to this thesis generally include pre-processing, data splitting, cross-validation, feature selection, classification, calculation of performance/generating ROC curves, and significant features as shown in Figure 4.8. The next section gives a brief description of the processes involved in the data analysis pipeline. These techniques have been used in a variety of clinical studies investigating the potential use of breath, faecal, and urinary VOC analysis for various disease detection [1,2,3,4,5].

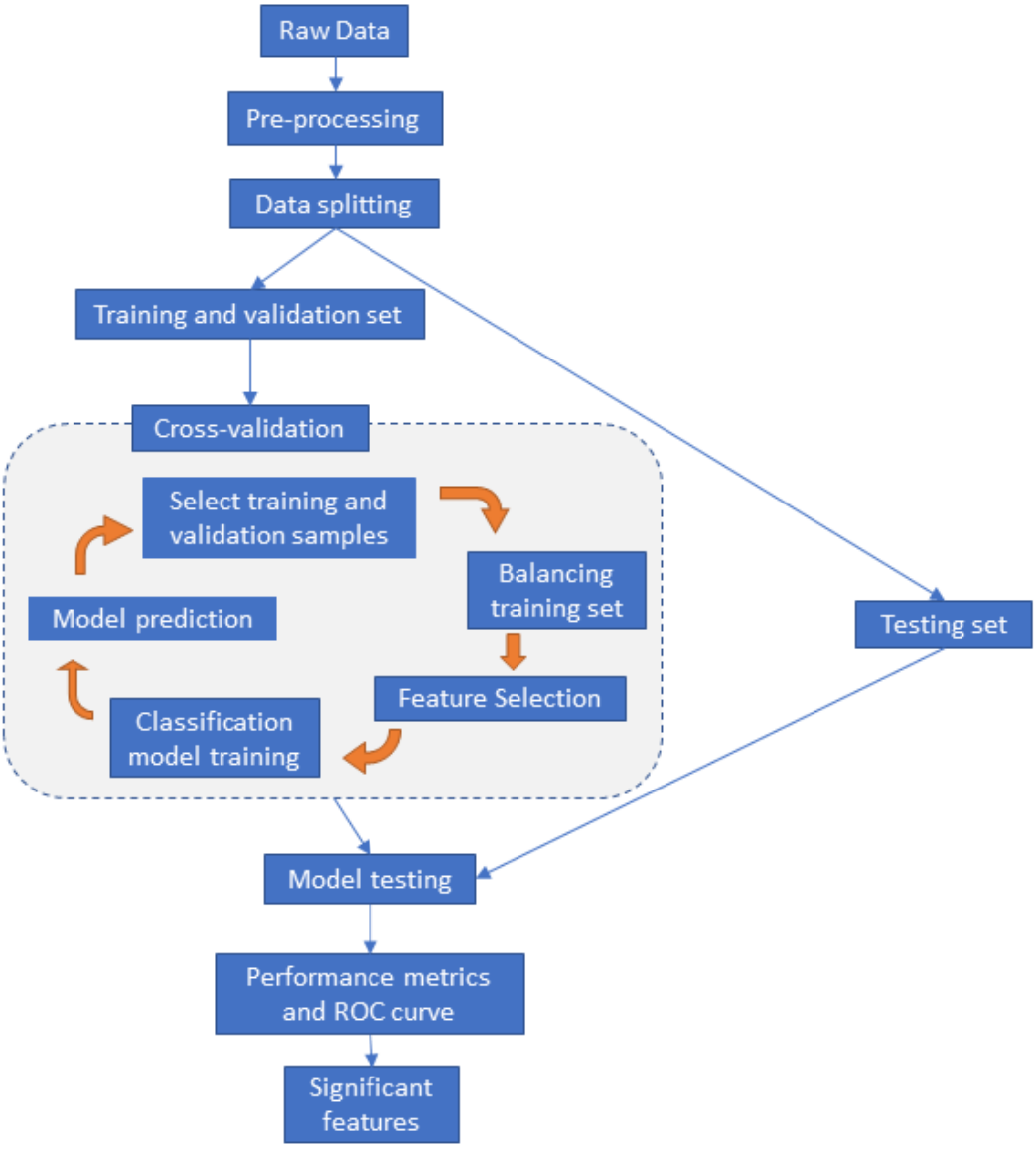


Figure 4.8. Diagram of VOC Data Analysis Pipeline

4.3.1. Pre-processing

The pre-processing objective is to reduce the dimensionality of the data by removing the data that does not hold chemical information. The raw GC-IMS output files are complicated and have high-dimensionality. These pre-processing procedures include RIP alignment, cropping, and thresholding.

WOLF E-nose

The objective of pre-processing step in analysing WOLF E-nose data is to extract robust, less redundant information from the response of the sensor array. To achieve this, we

use the baseline and maximum gas sensor response values as the most common and simplest E-nose feature extraction approach [6]. The maximum value represents the greatest change in the sensor response and indicates the final steady state of the whole dynamic response process, as shown in Figure 4.1. There are many approaches that may be used to calculate the difference between a baseline and a maximum. However, the model that was utilised in studies relating to this thesis is as simple as subtracting the maximum from minimum value, which is also known as the different model. The raw signal is analysed without the need to apply filtering and smoothing. By utilising a different model, each gas sensor in the WOLF E-nose array provides a single feature.

FAIMS

As shown in Figure 4.3, the majority of FAIMS data does not hold any chemical information, which is represented by the blue area in the output graph. Hence, the raw FAIMS data can be pre-processed to reduce the uninformative data. The pre-processing step in FAIMS includes cropping and applying Discrete Wavelet transform. After Mallat introduced image decomposition technique using multifrequency channel wavelet models in 1998, the Discrete Wavelet Transform (DWT) quickly gained popularity as a highly flexible signal pre-processing tool [7]. This technique utilises time-localised wave-like oscillation, which has zero mean to extract specific local spectral information from a dataset. In the studies related to this thesis, two-dimensional (2D) DWT was applied as a pre-processing step to filter, compress, and extract the unique signal in high dimensionality FAIMS data. We used imwd function from R packaged called wavethresh, which was built based on Mallat's pyramidal algorithm [8]. In essence, the 2D DWT is done by applying one-dimensional (1D) DWT on each row of a matrix, which acts as a horizontal filter, followed by another 1D DWT application on each column, which acts as a vertical filter [9]. As the 2D DWT works only to a squared 2^n matrix and the edge area of FAIMS matrix contains negligible signal, the 512 x 102 FAIMS (positive and negative ion) raw matrix was padded with zero columns and cropped to 128 x 128 matrix before 2D DWT was applied as shown in Figure 4.9. This procedure is a common practice when dealing with wavelets and does not alter the wavelet coefficient of interest. The wavelet filter type used in this step was Daubechies least-asymmetric wavelet. After applying 2D DWT, a standard deviation was calculated across all samples. A threshold of 0.5 was applied to the standard deviation value to exclude low variance data, with the same value applied to all the samples. This step decreases the redundant feature and increases the computational speed of analysis.

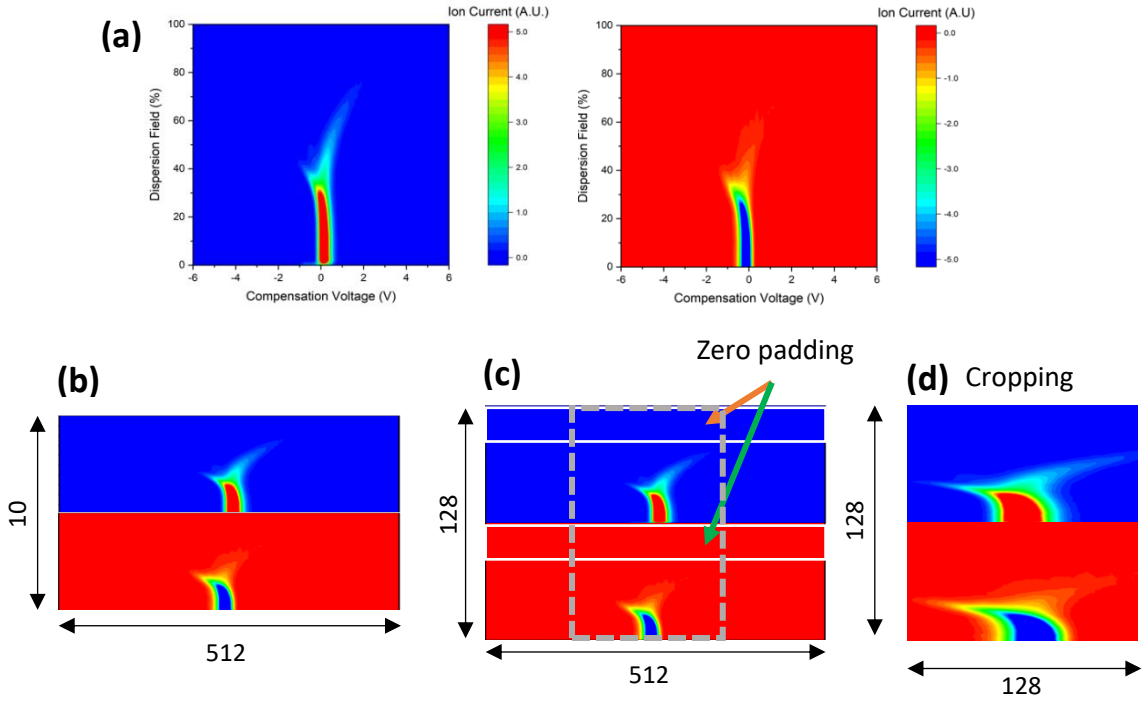


Figure 4.9. FAIMS Pre-Processing Step Involving: (a) Raw Positive and Negative Ions Data; (b) Combining Positive and Negative Ion Data Into 102 x 512 Matrix; (c) Padding the Data with Zero to Make 128 x 512 Matrix; (d) Cropping the Data Into 128 x 128 Matrix for 2D DWT.

In the beginning of data analysis pipeline development, similar pre-processing stages were also applied to GC-IMS dataset. One of the key benefits of GC-IMS compared to FAIMS is the potential of chemical discovery as each peak in the GC-IMS spectrum is associated with certain chemical compounds. Hence, the importance of the ability to track back important features into the original dataset is crucial for chemical biomarker discovery. The complexity of 2D DWT in very high dimensional dataset, in terms of remapping the features of interest into the original data, due to the scaling and location shifting of wavelet, made us decide to stop using this pre-processing technique. However, the pancreatic study (section 5.2.1) and IBD study (section 6.2) which both used FAIMS and were published in 2018, were still using 2D DWT and standard deviation threshold. On the other hand, the FAIMS study in detecting colorectal cancer (section 5.3.1), which was published in 2019, applied only a 0.5 standard deviation threshold.

GC-IMS

The goal of pre-processing in GC-IMS signal processing is to extract only the information that is required for the training and model's construction. The raw GC-IMS outputs are complex and contain a high degree of dimensionality, but much lower information

content. To extract the features, a variety of pre-processing procedures are needed, including RIP alignment, cropping, and thresholding.

As the GC-IMS operates through an ionisation process at atmospheric pressure, the nitrogen or air used as the carrier and drift gas in the IMS often contains water which is important for the ionisation process to occur. This water molecule reacts with fast electrons emitted by the beta radiation source, resulting in the formation of reactant ions. This reaction generates the RIP [10]. These ions are found in all spectra and may be utilised to align the samples. The RIP is found to remain consistent across all samples. However, due to the variation in ambient pressure level caused by weather conditions, it is noticed that the RIP position shifts slightly between samples. Thus, RIP alignment is performed as the first step of the pre-processing step by selecting 1 sample as a reference and moving all other RIP samples to match the position of the sample reference. This RIP correction is important to help the feature selection method to choose the most informative features as a classification model's input.

A typical GC-IMS output matrix consists of 11 million data points. This high dimensionality data does not align with the information content stored in the data. More than 50% of these data points are background noise, which can be seen as a blue area in Figure 4.5. To reduce this, the next steps of pre-processing are cropping out the area of interest and applying threshold to the dataset. These processes produce a pre-processed GC-IMS output matrix, which is made of zero data as a background and non-zero data points represent VOC peaks. Figure 4.10 illustrates the pre-processing steps as explained above.

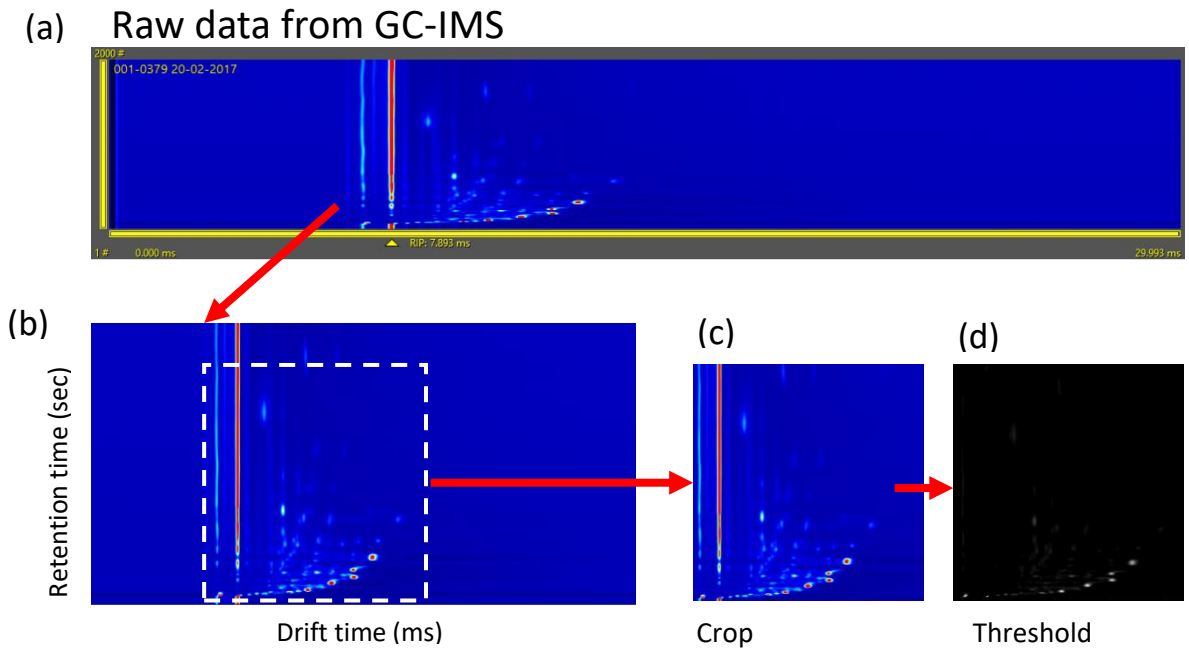


Figure 4.10. The Pre-Processing Steps of GC-IMS Dataset Involve: (a) Raw Data of GC-IMS; (b) RIP Alignment; (c) Cropping Area of Interest; (d) Thresholding to Remove the Noise.

GC-TOF-MS

In the case of GC-TOF-MS, there are three stages involved in pre-processing process. It begins by extracting important information, specifically a list of identified chemicals, the area of the peak, and its height for each of these chemicals. A new spreadsheet is then created with all the samples in which the rows contain all of samples and the column represent the chemicals detected in each sample. Finally, the chemicals that are present in only a few samples are eliminated to clean up the data and make it easier for further processing steps. This cleansing allows the number of chemicals to decrease the total number of chemicals found in all the samples from the 1000's to only 100's.

4.3.2. Data Splitting

Data splitting is a fundamental process that happens after raw data has been cleaned from noise and undesirable variability in the pre-processing step and before classification model construction. It is a technique for dividing a given dataset into at least two subsets, which are referred to as the training and test set. The training subset is used to create feature sets, develop, and compare the model to achieve a final model. The final model is then tested using the test subset to measure the model's performance. Ideally, the final model should not only be capable of making accurate predictions, but it should also have the capability of

making generalisations. The poor generalisation may be characterised by overtraining, which causes the model to learn the pattern too well, resulting in the model being unable to identify the sample pattern outside the pattern found in the training set [11]. These two conflicting goals, which are excellent prediction and good generalisation, are also known as the Bias and Variance dilemma [12]. Cross-validation is a method that is often used to solve the problem of bias and variance dilemmas. This technique will be discussed in more depth in the next section.

Sometimes, when a study has a big dataset, it is also possible to acquire an additional subset known as validation. This subset is a set that falls between the training and test set, which helps hyperparameter optimisation of the model. In the studies relating to this thesis, we utilised random data splitting to choose which data belongs to training, validation, or test set. Due to the fact that the total sample size of each study is not the same, we split the data according to certain rules. If a study had more than 50 samples in each group, we split the data into three sets: training, validation, and test set. The training and validation sets were used to train and tune the model, and the test set was used to evaluate how well the final model performed. Although some studies in this thesis contain more than 50 samples in each class, others have fewer than 50 samples per class since data collection in the medical field is costly and the availability of patients willing to participate in the study is often limited. In this particular instance, we simply split the data into training and validation sets, with both of them being subjected to cross-validation process.

4.3.3. Cross-Validation

As mentioned before, one of the most common methods to find the trade-off between bias and variance in model training process is cross-validation. It is a model validation approach that is used to assess how correctly a prediction model will perform in a real-world situation. The fundamental concept of cross-validation is to divide the dataset into several subsets in which some subsets (training set) are used to train the prediction model, and the other subset (called the validation set) are utilised to evaluate prediction error of the trained prediction model repeatedly using multiple split variation of the dataset. Then, the model's predictive performance is estimated using the averaged outcomes of the validation tests. The two most used cross-validation techniques are The Leave One Out Cross-Validation (LOOCV) and k-fold cross-validation.

The LOOCV method creates the training set by using the whole dataset as input, with the exception of a single observation. Training and testing a model are done using n-1

observations in which n is the total number of samples and the observation that is left out of the training set is then used for testing. This is then repeated until all of the observations have been used as part of the test, at which point the procedure is completed. It means one model needs to be constructed and assessed for each sample in training set. Having so many models fitted and assessed, results in a more robust assessment of model performance since each data has the chance to represent the whole test dataset, which leads to more accurate estimation. However, due to the fact that the model must be fitted n times as many as the number of samples available, it is computationally expensive. Therefore, LOOCV is not suitable for big dataset and already computationally tasking model such as neural network. The illustration of the LOOCV method can be seen in Figure 4.11.

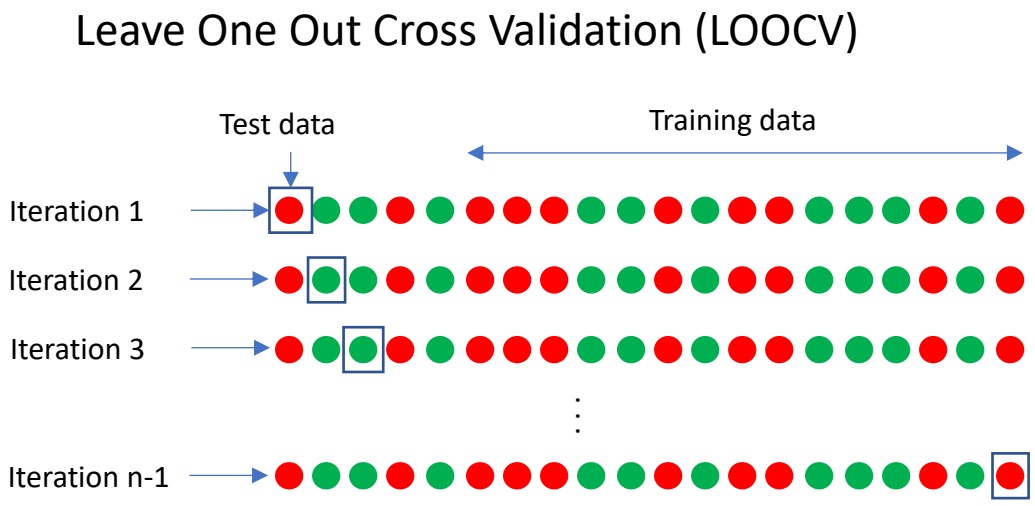


Figure 4.11. An Illustration of Leave One Out Cross Validation (LOOCV)

On the other hand, the k-fold cross validation makes use of a random splitting of the original dataset into roughly equal-sized subgroups of k to obtain its results. The most frequent values for k are 3, 5, and 10, with 10 being the most common value as a 10-fold cross-validation offers a good balance between low computational cost and minimal bias in estimating the model performance, when comparing to LOOCV and a simple single train-test split. Not only that, but a 10-fold cross-validation is also extremely common in the applied machine learning field. Hence, it is used in all of studies relating to this thesis. The ten randomly split subsets are then divided into two groups. The first group consists of nine subsets that are used as the training set, while another group (the one remaining subset) is retained for model assessment, as shown in Figure 4.12. This process is repeated until all of the subsets have been used as the test set for the final result.

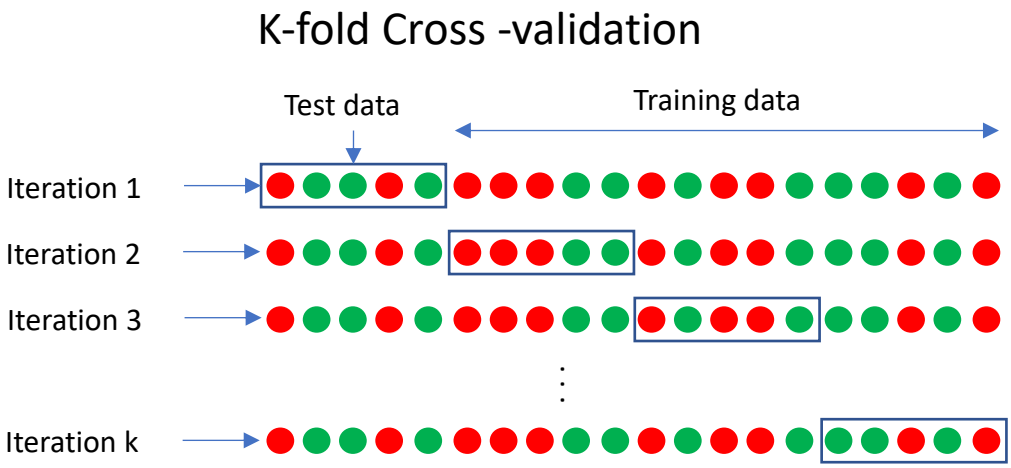


Figure 4.12. An Illustration of K-fold Cross-Validation

4.3.4. Imbalance Dataset

An issue that often arises in medical diagnostic research is imbalanced classification. In nearly every illness, a medical laboratory has more patients who do not have the disease than patients who do have it, especially in cancer [13]. In an imbalanced dataset, samples are distributed unevenly across the compared classes [14], with one class having fewer samples than the other classes. This imbalanced class distribution is likely to introduce extra problems into the classifier learning process. The assessment criteria that drive the learning process inside the classification algorithm may lead to the exclusion of minority classes (which are treated as noise), and as a result, the generated classifier may lose its classification capacity. Furthermore, this will lead to overly optimistic accuracy result as the classifier may attempt to optimise the accuracy of its classification rule by neglecting the minority class [15]. Thus, the classifier is therefore susceptible to overfitting the majority class. Many researchers have attempted to solve the imbalanced problem thru various techniques [16,17,18]. However, one of the most common techniques to approach this problem is resampling, which can be done by either undersampling the majority class or oversampling the minority class.

Undersampling is a technique that is often used to handle the imbalanced dataset by removing the samples in majority class to reduce the ratio of minority and majority class to desired ratio, like 50:50, 60:40 in binary classification or 40:30:30 in multiple classification problems. The simple undersampling technique reduces the majority of class samples randomly. Further development of undersampling technique involving Condensed Nearest Neighbour Rule (CNN) [19], TomekLinks [20], Edited Nearest Neighbor Rule (ENN) [21], or Neighbourhood Cleaning Rule (NCL) [22] in its technique to either choose which samples

from majority class that can be removed or which samples can be kept. Even though this technique is simple, it is only able to apply to a study in which the minority class have enough number of data to train the model. Due to the unavailability of data as a result of fewer patients who have the disease or the high cost of collecting more data, more often than not in medical research the overall dataset is already small. Another drawback of undersampling is this technique may eliminate information that might be useful.

On the other hand, the oversampling is a technique that addresses the imbalanced dataset problem by increasing the number of samples in minority classes to the desired ratio. The simple oversampling method works by replicating the minority class to match with the number of majority class. Although this technique does not reduce the valuable information from the original dataset and is more suitable for small medical dataset, employing this technique in the pipeline can increase the possibility of overfitting. Furthermore, this method does not give new information to the model during training which may lead to poor generalisation when new unseen data is tested. An illustration of undersampling and oversampling can be seen in Figure 4.13.

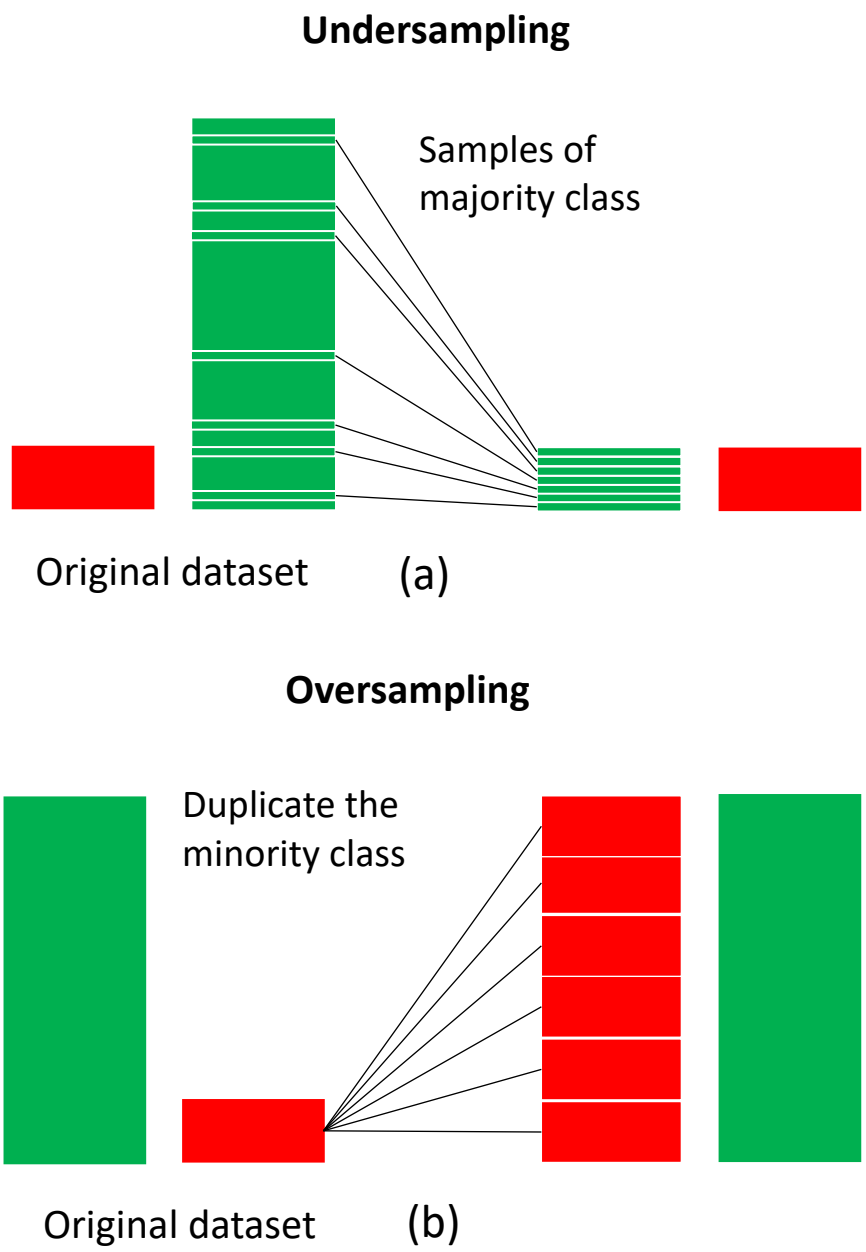


Figure 4.13. Illustration of (a) Undersampling in Which the Number or Majority Sample is Reduced to Match the Minority Sample; (b) Oversampling Where the Minority Sample is Duplicated Until it Matches with the Majority Sample.

Synthetic Minority Oversampling Technique (SMOTE) offers a better way of solving the above-mentioned problems. Instead of duplicating the sample from minority class randomly, it generates a new synthetic sample from the minority class data. It first randomly selects the samples that are close in the feature space from the minority class. Then, k of the nearest neighbours selected samples are identified (usually k =5). A randomly selected neighbour is picked, and a synthetic sample is generated at a randomly selected position

between the two samples in feature space (as illustrated in Figure 4.14) [23]. This method can be used to generate as much synthetic data for the minority class as necessary. This technique is an effective solution in reducing the possibility of overfitting while preserving the important information inside the dataset because the new synthetic data is created reasonably near in feature space to existing data from the minority class [23]. The drawback of this method is the new synthetic data is created without taking into account the majority of class, which may result in samples that are unclear when there is a significant overlap between the classes. In the studies relating to this thesis, we used a combination between SMOTE and undersampling as suggested in Nitesh Chawla, *et al.* [23]. It was performed inside 10-fold cross-validation using R package called unbalanced.

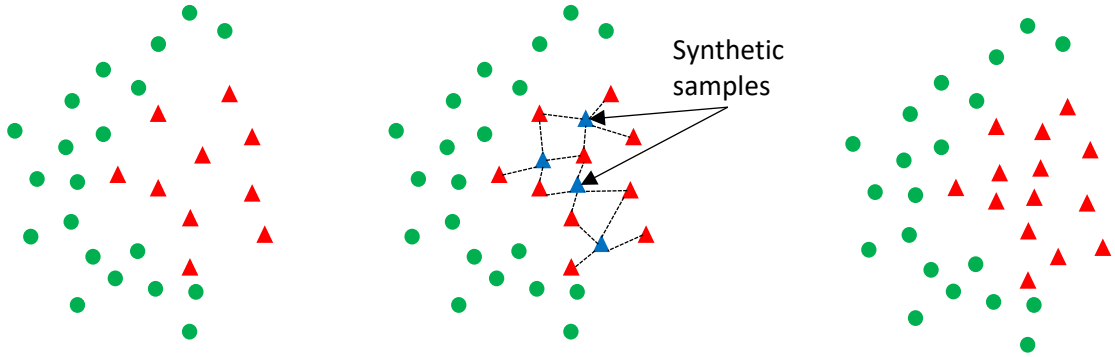


Figure 4.14. An illustration of SMOTE in creating synthetic samples from minority class

4.3.5. Feature Selection

As can be seen from Figure 4.9 and Figure 4.10, even though the pre-processing techniques have been applied, the dimension of FAIMS and GC-IMS data are still high (around 16384 and 1976000 data points in FAIMS and GC-IMS, respectively). We do not expect all of these features to be informative. Therefore, we further reduce the number of features by employing supervised feature selection technique in the pipeline. In this work, a Wilcoxon rank sum test is used as the supervised feature selection technique by comparing each feature between groups from pre-processed data (FAIMS or GC-IMS) in order to find the feature point with the lowest p-value among the others (as illustrated in Figure 4.15). It is a non-parametric version of the two-sample t-test that is often used for feature selection [24]. A non-parametric test is suitable for the nature of dataset as it does not make an assumption that the data follows a normal distribution.

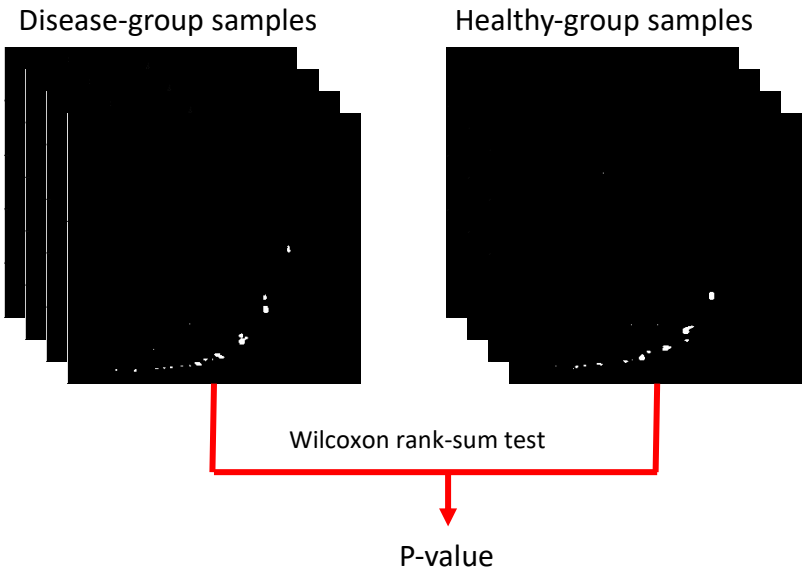


Figure 4.15. Illustrate the Feature Selection Method Using Wilcoxon Rank Sum Test by Comparing Each Feature Between Groups to Get the p-value.

Initially, we run the classification analysis using 2 to 100 lowest p-value features to find the optimum number of features to get the highest classification performance possible. In the first study of pancreatic cancer detection using FAIMS (section 5.2.1), we used the 44 most informative features. Figure 4.16 shows the performance of the random forest classification algorithm in detecting pancreatic cancer samples from healthy control samples in the variation of number of features used. However, as the work moved to process GC-IMS data, it was found that this procedure increased the computational cost and analysis time dramatically. Hence, we simplify the process by using the 20, 50 or 100 lowest p-value features as an input to the classification algorithms (or just 100 for very large sample numbers). If the study only covered a limited number of participants, 20 features are chosen. This is adequate information for the algorithms to learn and reduce the possibility of overfitting while still allowing for the discovery of significant features (later in the pipeline). Currently, the drawback of this method is the fact that features are identified only on the basis of statistics and not on the basis of any biological function.

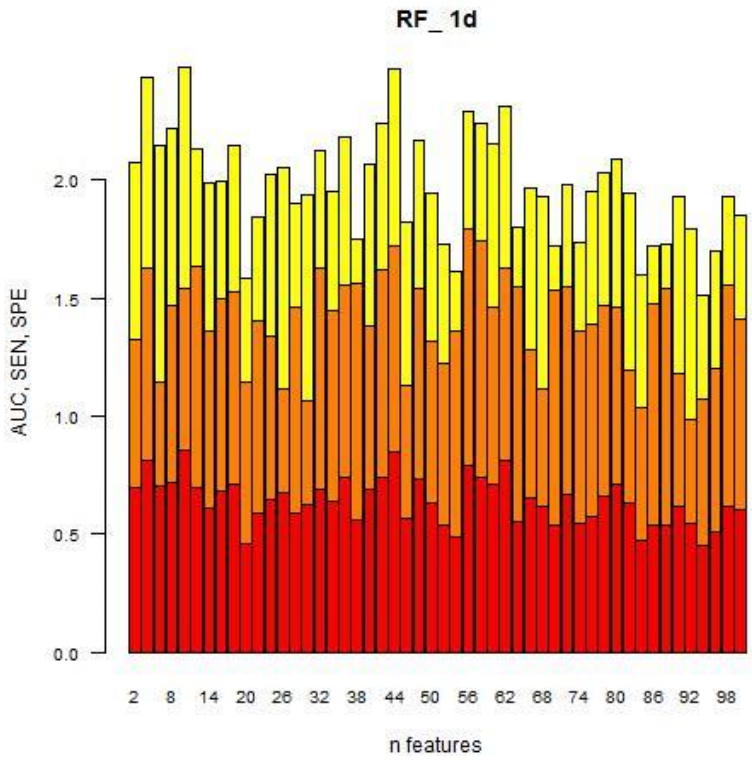


Figure 4.16. The Overall Performance of Random Forest Classifier is Represented as the Sum of AUC, Sensitivity, and Specificity for Variation of Features from 2-100 when Detecting Pancreatic Cancer from Healthy Control Using FAIMS.

4.3.6. Classification Models

The features that have been chosen by Wilcoxon rank sum test are then used as inputs for machine learning classification algorithms inside 10-fold cross-validation. Based on the review paper on the most classification algorithms used to analyse Breath VOC for disease detection, 4 classification algorithms were chosen [25]. They are Sparse Logistic Regression (SLR), Random Forest (RF), Support Vector Machine (SVM), and Neural Network (NN). We also included the Gaussian Process (GP) into our pipeline as the fifth classification algorithm. This algorithm is a sophisticated non-parametric machine learning classification method that is able to handle high dimensional data with limited samples [26] and is often used in medical diagnostic [27,28,29]. In certain cases, some of these methods outperform others depending on the dataset and technology used. Examples include a research done by Di Gilio *et al*., in which RF outperformed SVM in terms of discriminating between healthy controls and patients with Malignant Pleural Mesothelioma (MPM) by analysing breath VOCs using TD-GC-MS [30]. Another research by Bobak *et al*. also showed that RF surpassed SVM for distinguishing between tuberculosis and other lower respiratory illness in children using GC x GC-TOF-MS

breath analysis [31]. However, SLR successfully outperformed both SVM and RF in differentiating coeliac disease from irritable bowel syndrome (IBS) by urinary VOC analysis using FAIMS [32]. This is due to the fact that each algorithm has its own mathematical basis at its core. A short description of each algorithm is given below.

Sparse Logistic Regression

Sparse Logistic Regression (SLR) is a method that often used to solve classification problems in high dimensional dataset such as cancer classification based on gene expression [33,34]. It is because SLR has its own feature selection technique that is integrated into the method (usually called an "embedded method"), which means it performs feature selection and classification simultaneously [35]. This can be achieved by employing a Least Absolute Shrinkage and Selection Operator (LASSO) regularisation (L1 norm) into the classical logistic regression [36]. It makes the model only focus on small subset of non-zero parameters and makes it less prone to overfitting. SLR is suitable classification algorithm to apply in our pipeline since the dataset, especially generated by FAIMS and GC-IMS, are a high dimensional dataset and involve many zero coefficients in the feature.

Random Forest

The Random Forest (RF) method, developed by Breiman in 2001, is one of the most successful general-purpose algorithms to have been developed in recent history [37]. It is made up of hundreds or thousands of decision trees, each of which is built using a portion of the training data that has been randomly chosen with replacement, known as bootstrapping. It means some data will be used more than once in a single tree. Using these subsets as inputs, each tree is fitted to the input to learn the pattern and make a class prediction. Then, the class prediction of each tree inside the random forest model is combined and aggregated to create a single prediction. RF is well known for its accuracy as well as its capacity to cope with small sample sizes and high dimensional feature spaces [38]. Hence, RF is selected as one of the machine learning algorithms to be included in our pipeline.

Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm that works by identifying an optimal hyperplane with maximum margin distance in an n-dimensional space (where N is the number of features), which distinguishes between the different classes of data points. Hyperplane is a decision surface that helps differentiate data

points based on their classes and margin distance is the distance between the hyperplane and the closest data points. These data points are called support vectors, which have an impact on the location and orientation of the hyperplane. Increasing the margin distance to its maximum value offers some reinforcement, allowing subsequent data points to be categorised with more certainty [39]. When dealing with nonlinear problems, SVM uses a kernel function that turns low dimensional data into higher dimensional data to help the algorithm find the optimum hyperplane. SVM is a technique that is particularly well suited for binary classification problems with high dimensional datasets [40].

Gaussian Process

GP is a supervised probabilistic classification model, in which test predictions are expressed as class probabilities. To achieve a probabilistic classification, GP applies a prior probability distribution on a latent function, which is then squished via a link function [41]. The latent function is used to simplify model construction. However, the value of this function itself is not observable or relevant. Thus, it's often referred as nuisance function and will be eliminated during prediction. For binary classification, GP employs the logistic link function.

Neural Network

NN tries to mimic the learning process of biological brain, where it is made up of numerous complexed of linked "neurons" that work together to solve problem [42]. The neuron employed in the NN model is based on a non-linear function that produces an output depending on one or more inputs. Like synapses in a real brain, each neuron has the ability to send a signal (a real number) to other neurons. Each neuron has an associated weight, which can increase or decrease the strength of the input signal passing this neuron based on relative importance to the other neurons. NN is typically made of multiple layers of neurons, in which the final layer produces a single classification result.

The classification pipeline includes all five classifiers in order to compare their result. These analyses are conducted using R statistical programming language. The following machine learning packages are used in this study: SLR – glmnet; RF- randomForest, SVM – kernlab, GP – kernlab, NN – neuralnet. Hyperparameters tuning of each algorithm are done automatically using AutoTuner function from mlr3 package only to the training dataset. Here is the list of hyperparameters that are tuned:

- SLR – Family = binomial; alpha tuned in the range of 0 – 1.
- RF – ntree tuned in the range of 200 – 1000; mtry = $\sqrt{number\ of\ features}$.

- GP – Kernel = rbfdot; kernel spread (sigma) tuned in the range of 10^{-2} - 10^{7}.
- SVM – Kernel = rbfdot; C tuned in the range of 2^{-2} - 2^{2}; kernel spread between 10^{-2}-10^{2}.
- NN – hidden tuned between 1 to 5.

After the training process is finished inside the cross-validation, the final model is then applied to the testing set. However, this step was only applied to studies with a large dataset like the pancreatic cancer study using FAIMS (section 5.2), the colorectal cancer study using GC-IMS (section 5.5) and the faecal VOC IBD study using GC-IMS (section 6.5). Studies with small datasets took the result from the cross-validation as the final result. The performance of the models as the final result is represented by sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), p-value, Receiver Operating Characteristic (ROC) curve and Area Under the ROC curve (AUC). The details explanation of these metrics is explained in the next section.

4.3.7. Performance Metrics and ROC Curve

The final results are derived from the probabilities generated by the classifiers and averaged across all tests to provide performance metrics. The performance metrics are constituted based on a confusion matrix, as shown in Figure 4.17 [43].

		Predicted	
		Positive	Negative
Actual	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

Figure 4.17. Confusion Matrix

From the confusion matrix, the following information will be provided:

- True Positive (TP): The algorithm's predicted a positive result that matches the positive result for the patient.

- False Negative (FN): The algorithm is incorrectly classified the disease sample as not positive/control patient.
- True Negative (TN): The algorithm's predicted a negative result that matches a non-positive patient.
- False Positive (FP): The algorithm is incorrectly classified a negative patient as a positive patient.

Once the confusion matrix had been constructed, the performance of the classification algorithms was compared using parameters such as sensitivity, specificity, PPV, NPV, p-value, ROC Curve, and AUC . These parameters are determined as follows:

$$Sensitivity = \frac{TP}{TP + FN} \qquad PPV = \frac{TP}{TP + FP}$$

$$Specificity = \frac{TN}{TN + FP} \qquad NPV = \frac{TN}{TN + FN}$$

- Sensitivity (true positive rate) is the likelihood that a test will identify a positive patient as a patient with the disease.
- Specificity (true negative rate) is defined as the likelihood that individuals who are not positive will be classified as such.
- Positive Predictive Value (PPV) indicates the likelihood that a positive patient will present when the test is positive.
- Negative Predictive Value (NPV) indicates the likelihood that a negative patient will not present when the test is negative.

The calculation of the p-value during classification may be used to determine if the null hypothesis, i.e., that there are no differences between groups, is correct or not. A p-value of less than 0.05 has traditionally been regarded as an acceptable threshold for rejecting the null hypothesis [44]. This means that the presence of a low p-value suggests that an alternative null hypothesis is more likely to be correct. This may demonstrate that the chosen features can be used to differentiate between the groups to some extent.

Since the raw output from classification model is represented in probability, sensitivity and specificity are often not enough to measure the result. Ideally, the performance of a diagnostic test should be evaluated over the whole range of possible cut points. Hence, Receiver Operating Characteristic (ROC) curves are used in this pipeline to enable us to

clearly see the performance of the model prediction across all possible decision thresholds. In a ROC curve, the sensitivity is plotted as a function of 1-specificity in which each point corresponds to a pair of sensitivity and specificity values associated with a particular decision threshold [45]. The area under the ROC curve (AUC) is a useful method to assess the overall diagnostic accuracy of a test. The AUC value ranges from 0 to 1, with 0 representing a perfectly incorrect test, 0.5 representing the randomness (no discriminatory power), and 1 representing an accurate test. Typically, an AUC value of 0.7 or above indicates adequate separation, as illustrated in Figure 4.18 [46]. A 95% Confidence Interval (95% CI) is also used inside the brackets along with the AUC value. This shows the lower- and upper-range of the likelihood of the true model AUC value.

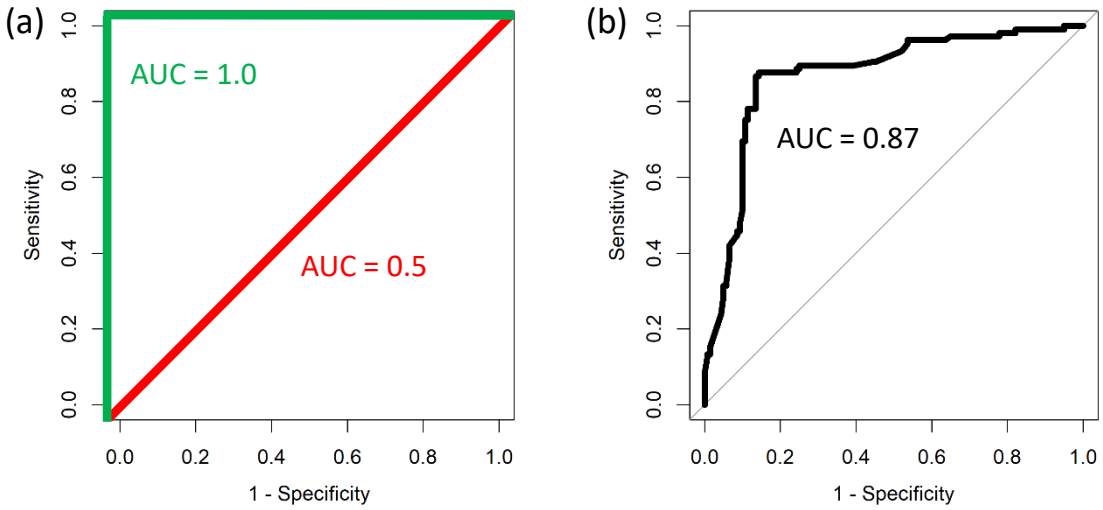


Figure 4.18. Example of ROC Curve with no, Good, and Perfect Separation Represented by Red, Black, and Green Line Respectively

4.3.8. Significant Feature

As mentioned in chapter 3, GC-IMS has the capability of identifying the chemical compounds correspond to the VOC signal peak in the data by matching the retention time and drift time of the peak to a NIST library. For GC-IMS, this library is embedded into the GCxIMS library search software (v1.0.1, G.A.S., Dortmund, Germany). Though this GCxIMS library search is not seen to be as sophisticated as the GC-MS library, it still has 83,000 compounds listed in the database [47]. When a retention time and drift time data of potential VOC signal peak is searched using the GCxIMS library search software, a list of possible matching compounds will be shown. This ability is important for biomarker discovery.

To achieve that, the selected features used for training the classification model are plotted back onto the pre-processed GC-IMS output. It is noticed that the selected feature datapoints 'cluster' around specific VOC peaks on the output plot when the classification models performed well (AUC of around 0.7 or higher). On the other hand, the features are scattered throughout the output plot when the models have poor performance, as shown in Figure 4.19. The white circle indicates the location of the features. Using this information, enable us to identify the VOC peaks of interest which help in biomarker discovery.

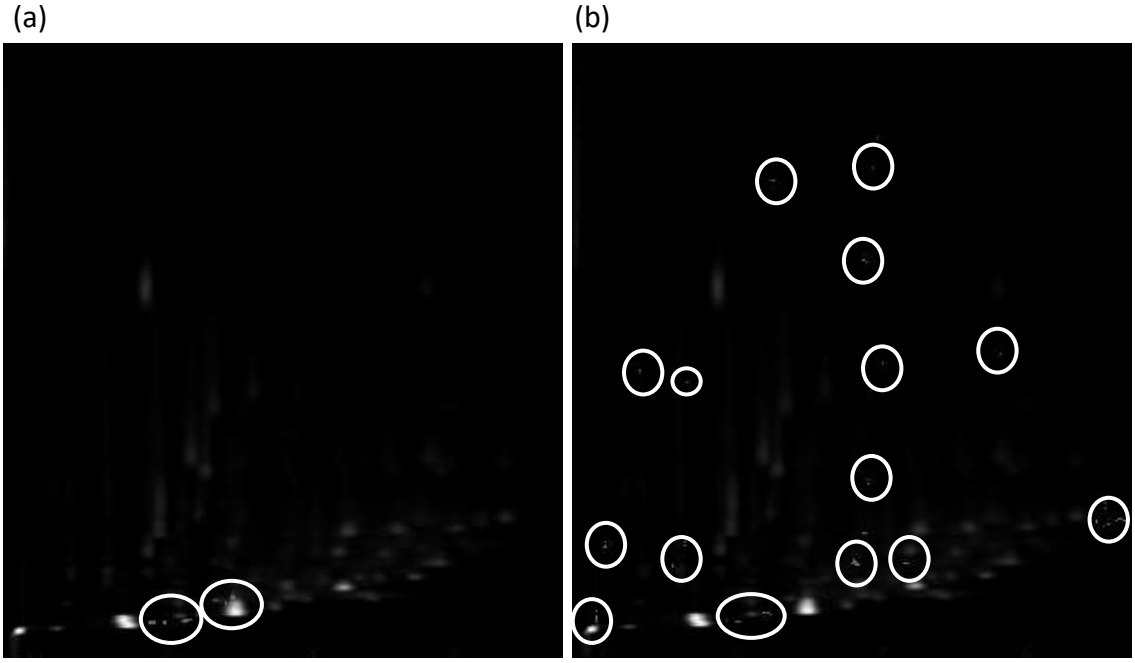


Figure 4.19. Selecting Features are Plotted Back onto the GC-IMS Output: (a) Excellent Separation AUC = 0.91; (b) Poor Separation AUC = 0.52.

4.4. Conclusion

In conclusion, the data analysis pipeline used in this study consist of pre-processing, data splitting, cross-validation, feature selection, classification, and model performance calculations. The pre-processing step is different for each instrument as every instrument has a different data characteristic. The data splitting is also different between studies in which the data from studies with smaller datasets was only divided into training and test sets, whilst the data from studies with large datasets was divided into three set, i.e., training, validation, and test set. Among all instruments, GC-TOF-MS and GC-IMS offer the capability of chemical identification through matching the data to a NIST library. Throughout this thesis, the techniques outlined in this chapter have been implemented, and they will be referred to in the corresponding method section in chapters 5 and 6. New clinical trials examining the urinary,

faecal, and breath VOC of patients with pancreatic cancer, colorectal cancer, and inflammatory bowel disease are described in the following two chapters (chapters 5 and 6), using the methods and instruments described in chapter 3 and this chapter.

4.5. Reference

[1] A. Tiele *et al.*, "Breath-based non-invasive diagnosis of Alzheimer's disease: A pilot study," *J. Breath Res.*, vol. 14, no. 2, p. 026003, Feb. 2020, doi: 10.1088/1752-7163/ab6016.

[2] S. Esfahani, A. Wicaksono, E. Mozdiak, R. P. Arasaradnam, and J. A. Covington, "Non-invasive diagnosis of diabetes by volatile organic compounds in urine using FAIMs and FOX4000 electronic nose," *Biosensors*, vol. 8, no. 4, 2018, doi: 10.3390/bios8040121.

[3] L. Lacey, E. Daulton, A. Wicaksono, J. A. Covington, and S. Quenby, "Detection of Group B Streptococcus in pregnancy by vaginal volatile organic compound analysis: a prospective exploratory study," *Transl. Res.*, vol. 216, pp. 23–29, 2020, doi: 10.1016/j.trsl.2019.09.002.

[4] E. Daulton, A. Wicaksono, J. Bechar, J. A. Covington, and J. Hardwicke, "The detection of wound infection by ion mobility chemical analysis," *Biosensors*, vol. 10, no. 3, pp. 1–9, 2020, doi: 10.3390/bios10030019.

[5] D. J. C. Berkhout *et al.*, "Preclinical Detection of Non-catheter Related Late-onset Sepsis in Preterm Infants by Fecal Volatile Compounds Analysis: A Prospective, Multi-center Cohort Study," *Pediatr. Infect. Dis. J.*, no. February, pp. 330–335, 2020, doi: 10.1097/INF.0000000000002589.

[6] J. Yan *et al.*, "Electronic nose feature extraction methods: A review," *Sensors (Switzerland)*, vol. 15, no. 11, pp. 27804–27831, 2015, doi: 10.3390/s151127804.

[7] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust.*, vol. 37, no. 12, pp. 2091–2110, 1989, doi: 10.1109/29.45554.

[8] G. Nason, "Package ' wavethresh ,'" 2016. https://cran.r-project.org/web/packages/wavethresh/wavethresh.pdf (accessed Jul. 20, 2021).

[9] A. Shahbahrami, "Algorithms and architectures for 2D discrete wavelet," pp. 1045–1064, 2012, doi: 10.1007/s11227-012-0790-x.

[10] C. J. Denawaka, I. A. Fowlis, and J. R. Dean, “Evaluation and application of static headspace-multicapillary column-gas chromatography-ion mobility spectrometry for complex sample analysis,” *J. Chromatogr. A*, vol. 1338, pp. 136–148, 2014, doi: 10.1016/j.chroma.2014.02.047.

[11] Z. Reitermanov, “Data Splitting,” in *WDS Proceeding*, 2010, pp. 31–36.

[12] I. Kononenko and M. Kukar, *Machine Learning and Data Mining: Introduction to Principles and Algorithms*. Horwood Publishing Limited, 2007.

[13] J. Diz, G. Marreiros, and A. Freitas, “Applying Data Mining Techniques to Improve Breast Cancer Diagnosis,” *J. Med. Syst.*, vol. 40, no. 9, 2016, doi: 10.1007/s10916-016-0561-y.

[14] Y. Sun, A. K. C. Wong, and M. S. Kamel, “Classification of imbalanced data: A review,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 4, pp. 687–719, 2009, doi: 10.1142/S0218001409007326.

[15] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, “A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 4, pp. 463–484, 2012, doi: 10.1109/TSMCC.2011.2161285.

[16] H. Zhang, H. Zhang, S. Pirbhulal, W. Wu, and V. H. C. D. Albuquerque, “Active balancing mechanism for imbalanced medical data in deep learning-based classification models,” *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 16, no. 1s, pp. 1–15, 2020, doi: 10.1145/3357253.

[17] C. Huang *et al.*, “Sample imbalance disease classification model based on association rule feature selection,” *Pattern Recognit. Lett.*, vol. 133, pp. 280–286, 2020, doi: 10.1016/j.patrec.2020.03.016.

[18] B. Sun, H. Chen, J. Wang, and H. Xie, “Evolutionary under-sampling based bagging ensemble method for imbalanced data classification,” *Front. Comput. Sci.*, vol. 12, no. 2, pp. 331–350, 2018, doi: 10.1007/s11704-016-5306-z.

[19] T. Nn, T. Bayes, T. Cnn, and T. Cnn, “The condensed nearest neighbor rule (Corresp),” *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 1966–1967, 1967.

[20] I. Tomek, “Tomek Link: Two Modifications of CNN,” *IEEE Trans. Syst. Man Cybern.*, vol. 6, no. 11, pp. 769–772, 1976, [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4309452.

[21] D. L. Wilson, “Asymptotic Properties of Nearest Neighbor Rules Using Edited Data,” *IEEE Trans. Syst. Man Cybern.*, vol. 2, no. 3, pp. 408–421, 1972, doi: 10.1109/TSMC.1972.4309137.

[22] J. Laurikkala, “Improving identification of difficult small classes by balancing class distribution,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2101, pp. 63–66, 2001, doi: 10.1007/3-540-48229-6_9.

[23] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[24] A. S. Martinez-Vernon *et al.*, “An improved machine learning pipeline for urinary volatiles disease detection: Diagnosing diabetes,” *PLoS One*, vol. 13, no. 9, pp. 1–20, 2018, doi: 10.1371/journal.pone.0204425.

[25] J. H. Leopold *et al.*, “Comparison of classification methods in breath analysis by electronic nose,” *J. Breath Res.*, vol. 9, no. 4, p. 46002, 2015, doi: 10.1088/1752-7155/9/4/046002.

[26] N. Zhang, J. Xiong, J. Zhong, and K. Leatham, “Gaussian Process Regression Method for Classification for High-Dimensional Data with Limited Samples,” in *2018 Eighth International Conference on Information Science and Technology (ICIST)*, Jun. 2018, pp. 358–363, doi: 10.1109/ICIST.2018.8426077.

[27] E. Challis, P. Hurley, L. Serra, M. Bozzali, S. Oliver, and M. Cercignani, “Gaussian process classification of Alzheimer’s disease and mild cognitive impairment from resting-state fMRI,” *Neuroimage*, vol. 112, pp. 232–243, 2015, doi: 10.1016/j.neuroimage.2015.02.037.

[28] C. Abi Nader, N. Ayache, P. Robert, and M. Lorenzi, “Monotonic Gaussian Process for spatio-temporal disease progression modeling in brain imaging data,” *Neuroimage*, vol. 205, no. May 2019, 2020, doi: 10.1016/j.neuroimage.2019.116266.

[29] R. Shashikant, U. Chaskar, L. Phadke, and C. Patil, “Gaussian process-based kernel as a diagnostic model for prediction of type 2 diabetes mellitus risk using non-linear heart rate variability features,” *Biomed. Eng. Lett.*, vol. 11, no. 3, pp. 273–286, 2021, doi: 10.1007/s13534-021-00196-7.

[30] A. Di Gilio *et al.*, “Breath analysis for early detection of malignant pleural mesothelioma: Volatile organic compounds (VOCs) determination and possible

biochemical pathways," *Cancers (Basel).*, vol. 12, no. 5, 2020, doi: 10.3390/cancers12051262.

[31] C. A. Bobak *et al.*, "Breath can discriminate tuberculosis from other lower respiratory illness in children," *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, 2021, doi: 10.1038/s41598-021-80970-w.

[32] R. P. Arasaradnam *et al.*, "Differentiating coeliac disease from irritable bowel syndrome by urinary volatile organic compound analysis - A pilot study," *PLoS One*, vol. 9, no. 10, pp. 1–9, 2014, doi: 10.1371/journal.pone.0107312.

[33] H. H. Huang, X. Y. Liu, and Y. Liang, "Feature selection and cancer classification via sparse logistic regression with the hybrid L1/2 +2 regularization," *PLoS One*, vol. 11, no. 5, pp. 1–15, 2016, doi: 10.1371/journal.pone.0149675.

[34] S. Wu, H. Jiang, H. Shen, and Z. Yang, "Gene selection in cancer classification using sparse logistic regression with L1/2 regularization," *Appl. Sci.*, vol. 8, no. 9, pp. 1–12, 2018, doi: 10.3390/app8091569.

[35] Z. Y. Algamal and M. H. Lee, "A two-stage sparse logistic regression for optimal gene selection in high-dimensional microarray data classification," *Adv. Data Anal. Classif.*, vol. 13, no. 3, pp. 753–771, 2019, doi: 10.1007/s11634-018-0334-1.

[36] M. Zanon, G. Zambonin, G. A. Susto, and S. McLoone, "Sparse logistic regression: Comparison of regularization and Bayesian implementations," *Algorithms*, vol. 13, no. 6, 2020, doi: 10.3390/A13060137.

[37] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[38] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016, doi: 10.1007/s11749-016-0481-7.

[39] W. S. Noble, "What is a support vector machine?," *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006, doi: 10.1038/nbt1206-1565.

[40] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[41] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[42] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. E. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey,"

Heliyon, vol. 4, no. 11, p. e00938, 2018, doi: 10.1016/j.heliyon.2018.e00938.

[43] A. M. Molinaro, "Diagnostic tests: How to estimate the positive predictive value," *Neuro-Oncology Pract.*, vol. 2, no. 4, pp. 161–165, 2015, doi: 10.1093/nop/npv030.

[44] J. Pereira *et al.*, "Breath analysis as a potential and non-invasive frontier in disease diagnosis: An overview," *Metabolites*, vol. 5, no. 1, pp. 3–55, 2015, doi: 10.3390/metabo5010003.

[45] J. N. Mandrekar, "Receiver operating characteristic curve in diagnostic test assessment," *J. Thorac. Oncol.*, vol. 5, no. 9, pp. 1315–1316, 2010, doi: 10.1097/JTO.0b013e3181ec173d.

[46] J. Pearce and S. Ferrier, "Evaluating the predictive performance of habitat models developed using logistic regression," *Ecol. Modell.*, vol. 133, no. 3, pp. 225–245, 2000, doi: 10.1016/S0304-3800(00)00322-7.

[47] G.A.S. Gesellschaft für analytische Sensorsysteme, "GCxIMS Library Search," 2016. http://www.gas-dortmund.de/index-gas.php?spath=464.

Chapter 5
APPLICATION ON CANCER DETECTION

5.1. Introduction

Cancer is currently a major public health problem globally with around 10 million cancer deaths worldwide in 2020 [1]. This trend is likely to continue, with the latest projections suggesting that annual cancer-related deaths will increase to 13 million by 2030 [1]. Thus, the burden of cancer will continue to exert tremendous physical, emotional, and financial strain on individuals, families, communities, and healthcare systems. Colorectal cancer and pancreatic cancer are the 2nd and 7th leading cause of cancer death worldwide (around 935.173 and 466.003 death, respectively) [1].

The most frequent pancreatic malignant tumour is pancreatic ductal adenocarcinoma (PDAC). PDAC represents around 85% of all reported pancreatic cancer cases [2]. It stubbornly resists our attempts to successfully target with current therapies, which are reflected in patient 5-year survival rates at below 8% [3]. This exceptionally poor prognosis is largely due to late diagnosis. However, if the disease is detected when the cancer is at an early stage (< 2 cm diameter) and still confined to the pancreas, the survival rate can increase, potentially up to 75% [4,5]. Current pancreatic cancer diagnostic techniques rely on imaging (e.g. CT and MRI), endoscopy ultrasonography and biopsy for grading tumour histology [6]. Unfortunately, these techniques are not very effective at detecting tumours that are smaller than 2 cm in diameter [7].

Colorectal cancer (CRC) is one of the three malignancies with the highest incidence in the industrialised world, with a 5-year survival rate of 64.4% and 66.6% for colon and rectum cancer, respectively [8]. The majority of CRC originates from dysplastic adenomatous polyps, so-called advanced adenomas (AA) [9]. Early detection and removal of these precancerous polyps are essential for the improvement of CRC course and prognosis [10]. The immunochemical faecal occult blood test (iFOBT) is currently used for population-based screening but lacks sensitivity, indicated by the missed diagnosis of CRC in 1-47% and AA in 43-61% of the tests [11]. In addition, specificity is suboptimal, as approximately 7% of the performed tests provide false positive results leading to the performance of unneeded colonoscopies. This emphasises the need for improvement of CRC/AA bowel screening tools.

This chapter describes work on new clinical studies investigating the potential of urinary and faecal VOC as a non-invasive biomarker to detect PDAC and CRC. This chapter consists of four studies. The first study describes the potential of urine VOC analysis for PDAC detection using FAIMS. Then, a follow-up study is discussed using GC-IMS and GC-

TOF-MS to understand more about the chemical biomarker in the urine that contributes to the separation of the previous study. The third and fourth studies are focused on colorectal cancer. It begins with the initial study using FAIMS and GC-IMS to detect CRC and adenoma from the VOC that is generated from the urine within the United Kingdom Bowel Cancer Screening Program (UK BCSP). The latter study discusses the use of faecal VOC analysis to detect CRC and adenoma from healthy controls using GC-IMS. It also describes the potential of faecal VOC analysis for secondary non-invasive follow-up after polypectomy. Finally, the result and comparison of these studies are then discussed separately based on cancer type, which is around the different instruments and different biological types used.

5.2. Non-invasive Diagnosis of pancreatic cancer through detection of volatile organic compounds in urine

The content of this section is based on a paper that has been published in peer-reviewed journals:

Arasaradnam, Ramesh P., Alfian Wicaksono, Harrison O'Brien, Hemant M. Kocher, James A. Covington, and Tatjana Crnogorac-Jurcevic. 2018. "Noninvasive Diagnosis of Pancreatic Cancer Through Detection of Volatile Organic Compounds in Urine." Gastroenterology 154 (3): 485-487.e1. https://doi.org/10.1053/j.gastro.2017.09.054.

5.2.1. Methods

In this work, urinary VOC was analysed using a commercial field asymmetric ion mobility spectroscopy instrument (FAIMS) (Lonestar, Owlstone, Cambridge, UK), which is fitted with an ATLAS (Owlstone) sampling system. It is highly sensitive (detecting ions up to parts per billion) and based on distinguishing different ions based on its physical properties, specifically its mobility; it does not identify per se the specific volatile organic compounds. The working principle of FAIMS has been described in Chapter 3. Patient information is summarised in Table 5.1. This information was provided by the clinical team.

Table 5.1. Demographic Information of All Study Participants

	Healthy Control (n = 81)	Pancreatic Cancer (n = 81)
Age (mean ± SD)	51.4 ± 10.6	64.3 ± 23.7
Male sex (%)	30.9	53.1

Stage (n)*		
I	NA	4
IIA	NA	5
IIB	NA	35
III	NA	24
IV	NA	12

NA, not applicable; SD, standard deviation. *One case could not be assessed

The PDAC patients were divided into 4 stages based on the AJCC TNM staging system (American Joint Committee on Cancer) (Tumor Nodes Metastasised): (i) Stage I, cancer has not spread to nearby lymph nodes or to distant sites (limited to the pancreas) and the size of it is less than 2 cm (Stage IA) or greater than 2 cm but no more than 4 cm (Stage IB); (ii) Stage II, the size of cancer is bigger than 4 cm and is either limited to the pancreas (Stage IIA) or it has spread to no more than 3 nearby lymph nodes (Stage IIB) and it has not spread to distant sites; (iii) Stage III, the cancer has spread to 4 or more nearby lymph nodes, but has not metastasised to distant sites; (iv) Stage IV, the cancer has spread to distant organs [12]. Based on this staging system, we grouped them into two, early-stage disease which includes stage I and II, and late-stage disease which includes stage III and IV. The healthy control group in this study consist of patients who did not have any cancer or problem in their pancreas but may have other disease.

The samples were frozen at –80°C within 4 hours of collection and 5-mL aliquots were transported on dry ice from Barts Cancer Institute to the School of Engineering at the University of Warwick. The samples were thawed overnight at 4°C in a laboratory fridge. The vial with the urine was placed in the ATLAS sampling system for 10 minutes at 40°C to maximise volatile organic compound chemical release. After 10 minutes, clean dry air is pushed over the surface of the urine and into the Lonestar. The flow rate over the sample was 500 mL/min with an additional 1500 mL/min of makeup air (to make 2 L/min) pushed into the Lonestar. Each sample was tested in triplicate. The machine was set with a dispersion field of 0% to 100% in 51 steps and –6 V to +6 V compensation voltage in 512 steps, with both positive and negative fields applied to the samples. This created a dataset of 52,224 points per sample, with each sample taking <60 seconds to capture.

Data Analysis

There were three classification tasks performed in this study: PDAC vs Healthy Control, Early-stage PDAC vs Healthy Control, and Early-stage PDAC vs Late-stage PDAC. These classification tasks were processed using our FAIMS data analysis pipeline which has been described in chapter 4. Using this pipeline, 44 statistically important features were used to predict the result of the test set (discussed in chapter 4). Four different classifiers were used for prediction: sparse logistic regression (SLR), random forest (RF), gaussian process (GP) classifier and support vector machine (SVM). Furthermore, the data was split into a training and validation set of 100 samples and a test set of 62 samples. The classifiers were trained and validated on the 100 samples and then the trained models were tested to the 62 unknown samples to get an unbiased measure of accuracy. Of note, previous studies from our research team have determined that age, diet, smoking, and sex are not confounders and no correction to the analyses was required [13,14,15].

5.2.2. Result

PDAC urine samples were detected with a sensitivity of 0.91 (95% confidence interval [CI], 0.83-0.96) and specificity of 0.83 (95% CI, 0.73-0.90), with an area under the curve (AUC) of 0.92 (95% CI, 0.88-0.96), using a SVM algorithm (Figure 5.1A). The results were validated by random data splitting into the training and test set achieving similar results with AUC of 0.92 (95% CI, 0.85-0.98), and sensitivity and specificity of 0.90 (95% CI, 0.74-0.98) and 0.81 (95% CI, 0.63-0.93), respectively. We also compared early-stage disease (I/II) with healthy individuals as well as early-stage disease with late-stage disease using the same analysis pipeline (Figure 5.1B, C). The best result (using SVM) is shown in Table 5.2, and the complete result can be found in Appendix Table 5.1a-c.

Table 5.2. Statistical Analysis of FAIMS Data Using SVM

Statistical parameter	All PDAC vs Healthy Control (All data set)	All PDAC vs Healthy Control (Testing data set)	Early-stage PDAC vs Healthy Control	Early-stage PDAC vs Late-stage PDAC
AUC	0.92 (0.88 – 0.96)	0.92 (0.85 – 0.98)	0.88 (0.79 – 0.97)	0.92 (0.86 – 0.97)
Sensitivity	0.91	0.90	0.84	0.82
Specificity	0.83	0.81	0.94	0.89
PPV	0.84	0.82	0.95	0.90
NPV	0.91	0.89	0.81	0.80
p-value	1.76×10^{-20}	2.41×10^{-10}	1.18×10^{-8}	4.35×10^{-13}

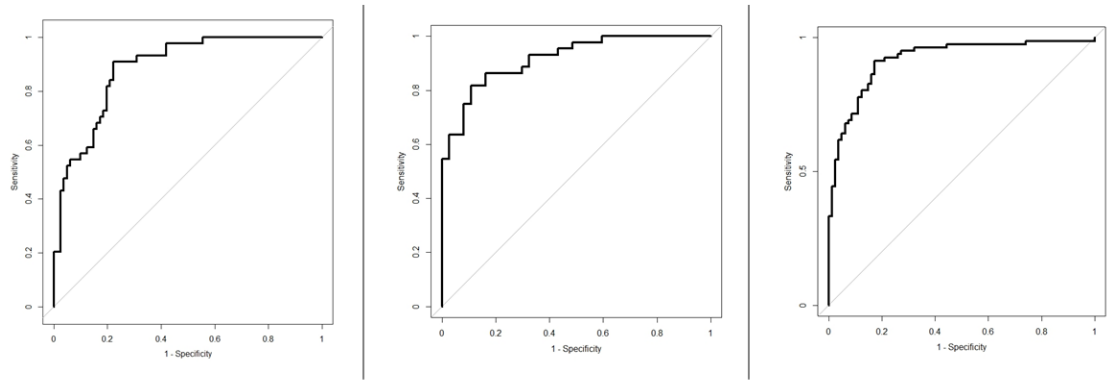


Figure 5.1. (A) Receiver operating characteristic curve showing performance of volatile organic compounds in differentiating healthy individuals from PDAC with an AUC of 0.92 (95% CI, 0.88-0.96), sensitivity of 0.91 (95% CI, 0.83-0.96), and specificity of 0.83 (95% CI, 0.73-0.90). (B) Healthy individuals from early stage I/II with an AUC of 0.89 (95% CI, 0.83-0.94), a sensitivity of 0.91 (95% CI, 0.78-0.97), and a specificity of 0.78 (95% CI, 0.69-0.86). (C) Early stage PDAC (I/II) could also be successfully separated from advanced stage PDAC (III/IV) with an AUC of 0.92 (95% CI, 0.86-0.97), sensitivity of 0.82 (95% CI, 0.67-0.92), and specificity of 0.89 (95% CI, 0.75-0.97).

5.3. Volatile Organic Compounds (VOCs) for the noninvasive detection of pancreatic cancer from urine

The above section investigated urinary VOCs using a Lonestar FAIMS (field asymmetric ion mobility spectrometry) instrument (Owlstone Medical, Cambridge, UK), which showed good discriminatory performance between PDAC and healthy control samples [26]. However, this instrument is unable to identify any specific biomarkers and the Lonestar unit is operated manually, testing one sample at a time, making its use more challenging in a clinical setting. To further explore this concept and, with a focus on identifying specific biomarkers of the disease, a second study was undertaken using more traditional a gas analysis approach and a second IMS instrument that may be more appropriate for use in a clinical setting.

The content of this section is based on a paper that has been published in peer-reviewed journals:

Daulton, Emma, Alfian N. Wicaksono, Akira Tiele, Hemant M. Kocher, Silvana Debernardi, Tatjana Crnogorac-Jurcevic, and James A. Covington. 2021. "Volatile Organic Compounds (VOCs) for the Noninvasive Detection of Pancreatic Cancer from Urine." Talanta 221 (August 2020): 121604. https://doi.org/10.1016/j.talanta.2020.121604.

5.3.1. Methods

As with the first study, urine samples were obtained from Barts Pancreas Tissue Bank, after patient consent and with ethical approval (reference number 13/SC/0593). All samples were stored at -80°C according to standard operating procedures compliant with tissue bank requirements under Human Tissue Act 2004. This study included patients with PDAC, chronic pancreatitis (CP) and healthy controls. Chronic pancreatitis is an inflammation of the pancreas and is a condition that is associated with an increased risk of developing pancreatic cancer [17]. This diagnostic group was included in the study to evaluate whether urinary analysis could be used to distinguish such patients at risk from the healthy and PDAC groups. The healthy controls group consist of patients who did not have any cancer or diease in their pancreas but may have other disease. In total, 123 urine samples were analysed: 33 healthy, 45 CP and 45 PDAC. The basic demographic information for the subject cohorts is summarised in Table 5.3.

Table 5.3. Demographics of Sample Groups and Cancer Stage

Group	Gender (Male: Female)	Av. age (range) years	Number of subjects (Stage S x Number)
PDAC	22M:23F	64.1 +/- 10.7 (29-77)	45 (SI x 1; SII x 21; SIII x 20; SIV x 3)
CP	28M:17F	52.9 +/- 12.2 (21-78)	45 (not applicable)
Healthy	15:M:18F	49.9 +/- 7.8 (30-66)	33 (not applicable)

The urine samples were analysed using two analytical methods: gas chromatography – ion mobility spectrometry (GC-IMS) and gas chromatography time-of-flight mass spectrometry (GC-TOF-MS). The working principle of both instruments has been described in Chapter 3. The GC-IMS system used was a G.A.S. GC-IMS (Dortmund, Germany) and was fitted with a 30 m, 0.32 mm inner diameter (ID) SE-54 column (CS Chromatographie Service, Langerwehe, Germany). For GC-IMS measurements, we have developed a standard method for testing urine samples, where we have optimised the sample temperature, method

temperatures and flow rates to maximise information content and improve reproducibility. The urine samples were shipped on dry ice to the University of Warwick and were stored at -20°C. Prior to analysis, the samples were thawed for four hours at room temperature. 5 ml samples of urine were aliquoted into 20 ml glass vials and sealed with a crimp cap and septa. Once sealed, the samples were agitated and heated to 40°C for 10 minutes. For sampling, a sterile needle, attached to the heated sample inlet of the GC-IMS, was inserted into the sample headspace, through the septa. The GC-IMS instrument sampled 2 ml of the headspace for analysis. The sampling and analysis were performed using the following settings: GC flow rate = 20 ml/min, drift tube flow rate = 150 ml/min, IMS temperature = 45°C, GC temperature (fixed) = 45°C, Sample loop = 45°C and inlet injector = 45°C. The analysis time for each sample was 8 minutes. For quality control, all flow rates, method temperatures and RIP (reactant ion peak) magnitude and location were checked for each sample, to be within unit tolerances. Furthermore, the output obtained from each sample was visually checked to ensure that they contained the expected level of information. Finally, air blanks were run either side of a test batch of samples (20 samples) to ensure that there was no machine drift.

GC-TOF-MS works in a similar way to traditional GC-MS methods, but instead of filtering ions by mass, the TOF utilises 'time of flight' and analyses all ions present. The GC-TOF-MS system consists of a TRACE 1300 GC (Thermo Fisher Scientific, Loughborough, UK), combined with a BenchTOF-HD TOF-MS (Markes Intl., Llantrisant, UK). The GC column used was a 20 m, 0.18 mm ID, Rxi-624Sil MS column (Thames Restek, Saunderton, UK). This system also includes a high-throughput autosampler and a thermal desorption unit, ULTRA-xr and UNITY-xr, respectively (both from Markes Intl.). For GC-TOF-MS analysis, urine samples were defrosted the same way with sample for GC-IMS analysis, with 10 ml of sample aliquoted into a 20 ml glass vial, which were then sealed with a specially adapted septa and crimp cap. A thermal desorption (TD) sorbent tube (C2-AXXX-5149, Markes Intl., Llantrisant, UK) was placed through the septa and into the headspace above the urine sample. The vial and sorbent tube were then placed into a heater block and heated to 40°C for 1 hour. Once completed, the tube was removed from the top of the vial and placed into the autosampler for analysis. The ULTRA-xr was set to run with a stand-by split of 150°C with an overlap (this allows the auto-sampler to reduce the overall run time), a GC run-time of 30 minutes, and a minimum carrier pressure of 5 psi. For each sample there is a pre-purge of 1 minute. The tube was desorbed for 10 minutes at 250°C, with the trap purge time set to 1 minute, and the trap cooled to -30°C. The trap was then purged for 3 minutes at a temperature of 300°C. The GC-TOF-MS method measured masses from 45 to 500 atu (atomic mass units). The transfer line and ion source are both heated to 250°C, with an ionisation voltage of -70.000V. The GC oven was heated to 280°C for 25 minutes. Upon completion of the GC-TOF-MS run-

time, peaks were identified using NIST list 2016. For quality control, the machine was calibrated in line with the manufacturer's recommendation. The method used was developed using healthy control urine samples from a previous study, where parameters were optimised to maximise separation and chemical content.

Data Analysis

There were four classification analysis performed in this study for each instrument. They were PDAC vs All (CP and Healthy Control), PDAC vs CP, PDAC vs Healthy Control, and CP vs Healthy Control. The data analysis of both GC-IMS and GC-TOF-MS have been described in chapter 4. Using this pipeline, unique features are selected and used for training three different classifiers, specifically random forest (RF), Gaussian process (GP), and sparse logistic regression (SLR). We have previously used these successfully in a range of medical VOC studies [18,19,20]. From GC-TOF-MS data analysis, we are also able to identify which features/chemicals hold discriminatory information.

5.3.2. Results

Three different analyses were undertaken to compare PDAC, healthy and CP groups. GC-IMS analysis indicated that there were differences between the three groups. The three classifiers resulted in very similar data, with the best results shown in Table 5.4 and the complete data in appendix Table 5.2a-d. ROC curves are shown in Figure 5.2. ROC for (a) PDAC vs Healthy and (b) PDAC vs CP using GC-IMS. The results indicate that GC-IMS (sensitivity 84%, specificity 94%, p-value >0.0001) can separate PDAC from healthy controls. However, the separation between PDAC and CP group is not as pronounced (sensitivity 51%, specificity 73%, p-value 0.11), indicating that the same biomarkers are involved in both conditions.

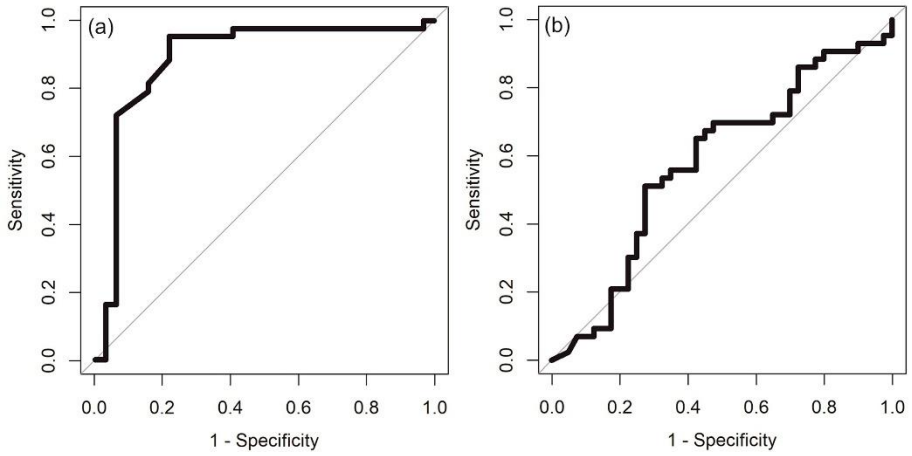


Figure 5.2. ROC for (a) PDAC vs Healthy and (b) PDAC vs CP using GC-IMS

Table 5.4. Statistical Analysis of GC-IMS Data (Best Results)

Statistical parameter	PDAC vs All	PDAC vs CP	PDAC vs Healthy	CP vs Healthy
Best classifier	SLR	GP	SLR	GP
AUC	0.69 (0.58 – 0.79)	0.58 (0.45 – 0.71)	0.88 (0.79 – 0.97)	0.86 (0.77 – 0.95)
Sensitivity	0.72	0.51	0.84	0.80
Specificity	0.60	0.73	0.94	0.91
PPV	0.52	0.67	0.95	0.91
NPV	0.78	0.58	0.81	0.78
p-value	4.39×10^{-4}	0.11	1.18×10^{-8}	9.79×10^{-9}

We undertook the same analytical methods for the data analysed by GC-TOF-MS. In this case, the statistical results are shown in Table 5.5, with the ROC curves in Figure 5.3. The complete result can be found in Appendix Table 5.3a-d. Here, we were also able to separate PDAC from healthy controls (sensitivity 72%, specificity 96%, p-value >0.0001). However, when comparing CP to PDAC (sensitivity 38%, specificity 88%, p-value 0.28), our results did not show a significant difference.

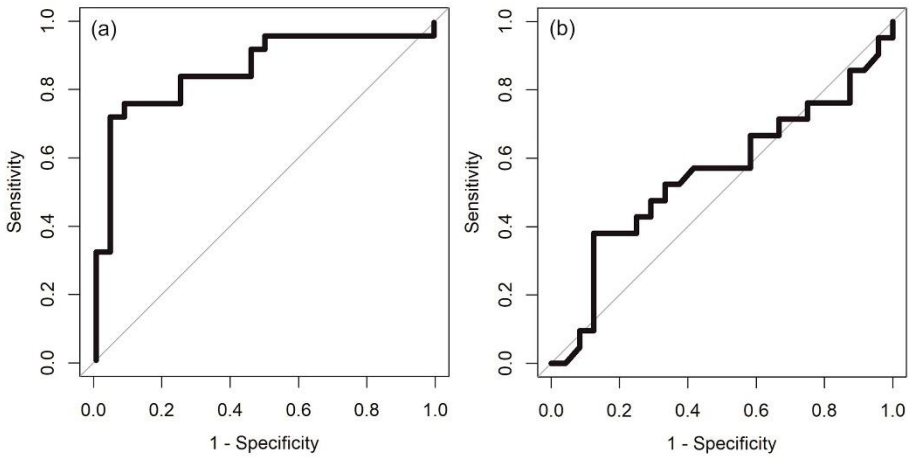


Figure 5.3. ROC for (a) PDAC vs Healthy and (b) PDAC vs CP using GC-TOF-MS

Table 5.5. Statistical Analysis of GC-TOF-MS Data (Best Results)

Statistical parameter	PDAC vs All	PDAC vs CP	PDAC vs Healthy	CP vs Healthy
Best classifier	SLR	RF	SLR	RF
AUC	0.75 (0.63 – 0.87)	0.55 (0.37 – 0.73)	0.86 (0.75 – 0.97)	0.67 (0.50 – 0.83)
Sensitivity	0.52	0.38	0.72	0.38
Specificity	0.96	0.88	0.96	0.96
PPV	0.96	0.73	0.95	0.89
NPV	0.51	0.62	0.77	0.65
p-value	3.11×10^{-4}	0.28	1.81×10^{-6}	2.75×10^{-2}

From the analysis, we were able to identify chemicals that held discriminatory properties. 15 chemicals for each comparison were identified. From this list, a rank-sum test of each chemical/feature was undertaken and the top three chemicals for each comparison are listed in Table 5.6, with the complete set-in appendix Table 5.4. As shown, we have identified common chemicals from each of the analyses, with 2,6-dimethyl-octane, 2-pentanone, nonanal and 4-ethyl-1,2-dimethyl-benzene being most frequent.

Table 5.6. Chemicals Used to Separate Sample Groups

PDAC vs CP	PDAC vs Healthy	CP vs Healthy
2-pentanone	2,6-dimethyl-octane	2-pentanone
nonanal	nonanal	benzene, 1-ethenyl-2-methyl-
4-ethyl-1,2-dimethyl-benzene	4-ethyl-1,2-dimethyl-benzene	4-ethyl-1,2-dimethyl-benzene

5.4. Discussion of Pancreatic Cancer Studies

The first study is a proof-of-concept study that demonstrates for the first time that urinary VOCs analysis using FAIMS can be used to discriminate healthy individuals from patients with PDAC. The result from running the analysis to all data sets showed an equal performance with the result when we split the data into training and test set and only used the test set to measure the final performance of the model. We are then furthering our investigation to see the ability of FAIMS to detect PDAC based on cancer's stages. We found that FAIMS could discriminate early-stage PDAC from healthy control with high sensitivity and specificity. In addition, FAIMS could also distinguish late-stage PDAC from early stage PDAC. These consistent results showed the potential of FAIMS to be an early diagnostic tool for detecting pancreatic cancer. However, the FAIMS is unable to identify any specific biomarkers and the Lonestar unit is operated manually, testing one sample at a time, making its use more challenging in a clinical setting.

Due to the limitation of the FAIMS, we did a follow-up study on urinary VOCs for the detection of PDAC with a focus on identifying specific biomarkers of the disease using a GC-IMS and GC-TOF-MS. According to the latest pancreatic cancer statistics, the prevalence of PDAC in the screening community is low (2.6% all over the world). However, due to the poor prognosis of pancreatic cancer, there were almost as many deaths (n=466 003) as there were cases (n= 495 773) in 2020 [1]. Therefore, the need for an early diagnostic tool which can be used in high volume samples input with simple sample preparation is crucial to tackle this problem. Both instruments, GC-IMS and GC-TOF-MS, are equipped with an autosampler, making it more appropriate to deploy in a clinical setting. In this follow-up study, we also included CP as it is associated with an increased risk of developing pancreatic cancer [17]. Analysing this will give us an understanding of how the instruments detect the progression of the disease. When detecting CP from healthy control, only the GC-IMS was able to accurately separate them, whilst the GC-TOF-MS showed inferior results. Comparing PDAC with CP for both instruments, the AUCs were around 0.6, showing a modest diagnostic performance. This limited ability to separate these two groups, could well have impacted the overall diagnostic performance of PDAC vs all other samples, with both instruments having an AUC of around 0.7. In statistical results, these variations are likely to be associated with the choice of column between the two analytical platforms and the sample capture process. The sorbent tubes used for the GC-TOF-MS analysis capture analytes from C3 (though more likely C4) and upwards, whilst the GC-IMS is analysing all the chemicals through direct headspace injection. Another difference between the platforms is that the GC-IMS molecular detection

depends on the proton affinity of the molecule. However, the overall statistical performance was similar.

Some of the discriminatory VOCs, identified in this study, have been suggested as potential biomarkers in breath or stool for other diseases. For example, 2-pentanone has been found to be associated with several diseases, such as non-alcoholic fatty liver disease [21], inflammatory bowel disease (ulcerative colitis and Crohn's disease) [22,23] and lung cancer [24]. Nonanal is a saturated fatty aldehyde formally arising from the reduction of the carboxy group of nonanoic acid. It has been observed as a discriminatory VOC for other cancers, such as ovarian and lung cancer [25,26]. This indicates that some of these compounds might be more generic markers of inflammation or illness.

In a recent review paper, Bax *et al*. [27] evaluated and compared cancer biomarker trends in urine as a new diagnostic pathway. Five studies [28,29,30,31,32] investigating urinary pancreatic cancer biomarkers were included in the review. The employed technologies were nuclear magnetic resonance (NMR) spectroscopy and liquid chromatography-tandem mass spectrometry (GeLC-MS/MS). Napoli *et al.* utilized NMR spectroscopy to investigate the urinary VOC profile of PDAC patients (n = 33) in comparison with healthy matched controls (n = 54), while Radon *et al.* conducted deeper investigating on different PDAC stages compared to healthy using GeLC-MS/MS involving 371 urinary specimens (87 HC, 92 CP, and 192 PDAC) from 18 patients (6 patients for each group) [28,31]. Both achieved a promising result with sensitivity of 75.8 % and specificity of 90.7% for NMR and specificity of 89.8% and sensitivity of 76.9% for GeLC-MS/MS when comparing PDAC with healthy control. When comparing early stage PDAC with healthy control, radon et. al reported an AUC of 0.93 (95% CI, 0.84–1.00). These results demonstrate similar performance to that of our investigation, indicating that urine analysis has the potential to be used as an early detection approach for pancreatic cancer.

To the best of our knowledge, this was the first urinary study to utilise FAIMS, GC-IMS and GC-TOF-MS technologies to investigate PDAC, CP and healthy controls. All instruments were able to separate PDAC from healthy controls with good sensitivity and specificity, with the FAIMS outperforming the GC-IMS and GC-TOF-MS.

5.5. Colorectal cancer and adenoma screening using urinary volatile organic compounds (VOC) detection: early result from a singles-centre bowel screening population (UK BCSP)

The content of this section is based on a paper that has been published in peer-reviewed journals:

Mozdiak, E., A. N. Wicaksono, J. A. Covington, and R. P. Arasaradnam. 2019. "Colorectal Cancer and Adenoma Screening Using Urinary Volatile Organic Compound (VOC) Detection: Early Results from a Single-Centre Bowel Screening Population (UK BCSP)." Techniques in Coloproctology 23 (4): 343–51. https://doi.org/10.1007/s10151-019-01963-6.

5.5.1. Methods

All enrolled patients were recruited from the Coventry and Warwickshire University Hospitals between April 2015 and November 2016. Regional ethical approval was granted by the Warwickshire Research and Development Department and Warwickshire Ethics Committee 09/H1211/38. Informed consent was obtained from the individual participants that took part in the study. This study was approved by the bowel cancer screening research committee only to approach those that had a positive FOBT test.

Patients were recruited from the nurse led BCSP clinics following a positive Faecal Occult Blood Test (FOBT) result. Consent and urine sample collection were carried out at the clinic prior to bowel preparation administration. A total of 181 patients were invited to participate and 163 consented to provide samples for the final analysis. Two 20 ml samples of urine were collected. Samples were immediately transferred to − 20 °C storage and then to − 80 °C within 24 h for long-term storage. Diagnostic outcome data were collected from the colonoscopy or computed tomography (CT) colonography result, histology was confirmed from the pathology report. Healthy control patients were defined as a patient with normal colonoscopy or CT colonography result in BCSP screening population. In addition, the sample preparation procedure used in pancreatic cancer study (section 5.2.1.1) was also used for analysis in this study.

Data analysis

Five classification models were used; each dataset was compared with each model to find the most accurate for each specific set of samples. The complete description of the FAIMS and GC-IMS data analysis pipeline can be found in Chapter 4. As CRC numbers were small within the screening population (incidence of 8–10%), a balancing technique was applied to the data to fairly match the non-CRC samples with the same number of CRC samples and avoid bias from an unbalanced healthy control group. Balancing involved the well-described synthetic minority over-sampling technique (SMOTE), where artificially generated points are plotted to represent the healthy control group as a whole and is used to provide a more fair representation [33]. SMOTE has also been described in chapter 4.

5.5.2. Results

FAIMS analysis

A total of 163 samples were analysed. 93 (57%) were males, the median age of patients was 67 years, 12 (7.4%) were current smokers. 41 (25.4%) were ex-smokers and 109 (67.2%) had never smoked. Patients were grouped into categories according to diagnosis for analysis. Diagnostic outcomes for study participants are listed in Table 5.7.

Table 5.7. Diagnostic Outcomes for Study Participants and Distribution of CRC by Site (Total of 13 Cancer Sites as One Patient Had a Synchronous Tumour)

Diagnosis		Number (%)
Cancer	Total	12 (7.6)
	Rectum	4 (2.4)
	Sigmoid	4 (2.4)
	Descending colon	0
	Transverse colon	1 (0.58)
	Ascending colon	2 (1.17)
	Cecum	2 (1.17)
Adenoma	Total	80 (49.1)
	High	17 (10.5)
	Intermediate	36 (21.1)
	Low	27 (17.5)

Diverticular disease		14 (8.2)
Healthy control		37 (19.3)
Haemorrhoids		5 (2.9)
Other		14 (8.2)^
Excluded		8 (4.7)*

*1 not fit enough for investigations, 7 declined investigations

^inflammatory bowel disease: n=7, rectal telangiectasia: n=2, rectal ulcer: n=1, radiation proctitis: n=1, inflammatory pseudo polyp: n=1, non-specific sigmoid inflammation: n=1, ischaemic sigmoid stricture: n=1

Group (a) CRC vs healthy control demonstrated the highest degree of separation with AUC 0.98 (95% CI 0.93–1.0) with 12 patients in each group. The corresponding ROC curve is shown in Figure 5.4a. Sensitivity and specificity were also high: 1.0 (95% CI 0.74–1) and 0.92 (95% 0.62–1), respectively (Table 5.8).

Table 5.8. Classification of BCSP Study Participants by Outcome Using FAIMS

Group	AUC	Sensitivity	Specificity	PPV	NPV	P-value
a) CRC (12) vs normal (12)	0.98 (0.93-1)	1	0.92	0.92	1	2.59×10^{-6}
b) CRC + all adenomas (93) vs normal (37)	0.64 (0.54-0.74)	0.48	0.89	0.92	0.41	6.3×10^{-3}
c) CRC + high risk adenomas (30) vs normal (37)	0.62 (0.48-0.76)	0.57	0.68	0.59	0.66	4.38×10^{-2}
d) CRC + high risk adenomas (30) vs other (70)	0.6 (0.47-0.73)	0.47	0.80	0.52	0.76	5.78×10^{-2}
e) CRC + all adenomas (93) vs other (70)	0.56 (0.47-0.65)	0.91	0.25	0.64	0.67	9.55×10^{-2}
f) Non cancer (113) vs normal (37)	0.61 (0.51-0.71)	0.56	0.68	0.83	0.35	2.57×10^{-2}
g) CRC (12) vs Adenoma (7) (hr)	0.92 (0.77-1)	0.83	1	1	0.78	6×10^{-4}
h) CRC (12) vs Adenoma (12) (ir)	0.84 (0.67-1)	0.83	0.75	0.77	0.82	1.5×10^{-3}
i) CRC (12) vs Adenoma (12) (lr)	0.83 (0.66-1)	0.75	0.92	0.90	0.79	2.26×10^{-3}

Using Sparse Logistic Regression and Gaussian Process

Corresponding 95% CIs are stated in brackets. Numbers in brackets in group column denote sample number

BSCP bowel cancer screening programme, FAIMS field asymmetric waveform ion mobility spectrometry, CRC Colorectal cancer, hr high risk, ir intermediate risk, lr low risk

In groups (b–e), CRC was grouped with adenomas and showed only modest AUC, sensitivity, and specificity results when attempting to classify groups according to diagnosis. The most accurate classification of the adenoma groups was seen in (b) CRC + all adenomas vs healthy control; here, sensitivity was low at 0.48 (95% CI 0.38–0.59), but specificity was high at 0.89 (95% CI 0.75–0.97). In (f), when CRC was excluded, separation was low with a sensitivity of 0.56 (95% CI 0.46–0.65).

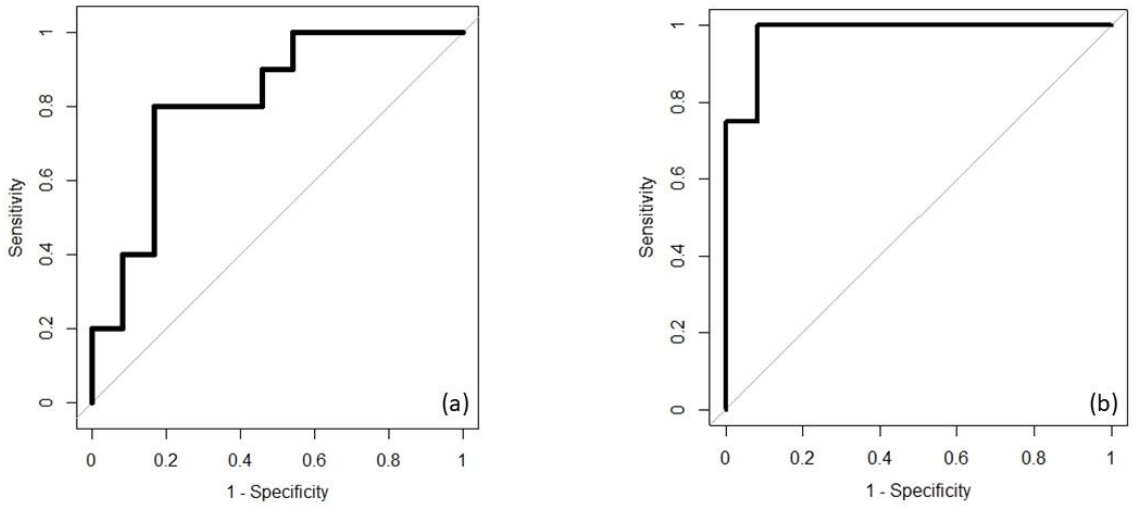


Figure 5.4. ROC curve for (a) classification of CRC vs healthy control in bowel cancer screening programme (BCSP) patients (balanced) using the sparse logistic regression classifier using FAIMS and (b) classification of CRC vs healthy control using GC-IMS with Gaussian process classifier

A further set of analyses were carried out to investigate the classification of the adenoma groups in more depth and to compare CRC with three categories of adenoma: (g) high risk, (h) intermediate risk and (i) low risk, according to the BSG guidelines [34], the results are displayed in Table 5.8. High sensitivity was demonstrated when each adenoma group was compared with CRC. The most accurate overall classification was seen in CRC vs high-risk adenoma with a sensitivity of 0.83 (95% CI 0.52–0.98) and specificity of 1 (95% CI 0.59–1).

GC–IMS analysis

One hundred and nine patient samples were analysed using the GC–IMS method. Five comparator groups were devised according to outcome (Table 5.9).

Table 5.9. Classification of BCSP Study Participants Using GC-IMS using Gaussian Process or Support Vector Machine

Group	AUC	Sensitivity	Specificity	PPV	NPV	P-value
a) Cancer (10) vs normal (24)	0.82 (0.67-0.97)	0.80	0.83	0.67	0.91	1.4×10^{-3}
b) Cancer + high risk adenomas (23) vs normal (24)	0.53 (0.36-0.70)	0.48	0.67	0.58	0.57	6×10^{-4}
c) Cancer (10) vs other (20)	0.77 (0.60-0.94)	1	0.57	0.5	1	1.05×10^{-2}
d) Cancer + all adenomas (65) vs other (42)	0.61 (0.49-0.72)	0.71	0.55	0.71	0.55	6.19×10^{-2}
e) All adenomas (55) vs normal (24)	0.61 (0.47-0.75)	0.58	0.62	0.78	0.39	2.88×10^{-2}

Corresponding 95% CI are in brackets. Study numbers are stated in the group column in brackets BCSP bowel cancer screening programme, GC-IMS gas chromatography coupled with ion mobility spectrometry, AUC area under the curve, CRC colorectal cancer, PPV positive predictive value, NPV negative predictive value

As with the analysis using FAIMS, when comparing CRC vs healthy control (group a), there was a high degree of separation with a sensitivity of 0.80 (95% CI 0.44–0.97) and specificity of 0.83 (95% CI 0.63–0.95). The corresponding ROC curve is seen in Figure 5.4b. CRC vs other diagnoses also had a high sensitivity of 1.0 (95% CI 0.66–1); however, specificity dropped to 0.57 (95% CI 0.34–0.78). When CRC samples were grouped with adenomas and compared with other groups (those with any diagnosis other than CRC or adenoma) the sensitivity dropped to a modest level of 0.71 (95% CI 0.58–0.81) with sensitivity 0.55 (0.39–0.70). Adenomas vs healthy control showed a low level of separation, with a sensitivity of only 0.58 (95% CI 0.44–0.71) and specificity 0.62 (95% CI 0.41–0.81) (Table 5.9).

5.6. Early detection and follow up of colorectal neoplasia based on faecal volatile organic compounds

In the above study, we investigate the possibility of urinary VOC analysis as a way to detect colorectal cancer. We further our investigation to see the possibility of faecal VOC analysis for colorectal cancer screening as the faecal collection has already been implemented in the bowel screening program. Several studies have focused on the application of the faecal volatolome as a biomarker for detection of CRC and AA, with promising results. However, none of the available studies has included patients with (low risk) adenomas to explore the specificity of VOC analysis. The aim of this study was to assess the potential of faecal VOC

as a non-invasive biomarker to detect colonic neoplasia and precursor lesions. In addition, we aimed to explore its potential for secondary non-invasive follow-up after polypectomy.

The content of this section is based on a paper that has been published in peer-reviewed journals:
Bosch, S., R. Bot, A. Wicaksono, E. Savelkoul, R. van der Hulst, J. Kuijvenhoven, P. Stokkers, *et al*. 2020. "Early Detection and Follow-up of Colorectal Neoplasia Based on Faecal Volatile Organic Compounds." Colorectal Disease, 0–1. https://doi.org/10.1111/codi.15009

5.6.1. Method

This multi-centre prospective case-control study was performed between February 2015 and November 2017 at the outpatient clinics of the gastroenterology and hepatology departments in one tertiary referral hospital (Amsterdam UMC, location VUmc) and two district hospitals (OLVG West, Amsterdam and Spaarne Gasthuis, Hoofdorp and Haarlem), all located in the Netherlands. This study was approved on 4 September 2014 by the Medical Ethical Review Committee (METc) of Amsterdam UMC (2014.404), and by local METcs of OLVG West and Spaarne Gasthuis. The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki as reflected in a priori approval by the institution's human research committee. Written informed consent was obtained from all participants.

All patients aged ≥ 18 years with a scheduled colonoscopy were asked to participate in this study, irrespective of the indication for endoscopy. Patients were divided into five subgroups based on observations during endoscopy, combined with histology reports in cases where biopsies or polypectomies were performed: (i) CRC, histologically confirmed carcinoma of the colon or rectum; (ii) AA, according to the European Society of Gastrointestinal Endoscopy (ESGE), that is, characterised by polyps ≥ 1 cm in diameter, or with villous histology, or high-grade dysplasia; (iii) large adenoma (LA), adenomas 0.5-1.0 cm in diameter without villous histology or high-grade dysplasia; (iv) small adenoma (SA), adenomas < 0.5 cm in diameter without villous histology or high-grade dysplasia; (v) healthy controls, characterised by no abnormalities observed during endoscopy (excluding small anal fibroma, haemorrhoids and/or diverticula), and in the case of mucosal biopsies on histopathological abnormalities [35]. Participants were asked to collect a faecal sample (Stuhlgef€aß 10 ml, Frickenhausen, Germany) prior to bowel preparation, store the sample in their own freezer at home within 1 h following the bowel movement and bring it to hospital on the day of their endoscopic assessment. The samples were stored at -24°C directly upon arrival at the hospital. Participants were asked to complete a questionnaire on the day of

sample collection, which included age, gender, body mass index (BMI), smoking habit, comorbidity, and medication use. Exclusion criteria were presence of a known underlying gastrointestinal disease, incomplete endoscopic assessment for various reasons (e.g., inadequate bowel cleansing, pain) and/or inability to collect or store sufficient faecal samples mass to perform VOC analysis.

Between May and November 2017, patients who underwent a successful polypectomy during endoscopy were asked to participate in the follow-up part of this study. Participants were excluded from this group in the case of incomplete removal of polyps. The remaining polypectomy patients were randomly matched to healthy controls in a 1:1 ratio. All patients included in the second part of this study were asked to collect a follow-up faecal sample and complete a second questionnaire (the same procedure as the first sample and questionnaire). These samples and questionnaires were collected by one of the researchers and the samples were transported to the hospital on dry ice where they were stored at -24°C upon arrival.

Endoscopy reports and histological outcomes of mucosal biopsies and/or polypectomy were checked using electronic patient files. These outcomes were used as standard reference for the localisation and total number of removed adenomas in this study. Endoscopies were either performed or supervised by a trained gastroenterologist. Histopathological reports were used as the standard reference for size, differentiation grade (e.g., hyperplasia, dysplasia), villous histology and type of CRC in this study. Mucosal biopsies were noted as 0.2 cm in size. The presence of sessile and/or serrated characteristics was noted for all non-AAs. When multiple polyps were present, classification was based on the most advanced of the largest lesion.

One frozen subsample of 500 mg per participant as weighted and transferred to a glass vial (20 ml headspace vial); Thames Restek, Saunderton, UK). Samples were transported to the University of Warwick (Coventry, UK) for faecal VOC analysis. Gas chromatography-ion mobility spectrometry (GC-IMS; Flavourspec; G.A.S., Dortmund, Germany) was used to measure the faecal VOC patterns. The working principle of GC-IMS can be found in Chapter 3. Prior to analysis, the samples were heated to 80°C for 8 min. After this, GC was performed at 40°C using nitrogen 99.9% (3.5 bar) as the carrier gas and IMS was performed at 45°C using nitrogen 99.9% as the drift gas. Flow rates were set at 150 ml/min (0.364 kPa) (IMS) and 20 ml/min (34 175 kPa) for 6 min (GC).

Data analysis

Using IBM SPSS Statistics (version 22), demographic data of each group were calculated and compared. One way ANOVA or Kruskall Wallis tests were used to compute differences in baseline demographics between groups. For the secondary part of this study, t-tests, Fisher's exact test, chi-square and Mann-Whitney U tests were used to compute differences between groups. GC-IMS data were processed using our data analysis pipeline. The complete description of the analysis pipeline can be found in Chapter 4. Data were split into three sets, 70% for training and validation and 30% as a test set. Wilcoxon rank sum test was used to find the 20, 50 and 100 most discriminatory features and subsequently Sparse Logistic Regression (SLR), Random Forest (RF), Gaussian Process (GP), Support Vector Machine (SVM) and Neural Net (NN) classification were used to provide statistical results (discussed in chapter 4). In the case of small subgroups of interest, undersampling of healthy controls was performed to avoid skewed analyses.

5.6.2. Results

Baseline characteristics

In total, 1039 patients collected a faecal sample prior to colonoscopy of which samples from 14 CRC patients, 64 AA, 69 LA and 127 SA were included in this study. 227 healthy Controls were included as they did not have any mucosal abnormalities during endoscopy. The baseline demographics of all study participants are depicted in Table 5.10. Age differed significantly between groups ($p<0.001$), with the controls displaying the lowest mean age (60 ± 11.8) and AA patients the highest (68.8 ± 6.7). Gender differed significantly between healthy controls, SA, LA, and AA ($p<0.0001$) but not between healthy control and CRC or any of the adenoma groups. There were no significant differences in BMI, smoking status and use of antibiotics between groups.

For the follow-up part of this study, all 32 patients undergoing a complete polypectomy and 32 healthy controls were included; these groups collected a second faecal sample three months after endoscopy. The baseline demographics of the follow-up study are given in Table 5.11. There were no significant differences in BMI and smoking status. Gender and age did differ significantly between groups ($p<0.014$, $p<0.001$, respectively). The use of antibiotics did differ between the polypectomy and healthy control group; however, the ratio of antibiotics users remained the same during follow-up period of this study (3:1).

Table 5.10. Demographics of All Study Participants

	Colorectal Cancer (n=14)	Advanced adenomas (n=64)	Large adenomas (n=69)	Small adenomas (n=127)	Healthy controls (n=227)
Age (mean, ± s.d.)	66.6 ± 8.7	68.8 ± 6.7	68.7 ± 7.2	63.7 ± 10.0	60 ± 11.8
Gender, f (n, [%])	6 [42.9]	17 [26.3]	21 [30.4]	44 [34.6]	129 [56.8]
BMI* (mean, ± s.d.)	25.9 ± 5.3	27.0 ± 4.1	26.6 ± 3.9	27.3 ± 6.5	26.6 ± 7.0
ABx 3 months prior to inclusion (n, [%])	0 [0]	6 [9.4]	3 [4.3]	21 [16.5]	31 [13.7]
Smoking Status (n, [%])					
Active	2 [14.3]	9 [14.1]	15 [21.7]	23 [18.1]	37 [16.3]
Quit	10 [71.4]	37 [57.8]	36 [52.2]	64 [50.4]	92 [40.5]
Never	2 [14.3]	18 [28.1]	17 [24.6]	40 [31.5]	98 [43.2]
Size largest adenoma (mean, ± s.d.)	0.7 ± 0.63	1.4 ± 0.64	0.7 ± 0.26	0.4 ± 0.32	NA
Localization of Largest adenoma† (n, [%])					
Caecum	0 [0]	4 [6.3]	7 [10.1]	12 [9.4]	NA
Colon Ascendens	1 [7.1]	8 [12.5]	12 [17.4]	32 [25.2]	NA
Flexura Hepatica	1 [7.1]	2 [3.1]	0 [0]	1 [0.8]	NA
Colon Transversum	0 [0]	3 [4.7]	15 [21.7]	18 [14.2]	NA
Flexura Lienalis	0 [0]	1 [1.6]	0 [0]	1 [0.8]	NA
Colon Descendens	1 [7.1]	4 [6.3]	6 [8.7]	9 [7.1]	NA
Sigmoid	5 [35.7]	30 [46.9]	19 [27.5]	24 [18.9]	NA
Rectosigmoid	2 [14.3]	3 [4.7]	1 [1.4]	4 [3.1]	NA
Rectum	2 [14.3]	6 [9.4]	6 [8.7]	14 [11.0]	NA
Terminal ileum	1 [7.1]	0 [0]	0 [0]	1 [0.8]	NA
CRC type (n, [%])					
Adenocarcinoma	12	NA	NA	NA	NA
Neuroendocrine	2	NA	NA	NA	NA
AA characteristics (largest adenoma) (n, [%])					
≥ 10 mm	NA	54 [84.4]	NA	NA	NA
Villous histology	NA	31 [48.4]	NA	NA	NA
HGD	NA	6 [9.4]	NA	NA	NA
Polyp characteristics (largest adenoma) (n, [%])					
No dysplasia	NA	1 [1.6]	9 [13.0]	9 [7.1]	NA
Hyperplasia	NA	2 [3.1]	4 [5.8]	14 [11.0]	NA
LGD	NA	55 [85.9]	55 [79.7]	100 [78.7]	NA
Sessile/serrated	NA	NA	8 [11.6]	10 [7.9]	NA
Total number adenomas removed (n, [%])					
1	6 [42.8]	11 [17.2]	18 [26.1]	59 [46.5]	NA
2	1 [7.1]	14 [21.9]	20 [29.0]	34 [26.8]	NA
3	3 [21.4]	11 [17.2]	14 [20.3]	14 [11.0]	NA
4-5	2 [14.3]	14 [21.9]	8 [11.6]	15 [11.8]	NA
6-10	1 [7.1]	13 [20.3]	9 [13.0]	4 [3.2]	NA
>10	0 [0]	1 [1.6]	0 [0]	0 [0]	NA

ABx, antibiotics; BMI, body mass index; CRC, colorectal cancer; HGD, high-grade dysplasia; LGD, low-grade dysplasia; NA, not applicable; SD, standard deviation. *Insufficient documentation of 2 colorectal cancers, 6 advanced adenomas, 6 large adenomas, 10 small adenomas and 11 healthy controls. †1 missing for CRC, 3 missing for AA, 3 missing for LA, 13 missing for SA

Table 5.11. Demographics of Participants Included in Follow-Up Study

	Polypectomy group (n=32)	Healthy controls (n=32)
Age (mean, ±SD)	71.0 ± 5.9	60.5 ± 11.3
Gender (n females, %)	5 [15.6]	15 [46.9]
BMI (mean, ±SD)	26.8 ± 4.2	26.3 ± 3.4
Smoking status (n, %)		
Active	6 [18.8]	4 [12.5]
Quit	18 [56.3]	19 [59.4]
Never	8 [25.0]	9 [28.1]
Indication for endoscopic assessment (n, %)		
Positive FIT	10 [31.3]	4 [12.5]
Rectal blood loss	4 [12.5]	1 [3.1]
Change in bowel habits	2 [6.3]	3 [9.4]
Surveillance	5 [15.6]	3 [9.4]
Abdominal Pain	2 [6.3]	11 [34.4]
Diarrhea	3 [9.4]	1 [3.1]
Weight Loss	1 [3.1]	0 [0]
Anaemia	0 [0]	2 [6.3]
Constipation	0 [0]	2 [6.3]
Family history CRC+	1 [3.1]	8 [25]
Monitoring previous diverticulitis/abscess	0 [0]	2 [6.3]
Other	4 [12.5]	5 [15.7]
ABx 3 months prior to inclusion	2 [6.3]	6 [18.8]
ABx 3 months prior to second sample	1 [3.1]	3 [9.4]
Size adenoma (mean, ±SD)	1.1 ± 0.5	NA
Localisation of adenoma (n, %)		
Caecum	3 [9.4]	NA
Colon Ascendens	6 [18.8]	NA
Flexura Hepatica	0 [0]	NA
Colon Transversum	2 [6.3]	NA
Flexura Lienalis	1 [3.1]	NA
Colon Descendens	2 [6.3]	NA
Sigmoid	11 [34.4]	NA
Rectosigmoid	1 [3.1]	NA
Rectum	4 [12.5]	NA
Ileocecal valve	1 [3.1]	NA
Adenoma characteristics (largest adenoma) (n, [%])		
≥ 10 mm	13 [40.6]	NA
Villous histology	9 [28.1]	NA
HGD	0 [0]	NA
No dysplasia	3 [9.4]	NA
Hyperplasia	0 [0]	NA
LGD	29 [90.6]	NA
Sessile/serrated	3 [9.4]	NA
Total number adenomas removed (n, [%])		
1	7 [21.9]	NA
2	11 [34.4]	NA
3	7 [21.9]	NA
4-5	3 [9.4]	NA
6-10	4 [12.6]	NA

ABx, antibiotics; BMI, body mass index; CRC, colorectal cancer; FIT, faecal immunochemical test; HGD, high-grade dysplasia; LGD, low-grade dysplasia; NA, not applicable; SD, standard deviation.

Table 5.12. Differences Between All Subgroups of Colorectal Neoplasia, Polyps and Healthy Controls Based on Faecal Volatile Organic Compounds

Comparison	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	P-value
CRC vs HC	0.96 (0.89 – 1.00)	1.00	1.00	0.89	1.00	<0.001
AA vs HC	0.96 (0.93 - 1.00)	0.97	0.94	0.94	0.97	<0.001
LA vs HC	0.96 (0.92 - 0.99)	0.99	0.91	0.92	0.98	<0.001
SA vs HC	0.96 (0.94 - 0.99)	0.96	0.93	0.93	0.96	<0.001
CRC vs AA	0.54 (0.38 - 0.70)	0.98	0.19	0.83	0.75	0.294
CRC vs LA	0.41 (0.31 - 0.51)	0.06	0.96	0.25	0.81	0.920
CRC vs SA	0.41 (0.38 - 0.45)	1.00	0	0.11	NA	0.965
AA vs LA	0.53 (0.43 - 0.63)	0.75	0.36	0.55	0.58	0.278
AA vs SA	0.58 (0.49 - 0.66)	0.72	0.44	0.40	0.76	0.039
T0 HC vs T0 pre-polypectomy	0.98 (0.95 – 1.00)	1.00	0.97	0.97	1.00	<0.001
T1 HC vs T1 post-polypectomy	0.55 (0.40 - 0.69)	0.91	0.25	0.55	0.83	0.256
T0 vs T1 pre- and post-polypectomy	0.94 (0.88 - 1.00)	0.91	0.91	0.91	0.91	<0.001
T0 vs T1 HC	0.58 (0.44 - 0.73)	0.94	0.27	0.58	0.80	0.139

AA, advanced adenoma; AUC, area under the curve; CI, confidence interval; CRC, colorectal cancer; LA, large adenoma; LGD, low-grade dysplasia; NPV, negative predictive value; PPV, positive predictive value; SA, small adenoma; T0, time before polypectomy; T1, 3 months after polypectomy.

Faecal volatile organic compound analysis

The results of VOC analyses by means of GC-IMS are shown in Table 5.12. Results from the Random Forest classifier based on the 20 most discriminative features are presented. Data generated based on all five classifiers using the 20, 50 and 100 most discriminative feature are given in appendix Table 5.5a-c. An example of GC-IMS output from faecal VOC patient with CRC, AA, and Healthy control is illustrated in Figure 5.5.

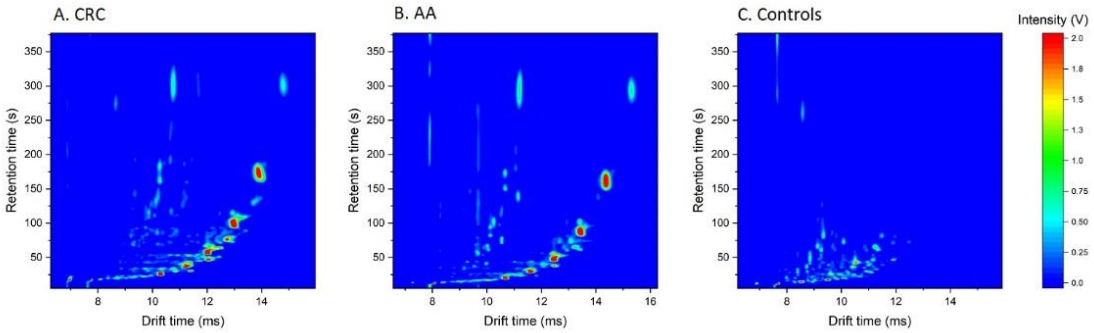


Figure 5.5. An example output of the GC-IMS instrument of the colorectal cancer group (CRC), advanced adenoma group (AA) and control group. The y-axis represents retention time from the gas chromatography column, and the x-axis represents drift time through the ion mobility spectrometer. Volatile organic compound levels in the sample are represented by colour intensity.

Detection of colorectal cancer, advanced adenomas, and non-advanced adenomas

Based on faecal VOC profiles, the CRC group was discriminated from healthy control with high diagnostic accuracy (AUC ± 95%CI: 0.96(0.89 – 1) (appendix Table 5.5a-c). Likewise, high diagnostic accuracy was observed for discrimination of AA, LA and SA when compared to healthy controls (AA 0.96(0.93 –1); LA 0.96(0.92 – 0.99); SA 0.96(0.94-0.99)) (appendix Table 5.5a-c). There were no significant differences between any of the CRC, AA, LA, and SA groups based on faecal VOCs (appendix Table 5.5a-c). Receiver operator characteristic (ROC) curves are constructed in Figure 5.6.

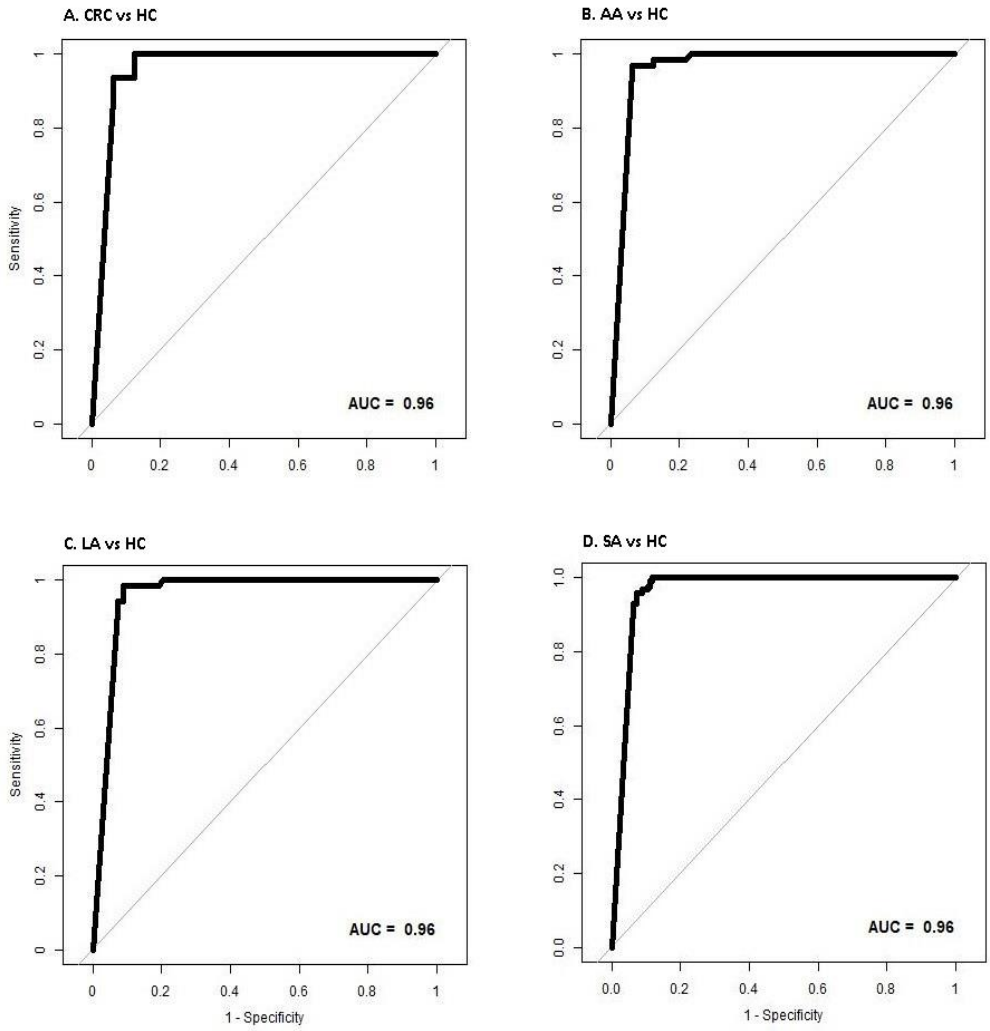


Figure 5.6. ROC curve for the comparison between colorectal cancer (CRC), advanced adenoma (AA), large adenoma (LA), small adenoma (SA) and healthy control (HC) (AUC, area under the curve)

Faecal volatile organic compounds for polypectomy follow-up

Faecal VOC profiles of patients with adenomas differed significantly from those of healthy controls before polyp removal (AUC ± 95% CI: 0.98 (0.95-1)). Remarkably, there was no difference between the faecal VOC profiles of patients who underwent a polypectomy and healthy controls three months after endoscopic intervention (AUC ± 95% CI: 0.55 (0.40-0.66)). There was a highly significant difference in the profiles of patients with adenomas before and after polypectomy, whereas no significant differences were present in faecal VOC profiles of healthy controls before and three months after endoscopy (T0 vs T1 polyps 0.94

(0.88-1); T0 vs T1 C 0.58 (0.44-0.73)). ROC-curves for the follow-up study are depicted in Figure 5.7.

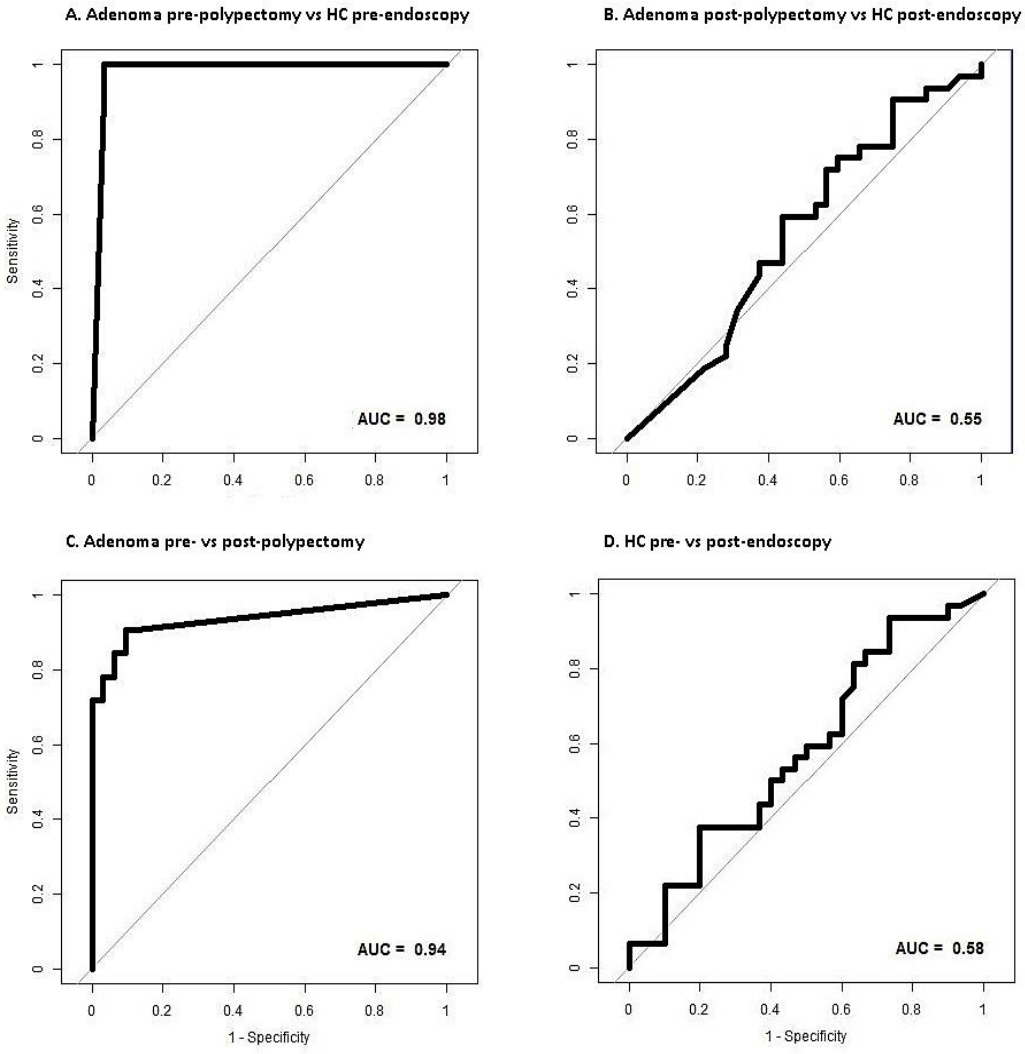


Figure 5.7. ROC curves for the polypectomy follow-up (AUC, area under the curve; HC, healthy control)

5.7. Discussion of Colorectal Cancer Studies

In these studies, we have successfully demonstrated the feasibility of utilising urine and faecal VOC based test to differentiate CRC from healthy control with high AUC, sensitivity, and specificity using both FAIMS and GC-IMS. The first study is the first reported research to demonstrate the application of GC-IMS in CRC detection [36], while the second study is

the first to show the potential of faecal VOC profiles for polypectomy follow-up [37]. A short summary of these studies is shown in Table 5.13.

Table 5.13. Summary of Two CRC Studies

	Original paper's title (date of published)	Sample type	Sample size	Instrument
Study 1	Colorectal cancer and adenoma screening using urinary volatile organic compound (VOC) detection: early results from a single-centre bowel screening population (UK BCSP) (April 2019)	Urine	CRC = 12 Adenoma = 80 HC = 37 Other Diseases = 33*	FAIMS and GC-IMS
Study 2	Early detection and follow-up of colorectal neoplasia based on volatile organic compounds (February 2020)	Faecal	CRC = 14 Adenoma = 260 HC = 227 Polypectomy group = 32	GC-IMS

*Inflammatory bowel disease: n = 7, rectal telangiectasia: n = 2, rectal ulcer: n = 1, radiation proctitis: n = 1, inflammatory pseudo polyp: n = 1, non-specific sigmoid inflammation: n = 1, ischaemic sigmoid stricture: n = 1, diverticular disease: n = 14, haemorrhoids: n = 5.

Based on the result of both studies, FAIMS and GC-IMS were consistently able to differentiate individuals with CRC from healthy controls, as summarised in Table 5.14. This was achieved regardless of sample type. Among these results, the urinary VOC analysis using GC-IMS performed slightly less accurate with AUC (95% CI) of 0.82 (0.67 – 0.93).

Table 5.14. Summary of Urinary and Faecal VOC Analysis Performance in Detecting CRC from Healthy Controls

Instrument | Sample Type	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
FAIMS | Urine	0.98 (0.93 - 1.00)	1.00	0.92	0.92	1.00	2.59×10^{-6}

GC-IMS | Urine	0.82 (0.67 - 0.93)	0.80	0.83	0.67	0.91	1.4×10^{-3}
GC-IMS | Faecal	0.96 (0.89 - 1.00)	1.00	1.00	0.89	1.00	<0.001

Both studies were also investigating the different urinary and faecal VOC of colorectal cancer, adenoma, and healthy control patients. Study 1 groups adenoma patients based on the British Society of Gastroenterology (BSG) guidelines into three group: high risk adenoma (hr), intermediate risk adenoma (ir), and low risk adenoma (lr) [34], whereas study 2 classify patients with adenoma into three group based on the European Society of Gastrointestinal Endoscopy (ESGE) guidelines: advanced adenoma (AA), large adenoma (LA), and small adenoma (SA) [38]. Although both societies have somewhat different criteria for adenoma classification, both take size, histology, and the number of polyps into account when classifying adenoma [38].

As demonstrated in Table 5.15, when attempting to distinguish CRC from adenoma, urinary VOC analysis consistently outperformed faecal VOC analysis. With an AUC of around 0.5, the faecal VOC pattern produced by GC-IMS shows little to no difference between CRC and adenoma. The same results were obtained when AA vs SA and AA vs LA faecal VOCs were compared. On the contrary, faecal VOC analysis surpassed urinary VOC analysis in discriminating adenoma from healthy control as shown in Table 5.16.

Table 5.15. Summary of CRC vs Adenoma

		AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
Study 1 (FAIMS | Urine)							
	CRC vs Adenoma (hr)	0.92 (0.77-1)	0.83	1	1	0.78	6 x 10-4
	CRC vs Adenoma (ir)	0.84 (0.67-1)	0.83	0.75	0.77	0.82	1.5 x 10-3
	CRC vs Adenoma (lr)	0.83 (0.66-1)	0.75	0.92	0.90	0.79	2.26 x 10-3
Study 2 (GC-IMS | Stool)							
	CRC vs AA	0.54 (0.38 - 0.70)	0.98	0.19	0.83	0.75	0.294
	CRC vs LA	0.41 (0.31 - 0.51)	0.06	0.96	0.25	0.81	0.920

	CRC vs SA	0.41 (0.38 - 0.45)	1.00	0	0.11	NA	0.965
	AA vs LA	0.53 (0.43 - 0.63)	0.75	0.36	0.55	0.58	0.278
	AA vs SA	0.58 (0.49 - 0.66)	0.719	0.444	0.40	0.76	0.039

Few studies have demonstrated the potential to discriminate CRC and AA from healthy controls based on faecal VOC analysis [39,40,41,42,43]. The current study outcomes are in line with this literature. However, reliable comparison with other studies is restricted due to the small number of included subjects, lack of knowledge on VOC profiles of AA and non-AA, and the use of different techniques and sampling protocols. Four previous studies have focused on detection of CRC using the analytical platform gas chromatography - mass spectrometry (GC-MS), which is considered the gold standard for the detection of specific metabolites [40,41,42,43]. In the most recent publication, a group of adenomatous polyp patients (n=56) was included in addition to CRC (n=21) and no neoplasia as healthy control (n=60). Multiple differences in metabolite levels were found. The highest diagnostic accuracy was found for the combination of Propan-2-ol and 3-methylbutanioc acid discriminating CRC samples from polyps and healthy controls (AUC 0.82). These differences specifically provided discrimination between CRC patients and other groups, whereas in the current study, VOC profiles differed between CRC and adenoma groups compared to healthy controls. Possible explanations for this are the differences in faecal VOC analysis technique, and in inclusion criteria per subgroup. In the publication by Bond et. al., polyp characteristics are not reported, which hampers reliable comparison to our study groups. All other studies using GC-MS reported relatively small groups of subjects, ranging from n=9-26 CRC and n=10-60 healthy controls. There were interesting similarities in study outcomes (e.g., increased levels of amino acids and short-chain fatty acids and decreased levels of polyhydric alcohols and bile acid), although, none of these metabolite levels were consequently altered. In one previous publication, VOC profiles of 40 CRC, 60 AA and 57 endoscopy controlled healthy controls were compared using pattern-recognition (eNose, Cyranose 320 ®) [39]. Based on faecal VOC patterns, CRC and AA were discriminated from healthy controls with sensitivity of 85% and 62%, and specificity of 87% and 86%, respectively. These test characteristics were below the characteristics found in the current study, however, using the Cyranose 320®, subgroups of CRC and AA were discriminated with moderate accuracy (sensitivity 75% and specificity 73%). A possibility for this apparent discrepancy is the number of CRC patients included in the study by de Meij *et al*. [39], increasing the power to find differences between groups. Another possibility is the instrumental difference between the Cyranose 320® and the GC-IMS system. The Cyranose 320® uses an array of nanocomposite sensors, whereas the GC-

IMS system is using drift time. Although GC-IMS is more sensitive and repeatable, it cannot measure molecules with low proton affinities. If differentiation between CRC and AA was based on these types of molecules, they could not have been detected in the current study.

Table 5.16. Summary of Adenoma vs Healthy Control

	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
Study 1 (GC-IMS | Urine)						
All adenoma vs HC	0.61 (0.47-0.75)	0.58	0.62	0.78	0.39	2.88×10^{-2}
Study 2 (GC-IMS | Stool)						
AA vs HC	0.96 (0.93 – 1.00)	0.97	0.94	0.94	0.97	<0.001
LA vs HC	0.96 (0.92 - 0.99)	0.99	0.91	0.92	0.98	<0.001
SA vs HC	0.96 (0.94 - 0.99)	0.96	0.93	0.93	0.96	<0.001

As mentioned in chapter 2, there are currently only 5 other studies investigating the urinary VOC for CRC detection, and none of them have included the comparison between adenoma vs healthy control. Hence, To the best of our knowledge this is the first study to specifically examine adenoma detection by urinary VOCs. In study 1, applying GC-IMS instrument to differentiate the adenoma group from healthy controls showed poor separation (see Table 5.16). Previous studies have reported advanced adenoma detection using breath VOCs [44,45] demonstrated more encouraging results (sensitivity 1.0). Our finding from study 2 in examining stool VOC for detecting AA, LA, and SA from HC were also provide promising result with AUC, sensitivity, and specificity were above 92% in all comparison. These conflicting results suggest that there is more work needed to establish the mechanism of Urinary VOC signature changes in the presence of colonic adenomas and other gastrointestinal disorders. This is vital as adenomas represent a pre-malignant process with adenoma detection intrinsically linked to CRC mortality [46,47,48].

In study 2, we also investigate the possibility of faecal VOC profile as an alternative noninvasive follow-up testing method after polypectomy. No previous studies have been performed on the potential of faecal VOC profiles for polypectomy follow-up. Surveillance of patients after polypectomy using FIT has been the subject of a previous study comparing stool haemoglobin levels with colonoscopy outcomes [49]. A total of 5225 participants

completed a first FIT one year after polypectomy, demonstrating sensitivity and specificity values of 27.6% and 94.1% at 40 μg/g, and 51.7% and 86.2% at 10 μg/g for CRC, respectively. For AA, sensitivity and specificity values were 17.0% and 95.1% for 40 μg/g and 33.0% and 88.0% for 10 μg/g, respectively. Replacing colonoscopy with FIT would reduce colonoscopies by 71% but would lead to 30-40% missed CRC cases and 40-70% missed AA cases. Observed discriminative accuracies for CRC and AA detection based on faecal VOC analysis exceed these reported accuracies [11,50]. In addition, accuracy to distinguish patients with adenomas from healthy control was high prior to polypectomy, whereas intra-individual profiles changed to physiological state three months following polypectomy, indicating the potential of faecal VOCs as a biomarker for timing of polyp follow-up, and tight control in high-risk populations (e.g., Lynch syndrome).

Alterations in urinary and faecal VOC patterns represent metabolic shifts that may be explained by various mechanisms (e.g., alterations in dietary intake, microbial dysbiosis, inflammatory processes, cancer degeneration). In a recent study, metabolic waste was retrieved from benign cells, colon cancer cells and breast cancer cells that were grown in vitro. It was observed that dogs were able to differentiate cancer cells from benign cells, but not the cell waste of breast from colon cancer, implying that both cancers phenotypes seem to share a common smell print [51]. This may also apply to different phenotypes of adenomas; adenoma and eventually CRC degeneration are possibly based on a shared (metabolic) pathway, explaining the similarities in VOC patterns observed in the current stool study. Apart from excretion of metabolic end-products, our findings may be explained by the presence of intestinal dysbiosis. Faecal microbiota have an important function in protection against invading pathogens and strong evidence exists for the association between microbial dysbiosis and polyp-associated tissue and colonic neoplasia [52]. Causality remains unclear; it is unknown whether this phenomenon is triggered or maintained by carcinogenesis. Intriguingly, in the present studies it was found that faecal VOC profiles of patients three months after polypectomy were altered to normalcy following this intervention, suggesting that the alleged faecal microbial dysbiosis returns to a physiological state after polyp removal.

A limitation of these studies was the small sample size for the CRC group, but this reflects the nature of the screening population with low CRC detection rates of around 8%. Machine learning algorithms that were used to analyse both sets of date always risk the possibility of overfitting of the data. This was minimised using a cross-validation technique, using two different technologies, different collection protocols, and two different sample types.

5.8. Conclusion

In conclusion, our result indicates that both PDAC and CRC could be detected by analysing the human waste's VOC of patients. In the first two studies, we investigated the use of urinary headspace volatiles to identify patients suffering from PDAC. Urinary headspace was analysed by FAIMS, GC-IMS and GC-TOF-MS, with all instruments showing potential in separating PDAC from healthy controls. FAIMS showed the best overall performance when detecting PDAC from healthy patients. However, due to the limitation in automation sampling of FAIMS, this instrument is less desirable to be used in clinical settings compared with other devices. When comparing PDAC from CP, GC-IMS and GC-TOF-MS showed only small differences, indicating that there is commonality between the VOCs produced by both conditions. This is supported by the high sensitivity of CP vs healthy controls. Chemical identification suggests that 2,6-dimethyl-octane, nonanal, 4-ethyl-1,2-dimethyl-benzene and 2-pentanone play at important role is separating the data. Further work is needed to validate these biomarkers in a larger study. However, we believe this approach might hold a potential as a completely non-invasive, detection tool for pancreatic cancer.

The third and fourth studies results shows that the detection of CRC though urinary and faecal VOC is feasible. CRC can be correctly classified from healthy control using both FAIMS and GC-IMS instruments. Only urinary VOC analysis was able to provide a clear separation between patient with CRC from adenoma. On the other hand, faecal VOC analysis showed better separation capability in discriminate adenoma patients among healthy controls compared with urinary VOC analysis. Additionally, intra-individual faecal VOC profiles of patients with adenomas altered towards a physiological state following polypectomy, emphasising its potential for intra-individual follow-up and timing of endoscopy.

5.9. References

[1] J. Ferlay *et al.*, "Global Cancer Observatory: Cancer Today.," *Lyon, France: International Agency for Research on Cancer.*, 2020. https://gco.iarc.fr/today (accessed Jan. 20, 2021).

[2] I. Aier, R. Semwal, A. Sharma, and P. K. Varadwaj, "A systematic assessment of statistics , risk factors , and underlying features involved in pancreatic cancer," *Cancer Epidemiol.*, vol. 58, no. September 2018, pp. 104–110, 2019, doi: 10.1016/j.canep.2018.12.001.

[3] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2016,” *CA. Cancer J. Clin.*, vol. 66, no. 1, pp. 7–30, 2016, doi: 10.3322/caac.21332.

[4] S. T. Chari *et al.*, “Early Detection of Sporadic Pancreatic Cancer: Summative Review,” *Pancreas*, vol. 44, no. 5, pp. 693–712, 2015, doi: 10.1097/MPA.0000000000000368.

[5] A. Kanno *et al.*, “Multicenter study of early pancreatic cancer in Japan,” *Pancreatology*, vol. 18, no. 1, pp. 61–67, 2018, doi: 10.1016/j.pan.2017.11.007.

[6] A. S. Takhar, P. Palaniappan, R. Dhingsa, and D. N. Lobo, “Recent developments in diagnosis of pancreatic cancer,” *Br. Med. J.*, vol. 329, no. 7467, pp. 668–673, 2004, doi: 10.1136/bmj.329.7467.668.

[7] G. Xie *et al.*, “Plasma metabolite biomarkers for the detection of pancreatic cancer,” *J. Proteome Res.*, vol. 14, no. 2, pp. 1195–1202, 2015, doi: 10.1021/pr501135f.

[8] J. Ferlay, D. M. Parkin, and E. Steliarova-Foucher, “Estimates of cancer incidence and mortality in Europe in 2008,” *Eur. J. Cancer*, vol. 46, no. 4, pp. 765–781, 2010, doi: 10.1016/j.ejca.2009.12.014.

[9] A. B. Ballinger and C. Anggiansah, “Colorectal cancer,” *Br. Med. J.*, vol. 335, no. 7622, pp. 715–718, 2007, doi: 10.1136/bmj.39321.527384.BE.

[10] S. Winawer *et al.*, “Colorectal cancer screening and surveillance: Clinical guidelines and rationale - Update based on new evidence,” *Gastroenterology*, vol. 124, no. 2, pp. 544–560, 2003, doi: 10.1053/gast.2003.50044.

[11] A. Katsoula, P. Paschos, A. B. Haidich, A. Tsapas, and O. Giouleme, “Diagnostic accuracy of fecal immunochemical test in patients at increased risk for colorectal cancer ameta-Analysis,” *JAMA Intern. Med.*, vol. 177, no. 8, pp. 1110–1118, 2017, doi: 10.1001/jamainternmed.2017.2309.

[12] M. B. Amin *et al.*, *AJCC Cancer Staging Manual*, 8th ed. Springer International Publishing, 2017.

[13] R. P. Arasaradnam *et al.*, “A Novel tool for noninvasive diagnosis and tracking of patients with inflammatory bowel disease,” *Inflamm. Bowel Dis.*, vol. 19, no. 4, pp. 999–1003, 2013, doi: 10.1097/MIB.0b013e3182802b26.

[14] R. P. Arasaradnam *et al.*, “Detection of colorectal cancer (CRC) by urinary volatile

organic compound analysis," *PLoS One*, vol. 9, no. 9, 2014, doi: 10.1371/journal.pone.0108750.

[15] R. P. Arasaradnam *et al.*, "Non-invasive distinction of non-alcoholic fatty liver disease using urinary volatile organic compound analysis: Early results," *J. Gastrointest. Liver Dis.*, vol. 24, no. 2, pp. 197–201, 2015, doi: 10.15403/jgld.2014.1121.242.ury.

[16] R. P. Arasaradnam, A. Wicaksono, H. O'Brien, H. M. Kocher, J. A. Covington, and T. Crnogorac-Jurcevic, "Noninvasive Diagnosis of Pancreatic Cancer Through Detection of Volatile Organic Compounds in Urine," *Gastroenterology*, vol. 154, no. 3, pp. 485-487.e1, 2018, doi: 10.1053/j.gastro.2017.09.054.

[17] J. Kirkegård, F. V. Mortensen, and D. Cronin-Fenton, "Chronic Pancreatitis and Pancreatic Cancer Risk: A Systematic Review and Meta-analysis," *Am. J. Gastroenterol.*, vol. 112, no. 9, pp. 1366–1372, 2017, doi: 10.1038/ajg.2017.218.

[18] L. Lacey, E. Daulton, A. Wicaksono, J. A. Covington, and S. Quenby, "Detection of Group B Streptococcus in pregnancy by vaginal volatile organic compound analysis: a prospective exploratory study," *Transl. Res.*, vol. 216, pp. 23–29, 2020, doi: 10.1016/j.trsl.2019.09.002.

[19] A. Tiele *et al.*, "Breath-based non-invasive diagnosis of Alzheimer's disease: A pilot study," *J. Breath Res.*, vol. 14, no. 2, p. 026003, Feb. 2020, doi: 10.1088/1752-7163/ab6016.

[20] M. D. Rouvroye *et al.*, "Faecal scent as a novel non-invasive biomarker to discriminate between coeliac disease and refractory coeliac disease: A proof of principle study," *Biosensors*, vol. 9, no. 2, 2019, doi: 10.3390/bios9020069.

[21] S. Van den Velde, F. Nevens, P. Van hee, D. van Steenberghe, and M. Quirynen, "GC-MS analysis of breath odor compounds in liver patients," *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 875, no. 2, pp. 344–348, 2008, doi: 10.1016/j.jchromb.2008.08.031.

[22] C. E. Garner *et al.*, "Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease," *FASEB J.*, vol. 21, no. 8, pp. 1675–1688, 2007, doi: 10.1096/fj.06-6927com.

[23] I. Ahmed, R. Greenwood, B. Costello, N. Ratcliffe, and C. S. Probert, "Investigation

of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease," *Aliment. Pharmacol. Ther.*, vol. 43, no. 5, pp. 596–611, 2016, doi: 10.1111/apt.13522.

[24] Y. Saalberg and M. Wolff, "VOC breath biomarkers in lung cancer," *Clin. Chim. Acta*, vol. 459, pp. 5–9, 2016, doi: 10.1016/j.cca.2016.05.013.

[25] H. Amal *et al.*, "Assessment of ovarian cancer conditions from exhaled breath," *Int. J. Cancer*, vol. 136, no. 6, pp. E614–E622, 2015, doi: 10.1002/ijc.29166.

[26] X.-A. Fu, M. Li, R. J. Knipp, M. H. Nantz, and M. Bousamra, "Noninvasive detection of lung cancer using exhaled breath," *Cancer Med.*, vol. 3, no. 1, pp. 174–181, 2014, doi: 10.1002/cam4.162.

[27] C. Bax, B. J. Lotesoriere, S. Sironi, and L. Capelli, "Review and comparison of cancer biomarker trends in urine as a basis for new diagnostic pathways," *Cancers (Basel).*, vol. 11, no. 9, 2019, doi: 10.3390/cancers11091244.

[28] C. Napoli, N. Sperandio, R. T. Lawlor, A. Scarpa, H. Molinari, and M. Assfalg, "Urine metabolic signature of pancreatic ductal adenocarcinoma by 1H nuclear magnetic resonance: Identification, mapping, and evolution," *J. Proteome Res.*, vol. 11, no. 2, pp. 1274–1283, 2012, doi: 10.1021/pr200960u.

[29] V. W. Davis, D. E. Schiller, D. Eurich, O. F. Bathe, and M. B. Sawyer, "Pancreatic ductal adenocarcinoma is associated with a distinct urinary metabolomic signature," *Ann. Surg. Oncol.*, vol. 20, no. 3 SUPPL., 2013, doi: 10.1245/s10434-012-2686-7.

[30] E. R. Lusczek *et al.*, "Urinary 1H-NMR metabolomics can distinguish pancreatitis patients from healthy controls," *JOP*, vol. 14, no. 2, pp. 161–170, Mar. 2013, doi: 10.6092/1590-8577/1294.

[31] T. P. Radon *et al.*, "Identification of a three-biomarker panel in urine for early detection of pancreatic adenocarcinoma," *Clin. Cancer Res.*, vol. 21, no. 15, pp. 3512–3521, 2015, doi: 10.1158/1078-0432.CCR-14-2467.

[32] J. Mayerle *et al.*, "Metabolic biomarker signature to differentiate pancreatic ductal adenocarcinoma from chronic pancreatitis," *Gut*, vol. 67, no. 1, pp. 128–137, 2018, doi: 10.1136/gutjnl-2016-312432.

[33] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE:

Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[34] S. R. Cairns *et al.*, "Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups (update from 2002)," *Gut*, vol. 59, no. 5, pp. 666–689, 2010, doi: 10.1136/gut.2009.179804.

[35] G. Vanbiervliet *et al.*, "Endoscopic management of superficial nonampullary duodenal tumors: European Society of Gastrointestinal Endoscopy (ESGE) Guideline," *Endoscopy*, vol. 53, no. 5, pp. 522–534, 2021, doi: 10.1055/a-1442-2395.

[36] W. Zhou, J. Tao, J. Li, and S. Tao, "Volatile organic compounds analysis as a potential novel screening tool for colorectal cancer: A systematic review and meta-analysis," *Medicine (Baltimore).*, vol. 99, no. 27, p. e20937, 2020, doi: 10.1097/MD.0000000000020937.

[37] F. Vernia *et al.*, "Are volatile organic compounds accurate markers in the assessment of colorectal cancer and inflammatory bowel diseases? A review," *Cancers (Basel).*, vol. 13, no. 10, pp. 1–20, 2021, doi: 10.3390/cancers13102361.

[38] N. Abu-Freha *et al.*, "Post-polypectomy surveillance colonoscopy: Comparison of the updated guidelines," *United Eur. Gastroenterol. J.*, vol. 9, no. 6, pp. 681–687, 2021, doi: 10.1002/ueg2.12106.

[39] T. G. De Meij *et al.*, "Electronic nose can discriminate colorectal carcinoma and advanced adenomas by fecal volatile biomarker analysis: Proof of principle study," *Int. J. Cancer*, vol. 134, no. 5, pp. 1132–1138, 2014, doi: 10.1002/ijc.28446.

[40] E. M. Song *et al.*, "Fecal Fatty Acid Profiling as a Potential New Screening Biomarker in Patients with Colorectal Cancer," *Dig. Dis. Sci.*, vol. 63, no. 5, pp. 1229–1236, 2018, doi: 10.1007/s10620-018-4982-y.

[41] T. L. Weir, D. K. Manter, A. M. Sheflin, B. A. Barnett, A. L. Heuberger, and E. P. Ryan, "Stool Microbiome and Metabolome Differences between Colorectal Cancer Patients and Healthy Adults," *PLoS One*, vol. 8, no. 8, 2013, doi: 10.1371/journal.pone.0070803.

[42] X. Wang, J. Wang, B. Rao, and L. I. Deng, "Gut flora profiling and fecal metabolite composition of colorectal cancer patients and healthy individuals," *Exp. Ther. Med.*, vol. 13, no. 6, pp. 2848–2854, 2017, doi: 10.3892/etm.2017.4367.

[43] A. Bond *et al.*, “Volatile organic compounds emitted from faeces as a biomarker for colorectal cancer,” *Aliment. Pharmacol. Ther.*, vol. 49, no. 8, pp. 1005–1012, 2019, doi: 10.1111/apt.15140.

[44] H. Amal *et al.*, “Breath testing as potential colorectal cancer screening tool,” *Int. J. Cancer*, vol. 138, no. 1, pp. 229–236, 2016, doi: 10.1002/ijc.29701.

[45] D. F. Altomare *et al.*, “The use of the PEN3 e-nose in the screening of colorectal cancer and polyps,” *Tech. Coloproctol.*, vol. 20, no. 6, pp. 405–409, 2016, doi: 10.1007/s10151-016-1457-z.

[46] G. Hoff *et al.*, “Polypectomy of adenomas in the prevention of colorectal cancer: 10 years’ follow-up of the Telemark Polyp Study I. A prospective, controlled population study,” *Scand. J. Gastroenterol.*, vol. 31, no. 10, pp. 1006–1010, 1996, doi: 10.3109/00365529609003121.

[47] W. S. Atkin, B. C. Morson, and J. Cuzick, “Long-Term Risk of Colorectal Cancer after Excision of Rectosigmoid Adenomas,” *N. Engl. J. Med.*, vol. 326, no. 10, pp. 658–662, Mar. 1992, doi: 10.1056/NEJM199203053261002.

[48] M. J. O’Brien *et al.*, “The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas.,” *Gastroenterology*, vol. 98, no. 2, pp. 371–379, Feb. 1990.

[49] A. J. Cross *et al.*, “Faecal immunochemical tests (FIT) versus colonoscopy for surveillance after screening and polypectomy: A diagnostic accuracy and cost-effectiveness study,” *Gut*, vol. 68, no. 9, pp. 1642–1652, 2019, doi: 10.1136/gutjnl-2018-317297.

[50] P. Rozen *et al.*, “Identification of colorectal adenomas by a quantitative immunochemical faecal occult blood screening test depends on adenoma characteristics, development threshold used and number of tests performed,” *Aliment. Pharmacol. Ther.*, vol. 29, no. 8, pp. 906–917, 2009, doi: 10.1111/j.1365-2036.2009.03946.x.

[51] I. S. Seo *et al.*, “Cross detection for odor of metabolic waste between breast and colorectal cancer using canine olfaction,” *PLoS One*, vol. 13, no. 2, pp. 1–9, 2018, doi: 10.1371/journal.pone.0192629.

[52] P. Louis, G. L. Hold, and H. J. Flint, “The gut microbiota, bacterial metabolites and

colorectal cancer," *Nat. Rev. Microbiol.*, vol. 12, no. 10, pp. 661–672, 2014, doi: 10.1038/nrmicro3344.

Chapter 6
APPLICATION ON INFLAMMATORY BOWEL DISEASE DETECTION

6.1. Introduction

Inflammatory bowel disease (IBD) is a chronic condition of unknown aetiology, characterized by periods of relapse and remission. It comprises of two conditions, specifically Crohn's disease (CD) and ulcerative colitis (UC) [1]. Both conditions involve inflammation of the gut and are particularly unpleasant. While UC only affects the colon (large intestine), CD can affect any part of the digestive system from the mouth to the anus [2]. IBD is a common condition in the Western world, affecting over 250,000 people in the UK and 28 million worldwide [3]. A worldwide increase in incidence and prevalence of IBD has also been observed. In the UK, a prevalence of 142 IBD cases per 10.000 patients has been reported [4]. IBD also presents in childhood in approximately 25% of de novo cases and its incidence has increased over the past decades to ten per 100,000 children [5,6]. The estimated annual cost of treatment, per patient, is approx. €30,000 with an average of 20% loss of working productivity [1,7]. This is due to the relapsing nature of the disease where there may be times when the symptoms are severe (flare-ups), followed by long periods when there are few or no symptoms at all (remission). Common symptoms include diarrhoea, cramping pains in the abdomen, vomiting, weight-loss and fatigue [8]. In addition to this, IBD has a damaging impact on psychosocial functioning, quality of life, and significant personal cost of delayed treatment [9]. A study from 2014, conducted in the UK, revealed that 10% of IBD patients are initially misdiagnosed with other gastrointestinal conditions, such as irritable bowel syndrome (IBS), and that 3% of misdiagnosed cases persisted for five or more years [10]. Misdiagnosis can have serious consequences for the patient; especially for those with CD, since delays are correlated with an increased risk of later bowel stenosis and CD-related intestinal surgery [10]. Early diagnosis of IBD remains a clinical challenge, with current tests being invasive and costly. Diagnostic tools for CD and UC include a thorough history, endoscopic investigations with histological examination, faecal inflammatory markers, capsule endoscopy and imaging [11]. Colonoscopy with histology is considered the "gold standard" to diagnose IBD [12]. This procedure is uncomfortable for the patient, often involves multiple biopsies, is expensive for the health service provider (such as the NHS) and has an associated morbidity. Currently, Faecal calprotectin (FCP) is the most commonly used non-invasive biomarker to detect IBD, which is characterized by a high sensitivity for mucosal inflammation (0.98; 95% confidence interval [CI], 0.95-0.99) but limited specificity (0.68; 95% CI, 0.50-0.86) [13]. Due to this low specificity, FCP often leads to false-positive results, resulting in unnecessary colonoscopies procedure.

Volatile organic compounds (VOCs) have emerged as a promising alternative. It has shown potential to serve as a diagnostic biomarker for a broad range of gastrointestinal

disease, in particular those linked to microbial dysbiosis, for example, *Clostridium difficile* infection, IBD, colorectal cancer and necrotizing enterocolitis [14,15,16,17]. VOCs are gaseous carbon-bound chemicals and are thought to represent both metabolic processes in the human body and the interaction between gut microbiota and hosts [18]. These molecular end-products can be found in all bodily excretion dependent on their volatility and sample temperature. Alteration in cellular metabolic processes (e.g. in disease state, microbiome metabolism, and microbiota-host interaction) are reflected by changes in the emitted VOC composition [19]. The most common sources of VOC that are used to diagnose diseases are urine, faecal and breath.

The potential of faecal, urinary and breath VOC profiles as non-invasive biomarkers has been described for various gastrointestinal diseases. Faecal VOC that are largely produced during bacterial fermentation in the gut, reflect microbial composition, function, and microbiota-host interaction [20]. Many studies of this VOC have demonstrated to allow for differentiation of a variety of diseases, including colorectal carcinoma and celiac disease [21,22]. However, collection of faeces has been described to evoke feelings of embarrassment and concerns about hygiene by patients [23]. In most countries, collection of urine samples is considered as more user friendly and, therefore, less of a burden for patients [24]. Furthermore, urine can be more easily provided on demand in situations where fast analysis is required to accelerate the diagnostic process. Yet, among all of these VOC sources, exhaled breath testing provides the easiest and the most convenient way for patients to collect the sample. It is known that human breath consists of over 3000 VOCs [25], which are products of host metabolic activity, and in some cases, specific biomarkers associated with a disease [26,27,28,29]. Previous studies have shown the potential of exhaled breath VOC to identify various diseases, such as cancer and diabetes [30,31]. However, exhaled breath analysis suffers from sample instability and the need of comprehensive analytical methods.

This chapter describes new clinical studies investigating the potential of faecal, urinary, and breath VOC analysis as a non-invasive way to diagnose IBD from healthy control patients. This chapter consists of four studies. The first study described the use of FAIMS in investigating the faecal VOC in differentiating between paediatric IBD from irritable bowel syndrome. Then, the potential of exhaled breath sample testing to detect IBD using E-nose and GC-IMS is described in the second study. The urinary and the faecal VOC assessment are also compared in the third study using GC-IMS. The last study talks about the potential of faecal VOC analysis to monitor the IBD activity in patients. Finally, the result and comparison of these studies are then discussed.

6.2. Differentiation Between Paediatric Irritable Bowel Syndrome and Inflammatory Bowel Disease Based on Faecal Scent (Study 1)

The content of this section is based on a paper that has been published in peer-reviewed journals:

Bosch, Sofie, Nora van Gaal, Roy P. Zuurbier, James A. Covington, Alfian N. Wicaksono, Maarten H. Biezeveld, Marc A. Benninga, Chris J. Mulder, Nanne K.H. de Boer, and Tim G.J. de Meij. 2018. "Differentiation between Paediatric Irritable Bowel Syndrome and Inflammatory Bowel Disease Based on Faecal Scent: Proof of Principle Study." Inflammatory Bowel Diseases 24 (11): 2468–75. https://doi.org/10.1093/IBD/IZY151.

6.2.1. Methods

This case-control study was performed at the outpatient clinics of the paediatric (gastroenterology) department of 2 tertiary centres, VU University Medical Center and Emma Children's Hospital, Academic Medical Centre (AMC), and 1 general hospital, OLVG Oost (all centres located in Amsterdam, the Netherlands). This study was performed between December 2013 and December 2016. This study was approved by the Medical Ethical Review Committee (METc) of the VU University Medical Centre under file number 2015.393 and by the local medical ethical committees of the other 2 participating centres. Written informed consent was obtained from all parents, and from the child in case of age over 12 years. The study Participants consists of three subgroups which were defined as

IBS and FAP-NOS

Children aged 4 to 17 years visiting the outpatient clinic in 1 of the 3 hospitals between August 2016 and December 2016 and fulfilling the ROME IV criteria for Irritable Bowel Syndrome (IBD) or Functional Abdominal Pain-Not Otherwise Specified (FAP-NOS) were eligible to participate [32]. During clinical appointments, patients were asked to participate in this study. Patients, for whom informed consent was obtained, were provided a stool container and a questionnaire on abdominal symptoms and defecation pattern, including consistency of stool using the Bristol stool chart, medication use, and medical history. Exclusion criteria were the use of anti-/probiotics or immunosuppressive therapy 3 months before inclusion, immunocompromised disease (i.e., leukaemia, human immunodeficiency virus), diagnosis of a gastrointestinal disease, proven infectious colitis in the month before presentation (determined by positive stool culture for *Salmonella spp*., *Shigella spp*., *Yersinia spp*., *Campylobacter spp*., *Clostridium spp*. Toxins, or parasites in stools), and a history of

gastrointestinal surgery (except appendectomy). From all IBS and FAP-NOS patients included in this study, faecal calprotectin levels were assessed to exclude IBD.

Inflammatory Bowel Disease (IBD)

Participants aged 4 to 17 years were extracted from an existing cohort consisting of de novo treatment naïve paediatric IBD patients (59 Crohn's disease [CD], 40 ulcerative colitis [UC]), included at the VU University medical centre and the Emma Children's Hospital (AMC) between December 2013 and October 2015 for study of diagnostic faecal biomarkers. All participants were instructed to collect a faecal sample before bowel cleaning, ileocolonoscopy, and esophagogastroduodenoscopy. The diagnosis of IBD was made according to the revised diagnostic Porto-criteria for paediatric IBD, including endoscopic, histologic, and radiologic findings by means of magnetic resonance enteroclysis (MRE) [33]. Localization and behaviour of disease were classified according to the Paris classification [34]. Clinical activity was determined at study inclusion based on the Physician Global Assessment (PGA score), levels of faecal calprotectin (FCP >250 ug/g was considered active disease), and C reactive protein (CRP). Exclusion criteria were similar to the IBS/FAP-NOS group, except for exclusion when diagnosed with IBD.

Healthy Controls (HC)

Children aged 4 to 17 years attending elementary and high schools in the province North-Holland, the Netherlands, were instructed to collect a faecal sample. Similar to the IBS/FAP-NOS group, all participants completed a questionnaire containing similar items. Exclusion criteria were functional gastrointestinal disorders according to the ROME IV criteria, diagnosis with a gastrointestinal or immunocompromised disease, history of gastrointestinal surgery (except appendectomy), or the use of pro- or antibiotics 3 months before inclusion.

Matching Procedure

A total of 15 IBD/FAP-NOS patients (9 IBS, 6 FAP-NOS) were strictly matched to 15 UC, 15 CD and 30 health controls (HCs), based on age and sex. For this, the following procedure was performed. First, from the 99 IBD patients (59 CD, 40 UC) of the existing cohort, all of the eligible subjects were strictly matched to IBS/FAP-NOS patients. Then, IBD patients were randomly included form the matched group in a 1:1:1 ratio (IBS/FAP-NOS to UC to CD). After this, 30 HCs recruited for this study were matched to the IBS/FAP-NOS group in a 2:1 ratio.

Sample Collection and Analysis

Patients were instructed to collect a fresh faecal sample in a stool container (Stuhlgefäß 10 mL, Frickenhausen, Germany) and store the sample in the refrigerator at home directly after bowel movement. The samples were transported to the hospital by 1 of the researchers, using cool elements and a cool bag. Here, samples were directly stored at -20°C until further handling.

Faecal volatile organic compounds analysis was performed using FAIMS (Lonestar, Owlstone, Ltd.), according to the protocol described in an earlier study by Bomers *et al*. [35]. In short, faecal samples were thawed to room temperature 10 minutes before VOC analysis. A mixture of 0.5 g faecal sample and 3.5 mL tap water was manually shaken to homogenize the sample. Compressed air (0.1 MPa) was used as carrier gas to transfer the sample headspace into the FAIMS device. The Lonestar was set up in a pressurized configuration with a flow rate of 2 L/ min. The temperatures were set at 35°C for the sample holder, 70°C for the lid, and 100°C for the filter region. After the procedure, the air in the Lonestar was refreshed by analysing the headspace of 10 mL tap water [36]. The dispersion field passed through 51 equal settings between 0% and 100% (in the ratio of high electric field to low electric field). The compensation voltage was set between +6V and –6V in 512 steps for each dispersion field [35]. Each faecal sample was analysed 3 times sequentially, producing 3 matrices in 540 seconds. For the statistical analysis, only the third matrix was used for optimal diagnostic potential [37].

Data Analysis

The demographic data of each group (IBS/FAP-NOS, UC, CD, and HC) were compared using the Kruskal-Wallis H test with addition of the Wilcoxon rank-sum test for continuous data. The Fisher exact test was performed for dichotomous data using IBM SPSS, version 22. The FAIMS data were processed using our data analysis pipeline which has been described in detail in Chapter 4. using this pipeline, 44 statistically important features were used. Four classification algorithms were applied, Sparse logistic regression (SLR), random forest (RF), Gaussian process (GP), and support vector machine (SVM).

6.2.2. Results

Baseline characteristics and disease specifics of the study subjects are displayed in Table 6.1. There were no significant differences in age, sex, and BMI between the IBS/FAP-NOS, IBD, and HC subgroups. Levels of FCP were below 250ug/g in the IBS/FAP-NOS

group, with the exception of 1 patient (476 ug/g), in whom it normalized after repeating the measurement, whereas the IBD group had a median FCP level (interquartile range [IQR]) of 1237 (580–1885) ug/g. At study inclusion, the majority of IBS/FAP-NOS patients had experienced abdominal symptoms for more than a year, with frequencies varying from once a week to daily. All of the children in the HC group were asymptomatic. Faecal frequency was higher in the IBS/ FAP-NOS group compared with the HC group, although this was not significant. In addition, no differences in faecal consistency based on the Bristol Stool Chart and way of delivery were found between IBS/FAP-NOS and HC.

Table 6.1. Baseline Characteristics

		Crohn's Disease (n=15)	Ulcerative Colitis (n=15)	IBS/FAP-NOS (n = 15 [9/6])	Healthy (n=30)
Age, years (median, IQR, minimum-maximum)		12.8 (5.0) (5.9 – 17.9)	11.8 (7.8) (3.2 – 17.8)	12.9 (8.4) (4.4-18.1)	12.7 (8.1) (4.1 – 17.9)
Male sex (n, %)		9 (60)	8 (53)	8 (53)	15 (50)
Storage time, month (median, IQR, minimum – maximum)		31.7 (25.3)[a] (8.2 – 54.5)	45.1 (36.2)[a] (15.0 – 59.4)	0.6 (0.6)[a] (0.2 - 2.9)	1.4 (0.3)[a] (0.5 – 4.5)
BMI (median, IQR)		NA	NA	16.7 (5)	17.0 (3)
Bristol stool chart (n, %)					
	Type 2	NA	NA	2 (14)[b]	4 (14)[b]
	Type 3	NA	NA	5 (36)	19 (66)
	Type 4	NA	NA	4 (29)	5 (17)
	Type 5	NA	NA	3 (21)	1 (3)
Stool frequency (n, %)					
	2 times/wk or less	NA	NA	2 (14)[b]	1 (4)[b]
	3-6 times/wk	NA	NA	1 (7)	9 (33)
	Once/d	NA	NA	5 (36)	14 (44)
	2-3 times/d	NA	NA	5 (36)	4 (15)
	4 times/d or more	NA	NA	1 (7)	1 (4)
Way of delivery					
	Cesarean section (n, %)	NA	NA	3 (23)[c]	2 (7)[b]
	Natural (n, %)	NA	NA	10 (77)	27 (93)
Frequency of symptoms (IBS/FAP) (n, %)					
	None	NA	NA	0 (0)	30 (100)
	Once/wk	NA	NA	4 (27)	0 (0)
	2 to 4 times/wk	NA	NA	10 (66)	0 (0)
	Every day	NA	NA	1 (7)	0 (0)
Duration of symptoms (n, %)					
	> 1 y	0 (0)b	1 (7)	10 (67)	NA
	2 to 12 mo	11 (73)	7 (47)	3 (20)	NA
	≤ 2 mo	3 (13)	7 (47)	2 (13)	NA

Physician Global Assessment					
	Quiescent	1	0	NA	NA
	Mild	0	3	NA	NA
	Moderate	5	5	NA	NA
	Severe	9	7	NA	NA
Feacal calprotectin, µg/g (median, IQR)		1214 (627 – 1860)	1260 (401 – 1950)	22 (4.8 – 133)	NA
CRP, mg/L (median, IQR)		21 (7-68)	4 (<2.5 – 7)	NA	NA
Crohn's disease localization[d]					
	Ileal (L1)	0	NA	NA	NA
	Colonic (L2)	6	NA	NA	NA
	Ileocolonic (L3)	9	NA	NA	NA
	Proximal disease (L4)	5	NA	NA	NA
Crohn's disease behaviour[d]					
	B1 (NSNP)	11	NA	NA	NA
	B1p (NSNP+p)	2	NA	NA	NA
	B2 (S)	0	NA	NA	NA
	B2p (S + p)	0	NA	NA	NA
	B3 (P)	0	NA	NA	NA
	B3p (P + p)	2	NA	NA	NA
Ulcerative colitis[d]					
	Proctitis (E1)	NA	3	NA	NA
	Left-sided (E2)	NA	2	NA	NA
	Extensive (E3)	NA	10	NA	NA

All values were obtained at study inclusion. Localization of IBD was obtained by ileocolonoscopy and esophagogastroduodenoscopy before treatment initiation and MR enteroclysis

Abbreviations: NA, not applicable; NSNP, nonstricturing nonpenetrating; S, stricturing; P, penetrating; p, perianal disease

[a] Significant differences between all subgroups (P < 0.001), analysed using Wilcoxon rank-sum tests.

[b] Missing data from 1 subject.

[c] Missing data 2 subjects.

[d] Based on Paris classification for inflammatory bowel disease

IBS/FAP-NOS vs IBD

The results of the VOC analysis by the FAIMS technique are shown in Table 6.2. For each analysis, the best performing of the 4 different applied classification models is shown. A complete overview of the data generated by the 4 classification models is given in Appendix Tables 6.1a-d. Faecal VOCs of IBS/FAP-NOS patients differed from IBD patients (AUC,

0.94; 95% CI, 0.88–1; 1, 0.87, 0.79, 1, 0.00000002613; sensitivity, specificity, PPV, NPV, P value, respectively). Corresponding receiver operating characteristic (ROC) curves are shown in Figure 6.1. An overview of the complete outcome of the 4 performed classifiers is displayed in Appendix Tables 6.1a and 6.1b. In addition, there were significant differences between the VOC profiles of IBS/FAP-NOS patients and both UC and CD subgroups (Table 6.2; Appendix Tables 6.1a-d). A complete overview of the data generated by the 4 classification models is given in Appendix Tables 6.1a-d.

Table 6.2. Performance Characteristics for the Discrimination of IBS/FAP-NOS, IBD, and HC by Faecal VOC Analysis

	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	P-value
IBS/FAP-NOS vs IBD	0.94 (0.88-1)	1	0.87	0.79	1	0.00000002613
IBS/FAP-NOS vs CD	0.87 (0.73-1)	0.93	0.82	0.83	0.92	0.0001617
IBS/FAP-NOS vs UC	0.96 (0.91-1)	1	0.8	0.83	1	0.000007501
IBS/FAP-NOS vs HC	0.59 (0.41 -0.77)	0.6	0.63	0.45	0.76	0.1667
IBS vs FAP-NOS	0.76 (0.44-1)	1	0.6	0.83	1	0.9504
IBD vs HC	0.96 (0.93-1)	0.93	0.97	0.97	0.94	0.0000000003982
UC vs HC	0.98 (0.94-1)	0.93	0.97	0.93	0.97	0.0000000005654
CD vs HC	0.95 (0.88-1)	0.93	0.93	0.88	0.97	0.0000001636
CD vs UC	0.67 (0.47 -0.88)	0.6	0.8	0.75	0.67	0.05799

Sensitivity, specificity, P-values, and AUCs are reported for the respective optimum cut-points

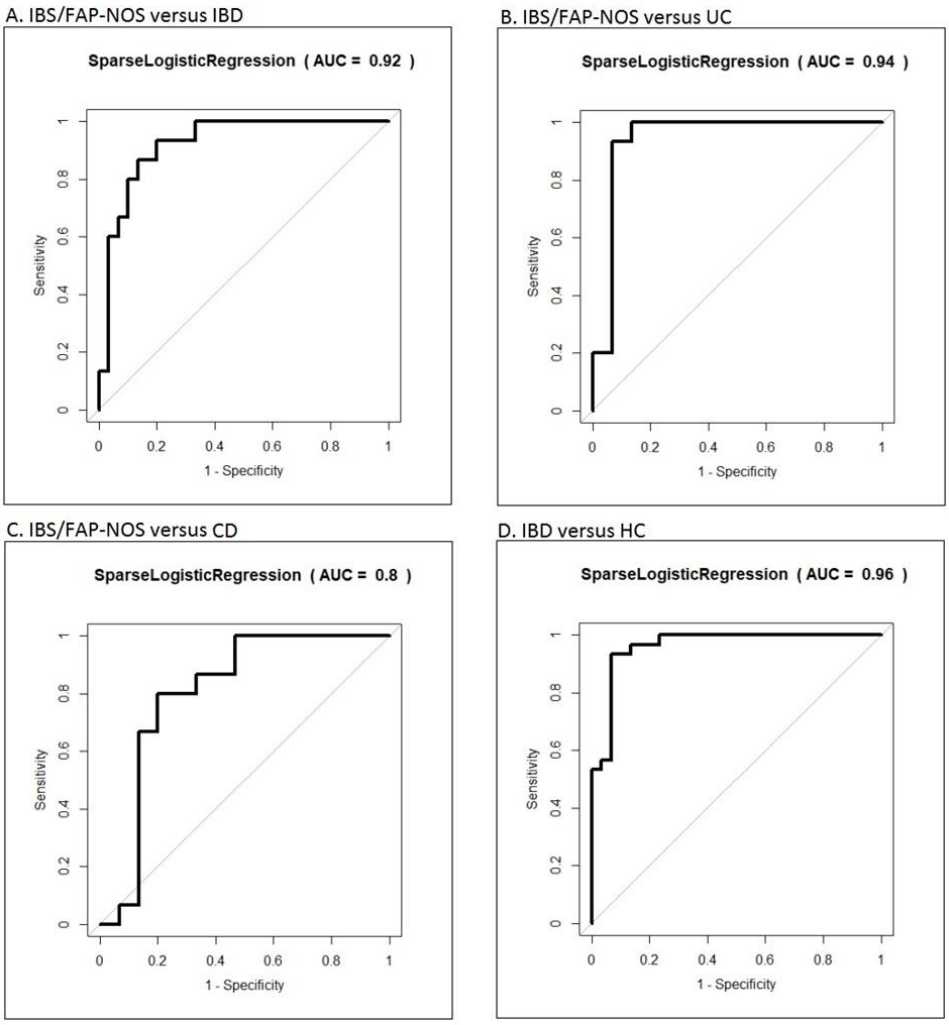


Figure 6.1. ROC for IBS/FAP-NOS vs IBD, UC, and CD and IBD vs HC. AUC are reported for the SLR analysis.

IBS/FAP-NOS vs HC

Children diagnosed with IBS/FAP could not be discriminated from HCs (AUC, 0.59; 95% CI, 0.41–0.77; P = 0.6, 0.63, 0.45, 0.76, 0.1667, sensitivity, specificity, PPV, NPV, respectively) (Table 6.2, Figure 6.1; Appendix Tables 6.1a-d).

IBD vs HC

Patients with IBD could be distinguished from HCs (AUC, 0.96; 95% CI, 0.9–1; P = 0.93, 0.97, 0.97, 0.94, 0.0000000003962, sensitivity, specificity, PPV, NVP, respectively) (Table 6.2, Figure 6.1; Appendix Tables 6.1a-d). Both IBD subtypes UC and CD could each be differentiated from HCs (Table 6.2; Appendix Tables 6.1a-d). Differentiation between CD and UC was not possible based on faecal VOC profiles (AUC, 0.67; 95% CI, 0.47–0.88; P =

0.6, 0.8, 0.75, 0.67, 0.05799, sensitivity, specificity, PPV, NPV, respectively) (Table 6.2; Appendix Tables 6.1a-d).

IBS vs FAP

Patient with IBS could not be discriminated from patients with FAP-NOS (AUC, 0.76; 95% CI, 0.44–1; P = 1, 0.6, 0.83, 1, 0.9504, sensitivity, specificity, PPV, NPV, respectively) (Table 6.2; Appendix Tables 6.1a-d).

Duration of Sample Storage

Duration of storage of the collected faecal samples did not differ between IBS/FAP-NOS and HCs. IBD samples were stored for a significantly longer period compared with both other subgroups (medium in months: CD, 31.7; UC, 45.1; IBS/FAP, 0.6; HC, 1.4; P < 0.001).

6.3. Breath Analysis Using E-nose and Ion Mobility Technology to Diagnose Inflammatory Bowel Disease – a Pilot Study (Study 2)

The content of this section is based on a paper that has been published in peer-reviewed journals:

Tiele, Akira, Alfian Wicaksono, Jiten Kansara, Ramesh P. Arasaradnam, and James A. Covington. 2019. "Breath Analysis Using E-nose and Ion Mobility Technology to Diagnose Inflammatory Bowel Disease — A Pilot Study." Biosensors 9 (2): 2019. https://doi.org/10.3390/bios9020055.

6.3.1. Methods

Subjects

A total of 39 subjects were recruited for this pilot study, as part of the larger 'Famished' study. Ethical approval was obtained from the Warwickshire research ethics committee (IRAS ref: 18717). 30 patients had a histologically confirmed IBD (14 CD, 16 UC), as well as 9 healthy control volunteers. IBD patients were recruited from dedicated IBD clinics at University Hospitals Coventry and Warwickshire (UHCW), UK. Details of medication and disease activity were recorded and simple colitis activity index (SCAI) for UC and Harvey Bradshaw index (HBI) for CD were calculated at the time of recruitment. Healthy controls were volunteers who did not report any overt gastrointestinal symptoms and were not on

routine oral medication or recovering from any recent illnesses. An overview of the demographic data of IBD patients and healthy controls is shown in Table 6.3. The mean age of the IBD cohort was 49.7 years (standard deviation 17.5) and there were 18 males and 12 females.

As shown in Table 6.3, inflammation parameters such as CRP and FCP were recorded for the IBD cohort. FCP is a good indicator of inflammation in the bowel [38]. FCP levels between CD and UC patients have been shown to differ by over 55 ug/g (higher in those with UC) [39]. In our IBD cohort, mean FCP between CD and UC patients differ by almost 300 ug/g. Mean scores of 117 and 414 ug/g, respectively, indicate that the CD group is in remission and the UC group has active disease. A box plot of FCP scores for CD and UC are shown in Figure 6.2.

Table 6.3. Demographic data of IBD patients and healthy controls

Parameter	Crohn's Disease (n=14)	Ulcerative Colitis (n=16)	Healthy (n=9)
Age (mean, SD)	46.1 (15.0)	52.8 (19.4)	30.9 (11.5)
Gender ration M:F	7:7	11:5	4:5
Smoking habits	1 active smoker 7 ex-smokers	1 active smoker 8 ex-smokers	4 ex-smokers
Alcohol, units/week (mean, SD)	4.3 (10.0)	8.1 (10.4)	3.7 (4.9)
BMI (mean, SD)	26.1 (8.0)	27.3 (4.9)	27.5 (9.0)
Medication	5-ASA1:8 AZA2: 4 Anti-TNFa3: 6 PPI4: 4 Salbutamol inhaled: 1 Inhaled steroid: 2	5-ASA1: 13 AZA2: 3 Anti-TNFa3: 3 PPI4: 2 inhaled steroid: 1	NSAID5: 1 COCP6: 1
HBI/SCAI score (mean, range, SD)	4.5 (0-29; 4.8)	1.5 (0-6; 2.25)	NA
Disease extent	Ileal disease: 2 Colonic: 3 Ileo-colonic: 9	Pancolitis: 7 Proctitis: 4 Quiescent: 1 Distal colitis: 3 Right-sided: 1	NA
CRP, mg/L (mean, SD)	7.8 (15.2) 1 unknown	5.3 (4.0) 1 unknown	NA
FCP, ug/g (mean, SD)	116.9 (112.8) 3 unknown	414.1 (315.5) 7 unknown	NA

NA, not applicable; SD, standard deviation; FCP, Faecal calprotectin; CRP, C reactive protein; HBI, Harvey Bradshaw index; SCAI, simple colitis activity index

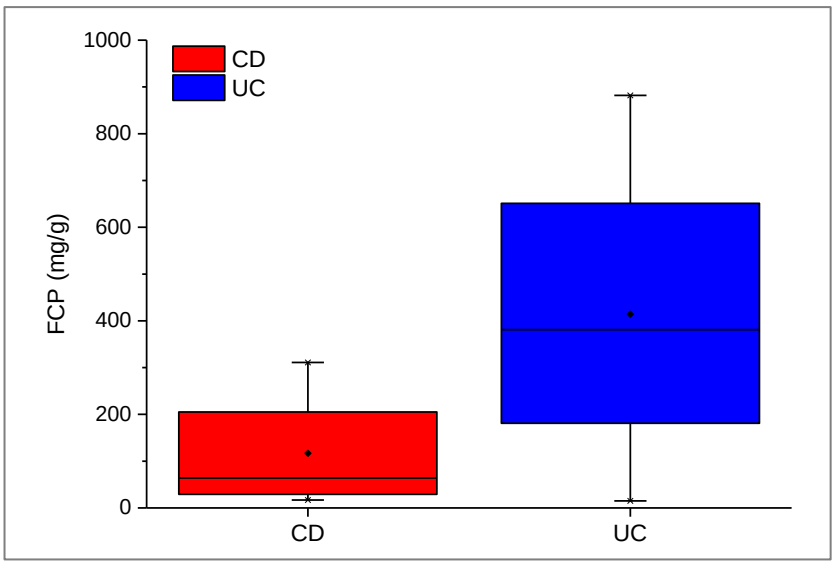


Figure 6.2. CD and UC boxplots for FCP scores

Electronic Nose (E-nose)

Warwick OLFaction (WOLF) was the E-nose system used in this section of work. The working principle of E-nose has been described in Chapter 3. Since the WOLF E-nose was designed and developed in-house, there is a high level of freedom allowing the instrument to be tailored to specific applications. A custom sample injection system was developed for this study, to enhance the capabilities of the WOLF E-nose to analyse breath samples.

Alveolar breath samples, for WOLF E-nose analysis, were collected using a commercially available breath sampling device, known as the Bio-VOC (Markes Int., Llantrisant, UK). Alveolar breath refers to the last portion (350 mL) of exhaled breath, expelled from within the lungs and the lower-airways, which have undergone gaseous exchange with the blood in the alveoli [40]. A healthy adult expires approximately 500 mL air with each breath, of which the first 150 mL consist of dead-space air (no transfer of oxygen) from the upper-air ways and nasopharynx [41]. Subjects were asked to perform a single slow vital capacity breath into a Bio-VOC unit, in order to trap the last 129 mL of exhaled breath [42]. Subjects were supplied with a disposable carboard mouthpiece, and the Bio-VOC was cleaned thoroughly using antibacterial, alcohol-free sanitary wipes after every sample. Collected breath samples were injected into the WOLF E-nose inlet port using a custom linear-actuator injection system. The injection system was manufactured using 5 mm acrylic sheets to support a 12 V, 200 mm linear actuator motor (JS-TGZ, Jianshun, Shenzhen, China). The

Bio-VOC was secured into the structure, as shown in Figure 6.3. Thereafter, the plunger was compressed automatically, at a constant rate, over a 30s injection period.

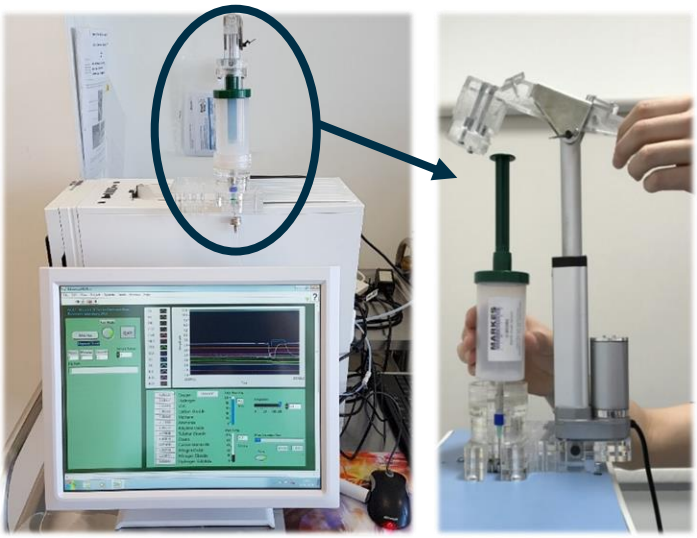


Figure 6.3. WOLF E-nose and custom Bio-VOC injection system

Gas Chromatography—Ion Mobility Spectrometry (GC-IMS)

A more recent, alternative approach to E-nose has been the use of portable GC-IMS analysers, which have demonstrated capabilities in medical diagnostics [43]. The BreathSpec GC-IMS (G.A.S., Dortmund, Germany) is a commercially available instrument that was used in this project. The working principle of GC-IMS has been described in Chapter 3. The BreathSpec is equipped with a SE54 mid-range polarity column. The IMS uses a drift tube where the time taken for molecules to traverse the tube against a buffer gas (in this case nitrogen) is measured. This buffer gas is generated using a Nitrostation 50LC (Leman Instruments, Geneva, Switzerland) with 99.999% purity. The gas slows down the ions resulting in larger ions being slowed more than smaller ones. Ions are then collected on the detector (Faraday plate), to deliver a time dependent signal that corresponds with ion mobility [44]. This technique can measure substances in the low parts-per-billion (ppb) range and delivers measurement results in less than 10 min.

The Bio-VOC was not required for G.A.S. BreathSpec GC-IMS analysis. Subjects were provided with a disposable plastic mouthpiece, which pushes into the mouthpiece holder/sample inlet and connects directly to the side-panel of the instrument. This sampling procedure also collects end-tidal breath, since only the last four seconds of exhaled breath are

collected for analysis [45]. The G.A.S. BreathSpec GC-IMS instrument is shown in Figure 6.4.

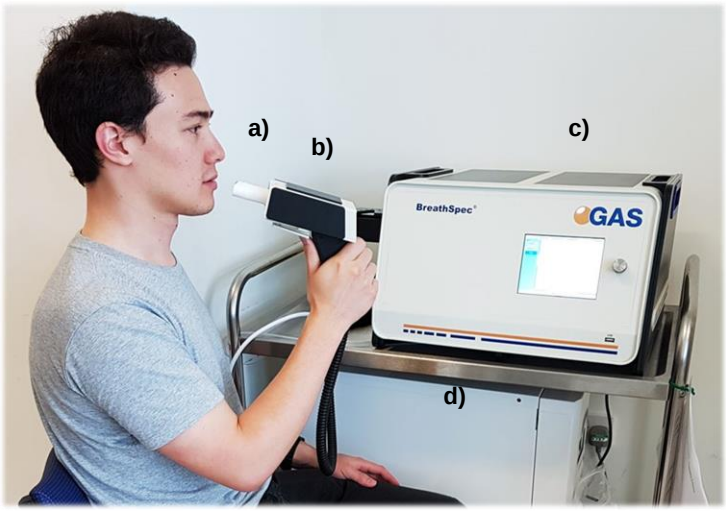


Figure 6.4. G.A.S Breathspec GC-IMS: (a) Plastic disposal mouthpiece; (b) mouthpiece holder; (c) GC-IMS instrument; (d) Nitrogen generator

Data Analysis

Both the WOLF E-nose and G.A.S. BreathSpec GC-IMS data were analysed using our standard pipeline which has been explained previously in Chapter 4. Class predictions and sensitivity/specificity calculations were performed using five classification algorithms, specifically: support vector machine (SVM), sparse logistic regression (SLR), Gaussian process, neural network (NN), and random forest (RF). To identify unknown compounds that contribute significantly to the classification analysis, GC-IMS files were loaded into the GCxIMS software and VOC identification is performed by simply clicking on the region of interest. The software then refers to the NIST database to generate a list of likely compound matches. A retention time range is provided, to indicate whether the suggested compound matches the expected retention time on the topographic map. Compounds with a close match in retention time and chemical structure were chosen as the identified compound.

Quality Assurance and Control

For quality assurance, the position and quality of the reactive ion peak (RIP) on the GC-IMS was regularly checked for signs of contamination. The RIP refers to the constant peak in the spectrum, which results from the carrier gas being always present in the measurement process. Moreover, samples were collected in the same setting, by the same operator,

throughout the entire study. This is an important factor to ensure consistent sampling procedures since the collection process is manually triggered by the operator, while the subject exhales through the mouthpiece. Furthermore, the GC-IMS instrument was normalised using a standard ketone mix (2-butanone, 2-pentanone, 2- hexanone, 2-heptanone, 2-octanone and 2-nonanone), to match the GC-IMS Library Search software with the equipped column. For WOLF E-nose calibration, the headspace gas from several chemical standards were tested (ketones, esters, alcohols, alkanes, and aromatics). These experiments revealed various relationships between sensor responses and concentrations. Furthermore, testing the different compounds individually produced responses from different sets of sensors, which confirms a degree of selectivity [46]. In addition to this, quality control procedures were implemented. This involved collecting regular room air samples to monitor changes in ambient air and identify possible exogenous VOCs, i.e., compounds which do not originate from within the body.

Confounding Factors

In a recent study, Blanchet *et al*. [47] explored factors that influence the VOC content in human breath and stated that any application of exhaled air for diagnostics should consider possible confounders. For this IBD study, the following confounders were considered: body mass index (BMI), smoking habits and gender. BMI categories include: underweight (<18.5 kg/m2), normal weight (18.5–24.9 kg/m2), overweight (25.0–29.9 kg/m2) and obese (>30.0 kg/m2) [48]. To simplify the analysis, underweight and normal weight were combined into a single category, as well as overweight and obese. Smokers can be broadly defined as individuals who have smoked at least 100 cigarettes in their lifetime [49]. Thus, never smokers are defined as adults who have never smoked or have smoked less than 100 cigarettes in their lifetime. These definitions were used to categorize smokers and non-smokers. Gender groups were divided into male and female—this factor is of particular importance, since it was not possible to attain a gender balanced UC group during recruitment. The confounding factor groups are shown in Table 6.4, with roughly evenly matched groups and various combinations of CD, UC and control subjects.

Table 6.4. Summary of confounding factor group

Factor	**Groups**	**CD**	**UC**	**Healthy**	**Total**
BMI	Under- & normal weight	9	5	6	20
BMI	Overweight & obese	5	11	3	19
Smoking	Smokers	8	9	4	21

Smoking	Never smokers	6	7	5	18
Gender	Male	7	11	4	22
Gender	Female	7	5	5	17

6.3.2. Results

Analysis results are presented as operating characteristic (ROC) curves. The associated area under curve (AUC) is a measure of how well parameters can distinguish between diagnostic groups. In our case, the groups were IBD vs controls, and CD vs UC. Generated ROC curves for G.A.S. BreathSpec GC-IMS and WOLF E-nose, IBD vs controls and CD vs UC, are shown in Figure 6.5 and Figure 6.6, respectively. The RF algorithm consistently performed best. The analysis results are summarized in Table 6.5 and Table 6.6.

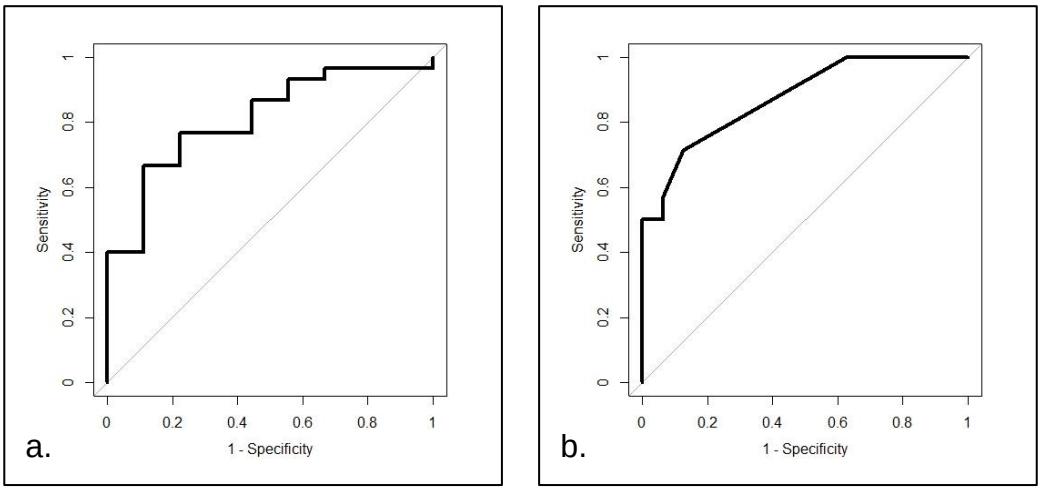


Figure 6.5. IBD analysis: G.A.S. Breathspec GC-IMS ROC curve; (a) IBD vs HC; (b) CD vs UC

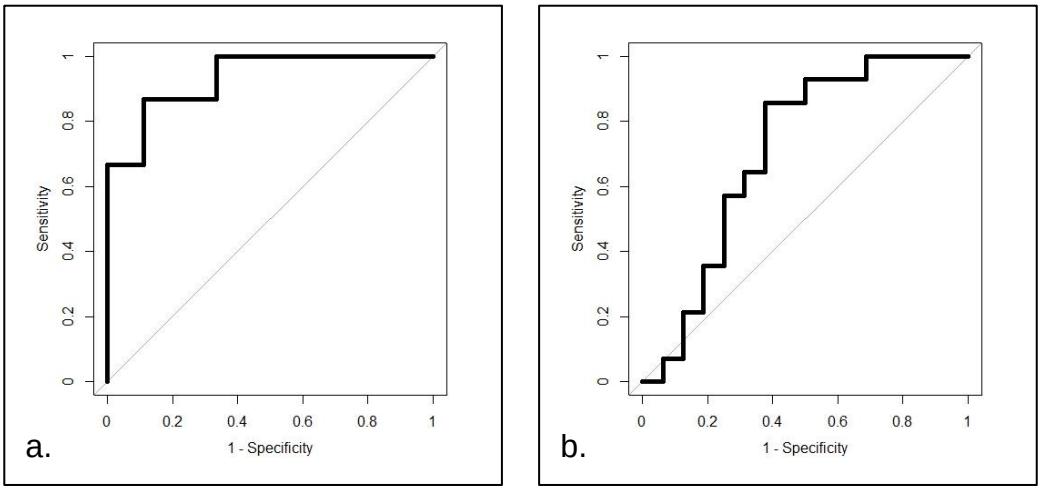


Figure 6.6. IBD analysis: WOLF E-nose ROC curves; (a) IBD vs HC; (b) CD vs UC

Table 6.5. IBD analysis G.A.S. Breathspec GC-IMS result

Test	**AUC ± 95%**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	***P*-value**
IBD vs HC	0.93 (0.85 – 1.00)	0.87	0.89	0.96	0.67	5.1×10^{-5}
CD vs UC	0.71 (0.51 – 0.91)	0.86	0.62	0.67	0.83	0.026

Table 6.6. IBD analysis WOLF E-nose result

Test	**AUC ± 95%**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	***P*-value**
IBD vs HC	0.81 (0.66 – 0.96)	0.67	0.89	0.95	0.44	0.0019
CD vs UC	0.88 (0.77 – 0.98)	0.71	0.88	0.83	0.78	0.0001

Chemical Identification

VOC analysis indicates that two compounds play a crucial role in distinguishing between IBD and controls. Chemical identification for the BreathSpec instrument, using the GC-IMS Library Search software, suggests that the best matches for the identified compounds include: butanoic acid (2-methyl-, propyl ester) and ethanoic acid (3-methyl-1-butyl ester).

Specific VOCs cannot be identified using the WOLF E-nose. However, in an attempt to identify the chemical groups, which contribute most to the separation between diagnostic groups, we have analysed the normalised average change in WOLF E-nose outputs, per group. A radar plot of the responses is shown in Figure 6.7.

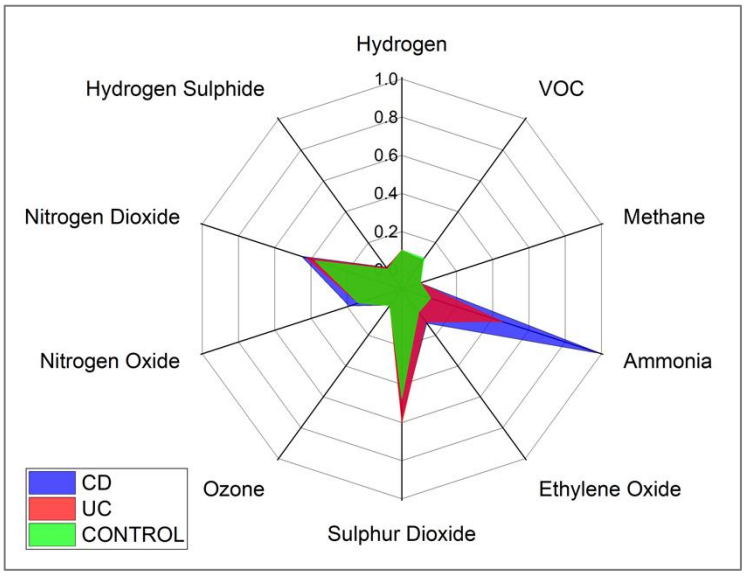


Figure 6.7. WOLF E-nose sensor response radar plot

Ammonia, sulphur dioxide (SO_2) and nitrogen dioxide (NO_2) sensors showed the greatest changes in sensor outputs and thereby contributed most to the separation between diagnostic groups. Significant changes were observed in ammonia by both UC and CD patients. Changes in NO_2 were marginally greater in CD over UC and controls, while UC is associated with increased sulphur dioxide. The other sensor outputs show small variations between groups, such as increased ethylene oxide in UC over controls.

Confounding Factors

The analysis previously conducted on IBD and control groups was repeated, using the same analytical techniques and algorithms, on the confounding factor groups, i.e., BMI: under- & normal weight vs overweight & obese; smoking: smokers vs never smokers; gender: male vs female. The analysis results are summarized in Table 6.7 and Table 6.8.

Table *6.7* and Table *6.8* demonstrate that the possible confounding factors of BMI, smoking and gender have insignificant influence on breath content. This is particularly true for BMI and smoking, as they achieve an AUC of around 50 for both technologies. Gender seems to have the most influence on breath content, with an AUC of around 60.

Table 6.7. Confounding factors: G.A.S. BreathSpec GC-IMS results

Factor	**AUC ± 95%**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	***P*-value**
BMI	0.47 (0.28 – 0.65)	0.63	0.50	0.55	0.59	0.6426
Smoking	0.58 (0.39 – 0.76)	0.57	0.67	0.67	0.57	0.2132
Gender	0.66 (0.48 – 0.84)	0.68	0.65	0.71	0.61	0.2298

Table 6.8. Confounding factors: WOLF E-nose results

Factor	**AUC ± 95%**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	***P*-value**
BMI	0.52 (0.33 – 0.71)	0.63	0.55	0.57	0.61	0.4284
Smoking	0.50 (0.32 – 0.69)	0.48	0.67	0.62	0.52	0.4944
Gender	0.61 (0.43 – 0.79)	0.73	0.47	0.64	0.57	0.5997

Feacal Calprotectin (FCP)

The demonstrated WOLF E-nose analysis for CD vs. UC was repeated using FCP as an additional feature, on a reduced dataset of 20 samples (11 CD, 9 UC) to match the availability of FCP scores. This analysis could not be repeated on the G.A.S. BreathSpec GC-IMS, because the features from this device are made up of clusters with numerous data points and are thus not directly compatible with single feature values, such as FCP.

The combined breath with FCP analysis was compared to breath without FCP, for the same dataset. The analysis results are summarised in Table 6.9. In addition to this, we investigated whether the combined analysis of breath with FCP could better distinguish between CD vs HC and UC vs HC, as shown in Table 6.10.

Table 6.9. CD vs UC (without and with FCP): WOLF E-nose results

Test	AUC ± 95%	Sensitivity	Specificity	PPV	NPV	*P*-value
Breath without FCP	0.85 (0.63 – 1.00)	1.00	0.67	0.79	1.00	0.0037
Breath with FCP	0.74 (0.5 – 0.98)	1.00	0.56	0.73	1.00	0.0311

Table 6.10. CD vs HC & UC vs HC (without and with FCP): WOLF E-nose results

FCP	Test	AUC ± 95%	Sensitivity	Specificity	PPV	NPV	*P*-value
Without FCP	CD vs HC	0.77 (0.54 – 0.99)	0.55	1.00	1.00	0.64	0.0232
	UC vs HC	0.72 (0.45 – 0.98)	0.89	0.67	0.73	0.86	0.0534
With FCP	CD vs HC	0.81 (0.61-1.00)	0.73	0.78	0.80	0.70	0.01
	UC vs HC	0.90 (0.75-1.00)	1.00	0.78	0.82	1.00	0.0023

6.4. Simultaneous Assessment of Urinary and Faecal Volatile Organic Compound Analysis in De Novo Paediatric IBD (Study 3)

The content of this section is based on a paper that has been published in peer-reviewed journals:

Manouni el Hassani, Sofia el, Sofie Bosch, Jesse P.M. Lemmen, Marina Brizzio Brentar, Ibrahim Ayada, Alfian N Wicaksono, James A Covington, Marc A Benninga, Nanne K.H. de Boer, and Tim G.J. de Meij. 2019. "Simultaneous Assessment of Urinary and Faecal Volatile Organic Compound Analysis in De Novo Paediatric IBD." Sensors 19 (20): 4496. https://doi.org/10.3390/s19204496.

6.4.1. Methods

Subjects

The current case-control pilot study was part of an ongoing cohort study, in which patients (aged 4–17 years) are included at the paediatric gastroenterology department outpatient clinic of a tertiary referral centre (Amsterdam University Medical Centre; locations: VU medical center (VUmc) and Amsterdammedical center (AMC)) [50,51,52]. For the current study, patients suspected of IBD in the period May 2017 to February 2019 were eligible to participate. Exclusion criteria were a proven bacterial or viral gastroenteritis during the month prior to inclusion, use of immunosuppressive therapy, antibiotic, or probiotic treatment in the last three months prior to inclusion, diagnosis with an immunocompromising disease, and insufficient ability to understand the Dutch language. Furthermore, participants who were not able to deliver both samples (i.e., faces and urine) were excluded. Participants, and in case of participants aged under 12 years, parents, were asked for informed consent after their first visit to the outpatient ward. The Medical Ethical Review Committee of the VUmc approved the study protocol (file number 2015.393, amendment number A2017.188).

All patients underwent diagnostic endoscopic evaluation because of suspected IBD based on clinical symptoms and/or biochemical abnormalities (e.g., elevated FCP). Participants were allocated to either the IBD or control group, based on the combination of endoscopy, histology, biochemical, and radiological findings, according to the currently applied international diagnostic criteria [5]. Disease localization and behaviour in IBD cases were assessed based on the Paris classification [34]. IBD cases were matched to controls based on age and gender.

Sample and Data Collection

Participants were asked to concurrently collect a faecal and urine sample in containers (Stuhlgefäb 10 mL, Frickenhausen, Germany), prior to bowel lavage. Participants were asked to store the samples in their freezer at home within one hour after collection and bring the samples to the hospital in cooled condition on their next regular appointment at the outpatient clinic. Samples were then directly stored at -20℃ until further handling. In addition to the sample collection, participants were asked to complete an online accessible and secured questionnaire (Castor EDC®) on dietary preferences, clinical symptoms, Bristol stool scale, bowel habits, extra-intestinal symptoms, residency, medical history, and medication use. Additional clinical information was collected from medical files, including IBD disease activity based on the physician global assessment (PGA) and laboratory values (CRP and FCP) [53,54].

Sample Preparation

Sample preparation was performed according to our standard operating procedure, previously described by Rouvroye *et al*. [22]. In short, a calibrated scale (Mettler Toledo, AT 261 Delta Range, Columbus, OH, USA) was used to weigh approximately 500 mg of faecal and frozen urine sample. Subsequently, the sample was transferred to a glass vial (20 mL, Frickenhausen, Germany), which was restored in a -20℃ freezer [50]. The faecal and urine samples were sent to the BioMedical Sensors Lab, School of Engineering, University of Warwick (Coventry, UK) on dry ice (-80℃) for VOC analyses.

Sample Analyses

Samples were analysed following the methods by Rouvroye *et al*. [22]. Samples were randomly analysed by means of GC-IMS (FlavourSpec®, G.A.S., Dortmund, Germany). The working principle of GC-IMS has been described in Chapter 3. The GC–IMS was connected to an automatic sampling system with a chiller allowing processing of a batch of 32 samples kept in cooled condition (4℃) until the start of the analyses to minimize sample degradation (PAL RSi, CTC Analytics AG, Zwingen, Switzerland). In the eight minutes before analysis, the samples' temperature was raised to 80℃. Then, a syringe transported the headspace from the vial into the injector port of the instrument and into the GC column. Nitrogen 99.9% (3.5 bar), served as a carrier gas at 40℃ for GC separation, and as drift gas for IMS at 45℃. GC flow rate was set at 20 mL/minute (34.175 kPa) for six minutes, while a 150 mL/min flow rate (0.364 kPa) was used for IMS.

Data Analysis

The Statistical Package for the Social Sciences (SPSS, IBM version 22.0) was used to perform the statistical analyses on demographics. Patient demographics and clinical characteristics were compared using a Fisher's exact test for dichotomous and ordinal data, and an independent t-test for parametric continuous data. In case of non-parametric continuous data, a Mann–Whitney U test was performed. A p-value < 0.05 was considered as a statistically significant difference. The GC-IMS data analysis was done using our data analysis pipeline as previously explained in Chapter 4. The 100 most discriminatory features were identified by Wilcoxon rank-sum test (undertaken within the fold) and then used to train five different classifiers: random forest (RF), gaussian process (GP), sparse logistic regression (SLR), support vector machine (SVM), and neural network (NN) algorithms.

6.4.2. Result

Baseline Characteristics

In total, ten IBD (five UC and five CD) and ten matched controls were included from the intention to diagnose cohort. The baseline characteristics are listed in Table 6.11. Faecal calprotectin was significantly higher (median 1208 g/g, p-value 0.009) in the IBD group compared to the control group (median 50 g/g). No differences were found for the remaining variables, as listed in Table 6.11.

Table 6.11. Baseline characteristics

	IBD (n=10)	Controls (n=10)	p-value
Sex *male* (n [%])	4 [40]	4 [40]	1.00
Age *years* (median [IQR])	15.0 [10.4-17.1]	14.2 [9.6-16.6]	0.71
BMI kg/m^2 (mean [SD])	18.9 [4.35]	21.3 [3.66]	0.20
Bristol stool scale (median [IQR])	6.0 [4.0-6.5]	6.0 [3.0-6.0]	0.97
FCP *µg/g* (median [IQR])	1208.0 [1023.5-3086.5]	50.0 [14.5-900]	0.01
CRP *mg/l* (median [IQR])	<2.5 [<2.5-39.5]	<2.5 [<2.5-4.35]	0.15
Sample weight feces *mg* (median [IQR])	500.5 [485.3-512.5]	502.5 [486.3-508.8]	0.88
Sample weight urine *mg* (median [IQR])	644.5 [532.8-726.0]	628.5 [592.0-663.3]	0.79
Storage time months (median[IQR])	6.5 [5.8-6.5]	10 [6.75-15.5]	0.14
Physician's Global Assessment			
Quiescent	0		
Mild	6		
Moderate	4		
Severe	0		
Crohn's Disease (n=5) localization			
Ileal (L1)	2		

Colonic (L2)	1	
Ileocolonic (L3)	2	
Proximal disease (L4)	1	
Crohn's Disease behaviour		
B1 (NSNP)	4	
B1p (NSNP +p)	1	
B2 (S)	0	
B2p (S+p)	0	
B3 (P)	0	
B3p (P+p)	0	
Ulcerative Colitis (n=5) localization		
Proctitis (E1)	0	
Left sided (E2)	2	
Extensive (E3)	3	

Values were obtained at inclusion. CD and UC localization and behaviour were determined using the Paris classification, based on findings during ileocolonoscopy and esophagogastroduodenoscopy and magnetic resonance enteroclysis before start of treatment [51]. **Abbreviations:** CRP, C-reactive protein; FCP, faecal calprotectin; IQR, interquartile range; SD, standard deviation; NSNP, non stricturing non-penetrating; S, stricturing; P, penetrating; p, perianal disease.

Table 6.12. Diagnoses Controls

Irritable bowel syndrome *n*	2
Functional abdominal pain *n*	1
Helicobacter pylori infection *n*	1
Juvenile polyp *n*	1
Multiple angiodysplasia *n*	1
IBD excluded without alternative diagnosis *n*	4

All diagnoses in the controls were established after diagnostic endoscopy. In four children, no diagnosis was established.

In Table 6.12, the diagnoses established after diagnostic endoscopy in the control group are listed. Three children were diagnosed with irritable bowel syndrome or functional abdominal pain, three had an alternative diagnosis, and in four children no alternative diagnosis was established. Three children in the IBD group used prescribed medication, opposed to six children in the control group (Table 6.13).

Table 6.13. Medication usage

IBD	**n**	**Controls**	**n**
Number of participants receiving medication	3		6

Type of medicine prescribed			
Ferrous fumarate	1	Macrogol	2
Ethinylestradiol/Desogestrel	1	Ibuprofen/Naproxen	2
Acrivastine	1	Formoterol/Beclamethason*	1
		Mebeverine	1
		Omeprazole	1
		Montelukast	1
		Ferrous fumarate	1
		Ondansetron	1
		Methylphenidate	1

All values were obtained at inclusion. In the control group, more medication usage was reported. Several participants were prescribed more than one type of medication. **Abbreviations:** n, number; IBD, inflammatory bowel disease; *, inhaler.

VOC Analysis

The discrimination of IBD from controls based on faecal VOC profiles, was statistically significant when training the algorithm using the neural network analysis (area under the curve (AUC) (95% confidence interval (CI)), p-value, sensitivity, specificity; 0.73 (0.47–0.99), 0.038, 0.70, 0.90), Figure 6.8A. For urinary VOC profiles, a similar accuracy was reached for the discrimination between IBD and controls using the sparse logistic regression classifier (0.78 (0.57–1), 0.028, 0.80, 0.70), Figure 6.8B. The results of the remaining statistical analyses are listed in Appendix Tables 6.2a and 6.2b. An example of GC–IMS output of faeces and urine, collected by the same participant, is displayed in Figure 6.9.

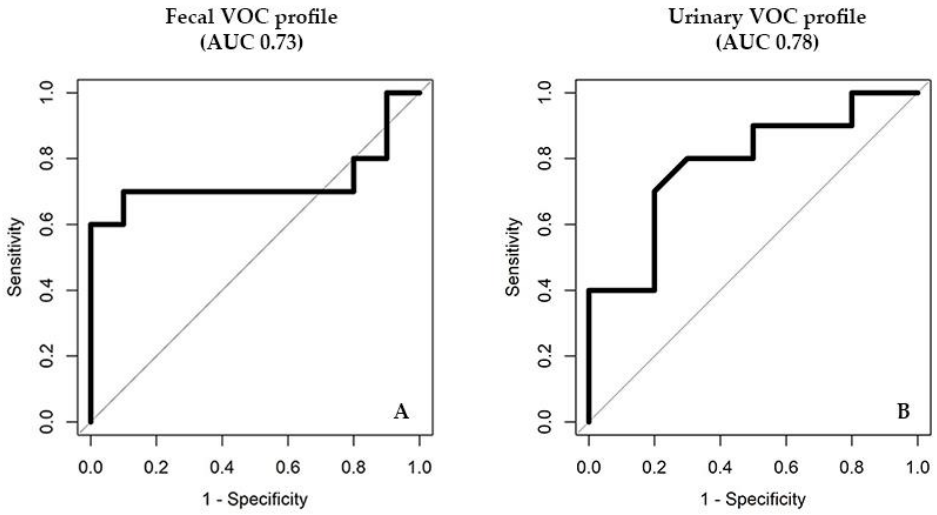


Figure 6.8. (A) Display the ROC curve for the faecal VOC profile, and (B) display the ROC curve for urinary VOC profile for the discrimination of IBD vs HC.

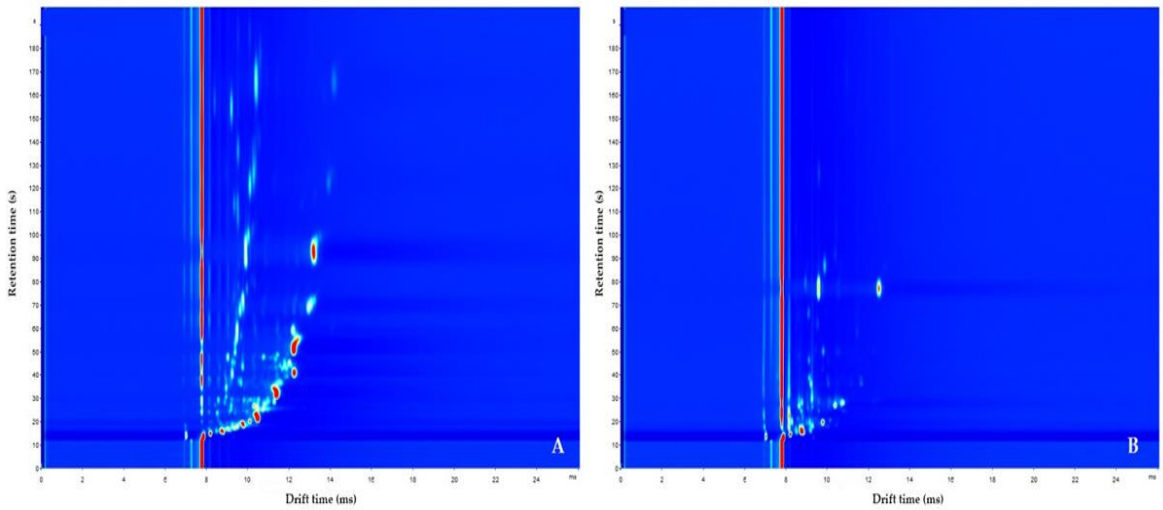


Figure 6.9. Examples of GC-IMS output. Here, the GC-IMS of feacal (A) and urinary (B) VOC profiles of one IBD patient are displayed

6.5. The faecal scent of inflammatory bowel disease detection and monitoring based on volatile organic compound analysis (Study 4)

The content of this section is based on a paper that has been published in peer-reviewed journals:

Bosch, Sofie, Dion S.J. Wintjens, Alfian Wicaksono, Johan Kuijvenhoven, René van der Hulst, Pieter Stokkers, Emma Daulton, *et al*. 2020. "The Faecal Scent of Inflammatory Bowel Disease: Detection and Monitoring Based on Volatile Organic Compound Analysis." Digestive and Liver Disease, no. xxxx. https://doi.org/10.1016/j.dld.2020.03.007.

6.5.1. Methods

This study was performed at the outpatient clinics of the Gastroenterology and Hepatology department in two tertiary referral hospitals (Amsterdam UMC, location VUmc, Amsterdam and Maastricht University Medical Centre (MUMC+) in Maastricht), and two district hospitals (OLVG West in Amsterdam and Spaarne Gasthuis (SG), location hoofddorp and Haarlem) all located In The Netherlands.

The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki (6th revision, 2008) as reflected in a prior approval by the Medical Ethical Review Committee (METc) of the Amsterdam UMC, location VUmc under file number 2016.135, by

the METc of the MUMC+ under file number NL24572.018.08, and by the local medical ethical committee of the OLVG West and Spaarne Gasthuis. Written informed consent was obtained from all study participants. Once sample collection was completed, all samples were shipped to the School of Engineering, University of Warwick (Coventry, UK) for VOC analysis.

Study participants

Inflammatory bowel disease patients

All patients aged 18 years or older with an established diagnosis of IBD based on clinical, endoscopic, histological and/or radiological criteria and with a scheduled consult at the outpatient clinic of one of the two tertiary referral hospitals were invited to participate in this study [55]. Patients were asked to collect a faecal sample and to complete a questionnaire, which included information on age, gender, BMI, smoking status, abdominal symptoms, medication use, comorbidity and clinical disease activity based on the Harvey Bradshaw Index (HBI) for CD patients and the Simple Clinical Colitis Activity Index (SCCAI) for UC patients [56][57]. Active disease was defined as FCP level of ≥250 mg/g. Remission was defined as FCP <100 mg/g combined with HBI <4 points of SCCAI <3 points. All IBD patients were included in the primary statistical analysis assessing the diagnostic potential of faecal VOCs to differentiate between IBD and HC. Only IBD patients with clearly defined disease activity based on FCP and HBI/SCCAI levels were included in the secondary analyses aiming to assess differences in faecal VOC pattern between active disease and remission. Demographic and clinical data (including Montreal classification and history of bowel surgery) were obtained from electronic patient files [58].

Healthy controls

All patients aged 18 years and order with a scheduled colonoscopy at the Amsterdam UMS, OLVG West and Spaarne Gasthuis were invited to participate in this study regardless of their endoscopy indication. They completed a questionnaire on demographics, smoking status, abdominal symptoms, bowel movement, dietary intake, comorbidity, and medication use. Patients without abnormalities detected during endoscopy were included in this study as healthy controls (except asymptomatic external haemorrhoids, asymptomatic diverticula and/or small anal fibromas). Mucosal biopsies were obtained to exclude microscopic inflammation and subjects were only included as HC if no histologic abnormalities were detected. Exclusion criteria included a history of bowel disease (e.g., celiac disease, IBD, CRC), failure to complete colonoscopy because of various reasons (e.g., inadequate bowel cleansing, pain) and/or collection of insufficient faecal sample mass to perform VOC analysis.

Sample Collection

Inflammatory bowel disease

Amsterdam University Medical Centres. Between February 2015 and November 2017, IBD patients were asked to collect two faecal samples (Stuhlgefäß 10ml, Frickenhausen, Germany) from the same bowel movement prior to the consult: one for FCP levels and one for VOC analysis. The FCP sample was sent to the hospital by mail. The sample for VOC analysis was the participant's own freezer within one hour following collection and transported to the hospital in cooled condition using ice packs and/or ice cubes on the day of their consultation. The samples were stored at -24°C directly upon arrival at the hospital.

Maastricht University Medical Centre. Between October 2009 and December 2010 patients were asked to collect stool from one bowel movement on the day of their consult and bring it fresh to the hospital. This stool sample was stress in the fridge (4°C) directly upon arrival at the hospital. From this bowel movement, two samples were prepared on the day of delivery. One for FCP measurements (using ELISA in Amsterdam UMC and FEIA in MUMC+) and one for research purpose. The second sample was stored in the freezer at -80°C.

Healthy Controls

Between February 2015 and November 2017, patients from the Amsterdam UMC, OLVG West and SG collected a faecal sample in a container (Stuhlgefäß 10ml, Frickenhausen, Germany) prior to bowel cleansing and endoscopic assessment. They were asked to store their sample in their own freezer within one hour after collection. These samples were transported to the hospital in cooled condition using ice packs and/or ice cubes on the day of their endoscopy. The samples were stored at -24°C directly upon arrival at the hospital.

Sample Preparation & Faecal Volatile Organic Compound Analysis

For the faecal VOC analysis, subsamples of 500mg per participant (with a maximum deviation 5%) were weighted on a calibrated scale (Mettler Toledo, AT 261 Delta Range, Ohio, United States). This was done whilst keeping the samples on dry ice to avoid thawing of the sample during preparation. Samples were then labelled and restored in glass vials (20ml headspace vial, Thames Restek, Saunderton, UK), in a -24°C freezer until further handling. As confirmed by our research team in a previous sampling method study on faecal VOC analysis, the 500mg weight was carefully chosen to provide an optimum amount of VOCs in the headspace of the sample [50]. The subsamples were shipped to the University of Warwick on dry ice for faecal VOC analysis. Faecal samples were analysed using gas chromatography

ion mobility spectrometry (GC-IMS, FlavourSpec, G.A.S., Dortmund, Germany). The working principle of GC-IMS has been described in Chapter 3.

Data Analysis

The GC-IMS data analysis was performed using our data analysis pipeline as previously described in Chapter 4. The data was split into three sets, 70% for training and validation and 30% as test set. A Wilcoxon rank-sum test was used to find the 20, 50 and 100 most discriminatory features and Sparse Logistic Regression (SLR), Random Forest (RF), Gaussian Process (GP), Support Vector Machine (SVM) and Neural Net (NN) classification were used to provide statistical results from the 30% test set, based on the training and validation sets.

6.5.2. Result

Baseline Characteristics

A total of 280 IBD patients (164 CD patients, 112 UC patients, 4 IBD-undetermined) were included. In total, 495 faecal IBD samples (292 CD, 197 UC, 6 IBD- U) were collected during the follow-up period of this study. Of these, 107 were active CD (CDa), 84 were CD in remission (CDr), 80 were active UC (UCa) and 63 were UC in remission (UCr) according to the previously mentioned criteria. The number of samples collected per individual varied as 159 patients provided one sample, 65 patients were sampled twice, 34 patients collected three samples, 10 patients collected four samples, 10 patients collected five samples, and two participants provided six and eight samples. Samples of IBD patients were compared to 227 HCs who all collected a single sample. Baseline demographics of all study participants are given in Table 6.14. There was no statistically significant difference in gender and smoking status between IBD patients and HC. The mean age of the IBD group was 46.1 (±29.8) compared to 60.6 (±11.8) for HC. The mean FCP levels for active disease and remission were 664.6 mg/g and 29.9 mg/g for CD and 1108.5 mg/g and 39.5 mg/g for UC, respectively (Table 6.15).

Table 6.14. Baseline characteristics

	CD (n=164)	UC (n=112)	IBD-U (n=4)	Healthy (n=227)
Female sex (n, %)	107 (65.2)	54 (48.2)	1 (25)	129 (56.8)
Age, year (mean ± SD)	45.3 (19-82)	51.1 (18-80)	49 (36-64)	60.6 ± 11.8
Smoking				
Current (n, %)	37 (22.6)	6 (5.4)	1 (25)	37 (16.3)
Past (n, %)	56 (34.1)	59 (52.7)	3 (75)	92 (40.5)

	Never (n, %)	67 (40.9)	43 (38.4)	0 (0)	98 (43.2)
Medication use at inclusion					
	Aminosalicylates (n, %)	22 (13.4)	54 (50.0)	1 (25)	N/A
	Corticosteroids (n,%)	26 (15.9)	14 (12.5)	0 (0)	N/A
	Immunosuppresives				
	Thiopurines (n, %)	58 (35.4)	28 (25.0)	2 (5)	N/A
	Methotrexate (n, %)	12 (7.3)	3 (2.7)	0 (0)	N/A
	Biologicals				
	Anti-TNF (n, %)	61 (37.2)	26 (23.2)	1 (25)	N/A
	Selective (n, %)	5 (3.0)	0 (0)	0 (0)	N/A
	Antibiotics 3 months prior to inclusion (n, %)	7 (4.3)	6 (5.4)	0 (0)	32 (14.1)
Surgery prior to inclusion (n, %)		72 (43.9)	15 (14.3)	0 (0)	NA
	Ileocecal resection (n, %)	57 (37.8)	2 (1.2)*	NA	NA
	(Partial) Colectomy (n, %)	31 (18.9)	15 (9.1)	NA	NA
	Small bowel resection (n, %)	5 (3.0)	0 (0)	NA	NA
Montreal Classification at inclusion					
	Age at diagnosis, year				
	A1 ≤ 16 (n, %)	20 (13.4)	4 (3.6)	1 (25)	NA
	A2 = 17 – 40 (n, %)	93 (56.7)	61 (54.5)	1 (25)	NA
	A3 ≥ 41 (n, %)	51 (31.1)	47 (42.0)	2 (50)	NA
	Localization Crohn's disease				
	L1 Ileal (n, %)		NA	NA	NA
	L2 colonic (n, %)		NA	NA	NA
	L3 ileocolic (n, %)		NA	NA	NA
	L4 proximal (n, %)		NA	NA	NA
	Behavior Crohn's disease**				
	B1 nonstricturing /nonpenetrating (n, %)	79 (48.1) 20(12.2)	NA	NA	NA
	B2 stricturing (n, %)	30 (18.3) 9(5.5)	NA	NA	NA
	B3 penetrating (n, %)	17 (10.4) 8(4.9)	NA	NA	NA
	Extent ulcerative colitis**				
	E1 Proctitis (n, %)	N/A	9 (8.0)	NA	NA
	E2 Left-sided (n, %)	N/A	53 (47.3)	NA	NA
	E3 Pancolitis (n, %)	N/A	49 (43.8)	NA	NA

Abbreviations: SD, standard deviation; CD, Crohn's Disease; UC, ulcerative colitis; IBD-U, inflammatory bowel disease undetermined; HC, healthy control; n, number of participants; NA, not

applicable. *Ileocecal resection in UC group was solely performed in combination with (partial) colectomy **Montreal classification of one participant missing.

Table 6.15. Inflammatory Bowel Disease Activity Scores

	CDa (n=107)	CDr (n=84)	UCa (n=80)	UCr (n=63)
FCP (mean ± s.d.)	664.6 ± 448.9	29.9 ± 24.4	1108.5 ± 952	39.5 ± 27.3
SCCAI (mean ± s.d.)	NA	NA	2.9 ± 0.38	0.8 ± 0.20
HBI (mean ± s.d.)	4.1 ± 0.42	1.1 ± 0.16	NA	NA
Medication use at time of sample collection				
Aminosalicylates (n, [%])	11 [10.3]	19 [22.6]	35 [43.8]	26 [41.3]
Corticosteroids (n, [%])	22 [20.6]	8 [8.3]	15 [18.8]	4 [6.3]
Thiopurines (n, [%])	34 [31.8]	32 [39.1]	21 [26.3]	33 [52.4]
Methotrexate (n, [%])	7 [6.5]	3 [3.6]	2 [2.5]	0 [0]
Anti-TNF (n, [%])	46 [43.0]	52 [61.9]	38 [47.5]	26 [41.3]
Selective biologicals (n, [%])	2 [1.9]	0 [0]	1 [1.3]	0 [0]
Antibiotics (n, [%])	1 [0.9]	1 [1.2]	4 [5.0]	1 [1.6]

Abbreviations: CDa, active Crohn's disease; CDr, Crohn's disease in remission; UCa, active ulcerative colitis; UCr, ulcerative colitis in remission; FCP, faecal calprotectin; SCCAI, simple clinical colitis activity index; HBI, Harvey Bradshaw Index. Active disease was defined as an FCP level of ≥250 mg/g, remission was defined as an FCP level of <100 combined with a HBI <4 points or SCCAI <3 points.

Faecal Volatile Organic Compound Analysis

The results of the VOC analysis by means of GC-IMS are shown in Table 6.16. For every comparison, the results from the Sparse Logistic Regression classification based on the 100 most discriminative features are presented. A complete overview of data generated using all five classifiers based on the 20, 50 and 100 most discriminative features are given in Appendix Table 6.3a-c. In addition, in Appendix Table 6.3d, the results of our post-hoc analysis comparing samples only collected in the similar time period are shown.

Table 6.16. Differences in Faecal Volatile Organic Compound Patterns

	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
Inflammatory bowel disease versus healthy controls						
IBD versus HC	0.96 (0.92 - 0.99)	0.97	0.92	0.98	0.87	<0.0001
CD versus HC	0.97 (0.95 - 1)	0.94	0.96	0.97	0.91	<0.0001
CDa versus HC	0.96 (0.94 - 0.99)	1	0.92	0.74	1	<0.0001
CDr versus HC	0.95 (0.93 - 0.98)	1	0.90	0.67	1	<0.0001
UC versus HC	0.95 (0.91 - 0.99)	1	0.91	0.88	1	<0.0001
UCa versus HC	0.96 (0.94 - 0.99)	1	0.92	0.74	1	<0.0001

UCr versus HC	0.95 (0.93 - 0.98)	1	0.88	0.52	1	<0.0001
Active disease versus remission						
IBDa versus IBDr	0.59 (0.51 - 0.67)	0.21	0.96	0.90	0.39	0.019
CDa versus CDr	0.52 (0.39 - 0.65)	0.72	0.43	0.71	0.45	0.645
UCa versus UCr	0.63 (0.44 - 0.82)	0.67	0.57	0.79	0.42	0.082
Crohn's disease versus ulcerative colitis						
CD versus UC	0.55 (0.50 - 0.60)	0.17	0.96	0.90	0.36	0.031
CDr versus UCr	0.52 (0.39 - 0.65)	0.95	0.18	0.67	0.67	0.607
CDa versus UCa	0.56 (0.37 - 0.75)	0.74	0.43	0.76	0.40	0.744

All the results of the VOC analysis are obtained using Sparse Logistic Regression classification based on the 100 most discriminative features. Sensitivities, specificities, p-values and AUC are reported for the respective optimum cut-off points. Abbreviations: IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; a, active disease state; r, remission; HC, healthy controls; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value. The 95% CI is in brackets.

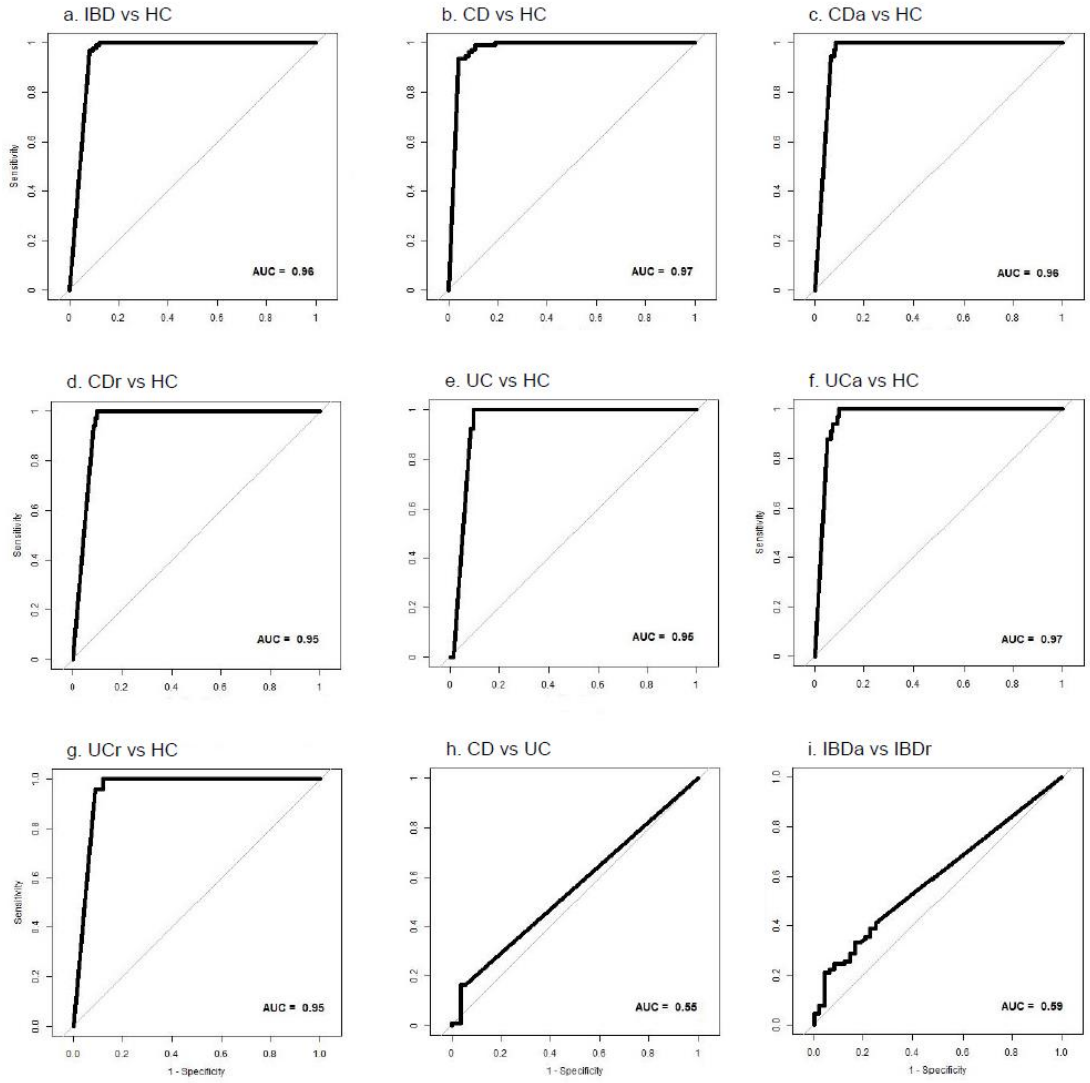


Figure 6.10. ROC curves for the differentiation between study groups based on SLR classification using the 100 most discriminative features.

Inflammatory Bowel Disease Versus Healthy Controls

IBD patients were discriminated from HC with a high diagnostic accuracy (AUC ±95%CI, sensitivity, specificity, PPV, NPV, P-values; 0.96 (0.92 –0.99), 0.97, 0.92, 0.98, 0.87, <0.0001) (Table 6.16, Appendix Table 6.3a-d). Likewise, high diagnostic accuracy was found for detection of CD during active state and remission (0.96 (0.94 –0.99), 1, 0.92, 0.74, 1, <0.0001 for active CD; 0.95 (0.93 –0.98), 1, 0.90, 0.67, 1, <0.0001 for CD in remission) (Table 6.16, Appendix Table 6.3a-d) . This was similar for the detection of UC both during active state and remission (0.96 (0.94–0.99), 1, 0.92, 0.74, 1, <0.0001 for UCa; 0.95 (0.93–0.98), 1, 0.88, 0.52, 1, <0.0001 for UCr) (Table 6.16, Appendix Table 6.3a-d). Corresponding Receiver Operating Characteristic (ROC) curves are visualized in Figure 6.10a-d.

Crohn's Disease Versus Ulcerative Colitis

Faecal VOC patterns of CD and UC differed significantly, though the diagnostic accuracy was very low (0.55 (0.50-0.60), 0.17, 0.96, 0.90, 0.36, 0.03) (Table 6.16, Figure 6.10h, Appendix Table 6.3a-d). Furthermore, there was no difference between UC and CD when comparing active disease and remission subgroups separately (Table 6.16).

Active Disease Versus Remission

There was a slight significant difference in faecal VOC patterns between active IBD (UC and CD combined) and remission (0.59 (0.51-0.67), 0.21, 0.96, 0.90, 0.39, 0.019) (Figure 6.10). However, when comparing active and remission state of CD and UC subgroups separately, this significance was not found (CD active vs CD in remission 0.52 (0.39-0.65), 0.72, 0.43, 0.71, 0.45, 0.645; UC active vs UC in remission 0.63(0.44-0.82), 0.67, 0.57, 0.79, 0.42, 0.08) (Table 6.16).

6.6. Discussion

Based on the result of all these studies, we have successfully shown the potential of faecal, urinary, and breath VOC analysis in detecting IBD from Healthy controls in adults and paediatric patients using E-nose, FAIMS, and GC-IMS approaches. A short summary of the 4 studies is shown in Table 6.17.

Table 6.17. Summary of Four Studies

	Original paper's title (date of published)	Sample type	Sample size	Instrument
Study 1	Differentiation Between Paediatric Irritable bowel syndrome and Inflammatory Bowel Disease Based on Faecal Scent: Proof of Principle Study (May 2018)	Faecal	CD = 15 UC =15 IBS/FAP-NOS=15 HC = 30	FAIMS
Study 2	Breath Analysis Using E-nose and Ion Mobility Technology to Diagnose Inflammatory Bowel Disease - A pilot study (April 2019)	Breath	CD = 14 UC = 16 HC = 16	E-nose and GC-IMS
Study 3	Simultaneous Assessment of Urinary and Faecal Volatile Organic Compound Analysis in De Novo Paediatric IBD (October 2019)	Faecal and Urine	IBD = 10 HC = 10	GC-IMS
Study 4	The faecal scent of inflammatory bowel disease: Detection and monitoring based on volatile organic compound analysis. (March 2020)	Faecal	CD = 164 UC = 112 IBD-U = 4 HC = 227	GC-IMS

Out of four studies, three studies investigated the potential of faecal VOCs in IBD detection using FAIMS and GC-IMS. When comparing classifier results, Gaussian Process (GP) performed best across these studies in discriminating IBD patients from healthy controls by faecal VOC profile analysis. The results of IBD vs HC analysis using faecal VOCs are summarised in Table 6.18.

Table 6.18. Summarized of GP Classifier Performance When Detecting IBD from HC in Faecal Samples.

	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
Study 1	0.95 (0.88 - 1)	0.93	0.93	0.93	0.93	<0.001
Study 3	0.69 (0.44 - 0.94)	0.9	0.5	0.64	0.83	0.435
Study 4	0.97 (0.95 - 0.99)	0.98	0.93	0.98	0.93	<0.001

Study 1 (FAIMS) and study 4 (GCIMS) showed high separation, whereas study 3 was only able to achieve AUC (95% CI) of 0.69 (0.44 - 0.94). However, this may be because study 3 included a small sample size, as shown previously in Table 6.17. The intention of study 3 was to compare the patient urinary VOC profile to theirs faecal VOC profile for detecting IBD. However, even with a small sample size, some differences were observed. When investigating UC, CD, and HC separately, both FAIMS and GCIMS were able to show significant differences in comparing UC vs HC and CD vs HC, but not for CD vs UC as shown in Table 6.19. Due to smaller sample size, we were unable to perform sub-analysis on study 3.

Table 6.19. The Summary of FAIMS and GC-IMS Performance in Differentiating UC vs HC, CD vs HC, and UC vs CD from Patient's Faecal VOC Profile Using Gaussian Process Classifier

		AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
Study 1 (FAIMS)							
	UC vs HC	0.98 (0.94 - 1)	0.93	0.93	0.88	0.97	<0.001
	CD vs HC	0.93 (0.86 - 1)	0.87	0.90	0.81	0.93	<0.001
	UC vs CD	0.46 (0.24-0.67)	1	0.067	0.52	1	0.6587
Study 4 (GCIMS)							
	UC vs HC	0.97 (0.93 - 1)	0.96	0.95	0.93	0.97	0.0001
	CD vs HC	0.98 (0.95 - 1)	0.96	0.93	0.95	0.95	0.0001
	UC vs CD	0.54 (0.49 - 0.60)	0.17	0.92	0.83	0.35	0.078

The potential of faecal VOC analysis in paediatric and adult IBD have been previously studied. Ahmed et. al analysed the faecal VOC samples of 117 CD, 100 UC, and 109 HC patients in the adult population using Gas Chromatography Mass Spectrometry (GC-MS) [59]. Although AUC values were not provided, CD and HC could be distinctly separated based on three unique metabolites. In contrast to our findings, study results of Ahmed et. al did not allow discrimination between UC patients compared to HC. Other studies have assessed the diagnostic potential of faecal VOCs for IBD using pattern-based techniques in paediatric

cohorts, such as E-nose instruments (CD 29, UC 26, HC 28) and FAIMS (23 CD, 13 UC, 24 HC) [60,51]. In the study using E-nose technology, similar accuracies to the current study were demonstrated for the detection of IBD. In addition, using an E-nose allowed clear separation of UC and CD phenotypes. Using FAIMS, CD patients were separated from HC with high accuracy and moderate separation was found for the discrimination of UC compared to both HC and CD samples (AUC of 0.74 and 0.67, respectively). The results in our studies also support previous finding that IBD can be differentiated from HC based on faecal VOC analysis with a high accuracy.

Differentiation between IBD and IBS can be difficult, especially in children, as biochemical diagnostic biomarkers are yet not available, the diagnostic relies on the symptom-based ROME IV criteria [32]. In study 1, the paediatric IBD/FAP-NOS group was included. We observed that faecal VOC profiles could differentiate between paediatric IBS/FAP-NOS and IBD patients with high accuracy, but not from HC. Studies on the potential of faecal VOC profiling to discriminate against paediatric IBS/FAP-NOS from IBD have not yet been performed. Ahmed *et al*. compared the faecal VOC profiles of 30 adult diarrhoea-predominant IBS (IBS-D) patients, with 62 active CD, 48 active UC, and 109 healthy subjects using gas chromatography–mass spectrometry (GC-MS) [61]. In that study, IBS-D could be discriminated from IBD based on 44 significantly different levels of metabolites. These metabolites were used to construct a discriminatory model with high diagnostic accuracy (AUC IBS-D vs CD, 0.97; AUC IBS-D vs UC, 0.96; $P = 0.001$). This diagnostic accuracy is comparable to that observed in our study. In addition, in the study by Ahmed and colleagues, significantly increased levels of 48 faecal metabolites were identified in adult IBS-D patients compared with HCs with AUC of 0.92 (p-value < 0.05). In the present study, however, VOC profiles of IBS/FAP-NOS were not significantly different compared with VOC profiles of HCs. This difference could possibly be explained by our relatively small sample size. Another explanation could be our heterogeneous IBS/FAP-NOS group, in which subjects experienced a variety of symptoms (diarrhoea, abdominal pain, bloating, constipation), whereas Ahmed *et al*. solely included patients with diarrhoea-predominant IBS type. However, we observed no significant differences in VOC profiles between the 2 subgroups IBS and FAP-NOS. In addition, the diagnostic accuracy could differ due to the fact that GC-MS and FAIMS analyse metabolite signals based on different techniques. However, as the diagnostic accuracy to differentiate between IBS/FAP-NOS and IBD is highly similar between these studies, we believe this had minimal influence on our study outcomes.

Active IBD was discriminated from remission in study 4 with a very weak accuracy and there was no difference between the active and inactive subgroups of CD and UC. The existing

literature of the differentiation between active and inactive IBD based on VOC profiles in adults is both scarce and contradictory. There is only one research on the assessment of IBD activity based on faecal VOC profile. Inactive CD was significantly separated from active CD based on faecal VOC profiles, whereas VOC profiles of active and inactive UC were similar [59]. Using a breath sample, separate research showed similar accuracy when comparing the breath VOC profile of 135 CD in remission with 140 active CD using GC-MS with AUC 0.98 [62]. In addition, a high diagnostic accuracy to differentiate between 62 active and 70 inactive UC breath VOC profiles was found by the same research group with AUC of 0.94 [63]. Using urine sample, another research group, also showed it was possible to discriminate between active IBD (24 CD, 24 UC) and remission (4 CD, 4 UC) using FAIMS, with moderate accuracies (AUC CD 0.66, UC 0.74 [64]. These results contradict with our result. However, this accuracy was based on different metabolite alternation. The reasons behind this can be explained as:

1. Alterations in the faecal VOC patterns may be explained by alteration in metabolic processes, like the secretion of inflammatory end products in the colon or alteration in dietary intake, by microbial dysbiosis or a combination of these. In a recent study on canine olfaction, in vitro breast cancer and colon cancer cells were grown and it was observed that dogs were able to differentiate between the metabolic waste retrieved from cancer cells and benign cells, but not between the cell waste of breast and colon cancer, implying that both cancers may share a common smell print [65]. The same might apply for inflammatory diseases like CD and UC of which it may be hypothesized that the VOC patterns of CD and UC patients are based on shared (metabolic) reaction, explaining the similarities in VOC patterns observed in the current study.

2. Another reason may be explained by the gut microbiota as one of the main sources of faecal VOC. The faecal microbiota of healthy individuals consist of over 400 different species, a large inter-individual diversity in the faecal microbiota composition of healthy individuals [66]. Nonetheless, the microbiome of an individual is remarkably stable, suggesting the presence of a core microbial community, which is dependent on host factors [67,68]. In multiple studies, this microbial stability has been found greater in healthy individuals compared to IBD patients of which the microbial composition is defined by more deviation over time and a decrease in diversity especially in abundance of *Firmicutes* in CD patients and decrease in butyrate-producing bacteria in UC patients (*Faecalibacterium prausnitzii, Roseburia hominis*) [69,70,71,72,73,74]. This could explain the ability to clearly discriminate between IBD patients and HC based on faecal VOC profiles in our study. Remarkably, this variability of microbial composition in IBD

patients does not correlate with disease activity. Fluctuations in microbiota composition have been observed both during exacerbation and clinical remission, which has hampered the identification of microbial changes related to the presence of flare-ups [69,70,75].

3. Lastly, this variation may be caused by the use of different techniques to analyse VOC profiles. The use of pattern recognition in our study provides a fast and relatively cheap manner to analyse faecal VOCs, which is highly adequate for clinical implementation. However, this technique is unable to identify specific metabolites. Differences between IBD and HC based on faecal VOCs have previously been demonstrated due to an altered composition of esters, short chain fatty acids (SCFAs) and cyclo-hexanecarboxylic acid, of which the first group is believed to be associated with bacterial dysbiosis [76]. The differences between active disease state and remission in faecal as well as breath VOC profiles originate from a different group of metabolites, mainly aldehydes and ketones [62,63,77]. These metabolites play a role in inflammatory processes as they are the metabolic products of tissue damage and oxidative stress and may therefore be the result of a more general host-response to inflammation rather than an IBD specific metabolic alteration. It is possible that the GC-IMS column chosen was sensitive to a range of metabolites for differentiating IBD from HC, but not to the metabolites produced in inflammatory processes.

Urinary VOC as the second type of sample in this chapter was only investigated using GCIMS in study 3 in which the potential of urinary VOC profiles is assessed for the first time in paediatric IBD. In adult IBD, urinary VOC analysis allowed for discrimination between IBD and HC by means of an E-nose device (Fox 4000; AlphaMOS, Toulouse, France) and FAIMS [64]. The result of study 3 and Arasaradnam et. al were compared in Table 6.20.

Table 6.20. The Result of Urinary VOC Analysis Using E-nose, FAIMS, and GC-IMS to Detect IBD from HC

	AUC	p-value
Arasaradnam et.al		
E-nose (FOX 4000)	0.88	<0.001
FAIMS	0.75	<0.001
Study 3		
GCIMS	0.78	0.028

The FAIMS analysis demonstrated similar diagnostic accuracy, while VOC profiles analysed by E-nose exceeded the accuracy as observed in the current study. The variation in diagnostic accuracies can be explained by several aspects. In the current study, a paediatric intention to diagnose cohort was established, whereas in the study by Arasaradnam *et al*., adult patients with an established IBD diagnosis which received treatment were included. Additionally, IBD cases were compared to healthy controls, rather than with symptomatic patients, in which it is expected that there is a larger difference between healthy controls and cases, than is seen between symptomatic non IBD patients compared to IBD cases. Furthermore, different analytical techniques were applied, Arasaradnam *et al*. used an E-nose device and FAIMS, whereas in the current study a novel technique, GC–IMS, was applied. Lastly, in the current study, a more diverse control group was included, which could possibly have led to a lower diagnostic accuracy.

The last sample type we investigated in this study were breath VOC profiles. Both E-nose and GCIMS were able to discriminate between adult IBD patients from HC by analysing their breath VOC profile regardless of disease activity as reflected by the FCP scores. Additionally, both technologies were able to provide some separation between UC and CD. The result indicates that GC-IMS is better suited towards distinguishing between IBD and HC, while E-nose can better separate between UC and CD. The sensor array deployed in the WOLF E-nose is focused more towards inorganic gas, which most likely accounted for the differences in diagnostic accuracy achieved by the employed technology. Ammonia, SO_2, and NO_2 sensors contributed significantly to the analysis of the WOLF E-nose. Ammonia has an established link to IBD breath [78,79,80], since it is one of the intermediaries generated from bacterial fermentation of proteins [81]. UC was associated with higher SO_2 levels in our study, which has been observed previously [82]. It has been suggested that residential exposures of SO_2 and NO_2 may increase the risk of early-onset CD and UC [83]. Increase in the aforementioned VOCs, butanoic acid and ethanoic acid, contributed significantly to the efficacy of our analysis for GC-IMS. These VOCs have been recently identified as important discriminatory volatile organic metabolites for IBD [61]. Short-chain fatty acids, such as butyric-, propionic- and acetic acids, are produced in the colon by fermentation of fibre. In particular, butanoic acids (also known as butyric acid) is a key component for health in the colon [84] and is the main energy substrate for colonocytes [85]. Butanoic acid has therefore been suggested to play an important role in the prevention and treatment of distal UC [86] and CD [87]. The variations in the identified compounds between IBD subjects and controls may be crucial for diagnostic purposes and need to be further investigated.

The analysis of the results for possible confounding factors in breath profile show that BMI and smoking habits have insignificant influence on breath content. In general, the effect of smoking is an obvious factor to influence breath. However, it is worth noting that 19 of the 21 "smoker" subjects consider themselves to be ex-smokers. It is therefore unsurprising that greater differentiation was not possible in this case. Gender showed a more significant influence on breath content than the other two factors (AUC: 0.66, sensitivity: 0.68, specificity: 0.65). While IBD generally affects men and women equally, some studies from North America show that UC is more common in men than women [88]. The unbalanced gender counts in the UC group (11M: 4F) could therefore be responsible for strengthening the separation between males and females for the confounding factor analysis. In addition to this, gender affects metabolism that can lead to differences in breath content [47]. However, this factor was not significant enough to create two distinct groups or undermine the IBD related analysis. Age was not considered in the confounding analysis; however, this factor is unknown to have limited effect on breath content, with some studies showing not statistically significant associations between age and common breath gas metabolites [89]. Moreover, unlike many other diseases, IBD can occur at any age (most likely between 15-35). Age is therefore unlikely to have had a significant effect on the conducted analysis. Lastly, the effect of medication could not be considered factor analysis, due to the number and variety of medications and treatment prescribed to each IBD subject. It is possible that the strong result distinguishing between IBD and CD, as well as CD vs UC, could be related to the effect of medication; however, the same class of drugs was proportionally present in both IBD groups, so this factor is less likely to be a confounder.

Results from the analysis combining breath analysis with FCP caused specificity, PPV, and AUC to reduce, compared to those without FCP. While the differences in mean FCP score between CD and UC is almost 300 ug/g (414.1-116.9 ug/g), this feature is prone to misclassification because all CD scores fall within the lower range of UC scores. Thus, including FCP with breath analysis does not improve the distinction between CD vs UC within this study. However, as shown in Table 9. FCP did improve separation of CD vs HC and UC vs HC. This is likely due to the significantly higher FCP score in both CD and UC, when compared with the normal reference range of a healthy individual (<50 ug/g) [90]. Nonetheless, this specific application of FCP has little practical or clinical value, since the same conclusions can be derived using FCP scores alone. Moreover, there would be added costs (£18 per test) to conduct FCP tests, alongside breath analysis. These additional costs cannot be justified without a significant improvement in diagnostic performance in distinguishing between CD and UC.

It has been shown that E-nose technology has some capability in identifying the chemical groups (based on sensor responses), which contribute most to the separation between diagnostic groups. E-nose also has a high level of freedom, allowing the instrument to be tailored to specific applications. However, other technologies, such as GC-IMS are associated with more advanced/sophisticated chemical identification methods. This can be conducted by matching the data to GC-IMS library search software, such as the NIST library. On the other hand, chemical identification in FAIMS technology is not possible, making it not suitable for biomarker discovery. Even though E-nose is the cheapest among these three instruments, FAIMS and GC-IMS provide the highest dimensionality data that can further enhance data analysis. Both technologies also measure the differences in the mobility of ionized VOC molecules in high-electric fields resulting in higher sensitivity and minimal drift.

Table 6.21. Discrimination Performance of IBD vs HC Based on Sample Type

Sample type	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
Stool	0.96 (0.92 – 0.99)	0.97	0.92	0.98	0.87	<0.0001
Urine	0.78 (0.57 - 1)	0.80	0.70	NA	NA	0.028
Breath	0.93 (0.85 – 1.00)	0.87	0.89	0.96	0.67	5.1 x 10-5

Nonetheless, VOC profile analysis provides a promising, non-invasive method in detecting IBD patients. Table 6.21 compare the best discrimination performance among three sample types in this thesis when comparing IBD vs HC. It showed that faecal VOC analysis gives the highest accuracy among other sample types. Stool samples are routinely collected to inform the management of many gastrointestinal diseases and infections [23]. However, patient compliance rarely exceeds 60%, due to embarrassment or concerns about results [91]. Additionally, collecting stool samples during disease activity (e.g., flare-up) is potentially uncomfortable for the patient. On the other hand, urine and breath samples can be provided anytime with virtually unlimited numbers of samples collection. However, there is still a lot of work that is needed in terms of standardisation for breath and urinary sampling compared to faecal.

Future research should compare faecal, urine, and breath VOC profiles from the same patient to get a deeper and better understanding of how metabolic processes play a role in IBD detection. Not only IBS/FAP-NOS, a broader range of gastrointestinal disorders should also be included into the future case-control research to be able to increase the accuracy and

sensitivity of this technique in detecting IBD. Since the pattern recognition model is a crucial part in VOC profile analysis for disease detection, future study should perform proper sample size calculations to make sure the model can pick the chemical fingerprint. Those, including more samples, are necessary, especially in urine and breath. VOC sampling conditions and data analysis processes should be also standardised to be able to make clear comparisons between studies. Finally, more studies should focus on monitoring of remission and active states of IBD as this would have significant clinical value. The IBD state monitoring will enable the clinician to predict the likelihood of a patient to relapse and provide an early treatment to maintain patient remission state and avoid irreversible bowel damage, such as fistulas and strictures that may lead to surgery.

6.7. Conclusion

In conclusion, we have shown that paediatric and adult patients with IBD can be distinguished from HC with a high diagnostic accuracy by faecal, urine and breath VOC analysis using E-nose, FAIMS and GC-IMS methods. The accuracy of faecal VOCs in detecting IBD was comparable to breath. The separation of UC and CD was only observed in the breath study. However, as faecal VOC patterns did not allow for differentiation between disease activity states, its potential for monitoring intra-individual course of IBD may be limited and should be assessed in a future study, enrolling an endoscopy-controlled cohort. When comparing faecal VOC profiles of IBD patients to other similar symptomatic diseases, such as IBS/FAP-NOS, high diagnostic accuracy was also observed. We also conclude that GC-IMS performed the best overall in detecting IBD. Compound analysis has identified two breath VOCs, which were likely to have direct link to IBD: butanoic acid and ethanoic acid. Analysis of possible confounding factors indicate that BMI, smoking habits and gender have insignificant influence on breath content. Thus, VOC analysis offers great potential as a non-invasive, fast, and cheap alternative assessment pathway for detecting IBD.

6.8. Reference

[1] A. Bassi, S. Dodd, P. Williamson, and K. Bodger, “Cost of illness of inflammatory bowel disease in the UK: A single centre retrospective study,” *Gut*, vol. 53, no. 10, pp. 1471–1478, 2004, doi: 10.1136/gut.2004.041616.

[2] M. Fakhoury, H. Al-Salami, R. Negrulj, and A. Mooranian, “Inflammatory bowel

disease: clinical aspects and treatments," *J. Inflamm. Res.*, vol. 16, no. 4, p. 113, Jun. 2014, doi: 10.2147/JIR.S65979.

[3] M. Schoultz, L. MacAden, and A. J. M. Watson, "Co-designing inflammatory bowel disease (Ibd) services in Scotland: Findings from a nationwide survey," *BMC Health Serv. Res.*, vol. 16, no. 1, pp. 1–9, 2016, doi: 10.1186/s12913-016-1490-7.

[4] K. Freeman, R. Ryan, N. Parsons, S. Taylor-Phillips, B. H. Willis, and A. Clarke, "The incidence and prevalence of inflammatory bowel disease in UK primary care: a retrospective cohort study of the IQVIA Medical Research Database," *BMC Gastroenterol.*, vol. 21, no. 1, p. 139, 2021, doi: 10.1186/s12876-021-01716-6.

[5] L. Birimberg-Schwartz *et al.*, "Development and validation of diagnostic criteria for IBD subtypes including IBdunclassified in children: A multicentre study from the pediatric IBD porto group of ESPGHAN," *J. Crohn's Colitis*, vol. 11, no. 9, pp. 1078–1084, 2017, doi: 10.1093/ecco-jcc/jjx053.

[6] M. J. Rosen, A. Dhawan, and S. A. Saeed, "Inflammatory Bowel Disease in Children and Adolescents," *JAMA Pediatr.*, vol. 169, no. 11, p. 1053, Nov. 2015, doi: 10.1001/jamapediatrics.2015.1982.

[7] M. Schoultz, I. M. Atherton, G. Hubbard, and A. J. M. Watson, "The use of mindfulness-based cognitive therapy for improving quality of life for inflammatory bowel disease patients: Study protocol for a pilot randomised controlled trial with embedded process evaluation," *Trials*, vol. 14, no. 1, pp. 1–9, 2013, doi: 10.1186/1745-6215-14-431.

[8] A. Sawczenko, "Presenting features of inflammatory bowel disease in Great Britain and Ireland," *Arch. Dis. Child.*, vol. 88, no. 11, pp. 995–1000, Nov. 2003, doi: 10.1136/adc.88.11.995.

[9] L. A. Graff *et al.*, "The Relationship of Inflammatory Bowel Disease Type and Activity to Psychological Functioning and Quality of Life," *Clin. Gastroenterol. Hepatol.*, vol. 4, no. 12, pp. 1491–1501, 2006, doi: 10.1016/j.cgh.2006.09.027.

[10] T. R. Card, J. Siffledeen, and K. M. Fleming, "Are IBD patients more likely to have a prior diagnosis of irritable bowel syndrome? Report of a case-control study in the general practice research database," *United Eur. Gastroenterol. J.*, vol. 2, no. 6, pp. 505–512, 2014, doi: 10.1177/2050640614554217.

[11] R. P. Arasaradnam *et al.*, “Non-invasive exhaled volatile organic biomarker analysis to detect inflammatory bowel disease (IBD),” *Dig. Liver Dis.*, vol. 48, no. 2, pp. 148–153, 2016, doi: 10.1016/j.dld.2015.10.013.

[12] Y. G. Kim and B. I. Jang, “The Role of Colonoscopy in Inflammatory Bowel Disease,” *Clin. Endosc.*, vol. 46, no. 4, p. 317, 2013, doi: 10.5946/ce.2013.46.4.317.

[13] P. F. Van Rheenen, E. Van De Vijver, and V. Fidler, “Faecal calprotectin for screening of patients with suspected inflammatory bowel disease: Diagnostic meta-analysis,” *BMJ*, vol. 341, no. 7765, p. 188, 2010, doi: 10.1136/bmj.c3369.

[14] M. Patel, D. Fowler, J. Sizer, and C. Walton, “Faecal volatile biomarkers of *Clostridium difficile* infection,” *PLoS One*, vol. 14, no. 4, pp. 1–15, 2019, doi: 10.1371/journal.pone.0215256.

[15] K. Van Malderen, B. Y. De Winter, J. G. De Man, H. U. De Schepper, and K. Lamote, “Volatomics in inflammatory bowel disease and irritable bowel syndrome,” *EBioMedicine*, vol. 54, p. 102725, 2020, doi: 10.1016/j.ebiom.2020.102725.

[16] W. Zhou, J. Tao, J. Li, and S. Tao, “Volatile organic compounds analysis as a potential novel screening tool for colorectal cancer: A systematic review and meta-analysis,” *Medicine (Baltimore).*, vol. 99, no. 27, p. e20937, 2020, doi: 10.1097/MD.0000000000020937.

[17] C. Probert *et al.*, “Faecal volatile organic compounds in preterm babies at risk of necrotising enterocolitis: The DOVE study,” *Arch. Dis. Child. Fetal Neonatal Ed.*, vol. 105, no. 5, pp. 474–479, 2020, doi: 10.1136/archdischild-2019-318221.

[18] D. K. Chan, C. L. Leggett, and K. K. Wang, “Diagnosing gastrointestinal illnesses using fecal headspace volatile organic compounds,” *World J. Gastroenterol.*, vol. 22, no. 4, pp. 1639–1649, 2016, doi: 10.3748/wjg.v22.i4.1639.

[19] M. P. van der Schee, T. Paff, P. Brinkman, W. M. C. van Aalderen, E. G. Haarman, and P. J. Sterk, “Breathomics in Lung Disease,” *Chest*, vol. 147, no. 1, pp. 224–231, Jan. 2015, doi: 10.1378/chest.14-0781.

[20] R. P. Arasaradnam, M. W. Pharaoh, G. J. Williams, C. U. Nwokolo, K. D. Bardhan, and S. Kumar, “Colonic fermentation - More than meets the nose,” *Med. Hypotheses*, vol. 73, no. 5, pp. 753–756, 2009, doi: 10.1016/j.mehy.2009.04.027.

[21] A. Bond *et al.*, “Volatile organic compounds emitted from faeces as a biomarker for colorectal cancer,” *Aliment. Pharmacol. Ther.*, vol. 49, no. 8, pp. 1005–1012, 2019, doi: 10.1111/apt.15140.

[22] M. D. Rouvroye *et al.*, “Faecal scent as a novel non-invasive biomarker to discriminate between coeliac disease and refractory coeliac disease: A proof of principle study,” *Biosensors*, vol. 9, no. 2, 2019, doi: 10.3390/bios9020069.

[23] D. M. Lecky, M. K. D. Hawking, and C. A. M. McNulty, “Patients’ perspectives on providing a stool sample to their GP: A qualitative study,” *Br. J. Gen. Pract.*, vol. 64, no. 628, pp. e684–e693, 2014, doi: 10.3399/bjgp14X682261.

[24] S. Esfahani *et al.*, “Variation in gas and volatile compound emissions from human urine as it ages, measured by an electronic nose,” *Biosensors*, vol. 6, no. 1, pp. 1–11, 2016, doi: 10.3390/bios6010004.

[25] J. Pereira *et al.*, “Breath analysis as a potential and non-invasive frontier in disease diagnosis: An overview,” *Metabolites*, vol. 5, no. 1, pp. 3–55, 2015, doi: 10.3390/metabo5010003.

[26] V. Besa *et al.*, “Exhaled volatile organic compounds discriminate patients with chronic obstructive pulmonary disease from healthy subjects,” *Int. J. COPD*, vol. 10, pp. 399–406, 2015, doi: 10.2147/COPD.S76212.

[27] D. F. Altomare *et al.*, “Exhaled volatile organic compounds identify patients with colorectal cancer,” *Br. J. Surg.*, vol. 100, no. 1, pp. 144–150, 2013, doi: 10.1002/bjs.8942.

[28] N. Alkhouri *et al.*, “Isoprene in the exhaled breath is a novel biomarker for advanced fibrosis in patients with chronic liver disease: A pilot study,” *Clin. Transl. Gastroenterol.*, vol. 6, no. 9, pp. e112-7, 2015, doi: 10.1038/ctg.2015.40.

[29] M. Xu, Z. Tang, Y. Duan, and Y. Liu, “GC-Based Techniques for Breath Analysis: Current Status, Challenges, and Prospects,” *Crit. Rev. Anal. Chem.*, vol. 46, no. 4, pp. 291–304, 2016, doi: 10.1080/10408347.2015.1055550.

[30] R. E. Amor, M. K. Nakhleh, O. Barash, and H. Haick, “Breath analysis of cancer in the present and the future,” *Eur. Respir. Rev.*, vol. 28, no. 152, pp. 1–10, 2019, doi: 10.1183/16000617.0002-2019.

[31] P. Trefz, J. Obermeier, R. Lehbrink, J. K. Schubert, W. Miekisch, and D. C. Fischer, "Exhaled volatile substances in children suffering from type 1 diabetes mellitus: results from a cross-sectional study," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019, doi: 10.1038/s41598-019-52165-x.

[32] J. S. Hyams, C. Di Lorenzo, M. Saps, R. J. Shulman, A. Staiano, and M. van Tilburg, "Childhood Functional Gastrointestinal Disorders: Child/Adolescent," *Gastroenterology*, vol. 150, no. 6, pp. 1456-1468.e2, May 2016, doi: 10.1053/j.gastro.2016.02.015.

[33] A. Levine *et al.*, "ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents," *J. Pediatr. Gastroenterol. Nutr.*, vol. 58, no. 6, pp. 795–806, 2014, doi: 10.1097/MPG.0000000000000239.

[34] A. Levine *et al.*, "Pediatric modification of the Montreal classification for inflammatory bowel disease: The Paris classification," *Inflamm. Bowel Dis.*, vol. 17, no. 6, pp. 1314–1321, 2011, doi: 10.1002/ibd.21493.

[35] M. K. Bomers *et al.*, "Rapid, Accurate, and on-site detection of *C. difficile* in stool samples," *Am. J. Gastroenterol.*, vol. 110, no. 4, pp. 588–594, 2015, doi: 10.1038/ajg.2015.90.

[36] R. P. Arasaradnam *et al.*, "Differentiating coeliac disease from irritable bowel syndrome by urinary volatile organic compound analysis - A pilot study," *PLoS One*, vol. 9, no. 10, pp. 1–9, 2014, doi: 10.1371/journal.pone.0107312.

[37] J. A. Covington, M. P. Van. Der Schee, A. S. L. Edge, B. Boyle, R. S. Savage, and R. P. Arasaradnam, "The application of FAIMS gas analysis in medical diagnostics," *Analyst*, vol. 140, no. 20, pp. 6775–6781, 2015, doi: 10.1039/c5an00868a.

[38] B. A. MacKalski and C. N. Bernstein, "New diagnostic imaging tools for inflammatory bowel disease," *Gut*, vol. 55, no. 5, pp. 733–741, 2006, doi: 10.1136/gut.2005.076612.

[39] A. Moniuszko, A. Wiśniewska, and G. Rydzewska, "Biomarkers in management of inflammatory bowel disease," *Prz. Gastroenterol.*, vol. 8, no. 5, pp. 275–283, 2013, doi: 10.5114/pg.2013.38728.

[40] W. Cao and Y. Duan, "Current status of methods and techniques for breath analysis," *Crit. Rev. Anal. Chem.*, vol. 37, no. 1, pp. 3–13, 2007, doi:

10.1080/10408340600976499.

[41] S. van den Velde, M. Quirynen, P. Van hee, and D. van Steenberghe, “Halitosis associated volatiles in breath of healthy subjects,” *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 853, no. 1–2, pp. 54–61, 2007, doi: 10.1016/j.jchromb.2007.02.048.

[42] D. Poli *et al.*, “Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatisation SPME-GC/MS,” *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 878, no. 27, pp. 2643–2651, 2010, doi: 10.1016/j.jchromb.2010.01.022.

[43] H. Borsdorf and G. A. Eiceman, “Ion mobility spectrometry: Principles and applications,” *Appl. Spectrosc. Rev.*, vol. 41, no. 4, pp. 323–375, 2006, doi: 10.1080/05704920600663469.

[44] J. I. Baumbach, “Process analysis using ion mobility spectrometry,” *Anal. Bioanal. Chem.*, vol. 384, no. 5, pp. 1059–1070, 2006, doi: 10.1007/s00216-005-3397-8.

[45] J. M. Lewis, R. S. Savage, N. J. Beeching, M. B. J. Beadsworth, N. Feasey, and J. A. Covington, “Identifying volatile metabolite signatures for the diagnosis of bacterial respiratory tract infection using electronic nose technology: A pilot study,” *PLoS One*, vol. 12, no. 12, pp. 1–10, 2017, doi: 10.1371/journal.pone.0188879.

[46] E. Westenbrink *et al.*, “Development and application of a new electronic nose instrument for the detection of colorectal cancer,” *Biosens. Bioelectron.*, vol. 67, pp. 733–738, 2015, doi: 10.1016/j.bios.2014.10.044.

[47] L. Blanchet *et al.*, “Factors that influence the volatile organic compound content in human breath,” *J. Breath Res.*, vol. 11, no. 1, 2017, doi: 10.1088/1752-7163/aa5cc5.

[48] NHS, “Healthy Weight.” https://www.nhs.uk/conditions/obesity/ (accessed Jun. 09, 2021).

[49] CDC, “Adult Tobacco Use Information.” https://www.cdc.gov/nchs/nhis/tobacco/tobacco_glossary.htm (accessed Jun. 09, 2021).

[50] S. Bosch *et al.*, “Optimized Sampling Conditions for Fecal Volatile Organic Compound Analysis by Means of Field Asymmetric Ion Mobility Spectrometry,”

Anal. Chem., vol. 90, no. 13, pp. 7972–7981, 2018, doi: 10.1021/acs.analchem.8b00688.

[51] N. Van Gaal *et al.*, "Faecal volatile organic compounds analysis using field asymmetric ion mobility spectrometry: Non-invasive diagnostics in paediatric inflammatory bowel disease," *J. Breath Res.*, vol. 12, no. 1, 2018, doi: 10.1088/1752-7163/aa6f1d.

[52] S. Bosch *et al.*, "Fecal Amino Acid Analysis Can Discriminate de Novo Treatment-Naïve Pediatric Inflammatory Bowel Disease from Controls," *J. Pediatr. Gastroenterol. Nutr.*, vol. 66, no. 5, pp. 773–778, 2018, doi: 10.1097/MPG.0000000000001812.

[53] J. S. Hyams *et al.*, "Development and validation of a pediatric Crohn's disease activity index.," *J. Pediatr. Gastroenterol. Nutr.*, vol. 12, no. 4, pp. 439–447, May 1991.

[54] D. Turner *et al.*, "Development, Validation, and Evaluation of a Pediatric Ulcerative Colitis Activity Index: A Prospective Multicenter Study," *Gastroenterology*, vol. 133, no. 2, pp. 423–432, 2007, doi: 10.1053/j.gastro.2007.05.029.

[55] A. Dignass *et al.*, "Second European evidence-based consensus on the diagnosis and management of ulcerative colitis Part 1: Definitions and diagnosis," *J. Crohn's Colitis*, vol. 6, no. 10, pp. 965–990, 2012, doi: 10.1016/j.crohns.2012.09.003.

[56] R. S. Walmsley, R. C. S. Ayres, R. E. Pounder, and R. N. Allan, "A simple clinical colitis activity index," *Gut*, vol. 43, no. 1, pp. 29–32, 1998, doi: 10.1136/gut.43.1.29.

[57] R. Harvey and M. Jane Bradshaw, "MEASURING CROHN'S DISEASE ACTIVITY," *Lancet*, vol. 315, no. 8178, pp. 1134–1135, May 1980, doi: 10.1016/S0140-6736(80)91577-9.

[58] M. S. Silverberg *et al.*, "Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology.," *Can. J. Gastroenterol.*, vol. 19 Suppl A, no. September, 2005, doi: 10.1155/2005/269076.

[59] I. Ahmed, R. Greenwood, B. Costello, N. Ratcliffe, and C. S. Probert, "Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease," *Aliment. Pharmacol. Ther.*, vol. 43, no. 5, pp. 596–611, 2016, doi:

10.1111/apt.13522.

[60] T. G. J. de Meij *et al.*, “Faecal gas analysis by electronic nose as novel, non-invasive method for assessment of active and quiescent paediatric inflammatory bowel disease: Proof of principle study,” *J. Crohn’s Colitis*, Sep. 2014, doi: 10.1016/j.crohns.2014.09.004.

[61] I. Ahmed, R. Greenwood, B. L. de Costello, N. M. Ratcliffe, and C. S. Probert, “An Investigation of Fecal Volatile Organic Metabolites in Irritable Bowel Syndrome,” *PLoS One*, vol. 8, no. 3, pp. 1–13, 2013, doi: 10.1371/journal.pone.0058204.

[62] A. G. L. Bodelier *et al.*, “Volatile organic compounds in exhaled air as novel marker for disease activity in Crohn’s disease: A metabolomic approach,” *Inflamm. Bowel Dis.*, vol. 21, no. 8, pp. 1776–1785, 2015, doi: 10.1097/MIB.0000000000000436.

[63] A. Smolinska *et al.*, “The potential of volatile organic compounds for the detection of active disease in patients with ulcerative colitis,” *Aliment. Pharmacol. Ther.*, vol. 45, no. 9, pp. 1244–1254, 2017, doi: 10.1111/apt.14004.

[64] R. P. Arasaradnam *et al.*, “A Novel tool for noninvasive diagnosis and tracking of patients with inflammatory bowel disease,” *Inflamm. Bowel Dis.*, vol. 19, no. 4, pp. 999–1003, 2013, doi: 10.1097/MIB.0b013e3182802b26.

[65] I. S. Seo *et al.*, “Cross detection for odor of metabolic waste between breast and colorectal cancer using canine olfaction,” *PLoS One*, vol. 13, no. 2, pp. 1–9, 2018, doi: 10.1371/journal.pone.0192629.

[66] C. Becker, M. F. Neurath, and S. Wirtz, “The intestinal microbiota in inflammatory bowel disease,” *ILAR J.*, vol. 56, no. 2, pp. 192–204, 2015, doi: 10.1093/ilar/ilv030.

[67] I. Martínez, C. E. Muller, and J. Walter, “Long-Term Temporal Analysis of the Human Fecal Microbiota Revealed a Stable Core of Dominant Bacterial Species,” *PLoS One*, vol. 8, no. 7, 2013, doi: 10.1371/journal.pone.0069621.

[68] J. J. Faith *et al.*, “The long-term stability of the human gut microbiota,” *Science (80-.).*, vol. 341, no. 6141, 2013, doi: 10.1126/science.1237439.

[69] C. Martinez *et al.*, “Unstable Composition of the Fecal Microbiota in Ulcerative Colitis During Clinical Remission,” *Am. J. Gastroenterol.*, vol. 103, no. 3, pp. 643–648, Mar. 2008, doi: 10.1111/j.1572-0241.2007.01592.x.

[70] J. Halfvarson *et al.*, “Dynamics of the human gut microbiome in inflammatory bowel disease,” *Nat. Microbiol.*, vol. 2, no. February, pp. 1–7, 2017, doi: 10.1038/nmicrobiol.2017.4.

[71] K. Machiels *et al.*, “A decrease of the butyrate-producing species *roseburia hominis* and *faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis,” *Gut*, vol. 63, no. 8, pp. 1275–1283, 2014, doi: 10.1136/gutjnl-2013-304833.

[72] C. Manichanh *et al.*, “Reduced diversity of faecal microbiota in Crohn’s disease revealed by a metagenomic approach,” *Gut*, vol. 55, no. 2, pp. 205–211, 2006, doi: 10.1136/gut.2005.073817.

[73] D. N. Frank, A. L. St. Amand, R. A. Feldman, E. C. Boedeker, N. Harpaz, and N. R. Pace, “Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 34, pp. 13780–13785, 2007, doi: 10.1073/pnas.0706625104.

[74] A. W. Walker *et al.*, “High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease,” *BMC Microbiol.*, vol. 11, no. 1, p. 7, 2011, doi: 10.1186/1471-2180-11-7.

[75] E. S. Wills, D. M. A. E. Jonkers, P. H. Savelkoul, A. A. Masclee, M. J. Pierik, and J. Penders, “Fecal microbial composition of ulcerative colitis and Crohn’s disease patients in remission and subsequent exacerbation,” *PLoS One*, vol. 9, no. 3, pp. 1–10, 2014, doi: 10.1371/journal.pone.0090981.

[76] K. Kajander *et al.*, “Elevated pro-inflammatory and lipotoxic mucosal lipids characterise irritable bowel syndrome,” *World J. Gastroenterol.*, vol. 15, no. 48, pp. 6068–6074, 2009, doi: 10.3748/wjg.15.6068.

[77] K. S. Fritz and D. R. Petersen, “An overview of the chemistry and biology of reactive aldehydes,” *Free Radic. Biol. Med.*, vol. 59, no. 1, pp. 85–91, Jun. 2013, doi: 10.1016/j.freeradbiomed.2012.06.025.

[78] N. Patel *et al.*, “Metabolomic analysis of breath volatile organic compounds reveals unique breathprints in children with inflammatory bowel disease: a pilot study,” *Aliment. Pharmacol. Ther.*, vol. 23, no. 1, p. n/a-n/a, Jul. 2014, doi: 10.1111/apt.12861.

[79] L. C. Hicks *et al.*, “Analysis of Exhaled Breath Volatile Organic Compounds in Inflammatory Bowel Disease: A Pilot Study,” *J. Crohns. Colitis*, vol. 9, no. 9, pp. 731–737, 2015, doi: 10.1093/ecco-jcc/jjv102.

[80] M. H. M. C. Van Nuenen, K. Venema, J. C. J. Van Der Woude, and E. J. Kuipers, “The metabolic activity of fecal microbiota from healthy individuals and patients with inflammatory bowel disease,” *Dig. Dis. Sci.*, vol. 49, no. 3, pp. 485–491, 2004, doi: 10.1023/B:DDAS.0000020508.64440.73.

[81] S. Kurada, N. Alkhouri, C. Fiocchi, R. Dweik, and F. Rieder, “Review article: Breath analysis in inflammatory bowel diseases,” *Aliment. Pharmacol. Ther.*, vol. 41, no. 4, pp. 329–341, 2015, doi: 10.1111/apt.13050.

[82] C. Maaser *et al.*, “European Crohn’s and Colitis Organisation Topical Review on Environmental Factors in IBD,” *J. Crohns. Colitis*, vol. 11, no. 8, pp. 905–920, 2017, doi: 10.1093/ecco-jcc/jjw223.

[83] G. G. Kaplan *et al.*, “The inflammatory bowel diseases and ambient air pollution: A novel association,” *Am. J. Gastroenterol.*, vol. 105, no. 11, pp. 2412–2419, 2010, doi: 10.1038/ajg.2010.252.

[84] T. Sakata, “Stimulatory effect of short-chain fatty acids on epithelial cell proliferation in the rat intestine: a possible explanation for trophic effects of fermentable fibre, gut microbes and luminal trophic factors,” *Br. J. Nutr.*, vol. 58, no. 1, pp. 95–103, 1987, doi: 10.1079/bjn19870073.

[85] W. E. W. Roediger, “Utilization of Nutrients by Isolated Epithelial Cells of the Rat Colon,” *Gastroenterology*, vol. 83, no. 2, pp. 424–429, 1982, doi: 10.1016/S0016-5085(82)80339-9.

[86] J. H. Cummings, “Short-chain fatty acid enemas in the treatment of distal ulcerative colitis,” *Eur. J. Gastroenterol. Hepatol.*, vol. 9, no. 2, 1997, [Online]. Available: https://journals.lww.com/eurojgh/Fulltext/1997/02000/Short_chain_fatty_acid_enemas_in_the_treatment_of.8.aspx.

[87] A. Di Sabatino *et al.*, “Oral butyrate for mildly to moderately active Crohn’s disease,” *Aliment. Pharmacol. Ther.*, vol. 22, no. 9, pp. 789–794, 2005, doi: 10.1111/j.1365-2036.2005.02639.x.

[88] E. V Loftus Jr, “Update on the Incidence and Prevalence of Inflammatory Bowel

Disease in the United States," *Gastroenterol. Hepatol. (N. Y).*, vol. 12, no. 11, pp. 704–707, Nov. 2016, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28035199.

[89] M. Hannemann *et al.*, "Influence of age and sex in exhaled breath samples investigated by means of infrared laser absorption spectroscopy," *J. Breath Res.*, vol. 5, no. 2, p. 27101, 2011, doi: 10.1088/1752-7155/5/2/027101.

[90] A. Banerjee *et al.*, "Faecal calprotectin for differentiating between irritable bowel syndrome and inflammatory bowel disease: a useful screen in daily gastroenterology practice," *Frontline Gastroenterol.*, vol. 6, no. 1, pp. 20–26, 2015, doi: 10.1136/flgastro-2013-100429.

[91] M. Von Euler-Chelpin, K. Brasso, and E. Lynge, "Determinants of participation in colorectal cancer screening with faecal occult blood testing," *J. Public Health (Bangkok).*, vol. 32, no. 3, pp. 395–405, 2010, doi: 10.1093/pubmed/fdp115.

Chapter 7
CONCLUSION AND FURTHER WORK

7.1. Introduction

The goal of this project was to utilise the VOCs emanating from human waste for disease diagnosis and monitoring using modern electronic nose technologies and analyse the data produced by them. To achieve this goal, we developed a new data analysis pipeline for the electronic nose (E-nose), field asymmetric ion mobility spectrometry (FAIMS), gas chromatography ion mobility spectrometry (GC-IMS), and gas chromatography time of flight mass spectrometry (GC-TOF-MS) methods, as well as conducting rigorously planned case-control studies. Eight novel clinical studies have been performed to explore the use of VOCs from human waste for non-invasive disease diagnosis, focussing on pancreatic cancer (PDAC), colorectal cancer (CRC), and inflammatory bowel disease (IBD) (Chapter 5 and 6, respectively). This chapter summarises the thesis findings and emphasises the originality and importance of these findings in the context of a broader study.

7.2. Data Analysis Pipeline

The data analysis pipeline utilised in this research includes pre-processing, data splitting and cross-validation, feature selection and classification, as well as model performance calculations. Each instrument requires a unique pre-processing phase due to its unique data characteristics. For the E-nose, pre-processing steps, in our case, only involves extracting the maximum value of sensor response for each sensor in the sensor array. While in FAIMS, the pre-processing processes include padding, cropping, and applying a threshold to reduce the number of uninformative data. Then, in GC-IMS, which has the highest dimensional raw data among these instruments, include RIP alignment, cropping, and thresholding in its data pre-processing steps. Lastly, GC-TOF-MS pre-processing involves extracting important information and removing the uninformative chemicals from the list.

Data splitting differs across studies, with data from studies with small datasets being separated simply into training and validation sets, while data from research with big datasets were divided into three sets, namely, training, validation, and test set. Then, these data are processed in cross-validation, features extraction, classification model training, and finally, the trained model produces the class prediction. Model performance is measured based on the confusion matrix and ROC curve. These are AUC (95% CI), sensitivity, specificity, PPV, NPV, and p-value. Only GC-TOF-MS and GC-IMS are able to identify chemical compounds by comparing the result with NIST library. The methods described in Chapter 4 have been used throughout this thesis (Chapter 5 and 6). Among the 5 different classification algorithms used in this thesis; Random Forest provide the most consistent performance throughout all the

studies based on the aggregate AUC value across all analysis (see Appendix Table 7.1 and 7.2). This conclusion is similar with the finding of Fernández-Delgado *et al.*, who found that random forest consistently outperformed other 179 classifiers while analysing 121 different datasets [1]. Random forest performed exceptionally well when dealing with high dimensionality dataset since it works with subsets of data. It is also robust to outliers and non-linear data. Due to the averaging step of all trees inside random forest, it generally has low bias and moderate variance. It also does a good job of handling missing data and unbalanced data. However, due to the ensemble of decision trees inside random forest, it suffers intrinsic interpretability. Additionally, the size of the trees might take up a significant amount of memory when dealing with very large datasets.

7.3. Pancreatic Cancer Studies

The pancreatic cancer (PDAC) studies in Chapter 5 utilised FAIMS, GC-IMS, and GC-TOF-MS technology to investigate the non-invasive diagnosis of PDAC (early- and late-stage), chronic pancreatitis (CP), and healthy controls. In addition to this, the study evaluated the possible biomarker that may associate with PDAC. The first study included 162 subjects: 81 healthy controls and 81 PDAC, which consisted of 44 early-stage PDAC and 36 late-stage PDAC. Whereas the second study included 123 subjects: 33 healthy controls, 45 PDAC, and 45 CP. The novelty of this work was that it was the first urinary study (to be best of our knowledge) to investigate PDAC, CP, and healthy using FAIMS, GC-IMS, and GC-TOF-MS technology [2]. The importance of our findings was:

a. Urinary VOC analysis could separate between PDAC and healthy control, early stage PDAC and healthy, early stage PDAC and late stage PDAC, and CP and healthy.
b. Urinary VOC analysis showed only small difference when comparing PDAC with CP.
c. Chemical identification suggested that 2,6-dimethyl-octane, nonanal, 4-ethyl-1,2-dimethyl-benzene, and 2-pentanone play an important role in the separation of PDAC and healthy.

The FAIMS results are summarised below

- PDAC vs Controls: AUC = 0.92, sensitivity = 0.90, specificity = 0.81.
- Early-stage PDAC vs Controls: AUC = 0.89, sensitivity = 0.91, specificity = 0.78.
- Late- vs Early-stage PDAC: AUC = 0.92, sensitivity = 0.82, specificity = 0.89.

The GC-IMS results are summarised below

- PDAC vs Controls: AUC = 0.88, sensitivity = 0.84, specificity = 0.94.

- PDAC vs CP&Controls: AUC = 0.69, sensitivity = 0.72, specificity = 0.60.
- CP vs Controls: AUC = 0.86, sensitivity = 0.80, specificity = 0.91.
- PDAC vs CP: AUC = 0.58, sensitivity = 0.51, specificity = 0.73.

The GC-TOF-MS results are summarised below

- PDAC vs Control: AUC = 0.86, sensitivity = 0.72, specificity = 0.96.
- PDAC vs CP&Controls: AUC = 0.75, sensitivity = 0.52, specificity = 0.96.
- CP vs Controls: AUC = 0.67, sensitivity = 0.38, specificity = 0.96.
- PDAC vs CP: AUC = 0.55, sensitivity = 0.38, specificity = 0.88.

7.4. Colorectal Cancer Studies

The colorectal cancer (CRC) studies in chapter 5 utilised FAIMS and GC-IMS technology to investigate whether urinary and faecal VOC analysis could be used to distinguish between CRC, adenoma, and healthy control patient. Furthermore, this study evaluates the possibility of faecal VOC analysis as a secondary follow-up after polypectomy. The first study that analysed urinary VOC included 163 subjects: 12 CRC, 80 adenoma, 37 healthy control, and 33 other gastrointestinal diseases. Whereas the second study analysed faecal VOCs included 565 subjects: 14 CRC, 260 adenoma (64 AA, 69 LA, 127 SA), 32 polypectomy patient, and 259 healthy. The novelty of this work was that it was the first study that demonstrate the potential of urinary VOC analysis in detecting CRC and adenoma using GC-IMS and also the first reported research to show the potential of faecal VOC analysis for polypectomy follow-up [3,4]. The importance of our findings were

a. The analysis of urinary and faecal VOC could differentiate between CRC and healthy.
b. Urinary VOC analysis could clearly separate between Adenoma and CRC, but not adenoma and control.
c. Faecal VOC analysis could clearly separate between Adenoma and control, but not Adenoma and CRC.
d. The faecal VOC analysis showed intra-individual VOC profile changed to physiological state three months following polypectomy which indicate the potential of faecal VOC analysis for polypectomy follow-up.

The FAIMS results analysing urinary VOC are summarised below

- CRC vs healthy: AUC = 0.98, sensitivity = 1.00, specificity = 0.92.
- CRC vs adenoma: AUC = 0.92, sensitivity = 0.83, specificity = 1.00.

The GC-IMS result analysing urinary VOC are summarised below

- CRC vs healthy: AUC = 0.82, sensitivity = 0.80, specificity = 0.83.

- Adenoma vs healthy: AUC = 0.61, sensitivity = 0.58, specificity = 0.62.

The GC-IMS result analysing faecal VOC are summarised below

- CRC vs healthy: AUC = 0.96, sensitivity = 1.00, specificity = 1.00.
- CRC vs adenoma: AUC = 0.54, sensitivity = 0.98, specificity = 0.19.
- Adenoma vs healthy: AUC = 0.96, sensitivity = 0.97, specificity = 0.94.
- T0 Healthy vs T0 pre-polypectomy: AUC = 0.98, sensitivity = 1.00, specificity = 0.97.
- T1 Healthy vs T1 post-polypectomy: AUC = 0.55, sensitivity = 0.91, specificity = 0.25.
- T0 vs T1 pre-and post-polypectomy: AUC = 0.94, sensitivity = 0.91, specificity = 0.91.
- T0 vs T1 Healthy: AUC = 0.58, sensitivity = 0.94, specificity = 0.27.

7.5. IBD Studies

The IBD studies in Chapter 6 utilised E-nose, FAIMS, and GC-IMS technology to investigate whether breath, urinary, and faecal VOC analysis could be used as a non-invasive alternative to separate between IBD and healthy control, including Crohn's disease (CD and ulcerative colitis (UC) as subtypes of IBD in paediatric and adult patients. In addition, the study compared the VOC profile of similar symptomatic disease like IBS/FAP-NOS to IBD profile and analysed possible compounds biomarker that contribute the most in separating IBD from healthy in breath sample. Lastly, the study evaluated whether the faecal VOC profiles could be used to monitor the state of IBD, i.e., active and remission both in UC and CD. The first study that analysed the faecal VOC of paediatric patient included 75 subjects: 30 with IBD (15 UC, 15 CD), 15 IBS/FAP-NOS, and 30 healthy. The second study, which analysed the breath VOC profiles, included 39 subjects: 30 IBD (14 CD, 16 UC) and 9 healthy. The third study, which analysed both urine and faecal VOC in paediatric, included 20 subjects: 10 with IBD and 10 healthy controls. The last study analysed the faecal VOC included 507 subjects: 280 with IBD (164 CD, 112 UC, 4 IBD-U) and 227 healthy controls. The novelties of these works were

- The first reported study that investigate the use of faecal VOC profiles of paediatric patient in separating between IBD and IBS/FAP-NOS.
- The first breath study that investigate CD, UC, and controls using E-nose and GC-IMS technology.
- The first reported study that assessed the potential of urinary VOC analysis in paediatric IBD.
- The first large sample, multicentre cohort faecal VOC analysis study that investigate IBD (UC and CD) and the active and remission stage of UC and CD using GC-IMS.

The importance of these studies were:

a. The analysis of breath, urine, and stool VOC could clearly separate between IBD (including UC and CD) and healthy controls in both paediatric and adult patients.
b. The analysis of faecal VOC could separate between paediatric patient with IBD and IBS/FAP-NOS.
c. The separation of UC and CD was only observed in breath VOC study.
d. In this study, the faecal VOC analysis did not show separation between IBD disease activity state.
e. Chemical identification suggested that butanoic acid and ethanoic acid play an important role in the separation of IBD and healthy in breath study.

The FAIMS results in paediatric faecal VOC analysis can be summarized below:

- IBD vs healthy: AUC = 0.96, sensitivity = 0.93, specificity = 0.97.
- UC vs healthy: AUC = 0.98, sensitivity = 0.93, specificity = 0.97.
- CD vs healthy: AUC = 0.95, sensitivity = 0.93, specificity = 0.93.
- IBD vs IBS/FAP-NOS: AUC = 0.94, sensitivity = 1.00, specificity = 0.87.
- CD vs UC: AUC = 0.67, sensitivity = 0.60, specificity = 0.80.

The GC-IMS results of breath VOC analysis in detecting IBD can be summarised below:

- IBD vs healthy: AUC = 0.93, sensitivity = 0.87, specificity = 0.89.
- CD vs UC: AUC = 0.71, sensitivity = 0.86, specificity = 0.62.

The E-nose results of breath VOC analysis in detecting IBD can be summarised below:

- IBD vs healthy: AUC = 0.81, sensitivity = 0.67, specificity = 0.89.
- CD vs UC: AUC = 0.88, sensitivity = 0.71, specificity = 0.88.

When comparing urinary and faecal VOC analysis in detecting IBD using GC-IMS, the results can be summarised below:

- Urine -> IBD vs healthy: AUC = 0.78, sensitivity = 0.80, specificity = 0.70.
- Stool -> IBD vs healthy: AUC = 0.73, sensitivity = 0.70, specificity = 0.90.

The GC-IMS results in faecal VOC analysis can be summarised below:

- IBD vs healthy: AUC = 0.96, sensitivity = 0.97, specificity = 0.92.
- UC vs healthy: AUC = 0.95, sensitivity = 1.00, specificity = 0.91.
- CD vs healthy: AUC = 0.97, sensitivity = 0.94, specificity = 0.96.
- CD vs UC: AUC = 0.55, sensitivity = 0.17, specificity = 0.96.
- IBDa vs IBDr: AUC = 0.59, sensitivity = 0.21, specificity = 0.96.

7.6. Further Work and Future Prospect

The use of VOC analysis for disease diagnosis is getting ever increasingly attractive with the advances in analytical instruments and machine learning. The studies in chapter 4 and 5 have demonstrated the potential use of breath, urine, and stool VOC analysis in detecting pancreatic cancer, colorectal cancer, and inflammatory bowel disease. The use of VOC analysis in disease diagnosis offers a fast, safe, non-invasive, inexpensive way of detection, making it an attractive alternative. Following up on the findings of these studies, future studies should be conducted with a larger cohort in a multicentre clinical trial setting. The inclusion of subjects suspected of have the disease (i.e., having similar symptoms) is also essential to further understand the VOC profile of targeted disease. Another crucial issue for future VOC studies to consider is the standardisation of sampling and sample storing procedure that is still lacking attention in VOC analysis, making it hard to compare the results between published studies. This is important topic to solve prior to clinical implementation as it will enable us to really monitor and assess the improvement in VOC analysis technology and eliminate the possible bias created by variation in sampling and sample storing procedure. This standardisation will also greatly help speed up the approval process for this technology to be implemented into clinical setting.

Future research should also focus on pattern recognition for the purpose of diagnostic testing, since once a disease-specific algorithm has been established, this allows for fast, easy-to-perform, high-throughput and low-cost analyses, underlining its suitability for application in clinical practice. Using machine learning (ML) algorithms for pattern recognition purposes to detect specific disease is desirable as ML performance improves continuously with increasing number of samples measured. Once the algorithm reaches satisfactory accuracy, the device may be implemented not only for population-based screening but also for intra-individual follow-up after polyp or cancer removal. In addition, the lack of standardisation on ML approaches needs to be addressed to be able to make significant improvement faster as researcher will easily compare the progress they make to previous study.

Chapter 4 and 5 have shown the capability of urinary VOC analysis in differentiating healthy, early-, and late-stage pancreatic cancer as well as the potential use of faecal VOC analysis in the follow-up of polypectomy and IBD activity monitoring. For future research on follow-up studies, larger cohorts with validated endoscopy outcomes need to be employed in the surveillance programme. In addition, other biological substances, such as breath in PDAC and CRC, have previously been shown to hold potential for the discrimination between these diseases and controls [5] [6]. These substances all have advantages and disadvantages with

regards to stability of VOC composition, ease of collection, stigma, and influence of comorbidity. Thus, it would be interesting to compare patient preferences, compliance, costs, and accuracy of all these substances in one large cohort with simultaneous measurements. Further studies on the origin of biomarker related disease also needs to be done to complete the knowledge gap in VOC analysis. Lastly, following the finding of potential use of faecal VOC analysis in polypectomy, it would be interesting to unravel the underlying entities causing the differences in faecal VOCs pre- and post-polypectomy.

7.7. Reference

[1] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *J. Mach. Learn. Res.*, vol. 15, pp. 3133–3181, 2014.

[2] B. T. Crosby, A. Ridzuan-Allen, and J. P. O'Neill, "Volatile organic compound analysis for the diagnosis of pancreatic cancer," *Ann. Pancreat. Cancer*, vol. 4, 2021, doi: 10.21037/apc-20-39.

[3] W. Zhou, J. Tao, J. Li, and S. Tao, "Volatile organic compounds analysis as a potential novel screening tool for colorectal cancer: A systematic review and meta-analysis," *Medicine (Baltimore).*, vol. 99, no. 27, p. e20937, 2020, doi: 10.1097/MD.0000000000020937.

[4] F. Vernia *et al.*, "Are volatile organic compounds accurate markers in the assessment of colorectal cancer and inflammatory bowel diseases? A review," *Cancers (Basel).*, vol. 13, no. 10, pp. 1–20, 2021, doi: 10.3390/cancers13102361.

[5] S. R. Markar, B. Brodie, S. T. Chin, A. Romano, D. Spalding, and G. B. Hanna, "Profile of exhaled-breath volatile organic compounds to diagnose pancreatic cancer," *Br. J. Surg.*, vol. 105, no. 11, pp. 1493–1500, 2018, doi: 10.1002/bjs.10909.

[6] D. F. Altomare *et al.*, "Chemical signature of colorectal cancer: case–control study for profiling the breath print," *BJS Open*, vol. 4, no. 6, pp. 1189–1199, 2020, doi: 10.1002/bjs5.50354.

Appendix

Table 5.1a. Statistical Analysis of FAIMS Data Using SLR for Pancreatic Cancer Study 1

Statistical parameter	All PDAC vs Healthy Control (All data set)	All PDAC vs Healthy Control (Testing data set)	Early-stage PDAC vs Healthy Control	Early-stage PDAC vs Late-stage PDAC
AUC	0.85 (0.79 – 0.91)	0.81 (0.70 – 0.93)	0.87 (0.81 – 0.94)	0.89 (0.81 – 0.97)
Sensitivity	0.77	0.87	0.68	0.82
Specificity	0.80	0.68	0.89	0.89
PPV	0.79	0.73	0.77	0.90
NPV	0.77	0.84	0.84	0.80
p-value	5.33×10^{-15}	4.83×10^{-6}	2.73×10^{-12}	1.64×10^{-11}

Table 5.1b. Statistical Analysis of FAIMS Data Using RF for Pancreatic Cancer Study 1

Statistical parameter	All PDAC vs Healthy Control (All data set)	All PDAC vs Healthy Control (Testing data set)	Early-stage PDAC vs Healthy Control	Early-stage PDAC vs Late-stage PDAC
AUC	0.90 (0.85 – 0.95)	0.82 (0.72 – 0.92)	0.91 (0.85 – 0.96)	0.89 (0.81 – 0.96)
Sensitivity	0.83	0.61	0.91	0.75
Specificity	0.84	0.94	0.74	0.95 (0.82 – 0.99)
PPV	0.84	0.90	0.66	0.94
NPV	0.84	0.71	0.94	0.76
p-value	6.93×10^{-19}	7.02×10^{-6}	3.04×10^{-14}	1.34×10^{-9}

Table 5.1c. Statistical Analysis of FAIMS Data Using GP for Pancreatic Cancer Study 1

Statistical parameter	All PDAC vs Healthy Control (All data set)	All PDAC vs Healthy Control (Testing data set)	Early-stage PDAC vs Healthy Control	Early-stage PDAC vs Late-stage PDAC
AUC	0.84 (0.78 – 0.90)	0.77 (0.65 – 0.89)	0.84 (0.77 – 0.91)	0.91 (0.84 – 0.97)
Sensitivity	0.75	0.84	0.86	0.82
Specificity	0.83	0.71	0.70	0.92
PPV	0.81	0.74	0.61	0.92
NPV	0.77	0.81	0.90	0.81
p-value	4.67×10^{-14}	7.29×10^{-5}	1.78×10^{-10}	2.86×10^{-11}

Table 5.2a. GC-IMS Result for PDAC vs All from Pancreatic Cancer Study 2

	Sparse Logistic Regression	Random Forest	Gaussian Process
AUC	0.68 (0.58 – 0.79)	0.62 (0.51 – 0.73)	0.66 (0.55 – 0.76)
Sensitivity	0.72	0.40	0.77
Specificity	0.60	0.83	0.53
PPV	0.52	0.59	0.49
NPV	0.78	0.70	0.79
P-Value	4.39×10^{-4}	1.73×10^{-2}	2.79×10^{-3}

Table 5.2b. GC-IMS Result for PDAC vs CP from Pancreatic Cancer Study 2

	Sparse Logistic Regression	Random Forest	Gaussian Process
AUC	0.50 (0.37 – 0.63)	0.53 (0.41 - 0.66)	0.58 (0.45 – 0.71)
Sensitivity	0.47	0.35	0.51
Specificity	0.68	0.80	0.73

PPV	0.61	0.65	0.67
NPV	0.54	0.53	0.58
P-Value	4.89×10^{-1}	3.06×10^{-1}	1.10×10^{-1}

Table 5.2c. GC-IMS Result for PDAC vs Healthy from Pancreatic Cancer Study 2

	Sparse Logistic Regression	Random Forest	Gaussian Process
AUC	0.88 (0.79 – 0.97)	0.82 (0.71 – 0.92)	0.84 (0.74 – 0.94)
Sensitivity	0.84	0.77	0.79
Specificity	0.94	0.91	0.94
PPV	0.95	0.92	0.94
NPV	0.81	0.74	0.77
P-Value	1.18×10^{-8}	1.71×10^{-6}	5.86×10^{-8}

Table 5.2d. GC-IMS Result for CP vs Healthy from Pancreatic Cancer Study 2

	Sparse Logistic Regression	Random Forest	Gaussian Process
AUC	0.82 (0.72 – 0.93)	0.81 (0.70 – 0.92)	0.86 (0.77 – 0.95)
Sensitivity	0.73	0.75	0.80
Specificity	0.91	0.91	0.91
PPV	0.91	0.91	0.91
NPV	0.73	0.74	0.78
P-Value	1.38×10^{-6}	3.32×10^{-6}	9.79×10^{-9}

Table 5.3a. Statistical Analysis Result of PDAC vs All Using GC-TOF-MS from Pancreatic Cancer Study 2

	Sparse Logistic Regression	Random Forest	Gaussian Process
AUC	0.75 (0.63 – 0.87)	0.71 (0.58 – 0.83)	0.50 (0.37 – 0.64)
Sensitivity	0.52	0.48	0.33
Specificity	0.96	0.92	0.92
PPV	0.96	0.92	0.88
NPV	0.51	0.48	0.42
P-Value	3.11×10^{-4}	2.58×10^{-3}	5.27×10^{-1}

Table 5.3b. Statistical Analysis Result of PDAC vs Healthy Using GC-TOF-MS from Pancreatic Cancer Study 2

	Sparse Logistic Regression	Random Forest	Gaussian Process
AUC	0.86 (0.75 – 0.97)	0.79 (0.66 – 0.92)	0.59 (0.66 – 0.92)
Sensitivity	0.72	0.60	0.60
Specificity	0.96	0.88	0.62
PPV	0.95	0.83	0.62
NPV	0.77	0.68	0.60
P-Value	1.81×10^{-6}	3.12×10^{-4}	1.43×10^{-1}

Table 5.3c. Statistical Analysis Result of PDAC vs CP Using GC-TOF-MS from Pancreatic Cancer Study 2

	Sparse Logistic Regression	Random Forest	Gaussian Process
AUC	0.46 (0.27 – 0.65)	0.55 (0.37 – 0.73)	0.48 (0.31 – 0.66)
Sensitivity	0.33	0.38	0.29
Specificity	0.92	0.88	0.83
PPV	0.78	0.73	0.60
NPV	0.61	0.62	0.57
P-Value	6.71×10^{-1}	2.77×10^{-1}	5.76×10^{-1}

Table 5.3d. Statistical Analysis Result of CP vs Healthy Using GC-TOF-MS from Pancreatic Cancer Study 2

	Sparse Logistic Regression	Random Forest	Gaussian Process
AUC	0.53 (0.36 – 0.70)	0.67 (0.50 – 0.83)	0.55 (0.38 – 0.72)
Sensitivity	0.62	0.38	0.62
Specificity	0.56	0.96	0.57
PPV	0.54	0.89	0.54
NPV	0.64	0.65	0.64
P-Value	6.42×10^{-1}	2.75×10^{-2}	7.23×10^{-1}

Table 5.4. All Chemicals Identified Between Groups from Pancreatic Cancer Study 2

PDAC vs CP		PDAC vs HC		CP vs Healthy	
RT (min)	Chemicals	RT (min)	Chemicals	RT (min)	Chemicals
1.45	Acetone	1.45	Acetone	1.45	Acetone
2.25	4-methyl-1-pentene	3.45	2-pentanone	3.45	2-pentanone
3.15	ethanethioamide	3.7	3 pentanone	3.7	3 pentanone
3.5	2 pentanone	4.5	3-hexanone	4.5	3-hexanone
4	methyl isobutylene ketone	5.35	4-heptanone	5.35	4-heptanone
4.5	3-hexanone	5.6	Allyl Isothiocyanate	5.6	Allyl Isothiocyanate
5.3	3-ethyl-2-Pentene	6.5	4-isothiocyanato-1-butene	6.5	4-isothiocyanato-1-butene
5.85	cyclohexanone	6.75	1,2,4-trimethyl-benzene	6.75	1,2,4-trimethyl-benzene
6.8	2,6-dimethyl-octane	6.8	2,6-dimethyl-octane	6.8	2,6-dimethyl-octane
6.9	2-ethyl-1-hexanol	6.9	2-ethyl-1-hexanol	6.9	Benzene, 1-ethenyl-2-methyl-
6.95	1-ethenyl-2-methyl-benzene	7	2-methyl-decane	7	2-methyl-decane
7.2	4-ethyl-1,2-dimethyl-Benzene	7.05	undecane	7.05	undecane
7.45	Nonanal	7.2	4-ethyl-1,2-dimethyl-Benzene	7.2	4-ethyl-1,2-dimethyl-Benzene
7.6	2-methyl-benzaldehyde	7.5	Nonanal	7.5	Nonanal
7.8	3,7,11-trimethyl-1-dodecanol	7.95	Dodecane	7.95	Dodecane
8.55	1,3,-bis(1,1-dimethylethyl)-benzene	8	3,7,11-trimethyl-1-dodecanol	8	3,7,11-trimethyl-1-dodecanol

Table 5.5a. Overview of the Data Generated Using All Five Classifiers Based on the 20 Most Discriminative Features in Colorectal Cancer Study 2

Classifier	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
CRC versus C						
Sparse logistic regression	0.94 (0.86-1)	1	1	0.88	1	<0.001
RandomForest	0.96 (0.89-1)	1	1	0.88	1	<0.001
GaussianProcess	0.93 (0.83-1)	1	1	0.84	1	<0.001
SupportVectorMachine	0.94 (0.85-1)	1	1	0.88	1	<0.001
NeuralNet	0.90 (0.77-1)	1	1	0.88	1	<0.001
AA versus C						
Sparse logistic regression	0.92 (0.86-0.98)	0.96	0.90	0.91	0.96	<0.001
RandomForest	0.96 (0.93-0.99)	0.96	0.93	0.93	0.96	<0.001
GaussianProcess	0.94 (0.90-0.99)	0.95	0.93	0.93	0.95	<0.001
SupportVectorMachine	0.95 (0.90-1)	0.98	0.92	0.92	0.98	<0.001

NeuralNet	0.94 (0.90-0.99)	1	0.90	0.91	1	<0.001
Large non-AA polyps (0.5-1.0 cm) versus C						
Sparse logistic regression	0.94 (0.90-0.99)	0.97	0.92	0.93	0.97	<0.001
RandomForest	0.95 (0.92-0.99)	0.98	0.91	0.91	0.98	<0.001
GaussianProcess	0.95 (0.91-0.99)	0.98	0.91	0.91	0.98	<0.001
SupportVectorMachine	0.93 (0.88-0.98)	0.98	0.91	0.91	0.98	<0.001
NeuralNet	0.95 (0.91-0.99)	0.98	0.99	0.90	0.98	<0.001
Small non-AA polyps (0.1-0.5 cm) versus C						
Sparse logistic regression	0.92 (0.89-0.96)	0.97	0.90	0.91	0.97	<0.001
RandomForest	0.96 (0.94-0.98)	0.96	0.92	0.93	0.95	<0.001
GaussianProcess	0.94 (0.91-0.98)	1	0.88	0.9	1	<0.001
SupportVectorMachine	0.94 (0.90-0.97)	0.96	0.92	0.92	0.96	<0.001
NeuralNet	0.93 (0.89-0.97)	0.98	0.90	0.91	0.98	<0.001
CRC versus AA						
Sparse logistic regression	0.61 (0.46-0.76)	0.53	0.75	0.89	0.28	0.077
RandomForest	0.54 (0.38-0.70)	0.98	0.18	0.82	0.75	0.294
GaussianProcess	0.43 (0.27-0.60)	0.93	0.12	0.81	0.33	0.226
SupportVectorMachine	0.44 (0.27-0.61)	0.87	0.18	0.81	0.27	0.756
NeuralNet	0.60 (0.46-0.74)	0.51	0.75	0.89	0.27	0.102
CRC versus Large non-AA polyps (0.5-1.0 cm)						
Sparse logistic regression	0.46 (0.31 – 0.61)	0.81	0.24	0.2	0.84	0.704
RandomForest	0.41 (0.31 – 0.51)	0.06	0.96	0.25	0.81	0.920
GaussianProcess	0.59 (0.44 – 0.74)	0.94	0.28	0.23	0.95	0.865
SupportVectorMachine	0.67 (0.53 – 0.80)	0.81	0.54	0.30	0.93	0.981
NeuralNet	0.59 (0.44 – 0.74)	0.94	0.26	0.23	0.95	0.857
CRC versus Small non-AA polyps (0.1-0.5 cm)						
Sparse logistic regression	0.64 (0.54-0.75)	1	0.373	0.168	1	0.972
RandomForest	0.41 (0.37-0.44)	1	0	0.113	NA	0.965
GaussianProcess	0.64 (0.51-0.76)	1	0.278	0.150	1	0.966
SupportVectorMachine	0.61 (0.49-0.74)	0.938	0.349	0.155	0.978	0.939
NeuralNet	0.65 (0.55-0.770)	0.875	0.476	0.175	0.968	0.973
AA versus Large non-AA polyps (0.5-1.0 cm)						
Sparse logistic regression	0.51 (0.41 - 0.61)	0.79	0.33	0.56	0.6	0.397
RandomForest	0.53 (0.43 - 0.63)	0.75	0.36	0.55	0.58	0.278
GaussianProcess	0.52 (0.42 - 0.62)	0.90	0.22	0.55	0.67	0.323
SupportVectorMachine	0.52 (0.42 - 0.62)	0.54	0.56	0.57	0.54	0.338
NeuralNet	0.47 (0.37 - 0.57)	0.56	0.5	0.54	0.52	0.729
AA versus Small non-AA polyps (0.5-1.0 cm)						
Sparse logistic regression	0.55 (0.47-0.644)	0.39	0.72	0.41	0.7	0.098
RandomForest	0.57 (0.49-0.662)	0.71	0.44	0.39	0.75	0.039
GaussianProcess	0.55 (0.46-0.640)	0.67	0.48	0.39	0.74	0.114
SupportVectorMachine	0.52 (0.43.616)	0.31	0.77	0.41	0.69	0.276
NeuralNet	0.55 (0.46-0.641)	0.70	0.46	0.39	0.75	0.112
T0 Controls versus T0 Polyp						
Sparse logistic regression	0.96 (0.907-1)	1	0.96	0.97	1	<0.001
RandomForest	0.98 (0.946-1)	1	0.96	0.97	1	<0.001
GaussianProcess	0.96 (0.907-1)	1	0.96	0.97	1	<0.001
SupportVectorMachine	0.96 (0.907-1)	1	0.96	0.97	1	<0.001
NeuralNet	0.96 (0.904-1)	0.96	0.96	0.96	0.96	<0.001
T1 Controls versus T1 Polyp						
Sparse logistic regression	0.58 (0.43-0.72)	0.28	0.96	0.81	0.56	0.135
RandomForest	0.54 (0.40-0.69)	0.90	0.25	0.54	0.82	0.256
GaussianProcess	0.54 (0.40-0.69)	0.25	0.96	0.88	0.56	0.259
SupportVectorMachine	0.48 (0.34-0.63)	0.65	0.43	0.53	0.56	0.575
NeuralNet	0.66 (0.53-0.80)	0.46	0.90	0.83	0.63	0.011
T0 Polyp versus T1 Polyp						
Sparse logistic regression	0.94 (0.88-0.99)	0.90	0.87	0.87	0.90	<0.001
RandomForest	0.93 (0.87-0.99)	0.90	0.90	0.90	0.90	<0.001
GaussianProcess	0.93 (0.86-0.99)	0.90	0.87	0.87	0.90	<0.001
SupportVectorMachine	0.89 (0.81-0.98)	0.87	0.90	0.90	0.87	<0.001
NeuralNet	0.95 (0.88-1)	0.90	0.96	0.96	0.91	<0.001
T0 Controls versus T1 Controls						
Sparse logistic regression	0.63 (0.49-0.77)	0.53	0.73	0.68	0.59	0.035
RandomForest	0.58 (0.43-0.72)	0.93	0.26	0.57	0.8	0.139

GaussianProcess	0.53 (0.38-0.68)	0.93	0.23	0.56	0.77	0.324
SupportVectorMachine	0.47 (0.32-0.62)	0.53	0.56	0.56	0.53	0.365
NeuralNet	0.62 (0.48-0.76)	0.40	0.86	0.76	0.57	0.043

Abbreviations: CRC, colorectal cancer; AA, advanced adenomas; LGD, low grade dysplasia; HC, healthy controls; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

Table 5.5b. Overview of the Data Generated Using All Five Classifiers Based on the 50 Most Discriminative Features in Colorectal Cancer Study 2

Classifier	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
CRC versus C						
Sparse logistic regression	0.96 (0.90-1)	1	0.87	0.88	1	<0.001
RandomForest	0.98 (0.96-1)	1	0.87	0.88	1	<0.001
GaussianProcess	0.96 (0.90-1)	1	0.87	0.88	1	<0.001
SupportVectorMachine	0.98 (0.95-1)	1	0.87	0.88	1	<0.001
NeuralNet	0.93 (0.82-1)	0.938	0.93	0.93	0.93	<0.001
AA versus C						
Sparse logistic regression	0.90 (0.84-0.96)	0.96	0.86	0.87	0.96	<0.001
RandomForest	0.95 (0.92-0.99)	0.95	0.93	0.93	0.95	<0.001
GaussianProcess	0.94 (0.9-0.99)	0.95	0.93	0.93	0.95	<0.001
SupportVectorMachine	0.94 (0.88-0.99)	0.95	0.93	0.93	0.95	<0.001
NeuralNet	0.92 (0.86-0.97)	0.95	0.93	0.93	0.95	<0.001
Large non-AA polyps (0.5-1.0 cm) versus C						
Sparse logistic regression	0.95 (0.91-0.99)	0.98	0.92	0.93	0.98	<0.001
RandomForest	0.97 (0.94-0.99)	0.98	0.91	0.91	0.98	<0.001
GaussianProcess	0.96 (0.92-1)	0.97	0.94	0.94	0.97	<0.001
SupportVectorMachine	0.95 (0.91-0.99)	0.98	0.92	0.93	0.98	<0.001
NeuralNet	0.93 (0.88-0.98)	0.98	0.91	0.91	0.98	<0.001
Small non-AA polyps (0.1-0.5 cm) versus C						
Sparse logistic regression	0.94 (0.90-0.98)	1	0.89	0.90	1	<0.001
RandomForest	0.97 (0.94-0.99)	1	0.88	0.9	1	<0.001
GaussianProcess	0.95 (0.92-0.98)	0.98	0.91	0.91	0.98	<0.001
SupportVectorMachine	0.95 (0.91-0.98)	0.96	0.92	0.93	0.96	<0.001
NeuralNet	0.93 (0.90-0.97)	0.98	0.91	0.91	0.98	<0.001
CRC versus AA						
Sparse logistic regression	0.60 (0.46-0.74)	0.40	0.87	0.92	0.27	0.916
RandomForest	0.53 (0.37-0.69)	1	0.18	0.83	1	0.319
GaussianProcess	0.47 (0.31-0.64)	0.96	0.12	0.81	0.5	0.389
SupportVectorMachine	0.58 (0.41-0.76)	0.90	0.31	0.84	0.45	0.560
NeuralNet	0.54 (0.37-0.70)	0.40	0.75	0.86	0.24	0.309
CRC versus Large non-AA polyps (0.5-1.0 cm)						
Sparse logistic regression	0.520 (0.36-0.67)	0.43	0.64	0.22	0.83	0.601
RandomForest	0.432 (0.32-0.54)	0.06	0.97	0.33	0.81	0.855
GaussianProcess	0.546 (0.39-0.69)	0.87	0.30	0.23	0.91	0.718
SupportVectorMachine	0.582 (0.41-0.74)	0.56	0.60	0.25	0.85	0.846
NeuralNet	0.521 (0.37-0.67)	0.87	0.26	0.21	0.9	0.604
CRC versus Small non-AA polyps (0.1-0.5 cm)						
Sparse logistic regression	0.67 (0.54-0.79)	0.938	0.48	0.18	0.98	0.987
RandomForest	0.40 (0.32-0.47)	1	0	0.11	NA	0.959
GaussianProcess	0.57 (0.44-0.71)	0.93	0.34	0.15	0.97	0.847
SupportVectorMachine	0.42 (0.28-0.57)	0.62	0.42	0.12	0.89	0.824
NeuralNet	0.47 (0.34-0.60)	0.93	0.20	0.13	0.96	0.622
AA versus Large non-AA polyps (0.5-1.0 cm)						
Sparse logistic regression	0.54 (0.44 – 0.64)	0.85	0.30	0.56	0.66	0.202
RandomForest	0.57 (0.48 – 0.67)	0.76	0.42	0.58	0.63	0.072
GaussianProcess	0.58 (0.49 – 0.68)	0.34	0.80	0.64	0.53	0.050
SupportVectorMachine	0.58 (0.49 – 0.68)	0.60	0.58	0.60	0.58	0.047
NeuralNet	0.55 (0.45 – 0.65)	0.72	0.42	0.57	0.59	0.167
T0 Controls versus T0 Polyp						

Sparse logistic regression	0.96 (0.90-1)	1	0.96	0.97	1	<0.001
RandomForest	0.98 (0.94-1)	1	0.96	0.97	1	<0.001
GaussianProcess	0.96 (0.90-1)	1	0.96	0.97	1	<0.001
SupportVectorMachine	0.96 (0.90-1)	1	0.96	0.97	1	<0.001
NeuralNet	0.94 (0.86-1)	0.969	0.96	0.96	0.96	<0.001
T1 Controls versus T1 Polyp						
Sparse logistic regression	0.56 (0.42-0.70)	0.87	0.40	0.59	0.76	0.182
RandomForest	0.56 (0.41-0.70)	0.90	0.28	0.55	0.75	0.200
GaussianProcess	0.53 (0.38-0.67)	0.68	0.46	0.56	0.6	0.326
SupportVectorMachine	0.60 (0.46-0.75)	0.81	0.50	0.61	0.72	0.072
NeuralNet	0.57 (0.43-0.72)	0.75	0.53	0.61	0.68	0.141
T0 Polyp versus T1 Polyp						
Sparse logistic regression	0.96 (0.91-1)	0.93	0.90	0.91	0.93	<0.001
RandomForest	0.96 (0.92-1)	0.96	0.87	0.88	0.96	<0.001
GaussianProcess	0.95 (0.90-1)	0.96	0.90	0.91	0.96	<0.001
SupportVectorMachine	0.92 (0.84-1)	0.93	0.93	0.93	0.93	<0.001
NeuralNet	0.96 (0.90-1)	0.93	0.96	0.96	0.94	<0.001
T0 Controls versus T1 Controls						
Sparse logistic regression	0.75 (0.63-0.87)	0.46	0.96	0.93	0.63	0.000
RandomForest	0.68 (0.54-0.81)	0.56	0.73	0.69	0.61	0.008
GaussianProcess	0.59 (0.45-0.74)	0.68	0.56	0.62	0.63	0.092
SupportVectorMachine	0.62 (0.48-0.76)	0.40	0.86	0.76	0.57	0.052
NeuralNet	0.72 (0.59-0.86)	0.71	0.83	0.82	0.73	0.001

Abbreviations: CRC, colorectal cancer; AA, advanced adenomas; LGD, low grade dysplasia; HC, healthy controls; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

Table 5.5c. Overview of the Data Generated Using All Five Classifiers Based on the 100 Most Discriminative Features in Colorectal Cancer Study 2

Classifier	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
CRC versus C						
Sparse logistic regression	0.97 (927-1)	1	0.87	0.88	1	<0.001
RandomForest	0.98 (0.96-1)	1	0.87	0.88	1	<0.001
GaussianProcess	0.96 (0.91-1)	1	0.87	0.88	1	<0.001
SupportVectorMachine	0.99 (0.97-1)	1	0.93	0.94	1	<0.001
NeuralNet	1 (1-1)	1	1	1	1	<0.001
AA versus C						
Sparse logistic regression	0.92 (0.87-0.98)	0.98	0.87	0.88	0.98	<0.001
RandomForest	0.95 (0.91-0.99)	0.95	0.93	0.93	0.95	<0.001
GaussianProcess	0.94 (0.88-0.99)	0.95	0.93	0.93	0.95	<0.001
SupportVectorMachine	0.93 (0.88-0.99)	0.95	0.93	0.93	0.95	<0.001
NeuralNet	0.91 (0.85-0.97)	0.95	0.93	0.93	0.95	<0.001
Large non-AA polyps (0.5-1.0 cm) versus C						
Sparse logistic regression	0.92 (0.87-0.98)	0.98	0.91	0.91	0.98	<0.001
RandomForest	0.96 (0.92-0.99)	0.98	0.91	0.91	0.98	<0.001
GaussianProcess	0.95 (0.91-1)	0.98	0.94	0.94	0.98	<0.001
SupportVectorMachine	0.93 (0.88-0.99)	0.98	0.94	0.94	0.98	<0.001
NeuralNet	0.95 (0.90-0.99)	0.98	0.92	0.93	0.98	<0.001
Small non-AA polyps (0.1-0.5 cm) versus C						
Sparse logistic regression	0.95 (0.92-0.98)	0.98	0.91	0.91	0.98	<0.001
RandomForest	0.97 (0.95-0.99)	0.95	0.95	0.95	0.95	<0.001
GaussianProcess	0.96 (0.93-0.99)	0.91	0.92	1		<0.001
SupportVectorMachine	0.96 (0.93-0.99)	0.97	0.94	0.94	0.97	<0.001
NeuralNet	0.94 (0.90-0.97)	0.98	0.92	0.93	0.98	<0.001
CRC versus AA						
Sparse logistic regression	0.58 (0.41-0.75)	0.48	0.68	0.86	0.25	0.153
RandomForest	0.57 (0.42-0.73)	0.57	0.62	0.86	0.27	0.167

GaussianProcess	0.48 (0.32-0.64)	0.87	0.18	0.81	0.27	0.407
SupportVectorMachine	0.49 (0.32-0.67)	0.87	0.25	0.82	0.33	0.517
NeuralNet	0.61 (0.45-0.77)	0.62	0.68	0.88	0.31	0.081
CRC versus Large non-AA polyps (0.5-1.0 cm)						
Sparse logistic regression	0.49 (0.33-0.65)	0.43	0.64	0.22	0.83	0.530
RandomForest	0.44 (0.32-0.56)	0.06	0.98	0.5	0.81	0.792
GaussianProcess	0.54 (0.39-0.69)	0.93	0.22	0.22	0.93	0.714
SupportVectorMachine	0.57 (0.41-0.74)	0.31	0.82	0.29	0.83	0.830
NeuralNet	0.46 (0.30-0.61)	0.5	0.57	0.21	0.83	0.692
CRC versus Small non-AA polyps (0·1-0·5 cm)						
Sparse logistic regression	0.66 (0.52-0.80)	0.68	0.66	0.20	0.94	0.986
RandomForest	0.43 (0.32-0.54)	0.18	0.82	0.12	0.88	0.852
GaussianProcess	0.46 (0.31-0.60)	0.87	0.19	0.12	0.92	0.704
SupportVectorMachine	0.56 (0.38-0.83)	0.5	0.68	0.16	0.91	0.783
NeuralNet	0.49 (0.34-0.65)	0.56	0.58	0.14	0.91	0.515
AA versus Large non-AA polyps (0.5-1.0 cm)						
Sparse logistic regression	0.45 (0.30-0.60)	0.81	0.23	0.2	0.84	0.704
RandomForest	0.41 (0.31-0.51)	0.06	0.95	0.25	0.81	0.920
GaussianProcess	0.58 (0.44-0.73)	0.93	0.27	0.23	0.95	0.865
SupportVectorMachine	0.66 (0.53-0.80)	0.81	0.54	0.29	0.92	0.981
NeuralNet	0.58 (0·43-0.73)	0.93	0.26	0.23	0.94	0.857
AA versus Small non-AA polyps (0.5-1.0 cm)						
Sparse logistic regression	0.61 (0·.53-0.69)	0.80	0.45	0.42	0.81	0.006
RandomForest	0.56 (0.47-0.65)	0.32	0.84	0.51	0.71	0.070
GaussianProcess	0.60 (0.52-0.68)	0.71	0.49	0.41	0.77	0.009
SupportVectorMachine	0.56 (0.47-0.65)	0.43	0.69	0.42	0.71	0.068
NeuralNet	0.54 (0.45-0.62)	0.57	0.59	0.42	0.73	0.181
T0 Controls versus T0 Polyp						
Sparse logistic regression	0.96 (0.90-1)	1	0.96	0.97	1	<0.001
RandomForest	0.97 (0.93-1)	1	0.96	0.97	1	<0.001
GaussianProcess	0.96 (0.90-1)	1	0.96	0.97	1	<0.001
SupportVectorMachine	0.96 (0.90-1)	1	0.96	0.97	1	<0.001
NeuralNet	0.94 (0.87-1)	0.96	0.96	0.96	0.96	<0.001
T1 Controls versus T1 Polyp						
Sparse logistic regression	0.56 (0.41-0.70)	0.87	0.31	0.56	0.71	0.208
RandomForest	0.54 (0.40-0.69)	0.90	0.25	0.54	0.72	0.260
GaussianProcess	0.55 (0.41-0.70)	0.71	0.46	0.57	0.62	0.210
SupportVectorMachine	0.58 (0.44-0.73)	0.81	0.46	0.60	0.71	0.113
NeuralNet	0.55 (0.40-0.69)	0.56	0.62	0.62	0.6	0.243
T0 Polyp versus T1 Polyp						
Sparse logistic regression	0.98 (0.95-1)	1	0.90	0.91	1	<0.001
RandomForest	0.98 (0.95-1)	1	0.84	0.86	1	<0.001
GaussianProcess	0.97 (0.95-1)	1	0.87	0.88	1	<0.001
SupportVectorMachine	0.93 (0.85-1)	0.96	0.90	0.81	0.96	<0.001
NeuralNet	0.98 (0.96-1)	0.96	0.96	0.96	0.96	<0.001
T0 Controls versus T1 Controls						
Sparse logistic regression	0.80 (0.69-0.91)	0.71	0.9	0.88	0.75	<0.001
RandomForest	0.85 (0.62-0.87)	0.59	0.86	0.82	0.66	<0.001
GaussianProcess	0.70 (0.57-0.83)	0.53	0.83	0.77	0.62	0.002
SupportVectorMachine	0.66 (0.53-0.80)	0.59	0.73	0.70	0.62	0.002
NeuralNet	0.77 (0.65-0.89)	0.59	0.93	0.90	0.68	<0.001

Abbreviations: CRC, colorectal cancer; AA, advanced adenomas; LGD, low grade dysplasia; HC, healthy controls; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

Table 6.1a Performance Characteristics with Corresponding Area Under the Curve, Sensitivity, Specificity, Positive and Negative Predictive Value of Faecal Volatile Organic Compound Analysis for the Discrimination of Irritable Bowel Syndrome, Functional Abdominal Pain-Not Otherwise Specified, Inflammatory Bowel Disease and Healthy Controls Using the Supervised Classification Method Sparse Logistic Regression

Analysis	p-value	AUC (± 95% CI)	Sensitivity (± 95% CI)	Specificity (± 95% CI)	PPV	NPV
IBS/FAP-NOS vs IBD	0.0000002113	0.92 (0.84 – 1)	0.93 (0.68 – 1)	0.8 (0.61 – 0.92)	0.7	0.96
IBS/FAP-NOS vs CD	0.001837	0.8 (0.63 - 0.98)	0.8 (0.52 - 0.96)	0.8 (0.52 - 0.96)	0.8	0.8
IBS/FAP-NOS vs UC	0.000002405	0.94 (0.84 - 1)	1 (0.78 - 1)	0.87 (0.6 - 0.98)	0.88	1
IBS/FAP-NOS vs HC	0.2191	0.57 (0.4 - 0.75)	0.8 (0.52 - 0.96)	0.4 (0.23 - 0.59)	0.4	0.8
IBS vs FAP-NOS	0.9504	0.76 (0.44 - 1)	1 (0.69 - 1)	0.6 (0.15 - 0.95)	0.83	1
IBD vs HC	0.000000000000560	0.96 (0.92 - 1)	0.93 (0.78 - 0.99)	0.93 (0.78 - 0.99)	0.93	0.93
UC vs HC	0.0000000005654	0.98 (0.94 – 1)	0.93 (0.68 - 1)	0.97 (0.83 - 1)	0.93	0.97
CD vs HC	0.00000001285	0.95 (0.9 - 1)	1 (0.78 - 1)	0.73 (0.54 - 0.88)	0.65	1
CD vs UC	0.05799	0.67 (0.47 – 0.88)	0.6 (0.32 – 0.84)	0.8 (0.52 – 0.96)	0.75	0.67

Table 6.1b Performance Characteristics with Corresponding Area Under the Curve, Sensitivity, Specificity, Positive and Negative Predictive Value of Faecal Volatile Organic Compound Analysis for the Discrimination of Irritable Bowel Syndrome, Functional Abdominal Pain-Not Otherwise Specified, Inflammatory Bowel Disease and Healthy Controls Using the Supervised Classification Method Random Forrest

Analysis	p-value	AUC (± 95% CI)	Sensitivity (± 95% CI)	Specificity (± 95% CI)	PPV	NPV
IBS/FAP-NOS vs IBD	0.000001351	0.93 (0.85 - 1)	0.93 (0.68 - 1)	0.83 (0.65 - 0.94)	0.74	0.96
IBS/FAP-NOS vs CD	0.000399	0.86 (0.72 - 0.99)	1 (0.78 - 1)	0.6 (0.32 - 0.84)	0.71	1
IBS/FAP-NOS vs UC	0.000007501	0.96 (0.91 - 1)	1 (0.78 - 1)	0.8 (0.52 - 0.96)	0.83	1
IBS/FAP-NOS vs HC	0.2817	0.55 (0.37 - 0.74)	0.8 (0.52 - 0.96)	0.4 (0.23 - 0.59)	0.4	0.8
IBS vs FAP-NOS	0.703	0.42 (0.039 - 0.8)	0.7 (0.35 - 0.93)	0.6 (0.15 - 0.95)	0.78	0.5
IBD vs HC	0.0000000003982	0.96 (0.9 - 1)	0.93 (0.78 - 0.99)	0.97 (0.83 - 1)	0.97	0.94
UC vs HC	0.0000001994	0.96 (0.91 - 1)	0.93 (0.68 - 1)	0.97 (0.83 - 1)	0.93	0.97
CD vs HC	0.000002138	0.92 (0.82 - 1)	0.93 (0.68 - 1)	0.87 (0.69 - 0.96)	0.78	0.96
CD vs UC	0.52 (0.3 – 0.73)	0.52 (0.3 – 0.73)	1 (0.79 – 1)	0.13 (0.017 – 0.4)	0.54	1

Table 6.1c Performance Characteristics with Corresponding Area Under the Curve, Sensitivity, Specificity, Positive and Negative Predictive Value of Faecal Volatile Organic Compound Analysis for the Discrimination of Irritable Bowel Syndrome, Functional Abdominal Pain-Not Otherwise Specified, Inflammatory Bowel Disease and Healthy Controls Using the Supervised Classification Method Gaussian Process

Analysis	p-value	AUC (± 95% CI)	Sensitivity (± 95% CI)	Specificity (± 95% CI)	PPV	NPV
IBS/FAP-NOS vs IBD	0.00000002613	0.94 (0.88 - 1)	1 (0.78 - 1)	0.87 (0.69 - 0.96)	0.79	1
IBS/FAP-NOS vs CD	0.0001617	0.87 (0.73 - 1)	0.93 (0.68 - 1)	0.8 (0.52 - 0.96)	0.82	0.92
IBS/FAP-NOS vs UC	0.000003275	0.94 (0.85 - 1)	1 (0.78 - 1)	0.87 (0.6 - 0.98)	0.88	1
IBS/FAP-NOS vs HC	0.1667	0.59 (0.41 - 0.77)	0.6 (0.32 - 0.84)	0.63 (0.44 - 0.8)	0.45	0.76
IBS vs FAP-NOS	0.5704	0.48 (0.14 - 0.82)	0.5 (0.19 - 0.81)	0.8 (0.28 - 0.99)	0.83	0.44
IBD vs HC	0.000000000009127	0.95 (0.88 - 1)	0.93 (0.78 - 0.99))	0.93 (0.78 - 0.99)	0.93	0.93
UC vs HC	0.0000000005654	0.98 (0.94 - 1)	0.93 (0.68 - 1)	0.93 (0.78 - 0.99)	0.88	0.97
CD vs HC	0.0000001173	0.93 (0.86 - 1)	0.87 (0.6 - 0.98)	0.9 (0.73 - 0.98)	0.81	0.93
CD vs UC	0.6587	0.46 (0.24 – 0.67)	1 (0.78 – 1)	0.067 (0.0017 – 0.32)	0.52	1

Table 6.1d Performance Characteristics with Corresponding Area Under the Curve, Sensitivity, Specificity, Positive and Negative Predictive Value of Faecal Volatile Organic Compound Analysis for the Discrimination of Irritable Bowel Syndrome, Functional Abdominal Pain-Not Otherwise Specified, Inflammatory Bowel Disease and Healthy Controls Using the Supervised Classification Method Support Vector Machine

Analysis	p-value	AUC (± 95% CI)	Sensitivity (± 95% CI)	Specificity (± 95% CI)	PPV	NPV
IBS/FAP-NOS vs IBD	0.000002783	0.89 (0.78 - 1)	1 (0.78 - 1)	0.8 (0.61 - 0.92)	0.71	1
IBS/FAP-NOS vs CD	0.01175	0.74 (0.54 - 0.94)	0.93 (0.68 - 1)	0.67 (0.38 - 0.88)	0.74	0.91
IBS/FAP-NOS vs UC	0.00001021	0.92 (0.79 - 1)	0.93 (0.68 - 1)	0.93 (0.68 - 1)	0.93	0.93
IBS/FAP-NOS vs HC	0.2878	0.45 (0.26 - 0.63)	0.53 (0.27 - 0.79)	0.57 (0.37 - 0.75)	0.38	0.71
IBS vs FAP-NOS	0.9354	0.74 (0.42 - 1)	0.9 (0.55 - 1)	0.6 (0.15 - 0.95)	0.82	0.75
IBD vs HC	0.0000000001787	0.92 (0.84 - 1)	0.93 (0.78 - 0.99)	0.9 (0.73 - 0.98)	0.9	0.93
UC vs HC	0.00000009573	0.93 (0.85 - 1)	0.93 (0.68 - 1)	0.93 (0.78 - 0.99)	0.88	0.97
CD vs HC	0.00000001636	0.95 (0.88 - 1)	0.93 (0.68 - 1)	0.93 (0.78 - 0.99)	0.88	0.97
CD vs UC	0.2062	0.59 (0.38 – 0.81)	0.6 (0.32 – 0.84)	0.73 (0.45 – 0.92)	0.69	0.65

Table 6.2a. Statistical Analyses of Faecal VOC Profiles for the Discrimination of IBD from Controls

	AUC [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	PPV	NPV	p-value	Threshold
SparseLogistic Regression	0.53 [0.25-0.81]	0.4 [0.12-0.74]	0.8 [0.44-0.97]	0.67	0.57	0.602	0.140
Random Forest	0.53 [0.24-0.82]	0.3 [0.067-0.65]	1.0 [0.69-1.00]	1.00	0.59	0.425	0.936
Gaussian Process	0.69 [0.44-0.94]	0.9 [0.55–1.00]	0.5 [0.19-0.81]	0.64	0.83	0.083	0.435
SupportVector Machine	0.60 [0.34-0.86]	0.8 [0.44-0.97]	0.5 [0.19-0.81]	0.62	0.71	0.218	0.618
Neural Network	0.73 [0.47-0.99]	0.7 [0.35-0.93]	0.9 [0.55-1.00]	0.88	0.75	0.038	0.499

Here, all the statistical analyses performed on the fecal VOC profiles are listed. For these analyses, 100 features were used in the 10 fold cross validation. The neural network analysis is identified as the best performing algorithm for the discrimination of IBD from controls based on fecal VOC profiles. Abbreviations**:** AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; IBD, inflammatory bowel disease; VOC, volatile organic compounds.

Table 6.2b. Statistical Analyses of Urinary VOC Profiles for the Discrimination of IBD from Controls (100 features 10 fold cross validation)

	AUC [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	PPV	NPV	p-value	Threshold
Sparse logistic regression	0.78 [0.57-1.00]	0.8 [0.44-0.97]	0.7 [0.35-0.93]	0.73	0.78	0.028	0.502
Random forest	0.69 [0.44-0.94]	1.0 [0.69–1]	0.5 [0.19-0.81]	0.67	1.00	0.070	0.876
Gaussian Process	0.35 [0.09-0.61]	0.6 [0.26-0.88]	0.5 [0.19-0.81]	0.55	0.56	0.879	0.542
Support Vector Machine	0.47 [0.19-0.75]	0.4 [0.12-0.74]	0.8 [0.44-0.97]	0.67	0.57	0.604	0.608
Neural Network	0.41 [0.14-0.68]	0.6 [0.26-0.88]	0.5 [0.19-0.81]	0.55	0.56	0.764	0.293

Here, all the statistical analyses performed on the urinary VOC profiles are listed. For these analyses, 100 features were used in the 10 fold cross validation. The Sparse logistic regression

analysis is identified as the best performing algorithm for the discrimination of IBD from controls based on fecal VOC profiles. Abbreviations**:** AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; IBD, inflammatory bowel disease; VOC, volatile organic compounds.

Table 6.3a. Overview of the Data Generated Using All Five Classifiers Based on the 20 most Discriminative Features Using All IBD patients

	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
IBD versus HC						
Sparse Logistic Regression	0.94 (0.90 – 0.98)	0.99	0.88	0.97	0.96	<0.0001
Random Forest	0.95 (0.92 - 0.99)	0.97	0.92	0.98	0.89	<0.0001
Gaussian Process	0.96 (0.92 – 0.99)	0.99	0.97	0.92	0.98	<0.0001
Support Vector Machine	0.96 (0.93 -0.99)	0.97	0.92	0.98	0.89	<0.0001
Neural Net	0.96 (0.93 – 0.99)	0.97	0.92	0.98	0.89	<0.0001
CD versus HC						
Sparse Logistic Regression	0.95(0.91-0.98)	0.97	0.91	0.94	0.96	<0.0001
Random Forest	0.95 (0.92 – 0.99)	0.98	0.89	0.93	0.97	<0.0001
Gaussian Process	0.95 (0.92 – 0.99)	0.97	0.92	0.95	0.96	<0.0001
Support Vector Machine	0.95 (0.92 – 0.99)	0.97	0.91	0.94	0.96	<0.0001
Neural Net	0.95 (0.90 – 0.99)	0.97	0.91	0.94	0.96	<0.0001
CDa versus HC						
Sparse Logistic Regression	0.95 (0.93 - 0.97)	1	0.90	0.71	1	<0.0001
Random Forest	0.97 (0.95 - 0.99)	1	0.95	0.82	1	<0.0001
Gaussian Process	0.97 (0.96 - 0.99)	1	0.95	0.82	1	<0.0001
Support Vector Machine	0.97 (0.96 - 0.99)	1	0.95	0.82	1	<0.0001
Neural Net	0.94 (0.91 - 0.97)	1	0.92	0.74	1	<0.0001
CDr versus HC						
Sparse Logistic Regression	0.95 (0.92 – 0.97)	1	0.90	0.67	1	<0.0001
Random Forest	0.97 (0.95 – 0.99)	1	0.92	0.71	1	<0.0001
Gaussian Process	0.97 (0.95 - 0.99)	1	0.94	0.76	1	<0.0001
Support Vector Machine	0.97 (0.95 - 0.99)	1	0.92	0.72	1	<0.0001
Neural Net	0.97 (0.95 - 0.99)	1	0.94	0.76	1	<0.0001
UC versus HC						
Sparse Logistic Regression	0.95 (0.92 – 0.99)	1	0.91	0.88	1	<0.0001
Random Forest	0.96 (0.94 – 0.99)	1	0.93	0.91	1	<0.0001
Gaussian Process	0.96 (0.93 – 0.99)	1	0.92	0.90	1	<0.0001
Support Vector Machine	0.96 (0.94 – 0.99)	1	0.93	0.91	1	<0.0001
Neural Net	0.96 (0.94 – 0.99)	1	0.93	0.91	1	<0.0001
UCa versus HC						
Sparse Logistic Regression	0.95 (0.93 - 0.98)	0.97	0.90	0.67	0.99	<0.0001
Random Forest	0.97 (0.95 - 0.99)	1	0.91	0.69	1	<0.0001
Gaussian Process	0.97 (0.95 - 0.99)	1	0.90	0.67	1	<0.0001
Support Vector Machine	0.96 (0.93 – 0.93)	0.99	0.97	0.68	0.99	<0.0001
Neural Net	0.97 (0.95 - 0.99)	1	0.90	0.67	1	<0.0001
UCr versus HC						
Sparse Logistic Regression	0.94 (0.91 - 0.96)	1	0.87	0.51	1	<0.0001
Random Forest	0.96 (0.94 - 0.98)	1	0.90	0.56	1	<0.0001
Gaussian Process	0.96 (0.94 - 0.98)	0.96	0.93	0.66	0.99	<0.0001
Support Vector Machine	0.96 (0.94 - 0.98)	0.96	0.93	0.66	0.99	<0.0001
Neural Net	0.96 (0.94 - 0.98)	0.96	0.93	0.66	0.99	<0.0001
IBDa versus IBDr						
Sparse Logistic Regression	0.57 (0.50 - 0.64)	0.27	0.8	0.8	0.39	0.034
Random Forest	0.57 (0.50 - 0.64)	0.27	0.8	0.8	0.39	0.035
Gaussian Process	0.57 (0.51 - 0.64)	0.27	0.8	0.8	0.39	0.025
Support Vector Machine	0.56 (0.49 - 0.63)	0.26	0.9	0.8	0.39	0.054
Neural Net	0.55 (0.48 - 0.62)	0.23	0.92	0.84	0.39	0.095
CDa versus CDr						
Sparse Logistic Regression	0.50 (0.93 - 0.60)	0.19	0.9	0.79	0.37	0.515
Random Forest	0.51 (0.41 - 0.62)	0.28	0.83	0.76	0.38	0.425
Gaussian Process	0.51 (0.40 - 0.61)	0.11	1	1	0.37	0.442
Support Vector Machine	0.51 (0.41 - 0.62)	0.28	0.83	0.76	0.38	0.425
Neural Net	0.51 (0.41 - 0.62)	0.26	0.87	0.79	0.38	0.406
UCa versus UCr						
Sparse Logistic Regression	0.58 (0.41 - 0.75)	0.58	0.64	0.79	0.39	0.194
Random Forest	0.64 (0.46 - 0.82)	0.73	0.57	0.8	0.47	0.062

Gaussian Process	0.61 (0.42 - 0.79)	0.64	0.64	0.81	0.43	0.126
Support Vector Machine	0.63 (0.44 - 0.82)	0.67	0.64	0.81	0.45	0.076
Neural Net	0.61 (0.44 - 0.78)	0.61	0.643	0.8	0.41	0.117
CD versus UC						
Sparse Logistic Regression	0.55 (0.51 - 0.60)	0.17	0.96	0.9	0.34	0.030
Random Forest	0.54 (0.49 - 0.60)	0.14	0.98	0.9375	0.36	0.069
Gaussian Process	0.54 (0.49 - 0.60)	0.09	1	1	0.35	0.077
Support Vector Machine	0.54 (0.49 - 0.60)	0.11	0.98	0.92	0.35	0.073
Neural Net	0.54 (0.49 - 0.60)	0.16	0.94	0.85	0.35	0.080
CDa versus UCa						
Sparse Logistic Regression	0.50 (0.37 - 0.63)	0.67	0.45	0.68	0.44	0.496
Random Forest	0.52 (0.39 - 0.65)	0.79	0.36	0.68	0.5	0.610
Gaussian Process	0.48 (0.35 - 0.61)	0.952	0.12	0.65	0.57	0.605
Support Vector Machine	0.51 (0.38 - 0.64)	0.67	0.45	0.68	0.44	0.463
Neural Net	0.48 (0.35 - 0.61)	0.60	0.51	0.68	0.43	0.605
CDr versus UCr						
Sparse Logistic Regression	0.5 (0.5 - 0.5)	1	0	0.71	N/A	1
Random Forest	0.58 (0.40 - 0.76)	0.29	0.93	0.91	0.35	0.811
Gaussian Process	0.53 (0.35 - 0.70)	0.5	0.71	0.81	0.37	0.624
Support Vector Machine	0.57 (0.39 - 0.75)	0.44	0.79	0.83	0.37	0.780
Neural Net	0.64 (0.46 - 0.82)	0.71	0.57	0.8	0.44	0.936

Abbreviations: IBD inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; a, active disease state; r, remission; HC, healthy controls; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

Table 6.3b. Overview of the Data Generated Using All Four Classifiers Based on the 50 Most Discriminative Features Using All IBD patients

	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
IBD versus HC						
Sparse Logistic Regression	0.95 (0.92 – 0.99)	0.99	0.91	0.98	0.96	<0.0001
Random Forest	0.97 (0.94 – 1)	0.97	0.93	0.98	0.88	<0.0001
Gaussian Process	0.97 (0.94 – 1)	0.97	0.92	0.98	0.88	<0.0001
Support Vector Machine	0.97 (0.95 – 1)	0.95	0.95	0.99	0.82	<0.0001
Neural Net	0.96 (0.92 – 0.99)	0.96	0.93	0.98	0.85	<0.0001
CD versus HC						
Sparse Logistic Regression	0.96 (0.93 – 0.99)	0.96	0.93	0.95	0.95	<0.0001
Random Forest	0.97 (0.95 – 1)	0.97	0.93	0.95	0.96	<0.0001
Gaussian Process	0.97 (0.94 – 1)	0.97	0.93	0.95	0.96	<0.0001
Support Vector Machine	0.97 (0.95 – 1)	0.98	0.92	0.95	0.97	<0.0001
Neural Net	0.96 (0.92 – 1)	0.97	0.93	0.95	0.96	<0.0001
CDa versus HC						
Sparse Logistic Regression	0.96 (0.94 - 0.98)	1	0.92	0.76	1	<0.0001
Random Forest	0.97 (0.96 - 0.99)	1	0.95	0.82	1	<0.0001
Gaussian Process	0.97 (0.96 - 0.99)	1	0.95	0.82	1	<0.0001
Support Vector Machine	0.97 (0.95 - 0.99)	1	0.94	0.80	1	<0.0001
Neural Net	0.97 (0.95 - 0.99)	1	0.94	0.80	1	<0.0001
CDr versus HC						
Sparse Logistic Regression	0.96 (0.94 - 0.98)	1	0.91	0.68	1	<0.0001
Random Forest	0.97 (0.94 – 0.99)	1	0.92	0.71	1	<0.0001
Gaussian Process	0.97 (0.94 - 0.99)	1	0.92	0.72	1	<0.0001
Support Vector Machine	0.97 (0.94 - 0.99)	1	0.93	0.74	1	<0.0001
Neural Net	0.97 (0.95 - 0.99)	1	0.93	0.74	1	<0.0001
UC versus HC						
Sparse Logistic Regression	0.95 (0.92 - 0.99)	1	0.91	0.88	1	<0.0001
Random Forest	0.97 (0.94 – 1)	1	0.93	0.91	1	<0.0001
Gaussian Process	0.97 (0.94 – 1)	1	0.92	0.90	1	<0.0001
Support Vector Machine	0.97 (0.94 – 1)	1	0.91	0.88	1	<0.0001
Neural Net	0.97 (0.94 – 1)	0.98	0.93	0.91	0.99	<0.0001
UCa versus HC						
Sparse Logistic Regression	0.97 (0.95 - 0.99)	1	0.90	0.66	1	<0.0001
Random Forest	0.97 (0.95 - 0.99)	1	0.92	0.70	1	<0.0001
Gaussian Process	0.97 (0.95 - 0.99)	1	0.91	0.69	1	<0.0001

Support Vector Machine	0.97 (0.95 - 0.99)	1	0.90	0.67	1	<0.0001
Neural Net	0.97 (0.95 - 0.99)	1	0.91	0.69	1	<0.0001
UCr versus HC						
Sparse Logistic Regression	0.95 (0.92 - 0.97)	1	0.88	0.52	1	<0.0001
Random Forest	0.97 (0.95 - 0.99)	0.96	0.95	0.70	0.99	<0.0001
Gaussian Process	0.97 (0.95 - 0.99)	0.96	0.95	0.70	0.99	<0.0001
Support Vector Machine	0.97 (0.95 - 0.99)	0.96	0.95	0.70	0.99	<0.0001
Neural Net	0.97 (0.95 - 0.99)	0.96	0.95	0.70	0.99	<0.0001
IBDa versus IBDr						
Sparse Logistic Regression	0.56 (0.50 - 0.63)	0.26	0.88	0.79	0.39	0.042
Random Forest	0.57 (0.50 - 0.65)	0.3	0.85	0.79	0.39	0.028
Gaussian Process	0.56 (0.49 - 0.63)	0.24	0.90	0.81	0.39	0.080
Support Vector Machine	0.57 (0.50 - 0.64)	0.26	0.90	0.82	0.39	0.040
Neural Net	0.53 (0.46 - 0.60)	0.23	0.88	0.78	0.38	0.215
CDa versus CDr						
Sparse Logistic Regression	0.53 (0.41 - 0.65)	0.18	0.93	0.83	0.37	0.292
Random Forest	0.58 (0.44 - 0.70)	0.52	0.63	0.73	0.41	0.116
Gaussian Process	0.56 (0.43 - 0.69)	0.51	0.67	0.74	0.42	0.171
Support Vector Machine	0.56 (0.44 - 0.69)	0.58	0.57	0.72	0.41	0.152
Neural Net	0.62 (0.50 - 0.74)	0.49	0.7	0.76	0.42	0.032
UCa versus UCr						
Sparse Logistic Regression	0.57 (0.39 – 0.75)	0.33	0.86	0.85	0.35	0.230
Random Forest	0.63 (0.45 – 0.81)	0.82	0.43	0.77	0.50	0.090
Gaussian Process	0.57 (0.36 – 0.78)	0.85	0.43	0.78	0.55	0.231
Support Vector Machine	0.56 (0.36 – 0.76)	0.85	0.43	0.78	0.55	0.261
Neural Net	0.54 (0.34 – 0.74)	0.82	0.43	0.77	0.50	0.350
CD versus UC						
Sparse Logistic Regression	0.55 (0.51 - 0.60)	0.17	0.96	0.9	0.36	0.030
Random Forest	0.54 (0.49 - 0.60)	0.16	0.96	0.89	0.36	0.072
Gaussian Process	0.54 (0.49 - 0.60)	0.17	0.92	0.83	0.35	0.078
Support Vector Machine	0.54 (0.49 - 0.59)	0.17	0.92	0.83	0.35	0.093
Neural Net	0.54 (0.49 - 0.59)	0.16	0.92	0.81	0.35	0.088
CDa versus UCa						
Sparse Logistic Regression	0.50 (0.37 - 0.63)	0.65	0.48	0.69	0.44	0.500
Random Forest	0.53 (0.40 - 0.67)	0.79	0.36	0.68	0.50	0.305
Gaussian Process	0.49 (0.36 - 0.63)	0.95	0.15	0.66	0.63	0.540
Support Vector Machine	0.51 (0.38 - 0.64)	0.42	0.70	0.71	0.41	0.427
Neural Net	0.52 (0.39 - 0.65)	0.60	0.58	0.71	0.45	0.374
CDr versus UCr						
Sparse Logistic Regression	0.43 (0.25 - 0.61)	0.5	0.57	0.74	0.32	0.773
Random Forest	0.45 (0.27 - 0.64)	0.09	1	1	0.31	0.691
Gaussian Process	0.55 (0.38 - 0.72)	0.5	0.79	0.85	0.39	0.715
Support Vector Machine	0.55 (0.36 - 0.73)	0.53	0.64	0.78	0.36	0.691
Neural Net	0.50 (0.31 - 0.69)	0.82	0.36	0.76	0.45	0.500

Abbreviations: IBD inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; a, active disease state; r, remission; HC, healthy controls; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

Table 6.3c. Overview of the Data Generated Using All Four Classifiers Based on the 100 Most Discriminative Features Using All IBD patients

	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
IBD versus HC						
Sparse Logistic Regression	0.96 (0.92 - 0.99)	0.97	0.92	0.98	0.87	<0.0001
Random Forest	0.97 (0.95 - 1)	0.97	0.95	0.99	0.88	<0.0001
Gaussian Process	0.97 (0.95 - 0.99)	0.98	0.93	0.98	0.93	<0.0001
Support Vector Machine	0.98 (0.95 - 0.99)	0.94	0.95	0.99	0.80	<0.0001
Neural Net	0.97 (0.93 - 0.99)	0.99	0.92	0.98	0.97	<0.0001
CD versus HC						
Sparse Logistic Regression	0.97 (0.95 - 1)	0.94	0.96	0.97	0.91	<0.0001
Random Forest	0.98 (0.95 - 1)	0.96	0.93	0.95	0.95	<0.0001
Gaussian Process	0.98 (0.95 - 1)	0.96	0.93	0.95	0.95	<0.0001
Support Vector Machine	0.97 (0.93 - 1)	0.97	0.92	0.95	0.96	<0.0001

Neural Net	0.96 (0.93 - 1)	0.98	0.92	0.95	0.97	<0.0001
CDa versus HC						
Sparse Logistic Regression	0.96 (0.94 - 0.99)	1	0.92	0.74	1	<0.0001
Random Forest	0.98 (0.96 - 1)	1	0.95	0.84	1	<0.0001
Gaussian Process	0.98 (0.96 - 1)	1	0.95	0.84	1	<0.0001
Support Vector Machine	0.98 (0.96 - 1)	1	0.95	0.84	1	<0.0001
Neural Net	0.97 (0.95 - 0.99)	1	0.94	0.80	1	<0.0001
CDr versus HC						
Sparse Logistic Regression	0.95 (0.93 - 0.98)	1	0.90	0.67	1	<0.0001
Random Forest	0.97 (0.95 - 0.99)	1	0.93	0.74	1	<0.0001
Gaussian Process	0.97 (0.95 - 0.99)	1	0.93	0.74	1	<0.0001
Support Vector Machine	0.97 (0.95 - 0.99)	1	0.94	0.76	1	<0.0001
Neural Net	0.97 (0.95 - 0.99)	1	0.94	0.76	1	<0.0001
UC versus HC						
Sparse Logistic Regression	0.95 (0.91 - 0.99)	1	0.91	0.88	1	<0.0001
Random Forest	0.97 (0.93 - 1)	1	0.91	0.88	1	<0.0001
Gaussian Process	0.97 (0.93 - 1)	0.96	0.95	0.93	0.97	<0.0001
Support Vector Machine	0.97 (0.94 - 1)	1	0.91	0.88	1	<0.0001
Neural Net	0.97 (0.94 - 1)	0.98	0.92	0.90	0.99	<0.0001
UCa versus HC						
Sparse Logistic Regression	0.96 (0.94 - 0.99)	1	0.92	0.74	1	<0.0001
Random Forest	0.98 (0.96 - 1)	1	0.95	0.84	1	<0.0001
Gaussian Process	0.98 (0.96 - 1)	1	0.95	0.84	1	<0.0001
Support Vector Machine	0.98 (0.96 - 1)	1	0.95	0.84	1	<0.0001
Neural Net	0.97 (0.95 - 0.99)	1	0.94	0.80	1	<0.0001
UCr versus HC						
Sparse Logistic Regression	0.95 (0.93 - 0.98)	1	0.88	0.52	1	<0.0001
Random Forest	0.97 (0.95 - 0.99)	0.96	0.95	0.72	0.99	<0.0001
Gaussian Process	0.97 (0.95 - 0.99)	0.96	0.95	0.72	0.99	<0.0001
Support Vector Machine	0.97 (0.95 - 0.99)	0.96	0.95	0.72	0.99	<0.0001
Neural Net	0.97 (0.95 - 0.99)	0.96	0.95	0.70	0.99	<0.0001
IBDa versus IBDr						
Sparse Logistic Regression	0.59 (0.51 - 0.67)	0.21	0.96	0.90	0.39	0.019
Random Forest	0.59 (0.51 - 0.68)	0.31	0.88	0.82	0.40	0.017
Gaussian Process	0.58 (0.49 - 0.66)	0.43	0.75	0.76	0.41	0.039
Support Vector Machine	0.58 (0.49 - 0.67)	0.43	0.75	0.76	0.41	0.038
Neural Net	0.57 (0.49 - 0.65)	0.29	0.90	0.84	0.40	0.066
CDa versus CDr						
Sparse Logistic Regression	0.52 (0.39 - 0.65)	0.72	0.43	0.71	0.45	0.645
Random Forest	0.49 (0.36 - 0.62)	0.33	0.77	0.731	0.38	0.562
Gaussian Process	0.54 (0.41 - 0.67)	0.30	0.83	0.771	0.38	0.291
Support Vector Machine	0.53 (0.39 - 0.66)	0.88	0.27	0.691	0.53	0.674
Neural Net	0.52 (0.39 - 0.64)	0.60	0.57	0.721	0.43	0.606
UCa versus UCr						
Sparse Logistic Regression	0.63 (0.44 - 0.82)	0.67	0.57	0.79	0.42	0.082
Random Forest	0.62 (0.44 - 0.81)	0.85	0.43	0.78	0.55	0.094
Gaussian Process	0.55 (0.34 - 0.77)	0.88	0.43	0.78	0.60	0.288
Support Vector Machine	0.50 (0.29 - 0.71)	0.85	0.36	0.76S	0.50	0.514
Neural Net	0.58 (0.40 - 0.76)	0.82	0.43	0.77	0.50	0.191
CD versus UC						
Sparse Logistic Regression	0.55 (0.50 - 0.60)	0.17	0.96	0.90	0.36	0.031
Random Forest	0.54 (0.49 - 0.60)	0.12	0.98	0.93	0.35	0.074
Gaussian Process	0.54 (0.49 - 0.60)	0.17	0.92	0.83	0.35	0.078
Support Vector Machine	0.54 (0.49 - 0.60)	0.17	0.92	0.83	0.35	0.089
Neural Net	0.54 (0.49 - 0.59)	0.16	0.92	0.81	0.35	0.090
CDa versus UCa						
Sparse Logistic Regression	0.52 (0.39 - 0.65)	0.95	0.18	0.67	0.67	0.607
Random Forest	0.55 (0.41 - 0.68)	0.54	0.67	0.74	0.46	0.228
Gaussian Process	0.53 (0.40 - 0.66)	0.89	0.27	0.68	0.60	0.311
Support Vector Machine	0.53 (0.41 - 0.66)	0.67	0.48	0.69	0.46	0.308
Neural Net	0.56 (0.43 - 0.69)	0.72	0.48	0.71	0.50	0.832
CDr versus UCr						
Sparse Logistic Regression	0.56 (0.37 - 0.75)	0.74	0.43	0.76	0.40	0.744
Random Forest	0.53 (0.33 - 0.72)	0.53	0.64	0.78	0.36	0.393
Gaussian Process	0.54 (0.36 - 0.71)	0.41	0.86	0.88	0.38	0.653
Support Vector Machine	0.58 (0.41 - 0.76)	0.26	1	1	0.36	0.199
Neural Net	0.59 (0.41 - 0.76)	0.59	0.71	0.83	0.42	0.844

Abbreviations: IBD inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; a, active disease state; r, remission; HC, healthy controls; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

Table 6.3d. Overview of the Data Generated Using All Four Classifiers Based on the 100 Most Discriminative Features Using Only the Amsterdam UMC Samples

	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	p-value
IBD versus HC						
Sparse Logistic Regression	0.95 (0.91 – 0.98)	0.98	0.92	0.92	0.98	<0.001
Random Forest	0.97 (0.95 – 0.99)	1	0.92	0.92	1	<0.001
Gaussian Process	0.96 (0.92 – 0.99)	0.98	0.94	0.93	0.98	<0.001
Support Vector Machine	0.96 (0.93 – 0.99)	1	0.91	0.91	1	<0.001
Neural Net	0.94 (0.90 – 0.98)	0.98	0.92	0.92	0.98	<0.001
CD versus HC						
Sparse Logistic Regression	0.96 (0.93 – 0.99)	1	0.92	0.85	1	<0.001
Random Forest	0.97 (0.94 – 0.99)	1	0.92	0.85	1	<0.001
Gaussian Process	0.95 (0.92 – 0.99)	1	0.92	0.85	1	<0.001
Support Vector Machine	0.95 (0.91 – 0.99)	1	0.92	0.85	1	<0.001
Neural Net	0.93 (0.88 – 0.97)	0.93	0.94	0.87	0.97	<0.001
UC versus HC						
Sparse Logistic Regression	0.96 (0.92 – 0.99)	1	0.90	0.81	1	<0.001
Random Forest	0.96 (0.93 – 0.99)	1	0.91	0.83	1	<0.001
Gaussian Process	0.96 (0.93 – 0.99)	0.96	0.93	0.85	0.98	<0.001
Support Vector Machine	0.96 (0.93 – 0.99)	1	0.91	0.83	1	<0.001
Neural Net	0.90 (0.85 – 0.96)	0.93	0.94	0.86	0.97	<0.001
CD versus UC						
Sparse Logistic Regression	0.63 (0.53 – 0.74)	0.64	0.62	0.63	0.63	0.008
Random Forest	0.58 (0.47 – 0.69)	0.63	0.55	0.58	0.59	0.075
Gaussian Process	0.57 (0.46 – 0.68)	0.59	0.65	0.63	0.61	0.106
Support Vector Machine	0.49 (0.38 – 0.60)	0.43	0.64	0.54	0.52	0.436
Neural Net	0.53 (0.42 – 0.63)	0.54	0.62	0.59	0.57	0.320

Abbreviations: IBD inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; a, active disease state; r, remission; HC, healthy controls; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

Table 7.1. Overview of the best algorithm based on AUC value for each analysis in every study in this thesis

STUDY	Analysis												
	1	2	3	4	5	6	7	8	9	10	11	12	13
PDAC study 1	SVM	SVM	RF	SVM									
PDAC study 2	SLR	GP	SLR	GP									
	SLR	RF	SLR	RF									
CRC study 1	SLR	SLR	GP	SLR	SLR	GP	RF	RF	GP				
	GP	GP	SVM	SVM	GP								
CRC study 2	RF	RF	RF, GP, NN	RF	SLR	SVM	NN	RF	RF	RF	NN	NN	SLR
IBD study 1	GP	GP	RF	GP	SLR	RF, SLR	GP, SLR	SLR, SVM	SLR				
IBD study 2	RF, SVM	SLR											
	RF	GP											

IBD study 3	NN	SLR											
IBD study 4	SVM	RF, GP	RF, GP, SVM	RF, GP, SVM, NN	RF, GP, SVM, NN	RF, GP, SVM	RF, GP, SVM, NN	SLR, RF	GP	SLR	SLR	NN	NN
	RF	RF	SLR, RF, GP, SVM	SLR									

The column with more than one algorithms indicate that those algorithms have the same AUC value and are the best compare to others.

Abbreviations: PDAC, Pancreatic Ductal Adenocarcinoma; CRC, Colorectal Cancer; IBD, Inflammatory Bowel Disease; SLR, Sparse Logistic Regression; GP, Gaussian Process; SVM, Support Vector Machine; NN, Neural Network.

Table 7.2. Overview of the frequency of each algorithm in all analysis

Classification Algorithm	Number of the highest AUC for every analysis
RF	25
SLR	22
GP	22
SVM	15
NN	10

Abbreviations: SLR, Sparse Logistic Regression; GP, Gaussian Process; SVM, Support Vector Machine; NN, Neural Network.