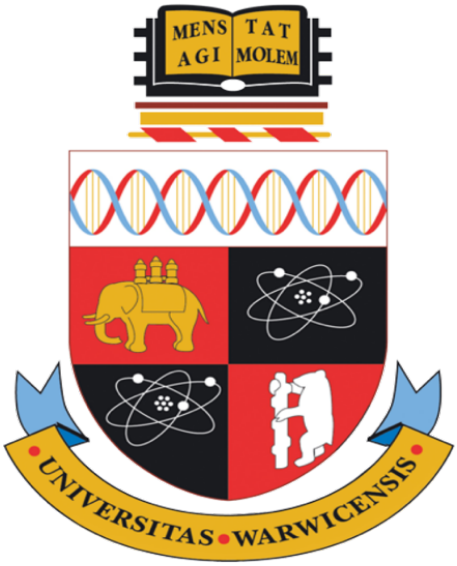


Gas Phase Biomarkers for Disease Diagnostics by Ion-Mobility

by

Heena Tyagi

A thesis presented for the degree of

Doctor of Philosophy in Engineering

Submitted to

School of Engineering

University of Warwick, United Kingdom

April 2023

Table of Content:

List of Tables:

List of figures

Table of Equations:

Acknowledgements

First of all, I would like to mention my sincere gratitude and thanks to my supervisor James A Covington, who has helped me throughout PhD and during hard times of COVID-19. His support, advice, encouragement, patience, and knowledge inspired me to develop my knowledge, encouraged me in the academic research. I am thankful for his invaluable supervision in reviewing the written work for the publications, university submissions, conferences, and this thesis.

I would like to mention thanks to Gavin Downs for his treasured support in circuit design and development which was really influential in shaping my experiment methods and results, Ian Griffith for his support with PCB design and helping me meet my deadlines during the hard times of Covid-19, and lastly to Frank Courtney for his help in 3D printing and mechanical requirements of the research. I am very grateful for them for bearing with me on the many occasions I sought their greater technical experience.

Finally, I would like to express gratitude to my family and friends and my partner, Oleseaden Ojuegba Lale, whose immense patience, support and love helped me to finish my research, cope during the hardships of Covid-19 and complete this thesis.

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 research papers listed. 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 School of Engineering of the University of Warwick.

Heena Tyagi

2018-2023

List of published works:

1. **Heena Tyagi**, E. Daulton, A. S. Bannaga, R. P. Arasaradnam and J. A. Covington, “Non-Invasive Detection and Staging of Colorectal Cancer Using a Portable Electronic Nose”, *Sensors,* vol. 21, no. 16, p. 5440, 2021.
2. **Heena Tyagi**, E. Daulton, A. S. Bannaga, R. P. Arasaradnam, and J. A. Covington, "Urinary Volatiles and Chemical Characterisation for the Non-Invasive Detection of Prostate and Bladder Cancers," *Biosensors,* vol. 11, no. 11, p. 437, 2021.
3. **Heena Tyagi**, E. Daulton, A. S. Bannaga, R. P. Arasaradnam, and J. A. Covington, "Electronic Nose for Bladder Cancer Detection," *Chemistry Proceedings,* vol. 5, no. 1, p. 22, 2021.
4. A. S. Bannaga, **Heena 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, no. 9, p. 2447, 2021.
5. M. McFarlane, R. P. Arasaradnam, B. Reed, E. Daulton, A. Wicaksono, **Heena Tyagi**, "Minimal Gluten Exposure Alters Urinary Volatile Organic Compounds in Stable Coeliac Disease," *Sensors,* vol. 22, no. 3, p. 1290, 2022.
6. Chandrapalan, Subashini, Sofie Bosch, **Heena Tyagi**, Emma Daulton, Joaquín Cubiella, Jordi Guardiola, Peter Kimani *et al*. "volatile organic compound analysis to improve faecal immunochemical testing in the detection of colorectal cancer-Authors' reply." *Alimentary pharmacology & therapeutics* 54, no. 4 (2021): 506-507.

Abstract

The detection of Volatile Organic Compounds (VOCs) has shown great potential as a non-invasive and rapid diagnostic tool for early cancer and disease detection. The use of VOCs for early disease detection has several advantages and can provide valuable insights into disease pathophysiology. This PhD research aimed to investigate the presence of VOCs in urine samples and the development of chemical fingerprints for various cancers, namely, bladder cancer, colorectal cancer, prostate cancer and hepatocellular cancer, and diseases such as fibrosis and urinary tract infection. This study investigated the possibility of using analytical techniques such as Gas Chromatography-Ion Mobility Spectrometry (GC-IMS), Gas Chromatography-Time of Flight-Mass Spectrometry (GC-TOF-MS), and Electronic Noses (eNoses) for the identification of these biomarkers.

The study findings demonstrate the presence of particular VOCs in urine samples. We found that GC-IMS and GC-TOF-MS was able to distinguish between some of the cancers with 100% sensitivity and 100% specificity. During this research, a total of 46 VOCs were identified as relevant for the identification of these cancer groups, with some VOCs being specific to each type of cancer. 13 VOCs with the highest concentration in urine samples of bladder cancer patients were identified in the course of this study. The electronic noses utilized in this research were the AlphaMOS FOX 4000 and PEN3 eNose. The AlphaMOS FOX 4000 was able to distinguish between some of the cancer groups with 100% sensitivity and 100% specificity. Although, PEN3 eNose did not exhibit 100% sensitivity in distinguishing between the different cancer groups, it still demonstrated high levels of sensitivity and specificity.

Overall, the study contributes to the advancement of research on the detection of VOCs and provides insights into the potential for using analytical techniques for the detection of VOCs in urine samples. However, there are some limitations to these technologies such as some of them require specialized training and expertise to operate and interpret results accurately and the cost of some of these technologies can be high, making them less accessible. Another limitation is that the detection of VOCs can be affected by factors such as sample collection, handling, and storage, which can impact on the accuracy of results.

The second aim of this research presents the development of an in-house device based on Photo Ionisation Detector (PID) for the detection of VOCs. The thesis focuses on the

evaluation of the performance of the developed PID+ based device for detecting ISB. Throughout this thesis, we have examined the electronics of PID+. Experiments were conducted to assess the performance of the PID+ at various flow rates, ISB (Isobutylene) gas concentrations, bias voltages of detector plates, and amplifier circuit gains.

The results of the study demonstrate the feasibility and efficacy of the developed device in detecting ISB with high response. It was observed that increasing the flow rate of the sample gas has the potential to improve the response time of the detector. Additionally, it was observed that the concentration of the target VOCs, in this case ISB gas, was a crucial factor that affected the response of the detector, with higher concentrations resulting in higher sensitivities. However, it was noted that elevating either the flow rate or the concentration of the target VOCs beyond a certain threshold could lead to the saturation of the output. During the testing of the two versions of PID+, the performance of the PID+ was significantly improved. Nevertheless, more extensive research is required to verify the validity of the results of this study with respect to various VOCs.

Abbreviations

AC	Alternating Current
ADC	Analog-To-Digital Converter
AFP	Alpha-Fetoprotein
ASCII	American Standard Code for Information Interchange
AUC	Area Under Curve
BCa	Bladder Cancer
BMI	Body Mass Index
BPH	Benign Prostatic Hypertrophy
BTA	Bladder Tumour Antigens
BTEX	Benzene, Toluene, Ethylbenzene and Xylenes
CAD	Computer Aided Design
CCFL	Common Cathode Florescent Lamp
CE	Conformité Européenne
CFU	Colony Forming Unit
CI	Chemical-Ionisation
CNT	Carbon Nano Tubes
COPD	Chronic Obstructive Pulmonary Disease
COVID	Corona Virus Disease
CP	Conducting Polymer
CPC	Conductive Polymer Nanocomposites
CRC	Colorectal Cancer
CT	Computed Tomography
CTU	Computed Tomography Urography
DC	Direct Current

DNA	Deoxyribonucleic Acid
EI	Electron Ionisation
EMI	Electro Magnetic Interference
FAIMS	High Field Asymmetric Waveform Ion Mobility Spectrometry
FDA	Food and Drug Administration
FID	Flame Ionisation Detector
FIT	Faecal Immunochemical Test
FR	Flow Rate
GC	Gas Chromatography
GLOBOCAN	Global Cancer Observatory
GP	Gaussian Process
HCC	Hepatocellular Cancer
IARC	International Agency for Research on Cancer.
IBD	Inflammatory Bowel Disease
IC	Integrated Chip
IDT	Interdigital Transducers
IEEE	Institute Of Electrical and Electronics Engineers
IMS	Ion Mobility Spectrometry
ISB	Isobutylene
JTAG	Joint Test Action Group
LC	Liquid Chromatography
LDA	Linear Discriminant Analysis
LP	Linear Progression
LPM	Litre Per Minute
LR	Logistic Regression
MCU	Micro Controller

MEMS	Micro-Electromechanical Systems
MOS	Metal Oxide Semiconductor
MOX	Metal Oxide
MS	Mass Spectrometry
NA	Nor Applicable
NHS	National Health Service
NIST	National Institute of Standards and Technology
NMP	Nuclear Matrix Protein
NMSC	Non-Melanoma Skin Cancers
NN	Neural Network
NPV	Negative Predictive Value
PAH	Polycyclic Aromatic Hydrocarbons
PCA	Principal Component Analysis
PCB	Printed Circuit Board
PEN	Portable Electronic Nose
PID	Photo Ionisation Detector
PLS	Partial Least Squares
PPB	Parts Per Billion
PPM	Parts Per Million
PPV	Positive Predictive Value
PSA	Prostate Specific Antigen
PTFE	Polytetrafluoroethylene
PWM	Pulse Width Modulation
QCM	Quartz Crystal Microbalance
RF	Radio Frequency
ROC	Receiver Operating Characteristic

SAW	Surface Acoustic Wave
SLR	Sparse Logistic Regression
SMD	Surface Mount Device
SPME	Solid Phase Micro Extraction
SSL	Split/Split less Injector
STD	Sexually Transmitted Diseases
SVM	Support Vector Machine
TD	Thermal Desorption
TIC	Total Ion Chromatogram
TNM	Tumour/Node/Metastasis
TOF	Time Of Flight
UHCW	University Hospital Coventry and Warwickshire
UK	United Kingdom
UTI	Urinary Tract Infection
UV	Ultra Violet
VOCs	Volatile Organic Compounds
WHO	World Health Organisation
XGB	Xtreme Gradient Boost

COVID-19 Impact Statement:

The COVID-19 pandemic has caused a significant disruption in research activities globally. As a PhD researcher in the field of Biomedical sensors, the pandemic has impacted my research project in various ways. This impact statement aims to describe the specific impacts of the pandemic on my research project.

1. Access to research facilities: Due to the lockdowns and restrictions imposed during the pandemic, I was unable to access the research facilities required for my study. I faced significant disruption in my academic pursuits due to the Covid-19 pandemic. The university was forced to close its doors in April 2020 in response to the outbreak. Unfortunately, the closure persisted until July 2021, which resulted in a significant delay in my academic progress. Even when the university reopened, the access was limited to only one day per week. Additionally, there was a considerable lack of technical support available, which made it difficult for me to carry out my research.
2. Delays in equipment procurement: The pandemic has also caused delays in the procurement of equipment required for my research, including various components required for circuit design and urine samples for VOCs analysis. This had led to delays in the calibration and validation of my research methods.
3. Challenges in conducting experiments: The pandemic has caused a delay in conducting experiments and testing the electronics PCBs under various conditions. Due to the restrictions imposed, it has been challenging to conduct experiments.

Overall, the COVID-19 pandemic has presented significant challenges to my research on VOCs biomarkers by photoionization. The pandemic has certainly caused unprecedented challenges and has significantly impacted the academic experiences of students like me. However, I remained committed to completing my research project and finding new ways to overcome these challenges in order to advance the field of biomarker research.

Chapter 1. Introduction

The early detection of diseases has always been a crucial aspect of the healthcare system with the use of biomarkers showing great promise in this regard. Among the various biomarkers that have been identified, Volatile Organic Compounds (VOCs) have emerged as a potentially powerful tool for early disease detection. This chapter delves into the background details surrounding VOCs and their potential as biomarkers. It provides a comprehensive overview of the biological basis of biomarkers and the history of VOCs research. The chapter also examines the generation of VOCs and their significance in detecting or recognizing diseases. The devices and technologies currently available for VOCs detection are also discussed, along with their advantages and disadvantages. Ultimately, the chapter culminates with an in-depth examination of the aims and objectives of this research, highlighting the potential impact that the detection of VOCs could have on early disease diagnosis and treatment.

1.1. Disease Diagnosis using Biomarkers:

Disease diagnosis and management have been the cornerstone of medical research for decades. The use of biomarkers has played a significant role in this field, enabling clinicians to identify and monitor disease progression in patients. Biomarkers can be defined as any measurable substance or process that provides information about the biological state of an individual and can be used to diagnose, monitor, or predict a disease's course [1]. The identification of specific biomarkers has revolutionized disease diagnosis and management, making it possible to detect diseases earlier and with greater accuracy than ever before [2]. Biomarkers can be used to monitor the response of the human body at different stages of the disease. It can be used to evaluate the cause, detection, progression, and reaction to the treatment and result of the treatment. VOCs emitted by the human body describe the various pathological processes and act as indicators of illness or disease or lack of illness [3][4]. VOCs biomarkers are being tested in a wide range of areas including the biomarkers for cardiovascular diseases [5, 6], cancer disease [7], psychiatric disorders [8], and diabetes [9] amongst many others. However, despite the significant advances in biomarker research, there are still significant challenges that need to be overcome.

The diagnosis of the majority of diseases relies heavily on the manifestation of clinical signs and symptoms, and despite significant advancements in medical science, many of the currently available screening methods demonstrate low sensitivity, particularly in the early stages of the disease. A prime example of this is cancer, which typically originates within organs and tissues, and progresses continuously until it develops into a tumour. While the survival rate for most cancers has improved since the mid-1970s, it remains the second most prevalent cause of death globally. Surgical or radiological interventions are often used to halt the advancement of cancer. However, the efficacy of these treatments is greatly reduced when the cancer is diagnosed at later stages [10, 11].

Therefore, early diagnosis and prognosis are crucial for effective disease management and can have a significant impact on patient outcomes. It can potentially increase the possibility of better outcomes, lower morbidity, improve survival chances, and propose effective and cheaper treatments. Biomarkers are used for the diagnosis of diseases at the early stages so that the prognosis and treatment selection becomes better [12]. Though the concept of early detection is one of the most promising methods to decrease the mortality rate, the technologies and screening methods required for it are challenging. With screening, the risk of overdiagnosis and overtreatment increases [13]. Screening tests should be inexpensive, minimally invasive, and should be highly effective to diagnose the disease in its early stages, as well as diagnose the stage of the disease. The technologies used until now are mostly invasive, expensive, and time-consuming. Therefore, it is important to search for a non-invasive and inexpensive method, which in part led to the discovery and use of biomarkers [14]. As a result, there is a continuous drive to develop analytical instruments that offer high sensitivity and specificity in the detection of diseases at their early stages by measuring biomarkers.

Numerous studies have been conducted to investigate the relationship between VOCs patterns and various diseases. Researchers have sought to identify specific VOCs biomarkers that could be used for disease diagnosis and monitoring. This was first reported when a dog started constantly sniffing at a mole on its owner's leg. After seeking medical advice, it was found that the lesion was a malignant melanoma [15]. However, this remained the hypothetical prediction and lacked experimental evidence. To investigate its potential, researchers trained six dogs for seven months to differentiate between urine from human

bladder cancer and controls. The dogs were able to recognize 22 cancer samples out of 54 samples proving that dogs have the ability to distinguish between cancer patients and healthy controls [16]. It was proposed that every disease has a 'smell-print', which could be used to differentiate them. In 1971, Pauling and co-workers reported a new method for the analysis of human breath and reported that it consisted of several hundreds of VOCs [17]. VOCs have been used to distinguish Inflammatory Bowel Disease (IBD) and healthy control patients [18]. VOCs can also be used to identify patients who suffered from gastrointestinal side effects after pelvic radiotherapy [19]. A total of 1840 VOCs have been identified in the healthy human body. Among these VOCs, 872 VOCs are associated with breath, 381 VOCs with faeces, 279 VOCs with urine, 532 VOCs with skin secretions, 154 VOCs with blood, and 359 VOCs with saliva [20]. It is worth mentioning that the high number of VOCs found in breath may reflect the significant amount of research and ease of sample availability rather than the chemical richness of each biological source.

Detection of VOCs for the diagnosis of cancer and other diseases is a promising method and potentially has high accuracy, specificity, and sensitivity. However, the sensitivity and specificity depend on the analytical technique and the biological samples used. Gas Chromatography coupled with Mass Spectrometry (GC-MS) is considered as the gold standard for the identification of VOCs because it has high sensitivity and specificity and is a very reliable technique for the identification of VOCs. It gives detailed information on the analytic compounds [21]. However, the high sensitivity of GC-MS often depends on the preconcentration of the samples. Preconcentrators are employed in GC-MS analysis to increase the concentration of the target analytes in the sample, can reduce the background noise and improve the signal-to-noise ratio of the GC-MS analysis. GC-Coupled with IMS (Ion Mobility Spectrometry) or IMS is another successful technique in the detection of VOCs. It works on the principle of ionisation of gases and detection of the molecules. In comparison to GC-MS, it is less expensive and gives results at a faster rate and mostly does not need pre-concentrators. The main disadvantage of this technique is it is less reliable in identifying unknown compounds and can have low selectivity depending on the chemicals [21, 22]. Though all these methods yield high sensitivity and are highly preferable for the identification of VOCs, they are very expensive, take a long time to undertake an analysis, involve highly complex technology and require specially trained operators. In comparison to these technologies, the electronic nose, or eNose, is less expensive, smaller in size, less

time-consuming, easy to manufacture and easily assessable to everyone. An eNose is made from an array of sensors resulting in high sensitivity and can identify VOCs without separation. However, eNoses may lack the sensitivity and selectivity of conventional analytical techniques, such as gas chromatography-mass spectrometry (GC-MS), resulting in the potential inability to distinguish between similar VOCs, leading to possible false positives or false negatives in VOCs detection. Moreover, regular calibration is necessary to maintain accuracy, as eNoses may experience drift over time. Changes in environmental conditions or sensor performance can compromise the device's reliability, necessitating frequent recalibration. Additionally, the presence of environmental factors, including temperature, humidity, and other odours, can exert interference on eNose performance and impact VOCs detection [23, 24]. Photo Ionisation detector (PID) is another small size, rapid and cheaper device used for the identification of VOCs. It works on the principle of ionisation with the use of high intensity UV-rays [25]. PID suffer from poor selectivity, similar to eNoses, but it offers very rapid analysis time in comparison to all the devices available including the eNose [26].

Despite the significant advancements in early disease diagnosis, challenges still exist. One of the major challenges is the identification of reliable biomarkers that are specific to a particular disease and demonstrate high sensitivity in detecting the disease at its earliest stages. Additionally, the development of accurate and non-invasive technology that can potentially detect the disease at its earliest stages is also challenging. In light of this, there is a need for a device that has the advantages of both the eNose as well as a PID, i.e., low cost, simpler technology, fast analysis, portable and offers the advantages of GC-MS devices such as high sensitivity, high selectivity, high accuracy, and repeatability. Hence, this has prompted the idea of developing a tool that can be used for the identification of VOCs for different diseases and malignancies. This device may be used as a screening test for diseases and may help to reduce the costs, mortalities, social pressure, and misdiagnosis from currently available tests. It is also important to generate VOCs chemical profiles for different diseases to understand the changes in VOCs patterns of diseases. If we know the chemical fingerprint of a disease, it can be used to identify the disease in its early stages. This would help to reduce mortality rates and severe physiological effects by detecting the disease before it develops.

1.2. Aim and Objectives:

The primary aim of this doctoral research was to identify and investigate biomarkers associated with various diseases, with a particular focus on cancer, and to develop an in-house diagnostic device for the early detection of these diseases. This study was conducted in collaboration between the School of Engineering, Warwick, and Owlstone Medical Ltd. The specific objectives of this thesis were as follows:

1. To employ advanced analytical techniques such as Gas Chromatography-Ion Mobility Spectrometry (GC-IMS), Gas Chromatography-Time of Flight-Mass Spectrometry (GC-TOF-MS) and electronic noses for VOCs analysis.
2. To explore the potential of urinary biomarkers as indicators for the early detection and diagnosis of cancer and other diseases. The aim was to identify unique VOCs patterns associated with each condition and to evaluate the sensitivity and specificity of these patterns as diagnostic biomarkers.
3. To investigate the efficacy of VOCs measurements in urine samples for disease detection and monitoring, considering their potential to facilitate early diagnosis and disease progression tracking.
4. To address the complexity and cost constraints associated with advanced analytical technologies by designing and developing an in-house diagnostic device based on the concept of ion mobility.
5. To employ the principle of photoionization in the design of the in-house device, thereby aiming to achieve comparable diagnostic capabilities with lower unit cost, ease of use, and portability.
6. To optimize the in-house device for the accurate detection and quantification of VOCs biomarkers in urine samples.

Overall, the overarching goal of this PhD thesis was to contribute to the advancement of non-invasive, cost-effective, and accurate diagnostic tools for the early detection and diagnosis of various diseases. By identifying and quantifying VOCs biomarkers in urine samples and developing a photo-ionization-based diagnostic device, this research has the potential to enhance patient outcomes and reduce healthcare expenses.

1.3. Thesis structure

Chapter 2: Literature Review. This chapter presents a comprehensive background on volatile organic compounds (VOCs), and their association with various cancers and diseases. It provides insight into the rationale for selecting VOCs as the focus of this research and includes a detailed review of published literature on the subject. Additionally, this chapter presents an extensive review of various cancers as well as diseases and the current technologies used for their detection with their advantages and disadvantages. It further provides a thorough overview of different analytical methods used for VOCs identification, along with a detailed discussion of their advantages and disadvantages.

Chapter 3: Analytical instruments for VOCs detection: This chapter provides a comprehensive overview of the methodology employed for the detection of the VOCs. It provides comprehensive details on the analytical instruments employed in this work. It elucidates the operating principles, sample preparation techniques, and methods used to operate each instrument. Moreover, the chapter features an illustrative example of the output generated by all the analytical instruments considered, along with an overview of the data analysis methods used for each instrument.

Chapter 4: Analysis of the cancer samples using GC-IMS and GC-TOF-MS: Chapter 4 consists of the biological testing and findings. The biological sample used in this research is urine. The main focus of this chapter is the presentation and interpretation of the results obtained from the GC-IMS and GC-TOF-MS analyses. We provide a detailed discussion of the chromatographic and spectrometric data obtained, including the identification and quantification of key metabolites and biomarkers associated with cancer. We also highlight any notable differences observed between cancer and control urine samples, as well as any potential limitations or sources of error in the analysis.

Chapter 5: Analysis of the cancer samples using eNose: Chapter 5 presents the results of our investigation into the use of electronic noses, specifically the eNose AlphaMOS FOX 4000 and the PEN3 eNose, for the detection of cancer samples from healthy controls. In contrast to the previous chapter, where the focus was on the identification and quantification of specific metabolites and biomarkers, the results presented in this chapter focus on the

overall performance of the eNose devices in distinguishing between cancer samples and healthy controls. We provide a detailed discussion of the statistical analysis used to evaluate the data, including the calculation of sensitivity, specificity, and other performance metrics.

Chapter 6: Analysis of Infectious disease VOCs: This chapter consists of the analysis of UTI urine samples using different analytical devices, specifically, we evaluated the sensitivity and specificity of the GC-IMS, GC-TOF-MS, and eNoses in detecting the presence of UTI in patient samples. The chapter provides a detailed discussion of the statistical analysis used to analyse the data and to determine the diagnostic accuracy of the different techniques. The results presented in this chapter focus on the overall performance of the analytical techniques in distinguishing between UTI positive and negative samples.

Chapter 7: Development of enhanced Photo Ionisation detector: Chapter 7 presents the details of the design and construction of an in-house device, an enhanced photo-ionization detector (PID+), for the detection of volatile organic compounds (VOCs). The chapter begins with an overview of the motivation behind the project, including the need for a low-cost, portable, and sensitive VOCs detection system. We then provide a detailed description of the design and construction of the PID+ prototype, including the selection of the key components and the design of the electronic circuitry. Different versions of the prototype were developed and tested to analyse the sensitivity and selectivity of the PID+.

Chapter 8: PID+ experimental results and discussions: Chapter 8 presents the experimental setup and the testing procedures used to evaluate the performance of the different versions of PID+ as well as the findings of these tests. The chapter provides a thorough discussion of the results, including the sensitivity and selectivity of the different PID+ prototypes.

Chapter 9: Conclusions and Further work: The different analytical devices used in this research as well as the findings of these devices are compared in this chapter. This chapter also considers the PID+ developments and comparisons. Further developments, limitations and advantages are also considered in this chapter.

1.4. References:

[1] K. Schmidt and I. Podmore, "Current challenges in volatile organic compounds analysis as potential biomarkers of cancer," *Journal of biomarkers,* vol. 2015, 2015.

[2] L. Richardson, "Sniffing and smelling," *Philosophical Studies,* vol. 162, no. 2, pp. 401-419, 2013.

[3] R. Mayeux, "Biomarkers: Potential uses and limitations," *NeuroRX,* vol. 1, no. 2, pp. 182-188, 2004/04/01 2004, doi: 10.1602/neurorx.1.2.182.

[4] M. Shirasu and K. Touhara, "The scent of disease: volatile organic compounds of the human body related to disease and disorder," *The Journal of Biochemistry,* vol. 150, no. 3, pp. 257-266, 2011, doi: 10.1093/jb/mvr090 %J The Journal of Biochemistry.

[5] E. J. N. E. J. o. M. Braunwald, "Biomarkers in heart failure," vol. 358, no. 20, pp. 2148-2159, 2008.

[6] R. E. Gerszten and T. J. Wang, "The search for new cardiovascular biomarkers," *Nature,* vol. 451, no. 7181, pp. 949-952, 2008/02/01 2008, doi: 10.1038/nature06802.

[7] R. Frank and R. Hargreaves, "Clinical biomarkers in drug discovery and development," *Nature Reviews Drug Discovery,* vol. 2, no. 7, pp. 566-580, 2003/07/01 2003, doi: 10.1038/nrd1130.

[8] I. Singh and N. Rose, "Biomarkers in psychiatry," *Nature,* vol. 460, no. 7252, pp. 202-207, 2009/07/01 2009, doi: 10.1038/460202a.

[9] 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," vol. 8, no. 4, p. 121, 2018. [Online]. Available: https://www.mdpi.com/2079-6374/8/4/121.

[10] M.-P. Raitanen *et al.*, "Differences Between Local and Review Urinary Cytology in Diagnosis of Bladder Cancer. An Interobserver Multicenter Analysis," *European Urology,* vol. 41, no. 3, pp. 284-289, 2002/03/01/ 2002, doi: https://doi.org/10.1016/S0302-2838(02)00006-4.

[11] P. D. Wagner, M. Verma, and S. J. A. o. t. N. Y. A. o. S. SRIVASTAVA, "Challenges for biomarkers in cancer detection," vol. 1022, no. 1, pp. 9-16, 2004.

[12] A. N. Bhatt, R. Mathur, A. Farooque, A. Verma, and B. J. I. J. M. R. Dwarakanath, "Cancer biomarkers-current perspectives," vol. 132, no. 2, pp. 129-149, 2010.

[13] T. Bevers *et al.*, "23 - Screening and Early Detection," in *Abeloff's Clinical Oncology (Sixth Edition)*, J. E. Niederhuber, J. O. Armitage, M. B. Kastan, J. H. Doroshow, and J. E. Tepper Eds. Philadelphia: Content Repository Only!, 2020, pp. 375-398.e7.

[14] A. D. Wilson, "Advances in Electronic-Nose Technologies for the Detection of Volatile Biomarker Metabolites in the Human Breath," vol. 5, no. 1, pp. 140-163, 2015. [Online]. Available: http://www.mdpi.com/2218-1989/5/1/140.

[15] H. Williams and A. J. T. L. Pembroke, "Sniffer dogs in the melanoma clinic?," vol. 333, no. 8640, p. 734, 1989.

[16] C. M. Willis *et al.*, "Olfactory detection of human bladder cancer by dogs: proof of principle study," vol. 329, no. 7468, p. 712, 2004, doi: 10.1136/bmj.329.7468.712 %J BMJ.

[17] L. Pauling, A. B. Robinson, R. Teranishi, and P. Cary, "Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography," vol. 68, no. 10, pp. 2374-2376, 1971, doi: 10.1073/pnas.68.10.2374 %J Proceedings of the National Academy of Sciences.

[18] R. P. Arasaradnam *et al.*, "A Novel Tool for Noninvasive Diagnosis and Tracking of Patients with Inflammatory Bowel Disease," *Inflammatory Bowel Diseases,* vol. 19, no. 5, pp. 999-1003, 2013, doi: 10.1097/MIB.0b013e3182802b26 %J Inflammatory Bowel Diseases.

[19] J. A. Covington *et al.*, "The detection of patients at risk of gastrointestinal toxicity during pelvic radiotherapy by electronic nose and FAIMS: a pilot study," (in eng), *Sensors (Basel),* vol. 12, no. 10, pp. 13002-13018, 2012, doi: 10.3390/s121013002.

[20] B. de Lacy Costello *et al.*, "A review of the volatiles from the healthy human body," in *Journal of Breath Research* vol. 8, ed: IOP Publishing, 2014, p. 14001.

[21] A. Ulanowska, T. Ligor, M. Michel, and B. Buszewski, "Hyphenated and unconventional methods for searching volatile cancer biomarkers," *Ecol. Chem. Eng,* vol. 17, no. 1, pp. 9-23, 2010.

[22] M. Westhoff, P. Litterst, L. Freitag, W. Urfer, S. Bader, and J.-I. Baumbach, "Ion mobility spectrometry for the detection of volatile organic compounds in exhaled breath of patients with lung cancer: results of a pilot study," in *Thorax* vol. 64, ed, 2009, pp. 744 LP - 748.

[23] J. Fitzgerald and H. Fenniri, "Cutting edge methods for non-invasive disease diagnosis using e-tongue and e-nose devices," *Biosensors,* vol. 7, no. 4, p. 59, 2017.

[24] A. D. Wilson and M. Baietto, "Applications and Advances in Electronic-Nose Technologies," *Sensors,* vol. 9, no. 7, pp. 5099-5148, 2009. [Online]. Available: https://www.mdpi.com/1424-8220/9/7/5099.

[25] S. O. Agbroko and J. Covington, "A novel, low-cost, portable PID sensor for the detection of volatile organic compounds," in *Sensors and Actuators B: Chemical* vol. 275, ed, 2018, pp. 10-15.

[26] Y. Kang, H. Jung, D. Baek, K. Lee, S. Pyo, and J. Kim, "Highly sensitive detection of VOC using impact ionization induced by photoelectron," in *2017 IEEE 30th International Conference on Micro Electro Mechanical Systems (MEMS)*, 22-26 Jan. 2017 2017, pp. 183-186, doi: 10.1109/MEMSYS.2017.7863371.

Chapter 2. Literature Review

In chapter 1, the use of VOCs for the detection of diseases, including cancer was discussed. This chapter aims to provide a comprehensive review of the use of VOCs for the detection of diseases, particularly cancer. It is divided into two parts. The first part focuses on the biological and pathological background of VOCs, their relevance in disease detection, and the sources of VOCs, including their advantages and disadvantages. Furthermore, the first part provides an overview of the cancers and diseases being investigated in this research, their pathological background, current diagnostic methods, and their limitations, as well as studies focused on the detection of VOCs for these cancers.

The second part of the chapter will provide a detailed analysis of the analytical instruments used for the detection of VOCs. This section will also discuss the advantages and disadvantages of these instruments and technologies. Additionally, the chapter will conclude with a discussion of Photoionization detectors (PIDs), including a brief review of commercially available PIDs and their advantages and disadvantages. The objective of this chapter is to provide a foundation for the research on the development of a novel VOCs detection system for early cancer diagnosis.

2.1. Medical Literature Review:

2.1.1. Background of VOCs:

VOCs are a diverse group of molecules that have been found to play an important role in medicine. The olfactory senses have historically been employed as a tool for disease diagnosis predating the advent of modern medicine [1, 2]. VOCs represent a heterogeneous class of organic molecules characterized by their low boiling points, which means that they easily evaporate. These compounds can originate from two sources: endogenous VOCs, which are produced through metabolic processes within the human body and influenced by the individual's physiological health status, and exogenous VOCs, which are derived from external sources such as environmental pollutants, tobacco smoke, dietary intake, and so on.

These external factors can impact the production of endogenous VOCs [3, 4]. VOCs families are commonly based on alcohols, hydrocarbons, ketones, acids, and aldehydes. Hydrocarbons, produced mainly through the peroxidation of polyunsaturated fatty acids, are found in cellular and subcellular membranes. Alcohols, which are absorbed into the bloodstream through the gastrointestinal tract, undergo metabolism by enzymes such as alcohol dehydrogenases. Aldehydes are generated from various sources, including metabolized alcohols, lipid peroxidation, tobacco smoke, and dietary intake etc. Ketones are produced by the liver from fatty acids. Finally, aromatic and nitrile VOCs, considered pollutants of exogenous sources, such as cigarette smoke and pollution, are stored in fatty tissues and are highly reactive, leading to peroxidative damage [5].

Due to the chemical characteristics of VOCs, real-time monitoring is possible with minimal sample preparation requirements. The study of VOCs in medicine has grown rapidly in recent years, with researchers exploring their potential as biomarkers for various diseases such as cancer, diabetes, and infectious diseases. The use of VOCs in medicine has several advantages, including non-invasiveness, low cost, and the ability to monitor disease progression and treatment efficacy in real time. The potential of VOCs to be used as non-invasive biomarkers in medicine for the diagnosis and monitoring of various medical conditions is discussed in the subsequent section.

2.1.2. Sources of VOCs for disease detection:

a. Urine:

VOCs are known to be transported throughout our body via the bloodstream, and subsequently, are filtered by the kidneys into urine, thus making urine an important source of VOCs [6, 7]. Urine VOCs belong to a diverse range of chemical classes, such as ketones, alcohols, aldehydes, carboxylic acids, amines, furans, pyrroles, hydrocarbons, and sulphur compounds, and are generally considered intermediate or end products of metabolic pathways.

Due to its non-invasive collection process and ability to be obtained in large volumes, urine has been utilized in several studies for VOCs analysis. A study conducted by Khalid *et al*

[8] used urine samples collected from 59 patients with prostate cancer and 43 healthy controls. They suggested that VOCs, namely, 2,6-dimethyl-7-octen-2-ol, pentanal, 3-octanone, and 2-octanone are VOCs fingerprints of prostate cancer. Another study showed the potential of urine used to distinguish between colorectal cancer from healthy controls with a sensitivity and specificity of 88% and 60% respectively [9].

While urine serves as an easily accessible source of VOCs, its components can be readily influenced by external factors, such as the ingestion of foods or drinks. Moreover, studies have shown that the VOCs in urine are susceptible to degradation due to improper storage procedures or inadequate storage conditions [10, 11].

b. Skin:

The analysis of skin VOCs has been explored as a potential tool for disease diagnosis and monitoring. Skin is one of the primary sources of odour compounds in humans and is home to bacteria, fungi and viruses which play an important role in the generation of skin VOCs [12]. The composition of VOCs generated by skin varies depending on several factors such as age, gender, ethnicity, diet, and lifestyle. These microorganisms can produce a range of VOCs, including short-chain fatty acids (SCFAs), alcohols, ketones, and esters, through metabolic processes [13, 14].

Several studies have reported that changes in the composition of skin VOCs can be used as biomarkers for various diseases such as cancer, diabetes, and skin disorders. For instance, a study conducted by Bernier *et al.* [15] collected VOCs from the palms and backs of hands. The researchers found that hundreds of compounds were released from the skin, most of which were organic acids, and that these acids were not volatile at body temperatures, meaning that they did not contribute significantly to the odour of human skin. This finding was noteworthy because it suggested that other VOCs present in smaller amounts may play a more significant role in determining the odour profile of human skin.

The use of skin VOCs is limited by the fact that the VOCs profile of the skin can be influenced by various external factors such as exposure to environmental pollutants and the use of personal care products.

c. Breath:

Breath sampling is a non-invasive technique that can be easily performed and is well received by patients. This makes it an attractive option for disease detection and monitoring, particularly for diseases that are difficult to diagnose using traditional methods [16]. Breath contains a complex mixture of VOCs that originates from various sources, including the airways, lungs, blood, and digestive system [17]. The concentration and composition of breath VOCs can vary depending on the metabolic state of the individual and the presence of diseases [18, 19]. For instance, it has been reported that individuals with diabetes have a distinct breath VOCs profile compared to healthy individuals. Similarly, changes in the composition of breath VOCs have been reported in individuals with respiratory disorders such as asthma and chronic obstructive pulmonary disease (COPD) [20].

The analysis of breath VOCs has been explored as a potential tool for disease diagnosis and monitoring. Several studies have reported that changes in the composition of breath VOCs can be used as biomarkers for various diseases. Phillips *et al.* [21] showed in their study that VOCs in the breath of women with breast cancer differed significantly from those in the breath of healthy women. The study was able to differentiate between women with and without breast cancer with an accuracy of 88% using a single breath sample.

However, Breath VOCs can be generated by endogenous metabolic processes or influenced by exogenous sources such as dietary intake and exposure to environmental pollutants. The composition of breath VOCs can vary depending on factors such as age, gender, diet, and lifestyle. For example, the composition of breath VOCs can be influenced by the type of diet consumed, with high-fat and low-carbohydrate diets resulting in the production of different VOCs profiles. Similarly, cigarette smoking and exposure to air pollutants can also influence the composition of breath VOCs.

d. Faeces:

Faeces are a rich source of VOCs that can provide valuable information about gut health, microbiome composition, and disease status. The analysis of VOCs in faecal samples can

provide a non-invasive method for disease diagnosis, monitoring, and treatment [22, 23]. As the human stool represents the end-product of diet, digestive and excretory processes, as well as colonic bacterial metabolism, the examination of faeces may be the best non-invasive way of diagnosing gastrointestinal diseases, including inflammatory bowel disease (IBD) [24], colorectal cancer (CRC) [25, 26], and irritable bowel syndrome (IBS)[27].

Despite the potential benefits, the use of faecal VOCs for disease detection also presents a number of limitations and drawbacks. For example, the sensitive nature of collecting faecal samples may make some individuals uncomfortable or reluctant to participate in the diagnostic process, though some cultures have fewer issues with this sample medium. Moreover, the use of faecal VOCs analysis as a diagnostic tool presents challenges and limitations due to limited research and a lack of standardized methods for sample collection, preparation, and analysis.

2.1.3. VOCs for disease detection:

Recent studies have focused on utilizing biomarkers discovered in biological samples as a diagnostic instrument for detecting illnesses and have shown that this approach can aid in the creation of cost-effective methods for screening diseases [28-30].

VOCs analysis performed on diseases, such as breast cancer [31-33], tuberculosis [34-36], inflammatory bowel diseases [37, 38], diabetes [39, 40] etc, suggests the potential of biomarkers that could be used for disease diagnosis. For instance, a study was conducted by Amal *et al* [41] to investigate the potential of gaseous VOCs obtained from cell cultures to differentiate between hepatocellular cancer patients from healthy controls. They compared and analysed four different groups, namely, all HCC and normal controls, low metastatic HCC and normal controls, high metastatic HCC and normal controls and high and low HCC using GC-MS and nanomaterial-based sensor arrays. They found four VOCs, namely, acetic acid, ethanol, 2,3 di-hydro-benzofuran and methane-sulfonyl chloride either significantly increased or decreased in HCC samples. They were also able to distinguish between different HCC samples and healthy controls with very high sensitivity and specificity. Their pilot study was able to deliver a proof-of-concept of the potential of headspace gaseous VOCs for the identification of HCC and its metastatic potential using nanomaterial-based

sensors. Another similar study conducted to diagnose and investigate the VOCs profile for lung cancer successfully identified four compounds 2-ethyl-1-hexanol, 2-pentanone, tetrahydrofuran, and 2-mehtylpyrazine as good indicators of lung cancer [42].

2.1.4. Advantages of VOCs for disease detection:

There are several advantages of using volatile organic compounds (VOCs) for disease detection:

1. Non-invasive: VOCs can be detected in various bodily fluids, including breath, urine, and faeces, which can be collected non-invasively and without causing discomfort to the patient [43].
2. Sensitivity and specificity: VOCs can provide early indicators of disease, as they can be present in trace amounts before clinical symptoms appear and VOCs profiles can be unique to certain diseases, which can aid in accurate diagnosis and treatment planning. [44].
3. Speed: VOCs analysis can provide rapid results, which can enable prompt treatment and improve patient outcomes [45].
4. Cost-effectiveness and portability: VOCs analysis is relatively inexpensive compared to other diagnostic methods, such as imaging or biopsies. VOCs analysis devices can be made small and portable, allowing for point-of-care testing in various settings [46].

Overall, the use of VOCs for disease detection holds great potential for improving patient outcomes through non-invasive, accurate, and cost-effective diagnostic methods.

2.1.5. Disadvantages of VOCs for disease detection:

While the analysis of VOCs has shown promising results as a non-invasive and rapid method for disease detection, there are also several disadvantages associated with this approach.

1. Identification: The identification of specific VOCs that are indicative of a particular disease or condition can be challenging. The composition of VOCs in bodily fluids can vary greatly between individuals and can also be affected by various external factors such as diet, medication, and environmental exposure. Therefore, the identification of reliable biomarkers that can be used for disease detection can be difficult [47, 48].
2. Cost implications: The analysis of VOCs requires sophisticated and expensive equipment, such as gas chromatography-mass spectrometry (GC-MS) or electronic nose devices, which can be cost-prohibitive for many healthcare settings. Additionally, the complexity of these instruments can also make them difficult to operate and they require specialized expertise to interpret the results [49].
3. Lack of standardized protocols: The interpretation of VOCs data can be challenging due to the large amount of data generated, as well as the lack of standardized protocols for VOCs analysis. There is also a need for comprehensive databases of VOCs profiles for different diseases and conditions, which can be time-consuming and expensive to develop [50].

In conclusion, while the analysis of VOCs has shown potential as a non-invasive and rapid method for disease detection, there are several challenges and limitations associated with this approach that need to be addressed.

2.2. Introduction to Cancer:

Early detection of cancer is important for its prevention and control. Recognising the signs of cancer and preventing its spread early can significantly improve the prognosis of patients. According to GLOBOCAN (Global Cancer Observatory), the estimated number of new

cases of cancer were 18.1 million and the number of deaths were 9.6 million in 2018 worldwide [51, 52]. The WHO stated in 2015, that the first or second cause of death below the age of 70, in most countries, was cancer and it has been predicted that the incident rate will increase to be as high as 29.5 million in another 20 years [53]. The reasons behind the increment in the incidence are aging, population growing, unhealthy lifestyles, physical inactivity etc [54].

Numerous organizations, both on a national and global scale, have exerted significant efforts towards managing cancer. The concept of early detection is one of the most promising methods in order to decrease its mortality rate [55]. However, screening and detection are important for the diagnosis of cancer, the technologies and methods required for it are challenging. In addition, with screening, the risk of over-diagnosis and overtreatment increases [56]. The main focus of researchers in the research for the best treatment of cancer, is discovering the cancer and finding it at an early stage.

Screening tests should be inexpensive, minimally invasive and should be highly effective in diagnosing the disease in its early stages. The technologies used until now are mostly invasive, expensive, and time-consuming. Therefore, it is important to search for a non-invasive, inexpensive and rapid biomarker [57]. An increasing number of researchers are focused on the diagnosis and monitoring of high-risk cancers, identifying their symptoms, and detecting them early through the evaluation of VOCs. The following sections provide a literature review of the cancers studied in this research.

2.2.1. Bladder Cancer (BCa):

Bladder cancer is one the most difficult cancers to manage and is the ninth most common cancer in the world. The diagnosis and treatment of bladder cancer strongly depends upon the stage of the cancer. Bladder cancer can be either a non-muscle-invasive cancer or muscle-invasive [58, 59]. Non-muscle invasive bladder cancer is relatively easy to treat and has higher survival rates and invasive bladder cancer is a high-grade cancer and is caused when the cancer cells invade in the bladder muscle layer [60].

Bladder cancer is associated with several risk factors including smoking, occupational exposure, genetics, age, and gender [61-64]. Presence of haematuria (blood in urine) is a common indicator of bladder cancer. The most common diagnostic methods for bladder cancer detection are Computed Tomography Urography (CTU) and Cystoscopy with cytology. Cystoscopy with Cytology is considered as the golden standard for the diagnosis of bladder cancer. However, this diagnostic method suffers from low sensitivity, high cost, its invasive nature, painful procedure and the sensitivity of cytology is very low for lower stages of cancer and a lot of it is affected by the experience of the examiner [65, 66]. Due to these reasons, this method cannot be used as a screening test. Similarly, a study showed that though CTU is effective for detection of bladder cancer with high sensitivity and specificity, the accuracy of CTU was low because of false-negative results and patients were exposed to high radiations during the diagnosis [67, 68].

Utilizing biomarkers as a diagnostic approach for detecting bladder cancer could potentially address the limitations of current methods, leading to an improvement in overall survival rates. Urine is considered the optimal source of VOCs for detecting bladder cancer. A study employing high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (HPLC-QTOFMS) was conducted on a cohort of 138 bladder cancer patients and 121 male healthy controls. The study successfully distinguished between the two groups with a high sensitivity of 91.3% and specificity of 92.5% using multivariate statistical analysis [69].

2.2.2. Prostate Cancer (PCa):

Prostate Cancer is the most common cancer in men in the UK and the sixth most common cancer in the world [70]. The primary factors associated with prostate cancer are age and ethnicity, genetic history, diet, obesity, sexually transmitted diseases (STDs), and smoking. It has been found that incidences of prostate cancer are most common in the age group of 75 and above [71-73]. Early detection of prostate cancer helps in the diagnosis and prognosis of prostate cancer. However, there are many factors that increase a patient's risk of prostate cancer. One of them is poor healthcare quality. Studies shown that poor healthcare quality influences the decisions of the men to undergo screening tests and delay screening. This serves as a barrier and also disrupts the collaboration between men and a physician [74].

PSA (Prostate Specific Antigen) based screening is one of the most common methods used for early detection. Elevated levels of PSA indicate the presence of prostate cancer. However, PSA based screening has a high risk of over diagnosis and over-treatment due to false results. The false test results lead to unnecessary biopsies which may end up causing fever, pain, bleeding, infection etc. to the patient [71-73]. Another study indicated that a PSA test alone is not accurate enough to be used as a screening procedure. It is useful when combined with rectal examination or ultrasonography [75]. Though PSA screening has shown a significant decrease in the mortality rate of prostate cancer, it has its own concerns and limitations.

Using a non-invasive method for prostate cancer detection not only overcomes the limitations of PSA screening, but it is also a fast, easy, and efficient method. Discovery of a VOC for the detection of cancer has emerged as a very promising approach for decreasing the mortality rate. A study was conducted to find out the VOCs from the headspace of urine from men with or without prostate cancer. A total of 37 urine samples consisting of 24 healthy controls and 13 patients with prostate cancer were used and 91 VOCs were identified. Out of these 91 VOCs, 21 VOCs distinguished between healthy and PCa with sensitivity and specificity of 92.3% and 96.3% respectively. However, the sample size for this study was relatively small [76]. Another study compared the VOCs in urine samples of prostate cancer patients and benign prostatic hypertrophy (BPH). 57 VOCs were identified, and it was found that urinary furan levels were higher in PCa patients than in BPH. They reported p-xylene as a good cancer biomarker and its presence in the urine of PCa patients [77]. In 2015, Khalid *et al.* [78] found that lower levels of 2, 6-dimethyl-7-octen-2-ol may be linked with the presence of prostate cancer.

2.2.3. Colorectal Cancer (CRC):

Colorectal Cancer is the second leading cause of the cancer-related deaths and the third leading cause of cancer-related deaths among men and women respectively in Europe [51]. The main causes of colorectal cancer are obesity, intake of red meat or alcohol, smoking and it has been found that physical activity, healthy body weight, dairy products, calcium supplements and smoking and alcohol cessation can reduce the risk of colorectal cancer [79-81]. A Colonoscopy is the most commonly used diagnostic technique for CRC screening. A

recent study was conducted to assess the rate of missed or newly detected CRC during colonoscopy and revealed that the rate of missed CRC diagnosis ranged from 2% to 6%. Secondly, the cancer might be too small and situated behind folds of the colon. Lastly, missing lesions could also occur due to the miss-handling of the physician or if the bowel preparation was not adequate enough etc [82, 83]. Another type of CRC screening is faecal immunochemical tests (FIT). This test is used to identify blood in the stool. Many studies reported high sensitivity and specificity for colorectal cancer detection using FITs for later stages of cancer, whereas they also found that the sensitivity of these tests is relatively low for early stages of cancer [84-87].

Therefore, VOCs profile analysis related to colorectal cancer is a novel non-invasive method. A study was conducted to identify VOCs analysis related to colorectal cancer using breath analysis [88]. They conducted two studies, one with 37 patients with colorectal cancer and 41 healthy controls and another with 15 CRC patients and 10 healthy controls and identified 58 VOCs. Out of these 58 VOCs, 15 were potentially discriminating VOCs between CRC patients and healthy controls [88]. A review paper compared the different technologies available and different studies conducted for the diagnosis of colorectal cancer using VOCs. They concluded that the detection of the colorectal cancer using VOCs present in urine, faecal or breath is a non-invasive, cheaper and a promising tool [89].

2.2.4. Hepatocellular Cancer (HCC):

Hepatocellular carcinoma is the cancer of the liver. It is the sixth most common occurring cancer in the world and the third most common reason for deaths in the world with 781,631 estimated deaths in 2018 according to GLOBOCAN 2018 [90]. The major causes of HCC are age, gender, consumption of alcohol, cirrhosis, and hepatitis B virus (HBV) or C (HBC) or both etc [91, 92]. HBV is a DNA virus, which can occur from sexual contact with an infected person, needles, transfusion or by birth. HBV causes mutation in the liver cells which leads to HCC [93, 94]. HCV contributes a high percentage of cirrhosis cases which contributes to HCC [95, 96]. Cirrhosis is the most significant risk factor and increases the chance of HCC development. There are more chances of patients suffering from cirrhosis to develop HCC, if not, then they are most likely to exhibit an undetectable tumour during their lifetime [97-99].

Finding an effective screening method for HCC is difficult as it is highly dependent on the stage of the cancer and there are no curative treatments available for patients with high grade HCC. Therefore, it is important to find HCC at its early stages to improve the survival rate [100]. The most common screening method is monitoring AFP (Alpha-Fetoprotein) levels [100]. One limitation of using AFP levels for screening is that monitoring AFP levels is not a standardized practice. This is because some patients with high-grade HCC may have normal AFP levels, while patients with cirrhosis or chronic hepatitis may have elevated AFP levels, leading to false positive results [101-103]. Ultrasound is a recommended method for HCC detection. It is a non-invasive process, but the sensitivity varies with operator experience [104]. A study showed that ultrasound is the best available technology for the early detection of HCC but its sensitivity is only 63% [105]. Another study showed that ultrasound examinations were insufficient for HCC surveillance. They proved that ultrasonography gives inadequate results for obese patients and those with excessive alcohol consumption [106].

The screening tests available for HCC detection are low in sensitivity and not appropriate for population screening. Therefore, use of VOCs for early detection of HCC is a promising diagnostic method. A study conducted in China concluded that there are three potential biomarkers obtained from breath analysis. They investigated that 3-hydroxy-2-butanone, styrene and decane exist as HCC biomarkers with sensitivity and specificity of 86.7 % and 91.7% respectively. From this, they concluded 3-hydroxy-2-butanone as the best marker for HCC with 83.3 % sensitivity and 91.7 % specificity [107]. Another study used blood biomarkers for the detection of liver cancer. They used SPME (Solid Phase Micro Extraction) with GC/MS and detected three highly sensitive biomarkers hexanal, 1-octen-3-ol and octane in the blood for the detection of liver cancer [108]. Another study used GC/MS coupled with head-space needle tap extraction and found 2-pentanone, 3-heptanone, 3-octanone, dimethyl sulphide, ethyl methyl sulphide, 3-methyl thiophene, 2-methyl-1-(methylthio)-propane, n-propyl acetate and 2-heptene as VOCs markers for HCC [109].

2.2.5. Overview of Urinary Tract Infection (UTI):

a. Introduction:

UTI is caused by the presence of bacteria in the urinary tract including the bladder (cystitis), urethra, or kidneys (pyelonephritis) and is expected to infect 35% of healthy women in their lifetime. Most of the bacterial infections of the urinary tract occur in either the renal parenchyma or the bladder [110]. When the presence of bacteria in the urine increases more than 10^5 bacteria/ml, than this can lead to UTI [111, 112]. The pathogens responsible for UTIs originate in the rectal flora and move from the urethra to the bladder and the kidney. Since the distance to the bladder is shorter in women, it makes them more prone to UTIs [113]. It is estimated that 60% of women suffer from UTIs once in their lifetime and have a 30-50% chance of reoccurrence [114]. The signs and symptoms of a UTI may include fever, chills, dysuria, urinary urgency, frequency, and cloudy or malodorous urine. Though UTIs are not considered fatal, they can cause high distress and require high levels of healthcare and costs [115].

UTIs can be characterised as either complicated or uncomplicated infections. Uncomplicated infections are the most common infections with no anatomical or functional abnormalities. Risk factors that can increase the chances of developing a UTI are con-genital abnormalities, urinary obstructions or non-secretion of specific blood-group antigens, pregnancy, diabetes etc. Pathogens responsible for uncomplicated UTIs are *E. Coli,* which is responsible for more than 70% of cases, *Proteus mirabilis, Staphylococcus saprophyticus, Enterococcus faecalis, Klebsiella pneumoniae and Pseudomonas aeruginosa* with less than 10% of cases [116-118]. *E. Coli* have the potential to increase the virulence and help bacteria to grow, injuring the host cells and tissues. A study was conducted by Vosti *et al.* [119] in which they studied 235 women over a period of 1.1 to 19.4 years and found 1018 confirmed infections. They also found that *E. Coli* was responsible for almost 69.3% of the infections. Complicated UTIs are mostly related to the anatomical and functional abnormalities of the renal tract [120]. When the intensity of the pathogens increases, it disturbs the normal host defence and results in the disturbance in the urine flow or other disruptions in the urine tract

[121]. With complicated UTIs, it is important to know the extent of the infection. For example, if a person is suffering from kidney failure and contracts a complicated UTI, it becomes hard to eradicate it and infection keeps reappearing [122]. The pathogens responsible for complicated UTIs include *E. Coli,* Gram-negative aerobic bacilli, and Gram-positive cocci [123]. A study showed that *Pseudomonas aeruginosa* (gram-negative aerobic bacilli) was an important pathogen causing complicated UTIs [124].

b. UTI Diagnostic methods:

The gold standard for the diagnosis of UTI is urine culture. Urine culture is used to test the presence of pathogens in the urine and is expressed in terms of the number of bacteria per high-power field. However, it is important to provide proper care while collecting, preserving and transporting the urine as any contamination can lead to death of the bacteria and hence give false results [125, 126]. Urine culture is an expensive and complex procedure. Furthermore, it is required to cultivate the bacteria in the medium for around 18 hours which leads to 24-48 hrs delay in treatment [127]. Another test used for UTI diagnosis is dip stick urinalysis. A dip stick is used to test the presence of nitrite, leukocyte esterase, protein, and blood (as a marker of inflammation). The presence of nitrite in urine is the result of metabolic activities of pathogens in urinary tracts. Urinalysis is cheaper and faster in comparison to urine culture, however, this testing method has low sensitivity, and the testing depends on the possibility of symptoms [128].

Therefore, there is need for a rapid, cheap and point of care test which can be used to identify UTIs and lead to swift treatment [129]. UTI detection can be achieved by the analysis of VOCs that originate due to the presence of bacterial pathogen. It has been proved by different studies that VOCs present great potential for diagnosis and prognosis of UTIs [130, 131].

2.3. Analytical Literature Review: Investigative tools for cancer and disease diagnosis:

Detection of VOCs requires highly sensitive and accurate instrumentation in order to correctly identify and distinguish them at very low levels. There are different techniques that have been reported for the measurement and analysis of VOCs, for instance, Gas Chromatography-Mass Spectrometry (GC-MS), Gas Chromatography-Ion mobility Spectrometry (GC-IMS), Gas Chromatography- Time of Flight- Mass Spectrometry (GC-TOF-MS), Electronics nose (eNose) and Photoionization Detector (PID) amongst others [132].

2.3.1. Gas Chromatography - Mass Spectrometry (GC-MS)

Gas Chromatography (GC) was first introduced by James and Martin in 1952. They reported the separation of volatile fatty acids by partition chromatography with nitrogen gas as the mobile phase [133-135]. GC-MS is considered as the gold standard for the analysis of VOCs due to its ability to analyse chemicals at trace levels, high sensitivity, accuracy and with extensive compound databases and experimental protocols [136, 137]. Gas chromatography is a method in which the chemicals are separated inside a capillary column. When a sample is introduced to GC, the chemicals move inside the capillary column at different rates based on their physical and chemical properties. This results in separation of the chemicals at the end of the column [133, 138]. The separated chemicals are then introduced to MS where they are first ionised using an ion source, and then passed through a drift tube under a high electric field. Once the sample is ionized, the ions are accelerated in an electric field and separated based on their mass-to-charge ratio using a mass analyser. There are several types of mass analysers, including time-of-flight (TOF), quadrupole, and ion trap. Each type of mass analyser works differently to separate the ions based on their mass-to-charge ratio [139, 140]. Quite a number of researchers have shown the capabilities of GC-MS as an effective analytical technique in different applications/industries, such as, the perfume

industry [141], the medical industry [142, 143] and the food industry [144, 145]. The advantages and disadvantages of GC-MS are listed in Table 2.3-1.

Table 2.3-1: The advantages and disadvantages of GC-MS.

Advantages	Disadvantages
Highly sensitive technique, capable of detecting trace amounts of analytes even in complex mixtures	Expensive to purchase, operate, and maintain, making them less accessible to some laboratories.
Excellent specificity, enabling the identification and quantification of individual compounds in a sample	GC-MS analysis requires skilled operators and expertise in data interpretation, making it less user-friendly for inexperienced users.
GC-MS can analyse a broad range of volatile and semi-volatile compounds, making it suitable for diverse applications.	Complex samples with numerous analytes.
Mass spectral libraries can be used to compare unknown spectra with reference spectra, facilitating compound identification.	GC-MS is a destructive technique, as the sample is consumed during analysis, preventing further analysis or retesting.
High-resolution separation of analytes, enhancing the accuracy of analysis.	The presence of complex matrices in samples can interfere with GC-MS analysis and affect the accuracy of results.

2.3.2. Gas Chromatography- Ion Mobility Spectrometry (GC-IMS):

GC combined with IMS provides a very efficient, fast response and high accuracy method for the separation and detection of volatile organic compounds. GC-IMS provides a 2-dimensional chromatograph based on retention time and drift time which provides better accuracy than GC-MS [146]. GC is used for the separation of chemicals and IMS is used for detection. Ion Mobility Spectrometry (IMS) was first introduced in 1970 under the name of plasma spectrometry and became popular in succeeding decades after some modifications in engineering design and drift tube [147, 148].

A study showed the use of GC-IMS for the detection of pancreatic cancer from urinary volatile organic compounds. They analysed a total of 123 urine samples, out of which, there were 33 control samples, 45 chronic pancreatitis and 45 pancreatic cancer samples. The results showed that GC-IMS was able to separate pancreatic cancer from healthy controls

with a sensitivity of 84% and specificity of 94% [149]. Another study showed that GC-IMS demonstrates the potential for it to be used as a diagnostic tool for clinical studies [150]. They used GC-IMS to identify chronic obstructive pulmonary disease (COPD) VOCs from exhaled breath [150]. The advantages and disadvantages of GC-IMS are listed in Table 2.3-2.

Table 2.3-2: The advantages and disadvantages of GC-IMS.

Advantages	**Disadvantages**
Fast analysis of volatile compounds.	It does not offer precise identification of individual compounds, which could be a limitation in some applications.
Excellent sensitivity, capable of detecting low concentrations of volatile compounds even in complex samples.	GC-IMS can be sensitive to matrix effects, leading to interference from complex sample matrices.
GC-IMS offers good selectivity.	High-quality GC-IMS instruments can be costly to acquire and maintain.
GC-IMS is a non-destructive technique, preserving the original sample for further analysis or confirmation using complementary methods.	GC-IMS may have limited resolution for highly complex samples, potentially resulting in overlapping peaks and reduced accuracy in compound identification.
Some GC-IMS instruments are portable, allowing on-site or field analysis	GC-IMS instruments require regular calibration to maintain accuracy, and drift may occur over time, necessitating periodic recalibration

2.3.3. Gas Chromatography - Time of Flight-Mass Spectrometry (GC-TOF-MS):

Time of Flight-Mass Spectrometry (TOF-MS) was introduced by A.E. Cameron and D.F. Eggers in 1948. They built a TOF instrument with a pulsed ion source, field-free drift space and an oscillograph for detection. TOF-MS was first published in 1953 [151] and this was followed by several developments to improve its separation and detection capabilities [152]. GC-TOF-MS is a GC coupled with TOF-MS where GC is used for separation and TOF-MS is used for detection [153-155].

Several studies showed that GC-TOF-MS provides a high separation capability. Joanna *et al.* [156] in their study used a gas chromatograph 7890A (Agilent, Waldbronn, Germany)

coupled with a spectrometer TruTOF (Leco, St. Joseph, MI, USA) equipped with CP-Porabond-Q (Varian Inc., Middelburg, The Netherlands) 25 m × 0.25 m × 3 μm column to analyse the exhaled VOCs from lung cancer patients to differentiate them from healthy controls. This study showed that propane, carbon disulfide, 2-propenal, ethylbenzene and isopropyl alcohol were the five compounds distinguishing lung cancer patients from healthy controls. Another study demonstrated the use of GC-TOF-MS to represent the contribution of CCl_4 in liver fibrosis. They used an Agilent GC system coupled with a Pegasus HT TOF-MS and analysed the resulting statistical data using Chroma TOF4.3X software [157]. The advantages and disadvantages of GC-TOF-MS are listed in Table 2.3-3.

Table 2.3-3: The advantages and disadvantages of GC-TOF-MS.

Advantages	**Disadvantages**
GC-TOF-MS provides high-resolution mass spectra, enabling accurate identification of compounds even in complex samples.	GC-TOF-MS may produce fragmented spectra, making it challenging to identify and characterize complex compounds accurately.
It can detect a broad range of volatile and semi-volatile compounds.	Interpretation of GC-TOF-MS data requires specialized knowledge and expertise.
GC-TOF-MS is highly sensitive, capable of detecting low levels of analytes.	GC-TOF-MS can be sensitive to matrix effects, leading to interference from complex sample matrices.
The technique generates comprehensive data, providing detailed information on the mass and abundance of compounds present in the sample.	Proper sample preparation is critical for GC-TOF-MS analysis, and it can be time-consuming and labour-intensive.
GC-TOF-MS can be used for quantitative analysis, providing accurate quantification of compounds in the sample.	Regular maintenance and calibration of the instrument are necessary to ensure reliable and accurate results.

2.3.4. Overview of Electronic Nose (eNose) Technology:

The Electronic Nose is a device that attempts to imitate the human olfactory system. In human olfactory, volatile molecules are detected by receptor cells. For an eNose, that function is performed by an array of gas sensors. In 1982, G.Dodd and K.Persuad of the Warwick Olfaction Research Group introduced their work on artificial olfaction [158]. A basic eNose system is shown in Figure 2.3-1.

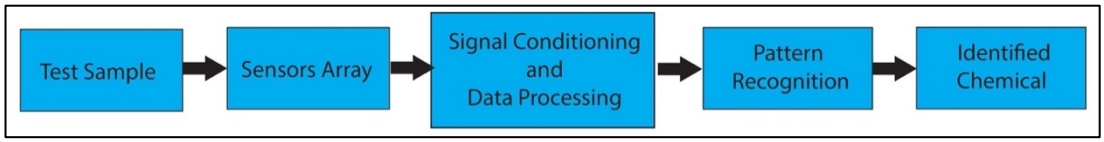


Figure 2.3-1: eNose System

ENose sensors can convert chemicals signals to electrical signals and respond to the concentration of chemicals, which include gases and VOCs, present in the test sample. Signals from different sensors can be combined into a single vector and used for quantitative and qualitative analysis of VOCs [159]. There are several different types of sensors that have been used in electronic nose systems.

A summary of the common ones is provided here:

a. Conducting Polymer Sensors

A conducting polymer (CP) gas sensor typically consists of a single layer of a conducting polymer coated onto an electrode. CP sensors became popular after the first electrochemical preparation and characterisation of polyaniline in 1862 [160]. Conducting polymers are organic compounds that can conduct electricity when they are oxidized or reduced. When a gas interacts with the surface of a conducting polymer, it can either donate or accept electrons, causing the polymer to become either more or less conductive [161]. They offer relatively low price, good durability, and broad-spectrum sensitivity whereas they suffer from oxidation drift which leads to large response time, effects of humidity and are temperature dependant. Figure 2.3-2 shows a diagram of CP sensors [162, 163].

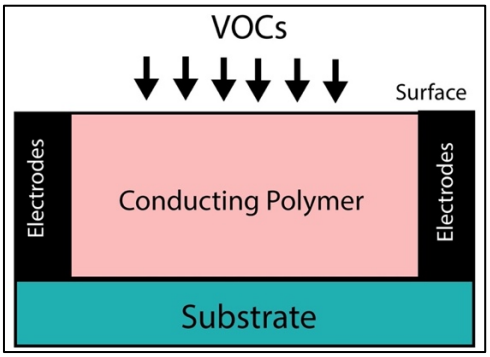


Figure 2.3-2: Conducting Polymer (CP) sensor.

The advantages and disadvantages of conducting polymer sensors are listed in Table 2.3-4.

Table 2.3-4: The advantages and disadvantages of Conducting Polymer sensors.

Advantages	**Disadvantages**
Conducting polymer sensors exhibit high sensitivity to changes in their environment, making them capable of detecting even small concentrations of target analytes.	conducting polymers have temperature limitations, which can affect their performance in extreme temperature conditions.
Conducting polymer eNose sensors typically respond rapidly to changes in the VOCs' environment.	These sensors can experience drift over time, requiring regular recalibration to maintain accuracy and consistency.
Compared to some other types of sensors, conducting polymer eNose sensors can be produced using relatively inexpensive materials and manufacturing processes, leading to cost-effective sensor arrays.	Conducting polymers may exhibit limited stability in harsh chemical environments, affecting the sensor's long-term reliability and performance.
Conducting polymer eNose sensors can be miniaturized, allowing for the development of portable and handheld devices	The response of conducting polymer eNose sensors may be influenced by environmental factors, interfering with the detection of target VOC.

b. Metal Oxide Semiconductor (MOS) sensors:

MOS sensors are the most commonly used sensors for eNose systems due to the widespread availability of the sensors. A metal oxide gas sensor typically consists of a thin film of metal oxide material deposited on a substrate, which is often a ceramic material, though more recently silicon is used. MOS sensors work on the principle of chemisorption, which occurs when gas molecules are adsorbed onto the surface of the metal oxide material and form chemical bonds with the surface atoms. This leads to the transfer of electrons between the

gas molecules and the metal oxide material, which changes the electrical conductivity of the material [164]. The advantages and disadvantages of MOS sensors are listed in Table 2.3-5.

Table 2.3-5: The advantages and disadvantages of MOS sensors.

Advantages	Disadvantages
MOS sensors offer fast response times, allowing real-time monitoring and quick detection of changes in the environment.	Regular calibration is often necessary to maintain accuracy, and MOS sensors may experience drift over time
MOS sensors can be tailored to respond selectively to specific gases or analytes, providing excellent selectivity in complex sample matrices.	Environmental conditions, such as temperature and humidity, can influence the sensor's response.
MOS sensors generally consume low power, making them energy efficient.	Some MOS sensors may exhibit reduced stability over prolonged use, resulting in a decline in sensitivity and selectivity.
MOS sensors can be miniaturized, making them suitable for integration into small and lightweight devices.	The sensitivity and performance of MOS sensors may change over time due to aging effects, impacting their long-term utility.

The sensitivity of these sensors is also affected by environmental factors [165, 166]. Figure 2.3-3 represents an illustration of a MOS sensor.

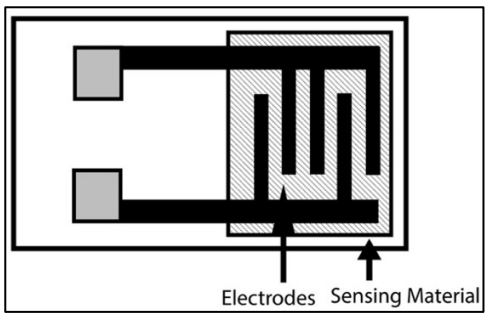


Figure 2.3-3: Metal Oxide Sensor

c. Quartz Crystal Microbalance (QCM) sensors:

QCM sensors consist of a quartz crystal that is equipped with thin-film gold electrodes on either side of it. An alternating current (AC) signal is applied to the electrodes, causing the crystal to vibrate at its resonant frequency, which is determined by its size, shape, and material properties. The change in frequency is proportional to the mass of the material that adheres to the surface of the sensor. Therefore, a QCM sensor can be used to measure the

mass of the deposited material [167]. There are almost limitless options for coating, allowing high sensitivity and a broad range of VOCs detection using QCM sensors. The main disadvantages of QCM sensors are environmental effects and poor signal to noise performance [168]. Figure 2.3-4 illustrates a QCM sensor.

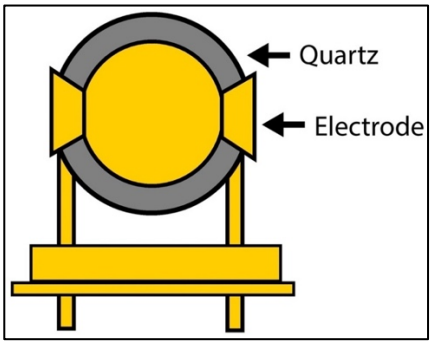


Figure 2.3-4: Quartz Crystal Microbalance Sensor.

The advantages and disadvantages of QCM sensors are listed in Table 2.3-6.

Table 2.3-6: The advantages and disadvantages of QCM sensors.

Advantages	**Disadvantages**
QCM sensors offer label-free detection, eliminating the need for additional chemical markers or labels, simplifying the assay process.	QCM sensors may lack inherent selectivity to specific analytes.
These sensors provide real-time monitoring capabilities, allowing dynamic and continuous measurements of molecular interactions or environmental changes.	Changes in temperature, humidity, and other environmental factors can impact the sensor's performance and introduce noise or interference.
QCM sensors require only small sample volumes, making them suitable for analysis with limited or precious samples.	Regular calibration is necessary to maintain accuracy and to correct for drift, which can add complexity to data analysis.
QCM sensors provide precise measurements, enabling accurate quantification and characterization of analytes.	High-quality QCM instruments can be costly to acquire.

d. Surface Acoustic Wave (SAW) sensors:

SAW sensors are comprised of piezoelectric crystals along with interdigital transducers (IDT). When an AC signal is applied to the transducer, an acoustic wave of a single

frequency is generated. This wave is generated by one electrode and measured by another and after interacting with VOCs, the mass of the crystal changes, which results in the change in the frequency of the wave [169]. The advantages and disadvantages of QCM sensors are listed in Table 2.3-7.

Table 2.3-7: The advantages and disadvantages of SAW sensors.

Advantages	**Disadvantages**
SAW sensors offer high sensitivity to changes in mass, viscosity, or other properties at the sensor surface, enabling the detection of small quantities of analytes.	The sensitivity of SAW sensors may vary for analytes with extremely high or low molecular weights, limiting their detection range in some cases.
SAW sensors provide label-free detection, eliminating the need for additional chemical markers or labels in the assay process.	Analysing SAW sensor data requires expertise in data interpretation and modelling to extract meaningful information about the interactions or processes being studied.
Some SAW sensors can operate wirelessly, allowing remote monitoring and integration into IoT (Internet of Things) systems.	The sensitivity of SAW sensors is affected by frequency shifts, which can complicate data analysis and interpretation.
These sensors enable real-time and continuous monitoring of interactions or changes in the environment, allowing dynamic data acquisition.	Contamination or fouling of the sensor surface can impact sensor performance and require cleaning or replacement.

Although SAW sensors are considered mass-sensitive, the high-mass sensitivity of SAW sensors decreases at high frequencies due to noise [170, 171]. Figure 2.3-5 shows an illustration of a SAW sensor.

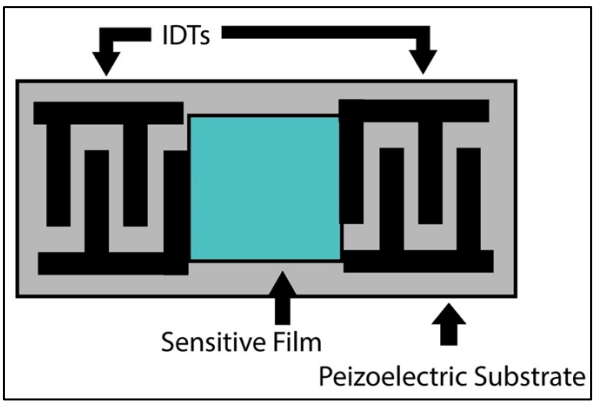


Figure 2.3-5: Figure shows a SAW sensor.

2.4. Photo-Ionisation Detector (PID) Technology:

Gas Chromatography and Mass spectrometry technologies, including GC-IMS and GC-TOF-MS, have presented very high separation capabilities. They have been one of the most employed analytical technologies in metabolic research due to their high robustness, high sensitivity, and selectivity. However, there are limitations to these technologies. They are very expensive, require high processing time, instrumentation complexity and need trained personal to operate them. Electronic noses are cheap and have easy to use sensors, but they are easily affected by the surrounding environment. Increase in temperature and humidity, the presence of oxygen, poor response time, non-linearities in the fabrication process etc limit their ability to be used for chemical detection. Therefore, there is a need for an analytical instrument which is portable, less costly, easy to use and does not require a lot of processing time.

One such technology is a Photo Ionisation Detector. A photoionization detector consists of a UV lamp which emits photons that are absorbed by the molecules of the gaseous sample and if the ionisation energies of the molecules present in the sample are less than the ionisation potential of the UV light, then this results in the ionisation of the molecules. These ions are then transported to the detector in the presence of the high electric field and produce electric current. The PID detector employed in this study was configured using Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS). FAIMS is a technique that exploits the variation in ion mobility of different ions when subjected to an asymmetric electric field. By applying an asymmetric waveform to the detector plates, ions with distinct mobilities encounter varying forces, resulting in selective transmission or filtration of ions. This capability enables the separation of ions based on their ion mobility, allowing FAIMS to selectively detect and identify specific compounds within complex mixtures. Through adjustment of the FAIMS parameters, targeted VOCs can be preferentially transmitted and detected, thereby improving the differentiation of various VOCs in the sample.

PID were found to be very sensitive to organic compounds in comparison to the available technologies in the 1960s, but were considered extremely unstable because they needed to

be operated at low pressure [172]. With the development of the PID with a sealed UV chamber by HNU Systems, Inc .(Newton Upper Falls, Massachusetts), PID have been extensively used as detectors for gas chromatography in the past few decades [173]. PID with a sealed UV chamber were tested and documented. Driscoll *et al.* [174] conducted a study that demonstrated improved sensitivity and dynamic range in a new design of PID (Photoionization Detector) with a 10.2 eV sealed UV source. They found that PID had 30 times more sensitivity than FIDs (Flame Ionisation Detectors). Another study conducted by Hobbs *et al.* [175] compared the ability of PID and electronic noses to distinguish between pig and chicken slurry based on the odour. They found that PID showed more sensitivity than an electronic nose Narayanan *et al.* [176] in their study reported a μPID compatible with μGC. They assembled a photoionizing gas detector with a micro-discharge and a remote electrode.

There is ongoing research on the use of PID for disease detection, particularly in the detection of volatile organic compounds (VOCs) that are associated with certain diseases. However, the use of PID as a standalone tool for disease detection is still in its early stages, and there are few published examples of its application in this field. One example of the use of a PID for disease detection can be found in a study by Gasparri *et al.* [177] which explored the potential of PID for the detection of urinary biomarkers associated with lung cancer. The study used GC-IMS and an electronic nose complemented by a photoionization detector to distinguish and identify VOCs in urine samples from lung cancer patients and healthy controls. The results of the study showed that the PID results were somewhat less effective than the GC-IMS results in distinguishing between individuals with early-stage lung cancer and those without, they still achieved an overall accuracy rate of 71.6% [177]. Another study by Sharma *et al.* [178] investigated the effectiveness of portable breath-based VOCs monitoring for the detection of COVID-19 infections. They utilized gas chromatography (GC) coupled with a photoionization detector (PID) to identify VOCs biomarkers in their study. Their findings revealed the presence of four VOCs - Benzene, Octane, Trimethyloctane, and Methyldecane - which demonstrated an ability to differentiate between COVID and non-COVID samples. Another study conducted by Oliva *et al.* [179] presented a novel approach for detecting VOCs using zeolite layers coupled with a PID. The author utilized a MiniRae 3000 PID consisting of an ionization chamber and a set of electrodes combined with a zeolite composite. This composite was used to function as a molecular

sieve, enabling the physical confinement of small molecules, which were later desorbed in an analysis chamber during photoionization detection. The study demonstrated the potential of this method for detecting VOCs in medical applications, as it was able to distinguish between different VOCs with good accuracy. Overall, this study highlighted the potential of the use of a PID coupled with zeolite layers for detecting VOCs for disease detection. PID sensors are considered to be highly sensitive in relative terms when compared to many other gas sensing technologies. They can detect a wide range of VOCs and other gases at very low concentrations. In many cases, PID sensors can detect VOCs in parts per billion (ppb) or even parts per trillion (ppt) levels. The sensitivity of PID sensors can vary depending on the specific model and design, as well as the characteristics of the target gases. Generally, PID sensors are capable of providing real-time, high-resolution measurements of gas concentrations, making them valuable tools for various applications, including environmental monitoring, industrial safety, and indoor air quality assessment. However, it is essential to note that the sensitivity of PID sensors can be influenced by factors such as the type of target gases, background levels of other gases, and the presence of interfering substances. Calibration and proper maintenance of the sensor are crucial for ensuring accurate and reliable measurements. Compared to some other gas sensing technologies, such as metal oxide sensors or catalytic bead sensors, PID sensors often offer superior sensitivity and selectivity, particularly for VOCs detection.

Miniaturized PID are extensions of conventional PID, utilizing microelectromechanical system (MEMS) technology to miniaturize the device. This class of PID has been deployed in various studies to test their efficacy in gas detection and analysis and has demonstrated varying levels of success. For instance, a study by Rezende *et al.* [180] developed a miniaturized photoionization detector (PID) that can be applied in various gas detection applications, including medical diagnostics and environmental monitoring. The device was designed using MEMS technology and comprises a micro gas chromatography column and a micro photoionization detector. They successfully demonstrated the feasibility of the device by detecting various VOCs that are commonly present in indoor air and concluded that the miniaturized PID could function as a standalone sensor or be integrated into a more complex analytical system. Though, miniaturized the PID offers several advantages over the traditional PID such as reduced size and weight, lower power consumption, and improved portability. However, they suffer from several disadvantages as well such as the small size

of the device which may result in limited sample throughput and reduced sample capacity, the potential for interference from background gases or other compounds in the sample matrix and in addition the use of MEMS technology in the fabrication of miniaturized PID devices can lead to increased complexity and cost compared to traditional PID instruments [181].

The basis of this research is inspired by the photoionization detector (PID) created by Owlstone Medical as a proof-of-concept prototype developed in-house for the purpose of detecting VOCs rather than a commercial product. The PID unit produced by Owlstone employs a 10.6 eV UV bulb for ionisation and necessitates a power supply ranging from +/- 5 V to 15 V @ 40 mA for multiple integrated circuits within the PID circuit, and 5 – 8 V @ 50 mA for the lamp supply. The unit generates two outputs within the +/- 3 V range, the first indicating the detected concentration and the second indicating the detected composition. The Owlstone Medical's PID unit had several disadvantages that prevented it from being commercialized. These included:

1. Requirement for frequent servicing and calibration: The unit had to be returned to the factory for servicing and calibration every three months, which was an inconvenience.
2. Voltage offset present on the concentration signal: Offsets within the internal circuitry caused voltage offset in the concentration signal, resulting in unstable data.
3. Complex maintenance: The PID unit required complex maintenance, which made it difficult to operate and maintain.

However, there are several PID which are commercially available. Table 2.4-1 contains the list of the commercially available PID.

Table 2.4-1: Commercial PID

S. No.	Commercial PID Product	PID Manufacturer
1	PID-A15, PID-AY5, PID-AH5, PID-AR5, PID-AG5, PID-AH2 and PID-A12	Alphasense Ltd., UK
2	PhoCheck Tiger, MiniPID2 sensor	Ion Science Ltd., UK
3	Mini RAE 3000, ppbRAE Plus, Ultra RAE 3000	RAE Systems Inc., USA

4	piD-TECH eVx	Baseline MOCON, USA
5	Gas-Pro	Crowcon
6	ToxiRAE Pro	Honeywell Ltd.,
7	VOC-TRAQ II	Myriad Industrial Solutions

a. PID Sensors by Alphasense Ltd., UK:

Alphasense offers a photoionization detector that can measure VOCs in the range of 0-10,000 ppm. Figure 2.4-1 shows the commercial PID-AH2 and PID-A12. All seven sensors require a voltage of between 3 - 3.6 V, consume 26-40 mA current and target a range of gases such as carbon monoxide, ethylene oxide, hydrogen chloride, hydrogen cyanide, VOCs etc [182]. Their compact size and light weight make them highly convenient for integration with a variety of systems, and also suitable for use with handheld devices. These products are either developed by Ion science or MOCON Ltd. but are rebranded by Alphasense.

The detection ranges for all the sensors listed in Table 2.4-2.

Table 2.4-2: Detection range of Alphasense PID

S. No.	PID Product	Detection Range
1	PID-A15	0-4000 ppm
2	PID-AY5	0-20 ppm
3	PID-AH5	0-40 ppm
4	PID-AR5	0-200 ppm
5	PID-AG5	0-10000 ppm
6	PID-AH2	1 ppb – 50 ppm
7	PID-A12	50 ppb - 6000 pm
8	SGX PID	Amphenol SGX Sensortech

The study conducted by Xiaobing *et al.* [183] showed the use of PID-AH to determine the potential of a commercial PID in comparison to the GC-Q-TOF-MS system. They used PID

as a standalone system as well as PID as a detector for a compact two-dimensional gas chromatography system to quantify different VOCs and reagents. Another study showed the potential of PID to be used for routine air quality testing [184]. They used PID-AH2 to test and verify the presence of benzene, ethylbenzene, isooctane, 2-propanol, toluene, and trichloroethylene in the air. They found that the PID showed a response to a wide variety of VOCs [184].

Figure 2.4-1: PID sensors from Alphasense Ltd [182].

b. MiniPID2 sensor, ION Science Ltd., UK:

MiniPID2 is another form of miniature PID sensor manufactured by ION science. They have 7 versions of MiniPI2 sensors with different detection ranges and capabilities [185]. They offer detection ranges from 0 to 10,000 ppm and response times of less than 3 sec and can be used in the detection of chemicals such as unsaturated fluorocarbons, formaldehyde, ethylene, chlorinated hydrocarbons and methanol [186]. PID require a supply voltage from 3.6 to 18V and the current consumption can range from 20 to 32 mA [187]. These are used in different industries such as oil and gas, food and beverage, defence, manufacturing, aerospace etc. The Ion science PID sensors are humidity resistant and provide long term stability. The small size of MiniPID2 sensors allow them to be easily applied within different systems, for example, the commercial mini-GC system PyxisGC BTEX, equipped with a MiniPID2 to test the preconcentrate material for BTEX (Benzene, Toluene, Ethylbenzene and Xylenes) which are harmful chemicals and can lead to various diseases such as cancer etc [188]. The most common use of a MiniPID2 is sensor Tiger, a handheld VOCs detector. Figure 2.4-2 shows an example of MiniPID2 from ION Science Ltd.

Figure 2.4-2: MiniPID2 sensors [187].

c. ppbRAE, RAE Systems Inc., USA:

A ppbRAE PID sensor is a 3rd generation PID/ VOCs monitor from RAE Systems which can detect more than 200 VOCs in a ppb range. It is used in many industries like oil and gas, defence, in hazardous environments, testing air quality etc. It provides real time sensor data with a 3 second response time [189]. Figure 2.4-3 shows an example of a ppbRAE PID sensor. Research conducted by V. Bocos-Bintintan *et al.* [190] used ppbRAE PID sensor to detect acetic anhydride and compared the results with an IMS system. They found that both PID and IMS were able to detect acetic anhydride successfully with different concentration levels. The main disadvantage of ppbRAE is the size. The bulb used in this PID alone is much bigger than the miniature PID sensors such as PID-AH2, PID-A12, MiniPID etc.

Figure 2.4-3: MiniPID2 sensors [189].

d. piD-TECH eVx, Baseline MOCON, USA:

MOCON piD-Tech eVx is an alternative PID sensor and can detect VOCs with as low as a 0.5 ppb detection range. The voltage supply required ranges from 3.2V to 5.5V with current consumption of 24mA to 36mA. It is used in various applications such as industrial hygiene and safety, soil contamination, leak detection, air quality etc [191]. Figure 2.4-4 shows the commercial PID sensors from Baseline MOCON.

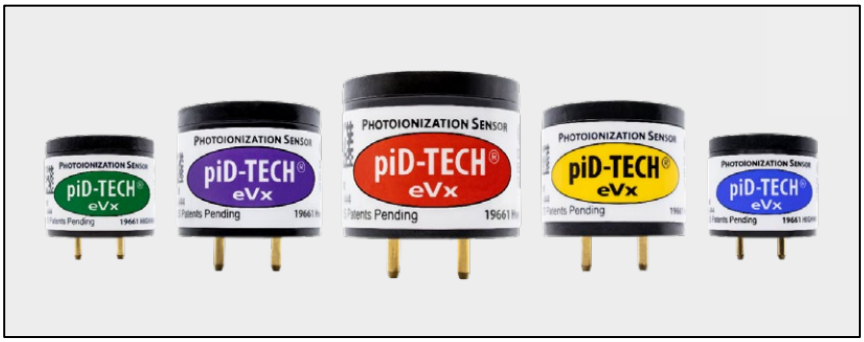


Figure 2.4-4: MiniPID2 sensors [191].

e. SGX PID, Amphenol SGX Sensortech:

Amphenol SGX Sensortech's PID sensors use UV lamps with different ionisation energies including 10.0eV, 10.6eV, and 11.7eV. The voltage supply required ranges from 3.6V to 10 V and can accurately detect VOCs with a sensitivity of up to 0.5ppb. The company's PID sensors are designed for use in a wide range of applications, including indoor air quality monitoring [192], environmental monitoring [192], and industrial safety [192]. They are available in a variety of sizes and configurations to suit different applications, and are highly sensitive and accurate, with fast response times and low power consumption.

Figure 2.4-5: SGX PID sensor [193].

2.5. Conclusion:

The detection and surveillance of volatile organic compounds (VOCs) have become crucial not only for disease detection but also for various other applications such as air quality monitoring, the defence industry, food, and beverage industries. Several analytical instruments are available for the detection of VOCs, but it is imperative to detect these compounds at trace levels, which necessitates the use of highly sensitive and accurate systems. Gas chromatography and mass spectrometry (GC-MS) coupled systems are widely regarded as the gold standard for VOCs detection due to their high sensitivity and selectivity capabilities. However, GC-IMS or GC-TOF-MS systems have their limitations. These instruments are costly, may require pre-concentration techniques, are complex, need skilled personnel to operate, and take longer to analyse, making them unsuitable for rapid medical diagnosis [194] [2, 195].

This highlights the importance of developing a point-of-care instrument that is simple, cost-effective, portable, and efficient for detecting a VOCs rapidly. One possible solution is the use of electronic nose (eNose) technology, which has evolved over time and employs various types of sensors such as MOS sensors, conducting polymer sensors, and Quartz Crystal Microbalance sensors. The eNose offers several benefits such as affordability, ease of integration, high versatility, reliability, fast response, and portability [196]. However, eNose technologies suffer from low selectivity, high operating temperatures, ageing effects, humidity etc. [197]. Another limiting factor for eNose technology is the inability to detect

individual VOCs in a gaseous mixture [198]. Due to these disadvantages, the eNose systems cannot be used as a standalone device for the diagnosis of diseases. Therefore, we need a device which can adapt the advantages of eNose and can overcome the disadvantages. A PID is one such device. It is a low-cost device with excellent sensitivity and a quick response [180, 199]. Commercially available PID provide advantages over eNose and GC-MS based instruments, but they suffer from limited sensitivity and linearity. A study was conducted to show if ppb-RAEs can be used detect VOCs. They used 3 unit of ppb-RAE to evaluate the repeatability and accuracy of the PID and to measure VOCs like toluene and ethyl acetate. They found that all three units produced different results and resulted in inconsistent performance [200]. Commercial instrumentation for online monitoring of gas with a PID is limited. This was the motivation for developing the PID+ prototype.

Therefore, the objective of this research is to develop a portable, low cost, easy to use and simple PID which may be able to overcome the disadvantages of commercially available PID and may be used to detect and diagnose diseases with better sensitivity and selectivity.

2.6. References:

[1] A. Pavlou and A. Turner, "Sniffing out the truth: clinical diagnosis using the electronic nose," 2000.

[2] S. Sethi, R. Nanda, and T. Chakraborty, "Clinical Application of Volatile Organic Compound Analysis for Detecting Infectious Diseases," *Clinical Microbiology Reviews,* vol. 26, no. 3, pp. 462-475, 2013, doi: doi:10.1128/CMR.00020-13.

[3] J. Y. H. Chang and S. Ladame, "Chapter 1.1 - Diagnostic, prognostic, and predictive biomarkers for cancer," in *Bioengineering Innovative Solutions for Cancer*, S. Ladame and J. Y. H. Chang Eds.: Academic Press, 2020, pp. 3-21.

[4] Y. Y. Broza, P. Mochalski, V. Ruzsanyi, A. Amann, and H. Haick, "Hybrid Volatolomics and Disease Detection," *Angewandte Chemie International Edition,* https://doi.org/10.1002/anie.201500153 vol. 54, no. 38, pp. 11036-11048, 2015/09/14 2015, doi: https://doi.org/10.1002/anie.201500153.

[5] Y. Y. Broza *et al.*, "Disease Detection with Molecular Biomarkers: From Chemistry of Body Fluids to Nature-Inspired Chemical Sensors," *Chemical Reviews,* vol. 119, no. 22, pp. 11761-11817, 2019/11/27 2019, doi: 10.1021/acs.chemrev.9b00437.

[6] Q. Gao *et al.*, "Application of Urinary Volatile Organic Compounds (VOCs) for the Diagnosis of Prostate Cancer," *Clinical Genitourinary Cancer,* vol. 17, no. 3, pp. 183-190, 2019/06/01/ 2019, doi: https://doi.org/10.1016/j.clgc.2019.02.003.

[7] A. Amann *et al.*, "The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva," *Journal of Breath Research,* vol. 8, no. 3, p. 034001, 2014/06/19 2014, doi: 10.1088/1752-7155/8/3/034001.

[8] T. Khalid *et al.*, "Urinary Volatile Organic Compounds for the Detection of Prostate Cancer," *PLOS ONE,* vol. 10, no. 11, p. e0143283, 2015, doi: 10.1371/journal.pone.0143283.

[9] R. P. Arasaradnam *et al.*, "Detection of Colorectal Cancer (CRC) by Urinary Volatile Organic Compound Analysis," *PLOS ONE,* vol. 9, no. 9, p. e108750, 2014, doi: 10.1371/journal.pone.0108750.

[10] M. L. Pelchat, C. Bykowski, F. F. Duke, and D. R. Reed, "Excretion and perception of a characteristic odor in urine after asparagus ingestion: a psychophysical and genetic study," *Chemical Senses,* vol. 36, no. 1, pp. 9-17, 2011.

[11] 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, p. 4, 2016.

[12] R. Jiang, E. Cudjoe, B. Bojko, T. Abaffy, and J. Pawliszyn, "A non-invasive method for in vivo skin volatile compounds sampling," *Analytica Chimica Acta,* vol. 804, pp. 111-119, 2013/12/04/ 2013, doi: https://doi.org/10.1016/j.aca.2013.09.056.

[13] P. Mochalski, J. King, K. Unterkofler, H. Hinterhuber, and A. Amann, "Emission rates of selected volatile organic compounds from skin of healthy volunteers," *Journal of Chromatography B,* vol. 959, pp. 62-70, 2014/05/15/ 2014, doi: https://doi.org/10.1016/j.jchromb.2014.04.006.

[14] H. S. Brown, D. R. Bishop, and C. A. Rowan, "The role of skin absorption as a route of exposure for volatile organic compounds (VOCs) in drinking water," *American Journal of Public Health,* vol. 74, no. 5, pp. 479-484, 1984/05/01 1984, doi: 10.2105/AJPH.74.5.479.

[15] U. R. Bernier, M. M. Booth, and R. A. Yost, "Analysis of Human Skin Emanations by Gas Chromatography/Mass Spectrometry. 1. Thermal Desorption of Attractants for the Yellow Fever Mosquito (Aedes aegypti) from Handled Glass Beads," *Analytical Chemistry,* vol. 71, no. 1, pp. 1-7, 1999/01/01 1999, doi: 10.1021/ac980990v.

[16] J. J. Haworth, C. K. Pitcher, G. Ferrandino, A. R. Hobson, K. L. Pappan, and J. L. D. Lawson, "Breathing new life into clinical testing and diagnostics: perspectives on volatile biomarkers from breath," *Critical Reviews in Clinical Laboratory Sciences,* vol. 59, no. 5, pp. 353-372, 2022/07/04 2022, doi: 10.1080/10408363.2022.2038075.

[17] L. Wallace, T. Buckley, E. Pellizzari, and S. Gordon, "Breath measurements as volatile organic compound biomarkers," *Environmental Health Perspectives,* vol. 104, no. suppl 5, pp. 861-869, 1996/10/01 1996, doi: 10.1289/ehp.96104s5861.

[18] T. Issitt, L. Wiggins, M. Veysey, S. T. Sweeney, W. J. Brackenbury, and K. Redeker, "Volatile compounds in human breath: critical review and meta-analysis," *Journal of Breath Research,* vol. 16, no. 2, p. 024001, 2022/02/23 2022, doi: 10.1088/1752-7163/ac5230.

[19] Y. Saalberg and M. Wolff, "VOC breath biomarkers in lung cancer," *Clinica Chimica Acta,* vol. 459, pp. 5-9, 2016/08/01/ 2016, doi: https://doi.org/10.1016/j.cca.2016.05.013.

[20] I. A. Ratiu, T. Ligor, V. Bocos-Bintintan, C. A. Mayhew, and B. Buszewski, "Volatile Organic Compounds in Exhaled Breath as Fingerprints of Lung Cancer, Asthma and COPD," *Journal of Clinical Medicine*, vol. 10, no. 1, doi: 10.3390/jcm10010032.

[21] M. Phillips, R. N. Cataneo, C. Saunders, P. Hope, P. Schmitt, and J. Wai, "Volatile biomarkers in the breath of women with breast cancer," *Journal of Breath Research,* vol. 4, no. 2, p. 026003, 2010/03/02 2010, doi: 10.1088/1752-7155/4/2/026003.

[22] M. Di Lena, F. Porcelli, and D. F. Altomare, "Volatile organic compounds as new biomarkers for colorectal cancer: a review," *Colorectal Disease,* https://doi.org/10.1111/codi.13271 vol. 18, no. 7, pp. 654-663, 2016/07/01 2016, doi: https://doi.org/10.1111/codi.13271.

[23] A. Bond *et al.*, "Volatile organic compounds emitted from faeces as a biomarker for colorectal cancer," *Alimentary Pharmacology & Therapeutics,* https://doi.org/10.1111/apt.15140 vol. 49, no. 8, pp. 1005-1012, 2019/04/01 2019, doi: https://doi.org/10.1111/apt.15140.

[24] R. P. Arasaradnam *et al.*, "Non-invasive exhaled volatile organic biomarker analysis to detect inflammatory bowel disease (IBD)," *Digestive and Liver Disease,* vol. 48, no. 2, pp. 148-153, 2016/02/01/ 2016, doi: https://doi.org/10.1016/j.dld.2015.10.013.

[25] S. Bosch *et al.*, "Early detection and follow-up of colorectal neoplasia based on faecal volatile organic compounds," *Colorectal Disease,* https://doi.org/10.1111/codi.15009 vol. 22, no. 9, pp. 1119-1129, 2020/09/01 2020, doi: https://doi.org/10.1111/codi.15009.

[26] M. McFarlane *et al.*, "Urinary volatile organic compounds and faecal microbiome profiles in colorectal cancer," *Colorectal Disease,* https://doi.org/10.1111/codi.14739 vol. 21, no. 11, pp. 1259-1269, 2019/11/01 2019, doi: https://doi.org/10.1111/codi.14739.

[27] C. E. Garner *et al.*, "Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease," *The FASEB Journal,* https://doi.org/10.1096/fj.06-6927com vol. 21, no. 8, pp. 1675-1688, 2007/06/01 2007, doi: https://doi.org/10.1096/fj.06-6927com.

[28] 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, p. e107312, 2014.

[29] H. Tyagi, E. Daulton, A. S. Bannaga, R. P. Arasaradnam, and J. A. Covington, "Non-Invasive Detection and Staging of Colorectal Cancer Using a Portable Electronic Nose," *Sensors,* vol. 21, no. 16, p. 5440, 2021.

[30] 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, no. 9, p. 2447, 2021.

[31] M. Phillips *et al.*, "Volatile Markers of Breast Cancer in the Breath," *The Breast Journal,* https://doi.org/10.1046/j.1524-4741.2003.09309.x vol. 9, no. 3, pp. 184-191, 2003/05/01 2003, doi: https://doi.org/10.1046/j.1524-4741.2003.09309.x.

[32] M. Phillips *et al.*, "Prediction of breast cancer using volatile biomarkers in the breath," *Breast Cancer Research and Treatment,* vol. 99, no. 1, pp. 19-21, 2006/09/01 2006, doi: 10.1007/s10549-006-9176-1.

[33] J. Giró Benet, M. Seo, M. Khine, J. Gumà Padró, A. Pardo Martnez, and F. Kurdahi, "Breast cancer detection by analyzing the volatile organic compound (VOC) signature in human urine," *Scientific Reports,* vol. 12, no. 1, p. 14873, 2022/09/01 2022, doi: 10.1038/s41598-022-17795-8.

[34] K. M. Banday *et al.*, "Use of Urine Volatile Organic Compounds To Discriminate Tuberculosis Patients from Healthy Subjects," *Analytical Chemistry,* vol. 83, no. 14, pp. 5526-5534, 2011/07/15 2011, doi: 10.1021/ac200265g.

[35] A. K. Pavlou, N. Magan, J. M. Jones, J. Brown, P. Klatser, and A. P. F. Turner, "Detection of Mycobacterium tuberculosis (TB) in vitro and in situ using an electronic nose in combination with a neural network system," *Biosensors and Bioelectronics,* vol. 20, no. 3, pp. 538-544, 2004/10/15/ 2004, doi: https://doi.org/10.1016/j.bios.2004.03.002.

[36] M. Phillips *et al.*, "Volatile biomarkers of pulmonary tuberculosis in the breath," *Tuberculosis,* vol. 87, no. 1, pp. 44-52, 2007/01/01/ 2007, doi: https://doi.org/10.1016/j.tube.2006.03.004.

[37] 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, 2020/04/01/ 2020, doi: https://doi.org/10.1016/j.ebiom.2020.102725.

[38] I. Ahmed, R. Greenwood, B. d. L. 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, p. e58204, 2013, doi: 10.1371/journal.pone.0058204.

[39] 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, p. 121, 2018.

[40] Hariyanto, R. Sarno, and D. R. Wijaya, "Detection of diabetes from gas analysis of human breath using e-Nose," in *2017 11th International Conference on Information & Communication Technology and System (ICTS)*, 31-31 Oct. 2017 2017, pp. 241-246, doi: 10.1109/ICTS.2017.8265677.

[41] H. Amal *et al.*, "The scent fingerprint of hepatocarcinoma: in-vitro metastasis prediction with volatile organic compounds (VOCs)," *International Journal of Nanomedicine,* vol. 7, pp. 4135-4146, 2012/12/01 2012, doi: 10.2147/IJN.S32680.

[42] Y. Hanai *et al.*, "Urinary volatile compounds as biomarkers for lung cancer," vol. 76, no. 4, pp. 679-684, 2012.

[43] K. Schmidt and I. Podmore, "Current challenges in volatile organic compounds analysis as potential biomarkers of cancer," *Journal of biomarkers,* vol. 2015, 2015.

[44] K. D. G. van de Kant, L. J. T. M. van der Sande, Q. Jöbsis, O. C. P. van Schayck, and E. Dompeling, "Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review," *Respiratory Research,* vol. 13, no. 1, p. 117, 2012/12/21 2012, doi: 10.1186/1465-9921-13-117.

[45] J. Zhou *et al.*, "Detection of volatile organic compounds using mid-infrared silicon nitride waveguide sensors," *Scientific Reports,* vol. 12, no. 1, p. 5572, 2022.

[46] A. D. Wilson, "Recent applications of electronic-nose technologies for the noninvasive early diagnosis of gastrointestinal diseases," in *Proceedings*, 2017, vol. 2, no. 3: MDPI, p. 147.

[47] J. Kwak and G. Preti, "Challenges in the investigation of volatile disease biomarkers in urine," ed: Elsevier: Amsterdam, The Netherlands, 2013, pp. 394-404.

[48] C. Davis and J. Beauchamp, *Volatile biomarkers: non-invasive diagnosis in physiology and medicine*. Newnes, 2013.

[49] L. Wu and X. Qu, "Cancer biomarker detection: recent achievements and challenges," *Chemical Society Reviews,* 10.1039/C4CS00370E vol. 44, no. 10, pp. 2963-2997, 2015, doi: 10.1039/C4CS00370E.

[50] G. A. Kwong, S. Ghosh, L. Gamboa, C. Patriotis, S. Srivastava, and S. N. Bhatia, "Synthetic biomarkers: a twenty-first century path to early cancer detection," *Nature Reviews Cancer,* vol. 21, no. 10, pp. 655-668, 2021/10/01 2021, doi: 10.1038/s41568-021-00389-3.

[51] J. Ferlay *et al.*, "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods," vol. 144, no. 8, pp. 1941-1953, 2019, doi: 10.1002/ijc.31937.

[52] G. 2018, "Estimated age-standardized incident rates in UK," IARC World Heath Organisation, 2018.

[53] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," vol. 68, no. 6, pp. 394-424, 2018, doi: 10.3322/caac.21492.

[54] A. Jemal, M. M. Center, C. DeSantis, and E. M. Ward, "Global Patterns of Cancer Incidence and Mortality Rates and Trends," vol. 19, no. 8, pp. 1893-1907, 2010, doi: 10.1158/1055-9965.EPI-10-0437 %J Cancer Epidemiology Biomarkers & Prevention.

[55] R. Etzioni *et al.*, "The case for early detection," *Nature Reviews Cancer,* vol. 3, no. 4, pp. 243-252, 2003/04/01 2003, doi: 10.1038/nrc1041.

[56] T. Bevers *et al.*, "23 - Screening and Early Detection," in *Abeloff's Clinical Oncology (Sixth Edition)*, J. E. Niederhuber, J. O. Armitage, M. B. Kastan, J. H. Doroshow, and J. E. Tepper Eds. Philadelphia: Content Repository Only!, 2020, pp. 375-398.e7.

[57] A. D. Wilson, "Advances in Electronic-Nose Technologies for the Detection of Volatile Biomarker Metabolites in the Human Breath," vol. 5, no. 1, pp. 140-163, 2015. [Online]. Available: http://www.mdpi.com/2218-1989/5/1/140.

[58] B. H. Bochner and V. P. J. O. Laudone, "Bladder cancer treatment: optimize, don't compromise," vol. 34, no. 1, 2020.

[59] A. Martini *et al.*, "The natural history of untreated muscle-invasive bladder cancer," *BJU International,* vol. 125, no. 2, pp. 270-275, 2020/02/01 2020, doi: 10.1111/bju.14872.

[60] A. Jung *et al.*, "Health-related quality of life among non-muscle-invasive bladder cancer survivors: a population-based study," vol. 125, no. 1, pp. 38-48, 2020, doi: 10.1111/bju.14888.

[61] M. G. Cumberbatch, M. Rota, J. W. F. Catto, and C. La Vecchia, "The Role of Tobacco Smoke in Bladder and Kidney Carcinogenesis: A Comparison of Exposures and Meta-analysis of Incidence and Mortality Risks," *European Urology,* vol. 70, no. 3, pp. 458-466, 2016/09/01/ 2016, doi: https://doi.org/10.1016/j.eururo.2015.06.042.

[62] N. D. Freedman, D. T. Silverman, A. R. Hollenbeck, A. Schatzkin, and C. C. Abnet, "Association Between Smoking and Risk of Bladder Cancer Among Men and Women," *JAMA,* vol. 306, no. 7, pp. 737-745, 2011, doi: 10.1001/jama.2011.1142 %J JAMA.

[63] R. C. Reulen, E. Kellen, F. Buntinx, M. Brinkman, and M. P. Zeegers, "A meta-analysis on the association between bladder cancer and occupation," *Scandinavian Journal of Urology and Nephrology,* vol. 42, no. sup218, pp. 64-78, 2008/01/01 2008, doi: 10.1080/03008880802325192.

[64] M. G. K. Cumberbatch, A. Cox, D. Teare, and J. W. F. Catto, "Contemporary Occupational Carcinogen Exposure and Bladder Cancer: A Systematic Review and Meta-analysis," *JAMA Oncology,* vol. 1, no. 9, pp. 1282-1290, 2015, doi: 10.1001/jamaoncol.2015.3209 %J JAMA Oncology.

[65] M.-P. Raitanen *et al.*, "Differences Between Local and Review Urinary Cytology in Diagnosis of Bladder Cancer. An Interobserver Multicenter Analysis," *European Urology,* vol. 41, no. 3, pp. 284-289, 2002/03/01/ 2002, doi: https://doi.org/10.1016/S0302-2838(02)00006-4.

[66] S. Bastacky, S. Ibrahim, S. P. Wilczynski, and W. M. Murphy, "The accuracy of urinary cytology in daily practice," vol. 87, no. 3, pp. 118-128, 1999, doi: 10.1002/(sici)1097-0142(19990625)87:3<118::Aid-cncr4>3.0.Co;2-n.

[67] C. G. T. Blick *et al.*, "Evaluation of diagnostic strategies for bladder cancer using computed tomography (CT) urography, flexible cystoscopy and voided urine cytology: results for 778 patients from a hospital haematuria clinic," vol. 110, no. 1, pp. 84-94, 2012, doi: 10.1111/j.1464-410X.2011.10664.x.

[68] B. W. TURNEY, J. M. G. WILLATT, D. NIXON, J. P. CREW, and N. C. COWAN, "Computed tomography urography for diagnosing bladder cancer," vol. 98, no. 2, pp. 345-348, 2006, doi: 10.1111/j.1464-410X.2006.06216.x.

[69] X. Jin, S. J. Yun, P. Jeong, I. Y. Kim, W.-J. Kim, and S. Park, "Diagnosis of bladder cancer and prediction of survival by urinary metabolomics," (in eng), *Oncotarget,* vol. 5, no. 6, pp. 1635-1645, 2014, doi: 10.18632/oncotarget.1744.

[70] M. B. Culp, I. Soerjomataram, J. A. Efstathiou, F. Bray, and A. Jemal, "Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates," *European Urology,* vol. 77, no. 1, pp. 38-52, 2020/01/01/ 2020, doi: https://doi.org/10.1016/j.eururo.2019.08.005.

[71] F. H. Schröder *et al.*, "Screening and Prostate-Cancer Mortality in a Randomized European Study," vol. 360, no. 13, pp. 1320-1328, 2009, doi: 10.1056/NEJMoa0810084.

[72] D. J. Rosario *et al.*, "Short term outcomes of prostate biopsy in men tested for cancer by prostate specific antigen: prospective evaluation within ProtecT study," vol. 344, p. d7894, 2012, doi: 10.1136/bmj.d7894 %J BMJ.

[73] J. Wade *et al.*, "Psychological impact of prostate biopsy: physical symptoms, anxiety, and depression," vol. 31, no. 33, pp. 4235-4241, 2013.

[74] M. E. Rezaee, C. E. Ward, E. F. Sverrisson, and L. M. Dagrosa, "Brief report: Impact of healthcare quality on prostate specific antigen screening for the early detection of prostate cancer," *Preventive Medicine Reports,* vol. 14, p. 100838, 2019/06/01/ 2019, doi: https://doi.org/10.1016/j.pmedr.2019.100838.

[75] W. J. Catalona *et al.*, "Measurement of Prostate-Specific Antigen in Serum as a Screening Test for Prostate Cancer," vol. 324, no. 17, pp. 1156-1161, 1991, doi: 10.1056/nejm199104253241702.

[76] S. Smith, P. White, J. Redding, N. M. Ratcliffe, and C. S. J. Probert, "Application of Similarity Coefficients to Predict Disease Using Volatile Organic Compounds," *IEEE Sensors Journal,* vol. 10, no. 1, pp. 92-96, 2010, doi: 10.1109/JSEN.2009.2035771.

[77] A. Jiménez-Pacheco, M. Salinero-Bachiller, M. C. Iribar, A. López-Luque, J. L. Miján-Ortiz, and J. M. Peinado, "Furan and p-xylene as candidate biomarkers for prostate cancer," *Urologic Oncology: Seminars and Original Investigations,* vol. 36, no. 5, pp. 243.e21-243.e27, 2018/05/01/ 2018, doi: https://doi.org/10.1016/j.urolonc.2017.12.026.

[78] T. Khalid *et al.*, "Urinary Volatile Organic Compounds for the Detection of Prostate Cancer," (in eng), *PloS one,* vol. 10, no. 11, pp. e0143283-e0143283, 2015, doi: 10.1371/journal.pone.0143283.

[79] M. Arnold *et al.*, "Recent trends in incidence of five common cancers in 26 European countries since 1988: Analysis of the European Cancer Observatory," *European Journal of Cancer,* vol. 51, no. 9, pp. 1164-1187, 2015/06/01/ 2015, doi: https://doi.org/10.1016/j.ejca.2013.09.002.

[80] A. J. Cornish *et al.*, "Modifiable pathways for colorectal cancer: a mendelian randomisation analysis," *The Lancet Gastroenterology & Hepatology,* vol. 5, no. 1, pp. 55-62, 2020/01/01/ 2020, doi: https://doi.org/10.1016/S2468-1253(19)30294-8.

[81] E. de Vries *et al.*, "Lifestyle changes and reduction of colon cancer incidence in Europe: A scenario study of physical activity promotion and weight reduction," *European Journal of Cancer,* vol. 46, no. 14, pp. 2605-2616, 2010/09/01/ 2010, doi: https://doi.org/10.1016/j.ejca.2010.07.040.

[82] B. Bressler, L. F. Paszat, Z. Chen, D. M. Rothwell, C. Vinden, and L. Rabeneck, "Rates of New or Missed Colorectal Cancers After Colonoscopy and Their Risk Factors: A Population-Based Analysis," *Gastroenterology,* vol. 132, no. 1, pp. 96-102, 2007/01/01/ 2007, doi: https://doi.org/10.1053/j.gastro.2006.10.027.

[83] J. H. Hancock and R. W. Talbot, "Accuracy of colonoscopy in localisation of colorectal cancer," *International Journal of Colorectal Disease,* vol. 10, no. 3, pp. 140-141, 1995/07/01 1995, doi: 10.1007/BF00298535.

[84] T. Morikawa, J. Kato, Y. Yamaji, R. Wada, T. Mitsushima, and Y. Shiratori, "A Comparison of the Immunochemical Fecal Occult Blood Test and Total Colonoscopy in the Asymptomatic Population," *Gastroenterology,* vol. 129, no. 2, pp. 422-428, 2005/08/01/ 2005, doi: https://doi.org/10.1053/j.gastro.2005.05.056.

[85] Z. Levi *et al.*, "A Quantitative Immunochemical Fecal Occult Blood Test for Colorectal Neoplasia," *Annals of Internal Medicine,* vol. 146, no. 4, pp. 244-255, 2007, doi: 10.7326/0003-4819-146-4-200702200-00003 %J Annals of Internal Medicine.

[86] F. A. OORT *et al.*, "Colonoscopy-controlled intra-individual comparisons to screen relevant neoplasia: faecal immunochemical test vs. guaiac-based faecal occult blood test," vol. 31, no. 3, pp. 432-439, 2010, doi: 10.1111/j.1365-2036.2009.04184.x.

[87] S. Tao, C. M. Seiler, U. Ronellenfitsch, and H. Brenner, "Comparative evaluation of nine faecal immunochemical tests for the detection of colorectal cancer," *Acta Oncologica,* vol. 52, no. 8, pp. 1667-1675, 2013/11/01 2013, doi: 10.3109/0284186X.2013.789141.

[88] D. F. Altomare *et al.*, "Exhaled volatile organic compounds identify patients with colorectal cancer," vol. 100, no. 1, pp. 144-150, 2013, doi: 10.1002/bjs.8942.

[89] M. Di Lena, F. Porcelli, and D. F. Altomare, "Volatile organic compounds as new biomarkers for colorectal cancer: a review," vol. 18, no. 7, pp. 654-663, 2016, doi: 10.1111/codi.13271.

[90] "Global Cancer Observatory," ed, 2019.

[91] J. A. Marrero, R. J. Fontana, S. Fu, H. S. Conjeevaram, G. L. Su, and A. S. Lok, "Alcohol, tobacco and obesity are synergistic risk factors for hepatocellular carcinoma," *Journal of Hepatology,* vol. 42, no. 2, pp. 218-224, 2005/02/01/ 2005, doi: https://doi.org/10.1016/j.jhep.2004.10.005.

[92] D. F *et al.*, "Hepatitis B and C virus infection, alcohol drinking, and hepatocellular carcinoma: A case-control study in Italy," *Hepatology,* vol. 26, no. 3, pp. 579-584, 1997, doi: 10.1002/hep.510260308.

[93] F. H. Pujol, M.-C. Navas, P. Hainaut, and I. Chemin, "Worldwide genetic diversity of HBV genotypes and risk of hepatocellular carcinoma," *Cancer Letters,* vol. 286, no. 1, pp. 80-88, 2009/12/01/ 2009, doi: https://doi.org/10.1016/j.canlet.2009.07.013.

[94] A. M. Di Bisceglie, "Hepatitis B and hepatocellular carcinoma," vol. 49, no. S5, pp. S56-S60, 2009, doi: 10.1002/hep.22962.

[95] M. S. Shiels, E. A. Engels, E. L. Yanik, K. A. McGlynn, R. M. Pfeiffer, and T. R. O'Brien, "Incidence of hepatocellular carcinoma among older Americans attributable to hepatitis C and hepatitis B: 2001 through 2013," vol. 125, no. 15, pp. 2621-2630, 2019, doi: 10.1002/cncr.32129.

[96] S. Bruno *et al.*, "Hepatitis C virus genotypes and risk of hepatocellular carcinoma in cirrhosis: A prospective study," vol. 25, no. 3, pp. 754-758, 1997, doi: 10.1002/hep.510250344.

[97] A. Amann, M. Corradi, P. Mazzone, and A. Mutti, "Lung cancer biomarkers in exhaled breath," *Expert Review of Molecular Diagnostics,* vol. 11, no. 2, pp. 207-217, 2011/03/01 2011, doi: 10.1586/erm.10.112.

[98] K. Kami *et al.*, "Metabolomic profiling of lung and prostate tumor tissues by capillary electrophoresis time-of-flight mass spectrometry," vol. 9, no. 2, pp. 444-453, 2013.

[99] J. Zhou, Z.-A. Huang, U. Kumar, and D. D. Y. Chen, "Review of recent developments in determining volatile organic compounds in exhaled breath as biomarkers for lung cancer diagnosis," *Analytica Chimica Acta,* vol. 996, pp. 1-9, 2017/12/15/ 2017, doi: https://doi.org/10.1016/j.aca.2017.09.021.

[100] M. Sherman, "Hepatocellular carcinoma: epidemiology, risk factors, and screening," in *Seminars in liver disease*, 2005, vol. 25, no. 02: Copyright© 2005 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New …, pp. 143-154.

[101] M. Corradi *et al.*, "Volatile hydrocarbons in exhaled air: preliminary data on the characteristic profile associated with lung tumors," vol. 25, no. 3, pp. 59-60, 2003.

[102] D. Poli *et al.*, "Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study," *Respiratory Research,* vol. 6, no. 1, p. 71, 2005/07/14 2005, doi: 10.1186/1465-9921-6-71.

[103] X. Chen *et al.*, "A study of the volatile organic compounds exhaled by lung cancer cells in vitro for breath diagnosis," vol. 110, no. 4, pp. 835-844, 2007, doi: 10.1002/cncr.22844.

[104] T. H.-H. Chen *et al.*, "Ultrasound screening and risk factors for death from hepatocellular carcinoma in a high risk group in Taiwan," vol. 98, no. 2, pp. 257-261, 2002, doi: 10.1002/ijc.10122.

[105] A. SINGAL *et al.*, "Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis," vol. 30, no. 1, pp. 37-47, 2009, doi: 10.1111/j.1365-2036.2009.04014.x.

[106] O. Simmons *et al.*, "Predictors of adequate ultrasound quality for hepatocellular carcinoma surveillance in patients with cirrhosis," vol. 45, no. 1, pp. 169-177, 2017, doi: 10.1111/apt.13841.

[107] T. Qin *et al.*, "The Screening of Volatile Markers for Hepatocellular Carcinoma," vol. 19, no. 9, pp. 2247-2253, 2010, doi: 10.1158/1055-9965.EPI-10-0302 %J Cancer Epidemiology Biomarkers & Prevention.

[108] R. Xue *et al.*, "Investigation of volatile biomarkers in liver cancer blood using solid-phase microextraction and gas chromatography/mass spectrometry," vol. 22, no. 8, pp. 1181-1186, 2008, doi: 10.1002/rcm.3466.

[109] P. Mochalski, A. Sponring, J. King, K. Unterkofler, J. Troppmair, and A. Amann, "Release and uptake of volatile organic compounds by human hepatocellular carcinoma cells (HepG2) in vitro," *Cancer Cell International,* vol. 13, no. 1, p. 72, 2013/07/17 2013, doi: 10.1186/1475-2867-13-72.

[110] B. J. Barnett and D. S. Stephens, "Urinary tract infection: an overview," *The American journal of the medical sciences,* vol. 314, no. 4, pp. 245-249, 1997.

[111] B. Foxman, "The epidemiology of urinary tract infection," *Nature Reviews Urology,* vol. 7, no. 12, pp. 653-660, 2010/12/01 2010, doi: 10.1038/nrurol.2010.190.

[112] T. L. Griebling, "Urinary tract infection in women," *Urologic diseases in America,* vol. 7, pp. 587-619, 2007.

[113] J. B. Lee and G. H. Neild, "Urinary tract infection," *Medicine,* vol. 35, no. 8, pp. 423-428, 2007.

[114] X. Albert, I. Huertas, I. Pereiro, J. Sanfélix, V. Gosalbes, and C. Perrotta, "Antibiotics for preventing recurrent urinary tract infection in non-pregnant women," *Cochrane Database of Systematic Reviews,* no. 3, 2004.

[115] B. Foxman, R. Barlow, H. D'Arcy, B. Gillespie, and J. D. Sobel, "Urinary tract infection: self-reported incidence and associated costs," *Annals of epidemiology,* vol. 10, no. 8, pp. 509-515, 2000.

[116] A. Moura, A. Nicolau, T. Hooton, and J. Azeredo, "Antibiotherapy and pathogenesis of uncomplicated UTI: difficult relationships," *Journal of Applied Microbiology,* https://doi.org/10.1111/j.1365-2672.2008.04115.x vol. 106, no. 6, pp. 1779-1791, 2009/06/01 2009, doi: https://doi.org/10.1111/j.1365-2672.2008.04115.x.

[117] T. M. Hooton, "Pathogenesis of urinary tract infections: an update," *Journal of Antimicrobial Chemotherapy,* vol. 46, no. suppl_1, pp. 1-7, 2000, doi: 10.1093/jac/46.suppl_1.1.

[118] A. Ronald, "The etiology of urinary tract infection: traditional and emerging pathogens," *The American Journal of Medicine,* vol. 113, no. 1, pp. 14-19, 2002, doi: 10.1016/S0002-9343(02)01055-0.

[119] K. L. Vosti, "Infections of the Urinary Tract in Women: A Prospective, Longitudinal Study of 235 Women Observed for 1–19 Years," *Medicine,* vol. 81, no. 5, 2002. [Online]. Available: https://journals.lww.com/md-journal/Fulltext/2002/09000/Infections_of_the_Urinary_Tract_in_Women__A.3.aspx.

[120] P. Lichtenberger and T. M. Hooton, "Complicated urinary tract infections," *Current infectious disease reports,* vol. 10, no. 6, pp. 499-504, 2008.

[121] A. M. Dobardzic and R. Dobardzic, "Epidemiological features of complicated UTI in a district hospital of Kuwait," *European Journal of Epidemiology,* vol. 13, no. 4, pp. 465-470, 1997/06/01 1997, doi: 10.1023/A:1007365317045.

[122] L. E. Nicolle, "Urinary Tract Pathogens in Complicated Infection and in Elderly Individuals," *The Journal of Infectious Diseases,* vol. 183, no. Supplement_1, pp. S5-S8, 2001, doi: 10.1086/318844.

[123] M. D. Melekos and K. G. Naber, "Complicated urinary tract infections," *International Journal of Antimicrobial Agents,* vol. 15, no. 4, pp. 247-256, 2000/08/01/ 2000, doi: https://doi.org/10.1016/S0924-8579(00)00168-0.

[124] K. Shigemura, S. Arakawa, Y. Sakai, S. Kinoshita, K. Tanaka, and M. Fujisawa, "Complicated urinary tract infection caused by Pseudomonas aeruginosa in a single institution (1999–2003)," *International Journal of Urology,* https://doi.org/10.1111/j.1442-2042.2006.01359.x vol. 13, no. 5, pp. 538-542, 2006/05/01 2006, doi: https://doi.org/10.1111/j.1442-2042.2006.01359.x.

[125] G. Schmiemann, E. Kniehl, K. Gebhardt, M. M. Matejczyk, and E. Hummers-Pradier, "The diagnosis of urinary tract infection: a systematic review," *Dtsch Arztebl Int,* vol. 107, no. 21, p. 361, 2010.

[126] T. M. Hooton and W. E. Stamm, "Diagnosis and treatment of uncomplicated urinary tract infection," *Infectious Disease Clinics,* vol. 11, no. 3, pp. 551-581, 1997.

[127] S. Najeeb, T. Munir, S. Rehman, A. Hafiz, M. Gilani, and M. Latif, "Comparison of urine dipstick test with conventional urine culture in diagnosis of urinary tract infection," *J Coll Physicians Surg Pak,* vol. 25, no. 2, pp. 108-10, 2015.

[128] S. Bent, B. K. Nallamothu, D. L. Simel, S. D. Fihn, and S. Saint, "Does This Woman Have an Acute Uncomplicated Urinary Tract Infection?," *JAMA,* vol. 287, no. 20, pp. 2701-2710, 2002, doi: 10.1001/jama.287.20.2701.

[129] P. Whiting, M. Westwood, I. Watt, J. Cooper, and J. Kleijnen, "Rapid tests and urine sampling techniques for the diagnosis of urinary tract infection (UTI) in children under five years: a systematic review," *BMC Pediatrics,* vol. 5, no. 1, p. 4, 2005/04/05 2005, doi: 10.1186/1471-2431-5-4.

[130] V.-M. Dospinescu, A. Tiele, and J. A. Covington, "Sniffing Out Urinary Tract Infection—Diagnosis Based on Volatile Organic Compounds and Smell Profile," *Biosensors,* vol. 10, no. 8, p. 83, 2020. [Online]. Available: https://www.mdpi.com/2079-6374/10/8/83.

[131] K. Hewett *et al.*, "Towards the Identification of Antibiotic-Resistant Bacteria Causing Urinary Tract Infections Using Volatile Organic Compounds Analysis—A Pilot Study," *Antibiotics,* vol. 9, no. 11, p. 797, 2020. [Online]. Available: https://www.mdpi.com/2079-6382/9/11/797.

[132] A. D. Wilson, "Noninvasive Early Disease Diagnosis by Electronic-Nose and Related VOC-Detection Devices," *Biosensors,* vol. 10, no. 7, p. 73, 2020. [Online]. Available: https://www.mdpi.com/2079-6374/10/7/73.

[133] H. M. McNair, J. M. Miller, and N. H. Snow, *Basic gas chromatography*. John Wiley & Sons, 2019.

[134] R. L. Grob and E. F. Barry, *Modern practice of gas chromatography*. John Wiley & Sons, 2004.

[135] K. D. Bartle and P. Myers, "History of gas chromatography," *TrAC Trends in Analytical Chemistry,* vol. 21, no. 9-10, pp. 547-557, 2002.

[136] H.-J. Hübschmann, *Handbook of GC-MS: fundamentals and applications*. John Wiley & Sons, 2015.

[137] A. Garcia and C. Barbas, "Gas Chromatography-Mass Spectrometry (GC-MS)-Based Metabolomics," in *Metabolic Profiling: Methods and Protocols*, T. O. Metz Ed. Totowa, NJ: Humana Press, 2011, pp. 191-204.

[138] A. B. Littlewood, *Gas chromatography: principles, techniques, and applications*. Elsevier, 2013.

[139] G. L. Glish and R. W. Vachet, "The basics of mass spectrometry in the twenty-first century," *Nature Reviews Drug Discovery,* vol. 2, no. 2, pp. 140-150, 2003/02/01 2003, doi: 10.1038/nrd1011.

[140] E. De Hoffmann and V. Stroobant, *Mass spectrometry: principles and applications*. John Wiley & Sons, 2007.

[141] A. van Asten, "The importance of GC and GC-MS in perfume analysis," *TrAC Trends in Analytical Chemistry,* vol. 21, no. 9, pp. 698-708, 2002/09/10/ 2002, doi: https://doi.org/10.1016/S0165-9936(02)00807-5.

[142] M. Monteiro *et al.*, "GC-MS metabolomics-based approach for the identification of a potential VOC-biomarker panel in the urine of renal cell carcinoma patients," *Journal of Cellular and Molecular Medicine,* https://doi.org/10.1111/jcmm.13132 vol. 21, no. 9, pp. 2092-2105, 2017/09/01 2017, doi: https://doi.org/10.1111/jcmm.13132.

[143] W. J. Griffiths and Y. Wang, "Analysis of neurosterols by GC–MS and LC–MS/MS," *Journal of Chromatography B,* vol. 877, no. 26, pp. 2778-2805, 2009/09/15/ 2009, doi: https://doi.org/10.1016/j.jchromb.2009.05.017.

[144] U. Roessner, C. Wagner, J. Kopka, R. N. Trethewey, and L. Willmitzer, "Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry," *the plant journal,* vol. 23, no. 1, pp. 131-142, 2000.

[145] P. M. Medeiros and B. R. Simoneit, "Analysis of sugars in environmental samples by gas chromatography–mass spectrometry," *Journal of Chromatography A,* vol. 1141, no. 2, pp. 271-278, 2007.

[146] C. Kwan *et al.*, "Chemical Agent Detection Using GC-IMS: A Comparative Study," *IEEE Sensors Journal,* vol. 10, no. 3, pp. 451-460, 2010, doi: 10.1109/JSEN.2009.2038128.

[147] H. Borsdorf, T. Mayer, M. Zarejousheghani, and G. A. Eiceman, "Recent Developments in Ion Mobility Spectrometry," *Applied Spectroscopy Reviews,* vol. 46, no. 6, pp. 472-521, 2011/08/01 2011, doi: 10.1080/05704928.2011.582658.

[148] A. B. Kanu, P. Dwivedi, M. Tam, L. Matz, and H. H. Hill Jr, "Ion mobility–mass spectrometry," *Journal of Mass Spectrometry,* https://doi.org/10.1002/jms.1383 vol. 43, no. 1, pp. 1-22, 2008/01/01 2008, doi: https://doi.org/10.1002/jms.1383.

[149] E. Daulton *et al.*, "Volatile organic compounds (VOCs) for the non-invasive detection of pancreatic cancer from urine," *Talanta,* vol. 221, p. 121604, 2021/01/01/ 2021, doi: https://doi.org/10.1016/j.talanta.2020.121604.

[150] M. Allers *et al.*, "Measurement of exhaled volatile organic compounds from patients with chronic obstructive pulmonary disease (COPD) using closed gas loop GC-IMS and GC-APCI-MS," *Journal of Breath Research,* vol. 10, no. 2, p. 026004, 2016/04/08 2016, doi: 10.1088/1752-7155/10/2/026004.

[151] B. A. Mamyrin, "Time-of-flight mass spectrometry (concepts, achievements, and prospects)," *International Journal of Mass Spectrometry,* vol. 206, no. 3, pp. 251-266, 2001/03/22/ 2001, doi: https://doi.org/10.1016/S1387-3806(00)00392-4.

[152] U. Boesl, "Time-of-flight mass spectrometry: Introduction to the basics," *Mass Spectrometry Reviews,* https://doi.org/10.1002/mas.21520 vol. 36, no. 1, pp. 86-109, 2017/01/01 2017, doi: https://doi.org/10.1002/mas.21520.

[153] R. J. Cotter, "Time-of-flight mass spectrometry: An increasing role in the life sciences," *Biomedical & environmental mass spectrometry,* vol. 18, no. 8, pp. 513-532, 1989.

[154] U. Boesl, "Time-of-flight mass spectrometry: introduction to the basics," *Mass spectrometry reviews,* vol. 36, no. 1, pp. 86-109, 2017.

[155] M. Guilhaus, "Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts," *Journal of mass spectrometry,* vol. 30, no. 11, pp. 1519-1532, 1995.

[156] J. Rudnicka, T. Kowalkowski, T. Ligor, and B. Buszewski, "Determination of volatile organic compounds as biomarkers of lung cancer by SPME–GC–TOF/MS and chemometrics," *Journal of Chromatography B,* vol. 879, no. 30, pp. 3360-3366, 2011/11/15/ 2011, doi: https://doi.org/10.1016/j.jchromb.2011.09.001.

[157] Y. Zhang *et al.*, "Serum metabonomics study of the hepatoprotective effect of amarogentin on CCl4-induced liver fibrosis in mice by GC-TOF-MS analysis," *Journal of Pharmaceutical and Biomedical Analysis,* vol. 149, pp. 120-127, 2018/02/05/ 2018, doi: https://doi.org/10.1016/j.jpba.2017.10.029.

[158] 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, 1982/09/01 1982, doi: 10.1038/299352a0.

[159] A. D. Wilson and M. Baietto, "Applications and Advances in Electronic-Nose Technologies," *Sensors,* vol. 9, no. 7, pp. 5099-5148, 2009. [Online]. Available: https://www.mdpi.com/1424-8220/9/7/5099.

[160] H. Letheby, "XXIX.—On the production of a blue substance by the electrolysis of sulphate of aniline," *Journal of the Chemical Society,* vol. 15, pp. 161-163, 1862.

[161] A. Ramanavičius, A. Ramanavičienė, and A. Malinauskas, "Electrochemical sensors based on conducting polymer—polypyrrole," *Electrochimica acta,* vol. 51, no. 27, pp. 6025-6037, 2006.

[162] J. J. Miasik, A. Hooper, and B. C. Tofield, "Conducting polymer gas sensors," *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases,* vol. 82, no. 4, pp. 1117-1126, 1986.

[163] M. H. Naveen, N. G. Gurudatt, and Y.-B. Shim, "Applications of conducting polymer composites to electrochemical sensors: A review," *Applied materials today,* vol. 9, pp. 419-433, 2017.

[164] C. Wang, L. Yin, L. Zhang, D. Xiang, and R. Gao, "Metal Oxide Gas Sensors: Sensitivity and Influencing Factors," *Sensors,* vol. 10, no. 3, 2010, doi: 10.3390/s100302088.

[165] K. Govardhan and A. N. Grace, "Metal/metal oxide doped semiconductor based metal oxide gas sensors—A review," *Sensor letters,* vol. 14, no. 8, pp. 741-750, 2016.

[166] P. T. Moseley, "Progress in the development of semiconducting metal oxide gas sensors: a review," *Measurement Science and Technology,* vol. 28, no. 8, p. 082001, 2017.

[167] A. Alassi, M. Benammar, and D. Brett, "Quartz Crystal Microbalance Electronic Interfacing Systems: A Review," *Sensors,* vol. 17, no. 12, p. 2799, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/12/2799.

[168] F. Pascal-Delannoy, B. Sorli, and A. Boyer, "Quartz Crystal Microbalance (QCM) used as humidity sensor," *Sensors and Actuators A: Physical,* vol. 84, no. 3, pp. 285-291, 2000/09/01/ 2000, doi: https://doi.org/10.1016/S0924-4247(00)00391-5.

[169] R. B. Priya, T. Venkatesan, G. Pandiyarajan, and H. M. Pandya, "A short review of saw sensors," *J. Environ. Nanotechnol,* vol. 4, no. 4, pp. 15-22, 2015.

[170] J. Devkota, P. R. Ohodnicki, and D. W. Greve, "SAW Sensors for Chemical Vapors and Gases," *Sensors,* vol. 17, no. 4, p. 801, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/4/801.

[171] H. K. Patel and M. J. Kunpara, "Electronic nose sensor response and qualitative review of e-nose sensors," in *2011 Nirma University International Conference on Engineering*, 2011: IEEE, pp. 1-6.

[172] A. N. Freedman, "The photoionization detector: Theory, performance and application as a low-level monitor of oil vapour," *Journal of Chromatography A,* vol. 190, no. 2, pp. 263-273, 1980.

[173] M. L. Langhorst, "Photoionization Detector Sensitivity of Organic Compounds*," *Journal of Chromatographic Science,* vol. 19, no. 2, pp. 98-103, 1981, doi: 10.1093/chromsci/19.2.98.

[174] J. N. Driscoll, "Evaluation of a new photoionization detector for organic compounds," *Journal of Chromatography A,* vol. 134, no. 1, pp. 49-55, 1977/04/01/ 1977, doi: https://doi.org/10.1016/S0021-9673(00)82568-6.

[175] P. J. Hobbs, T. H. Misselbrook, and B. F. Pain, "Assessment of Odours from Livestock Wastes by a Photoionization Detector, an Electronic Nose, Olfactometry and Gas Chromatography-Mass Spectrometry," *Journal of Agricultural Engineering Research,* vol. 60, no. 2, pp. 137-144, 1995/02/01/ 1995, doi: https://doi.org/10.1006/jaer.1995.1007.

[176] S. Narayanan, G. Rice, and M. Agah, "A micro-discharge photoionization detector for micro-gas chromatography," *Microchimica Acta,* vol. 181, no. 5, pp. 493-499, 2014/04/01 2014, doi: 10.1007/s00604-013-1146-9.

[177] R. Gasparri *et al.*, "Volatolomic urinary profile analysis for diagnosis of the early stage of lung cancer," *Journal of Breath Research,* vol. 16, no. 4, p. 046008, 2022/09/02 2022, doi: 10.1088/1752-7163/ac88ec.

[178] S. Ruchi *et al.*, "Portable Breath-Based Volatile Organic Compound Monitoring for the Detection of COVID-19: Challenges of Emerging Variants," *medRxiv,* p. 2022.09.06.22279649, 2022, doi: 10.1101/2022.09.06.22279649.

[179] G. Oliva and A. S. Fiorillo, "Detection of Volatile Organic Compounds Adsorbed Onto Zeolite Layers," in *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 22-24 June 2022 2022, pp. 1-5, doi: 10.1109/MeMeA54994.2022.9856403.

[180] G. Coelho Rezende, S. Le Calvé, J. J. Brandner, and D. Newport, "Micro photoionization detectors," *Sensors and Actuators B: Chemical,* vol. 287, pp. 86-94, 2019/05/15/ 2019, doi: https://doi.org/10.1016/j.snb.2019.01.072.

[181] H. Zhu *et al.*, "Flow-through microfluidic photoionization detectors for rapid and highly sensitive vapor detection," *Lab on a Chip,* 10.1039/C5LC00328H vol. 15, no. 14, pp. 3021-3029, 2015, doi: 10.1039/C5LC00328H.

[182] Alphasense. https://www.alphasense.com/products/pid/ (accessed.

[183] X. Pang, H. Nan, J. Zhong, D. Ye, M. D. Shaw, and A. C. Lewis, "Low-cost photoionization sensors as detectors in GC × GC systems designed for ambient VOC measurements," *Science of The Total Environment,* vol. 664, pp. 771-779, 2019/05/10/ 2019, doi: https://doi.org/10.1016/j.scitotenv.2019.01.348.

[184] J. Bílek, P. Maršolek, O. Bílek, and P. Buček, "Field Test of Mini Photoionization Detector-Based Sensors—Monitoring of Volatile Organic Pollutants in Ambient Air," *Environments,* vol. 9, no. 4, p. 49, 2022. [Online]. Available: https://www.mdpi.com/2076-3298/9/4/49.

[185] Ionscience. https://ionscience.com/en/sensors-and-components/oem-sensors/ (accessed.

[186] K. Thurow and S. Neubert, "Innovative Sensor Technology for Emergency Detection in Life Science Laboratories," in *Accident and Emergency Informatics*: IOS Press, 2022, pp. 62-87.

[187] I. S. Ltd. https://ionscience.com/en/products/minipid-2-ppm-gas-sensor/#:~:text=The%20MiniPID%202%20PPM%20VOC,VOCs)%20across%20many%20diverse%20applications. (accessed.

[188] A. Rozzi *et al.*, "Cavitand Decorated Silica as a Selective Preconcentrator for BTEX Sensing in Air," *Nanomaterials,* vol. 12, no. 13, p. 2204, 2022. [Online]. Available: https://www.mdpi.com/2079-4991/12/13/2204.

[189] R. S. Ltd. https://www.environst.com/ppbrae3000#:~:text=Using%20a%203rd%20gen%20photoionization,a%20three%2Dsecond%20response%20time. (accessed.

[190] V. Bocos-Bintintan, G.-B. Ghira, M. Anton, A.-V. Martiniuc, and I.-A. Ratiu, "Sensing Precursors of Illegal Drugs—Rapid Detection of Acetic Anhydride Vapors at Trace Levels Using Photoionization Detection and Ion Mobility Spectrometry," *Molecules,* vol. 25, no. 8, p. 1852, 2020. [Online]. Available: https://www.mdpi.com/1420-3049/25/8/1852.

[191] B. MOCON. https://www.ametekmocon.com/products/oemphotoionization/pidtechevxphotoionizationdetector (accessed.

[192] T. H. Frampton, A. Tiele, and J. A. Covington, "Development of a personalised environmental quality monitoring system (PONG)," *IEEE Sensors Journal,* vol. 21, no. 13, pp. 15230-15236, 2021.

[193] *PID-10.6eV 3*, 2022. [Online]. Available: https://www.mouser.co.uk/datasheet/2/18/1/Amphenol_7_18_2022_DS_0451_PID_10_6eV_3-3000419.pdf.

[194] K. M. Tripathi, T. Kim, D. Losic, and T. T. Tung, "Recent advances in engineered graphene and composites for detection of volatile organic compounds (VOCs) and non-invasive diseases diagnosis," *Carbon,* vol. 110, pp. 97-129, 2016/12/01/ 2016, doi: https://doi.org/10.1016/j.carbon.2016.08.040.

[195] H. Haick, Y. Y. Broza, P. Mochalski, V. Ruzsanyi, and A. Amann, "Assessment, origin, and implementation of breath volatile cancer markers," *Chemical Society Reviews,* 10.1039/C3CS60329F vol. 43, no. 5, pp. 1423-1449, 2014, doi: 10.1039/C3CS60329F.

[196] A. D. Wilson, "Advances in Electronic-Nose Technologies for the Detection of Volatile Biomarker Metabolites in the Human Breath," *Metabolites,* vol. 5, no. 1, pp. 140-163, 2015. [Online]. Available: https://www.mdpi.com/2218-1989/5/1/140.

[197] J.-B. Sanchez, F. Berger, M. Fromm, and M.-H. Nadal, "A selective gas detection micro-device for monitoring the volatile organic compounds pollution," *Sensors and Actuators B: Chemical,* vol. 119, no. 1, pp. 227-233, 2006/11/24/ 2006, doi: https://doi.org/10.1016/j.snb.2005.12.028.

[198] I. Mota, R. Teixeira-Santos, and J. Cavaleiro Rufo, "Detection and identification of fungal species by electronic nose technology: A systematic review," *Fungal Biology Reviews,* vol. 37, pp. 59-70, 2021/09/01/ 2021, doi: https://doi.org/10.1016/j.fbr.2021.03.005.

[199] S. O. Agbroko and J. Covington, "A novel, low-cost, portable PID sensor for the detection of volatile organic compounds," *Sensors and Actuators B: Chemical,* vol. 275, pp. 10-15, 2018/12/01/ 2018, doi: https://doi.org/10.1016/j.snb.2018.07.173.

[200] D. Choi *et al.*, "Laboratory Evaluation of the Accuracy, Precision, and Inter-instrumental Variance of a Portable Photoionization Detector," *Journal of Korean Society of Occupational and Environmental Hygiene,* vol. 22, no. 3, pp. 200-208, 2012.

Chapter 3. Analytical Instruments for VOCs Detection

This chapter introduces the primary instrumentation, methodology, and data analysis techniques employed for the various analytical instruments in the current research. The analytical technologies utilized in this study comprise GC-IMS (Gas Chromatography - Ion Mobility Spectrometry), GC-TOF-MS (Gas Chromatography - Time of Flight - Mass Spectrometry), and electronic noses. These instruments were employed for analysing biological samples to detect VOCs and to obtain qualitative and quantitative data pertaining to VOCs associated with various cancers and illnesses.

3.1. Gas Chromatography-Ion Mobility Spectrometry (GC-IMS):

3.1.1. Introduction:

The GC-IMS technique is a very powerful approach that uses the advantages of highly selective GC technology combined with highly sensitive IMS. GC-IMS is used for the separation and detection of Volatile Organic Compounds [1, 2]. Figure 3.1-1 illustrates a pictorial representation of the operation of GC-IMS.

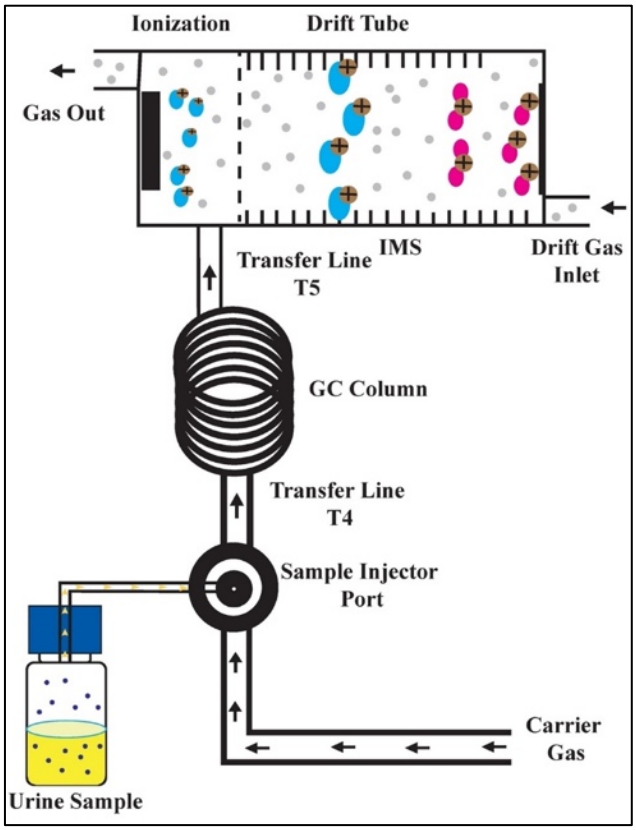


Figure 3.1-1: Principal operation of GC-IMS.

The GC-IMS instrument used for this research was a FlavourSpec device (G.A.S., Dortmund, Germany) shown in Figure 3.1-2. The FlavourSpec is an analytical instrument used for the analysis of volatiles in both liquid and solid headspaces. It consists of an automatic headspace injector that decreases the difficulty and time-consuming handling of samples. The results obtained from the FlavourSpec system can be analysed by G.A.S. VOCal software.

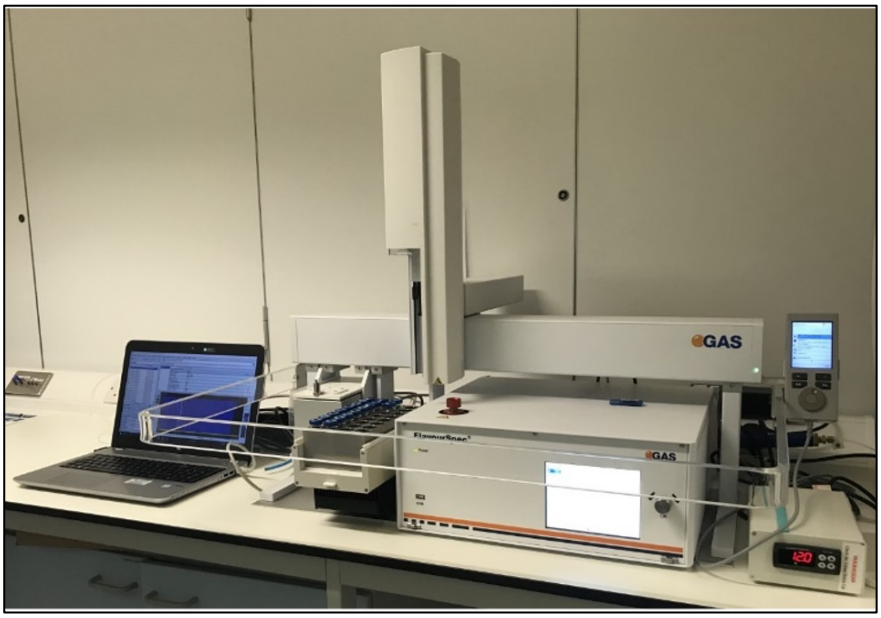


Figure 3.1-2: GC–IMS (FlavourSpec®; G.A.S., Dortmund, Germany) with PAL RSI (CTC Analytics) autosampler.

a. Principle of Operation: GC System:

The purpose of GC is to separate the individual chemical components of a complex sample mixture [3]. Gas chromatography operates on the fundamental principle of separating the chemical components of a sample using two phases. The stationary phase, which is a liquid, is coated on the interior of the column, and the gas phase serves to transport the sample through the column [4].

The length, temperature, and coating material inside the column are critical factors that influence separation in gas chromatography. The sample is introduced into the column via a syringe and transported along the column by a mobile phase (i.e., a carrier gas). For gas chromatography, hydrogen, helium, nitrogen, or argon are typically used as these gases are inert and do not interact with the compounds being analysed. The carrier gas carries the compounds across the column, and the purity of the gas flow is crucial for column efficiency. Maintaining a proper linear gas flow throughout the system is important to ensure effectiveness and enable qualitative analysis. Temperature is also a vital factor in optimizing the efficiency of GC-IMS. It helps to enhance analysis, improve selectivity, and achieve good separation [5].

Since different compounds have different distribution/separation constant, some of the compounds are retained by the coating on the column or stationary phase for different lengths of time. Compounds that have stronger interactions with the stationary phase will be retained in the column for a longer time. They "stick" to the stationary phase more effectively, resulting in slower movement through the column. Compounds that have weaker interactions with the stationary phase will be less retained and spend less time in the column. They move more quickly through the column with the carrier gas. This results in the separation of the compounds. The time taken by the chemicals to elute from the column is called the retention time [6].

b. Principle of Operation: Ion Mobility Spectrometry (IMS):

Ion Mobility Spectrometry (IMS) is a technique used for ion separation and detection based on the mobility of the ionised compounds. The first step for an IMS system is sample introduction. There are two types of IMS techniques as follows:

1. Drift-tube IMS (DTIMS): This is the most common type of IMS, which uses a long tube filled with a buffer gas, such as nitrogen or helium, and an electric field to separate ions. The ions are introduced into the tube and are separated based on their drift time through the buffer gas.
2. FAIMS (high-field asymmetric waveform ion mobility spectrometry): This type of IMS uses asymmetric waveform electric fields to separate ions based on their mobility. FAIMS is often used as a front-end separation technique in mass spectrometry to improve selectivity and reduce sample complexity.

In this study, a Drift-tube IMS was considered. The analytes from GC are introduced to the drift tube of the IMS. The approach used for sample introduction is unidirectional, i.e., the analytes are introduced in the drift tube directly into the buffer gas flow that flows in the opposite direction and swept out depending upon the electric field strength and the number of collisions with the buffer gas [7, 8]. The ionisation method in the G.A.S. FlavourSpec is soft chemical-ionisation (CI). Soft chemical-ionisation generates ions by a cascade of reactions following the collision of a fast electron emitted from the β-radiator H3. This

transfer of charge carrier from a charged electron to the reactant ion results in low internal energy ions.

In this IMS system, the drift tube is the main part of the IMS system. It provides an electric region for the ions to be migrated towards the detector under the influence of an electric field. The drift tube is filled with a neutral gas which flows in the opposite direction to slow down the ions based upon their mass-to-charge (m/z) ratio. When analytes are introduced to the drift tube, a force is applied to the ions directly proportional to the magnitude of the electric field [9, 10]. Figure 3.1-3 represents a drift tube for IMS.

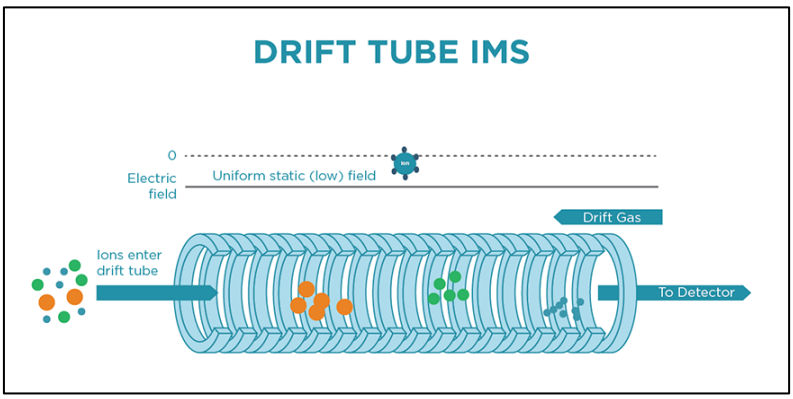


Figure 3.1-3: Drift Tube IMS [11].

The time taken by the ions to drift to the electrodes is called the drift time [12]. The drift time depends on several factors such as the density of the drift gas, the number of collisions, the electric field applied etc [13]. Also, the velocity with which ions migrate towards the detector under the electric field is called drift velocity v_d (m/s), which is given by:

$$v_d = \mu E \qquad \textit{Equation 1}$$

where μ is mobility ($m^2/(V \cdot s)$) and E (V/m) is the electric field applied.

The main function of the ion detector is to convert an ion current into an electrical current. The electrometer converts an ion current into an electric current, which is then amplified and converted into a voltage. These voltages are measured by the detector at different drift times creates a plot called a chromatogram, which contains all the quantitative information

about the ion based on its separation with respect to its drift velocity in the form of peak heights and area [14].

c. Parameter Settings for the G.A.S. FlavourSpec:

An autosampler headspace injector (PAL RSI from CTC Analytics) was used to introduce samples into the FlavourSpec device. It consisted of an incubator, agitator, and gastight syringe where the incubator stored and heated the current sample under analysis, the agitator shook the sample to increase the sample headspace concentration and a gastight syringe was used to transfer the sample into the FlavourSpec. The incubator temperature was set at 40°C and the agitation time to 10 min. These parameters were set after several optimizations and are limited to water-based samples only.

The carrier gas used throughout the FlavourSpec experiments was Nitrogen. A nitrogen generator (G.A.S, Dortmund, Germany) was used to provide gas through the inlet, the column, and the detector at a speed of 150 ml/min. The carrier gas used was 99.995% pure. The temperature setup for the various components of the G.A.S. FlavourSpec was as follows: Temperature of IMS was heated to 45°C (T1), the GC to 40°C (T2), the injector to 80°C (T3), the T4 transfer line to 80°C and the T5 transfer line to 45°C.

The GC Column used was SE-54. SE-54 is a non-bonded or cross-linked column based on a poly silicone (94%methyl/ 5%phenyl/ 1%vinyl). The length of SE-54 is 30 m, and its internal diameter is 0.25 mm.

The length of the drift tube for the G.A.S FlavourSpec is 53 mm, and a constant voltage of 500 V was applied. The ionisation of the molecules in the G.A.S. FlavourSpec was initiated by Radioactive - Tritium H3 (ß— Radiation).

d. Sample Preparation:

Urine samples used in this research were collected directly from the patient and stored at -80°C in standard universal sterile specimen containers according to standard operating

procedures complaint with tissue bank requirements under the Human Tissue Act 2004. This study was approved by Coventry and Warwickshire and North-East Yorkshire NHS Ethics Committees (Ref 18717 and Ref 260179). The samples were then transferred to University of Warwick and briefly stored at -20°C. For sample analysis, the urine samples were defrosted at room temperature. The defrosted samples were transferred from sample bottles to 20 mL glass vials with crimp caps. 5 ml of each sample were then transferred into 20 ml glass vials which were sealed with Blue BiMetal Seal, 3mm Thick PTFE/Blue Silicone, Ultra Low Bleed crimp caps (Thames Restek, FI5150BMUL-20B) using an appropriate crimp tool.

The glass vials containing the samples were transferred to the autosampler (PAL RSI from CTC Analytics). The sample tray was chilled to 4°C to reduce degradation of the samples while waiting for the analysis. Every glass vial containing the sample was transferred to the agitator and agitated for 10 minutes at a temperature of 40°C before sampling. 0.5 mL of sample was transported to the device using a gastight syringe. The machine settings for the analysis are given in Table 3.1-1. The total run time per sample was set to 10 minutes.

Table 3.1-1: Parameter settings of G.A.S. FlavourSpec.

S. No.	Parameter	Parameter Setting
1	The temperature of GC Column	80°C
2	The temperature of injector	70°C
3	The temperature of drift tube for IMS	45°C
4	The flow rate of Carrier gas	20 mL/min
5	The flow rate for drift gas	150 mL/min

e. Qualitative and Quantitative Analysis:

The output obtained from the urine samples using G.A.S. FlavourSpec is in the form of a chromatogram as shown in Figure 3.1-4. The x-axis represents the drift time spectra from the IMS and is measured in milliseconds, and the y-axis represents the retention time of the GC and is measured in seconds. In the figure, the ‘dots’ are the chemicals detected by the IMS and the intensity of the peak represents the number of ions. The ‘dots’ in red are the

most intense. 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 software used for viewing and processing the FlavourSpec data was VOCal (v0.1.3, G.A.S., Dortmund, Germany). FlavourSpec data were extracted from a .mea file to a .csv file format which was further pre-processed using a bespoke LabView program.

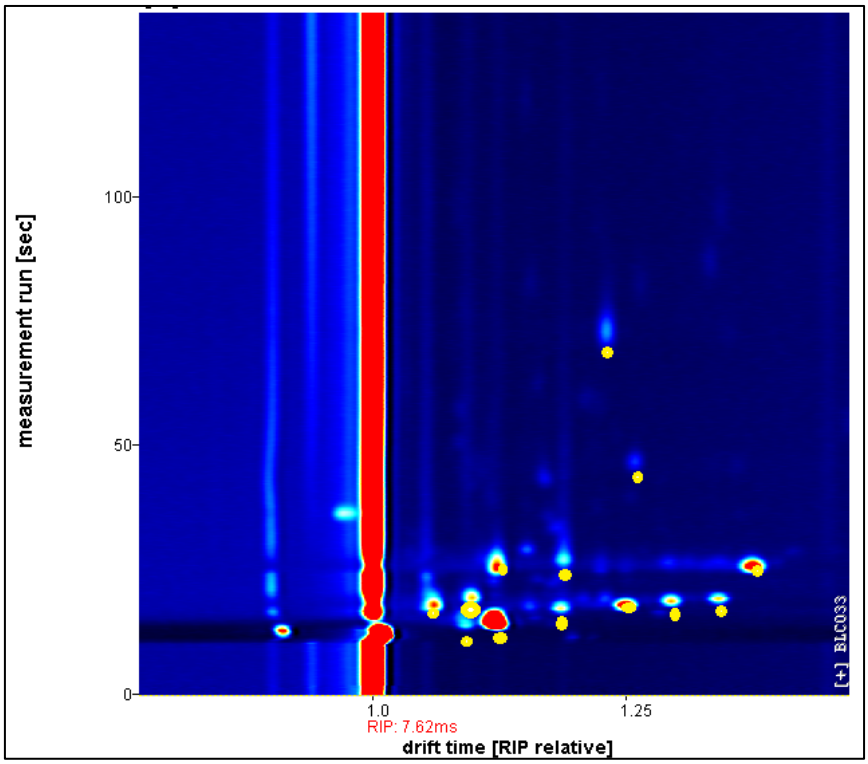


Figure 3.1-4: Typical output plot from the GC-IMS instrument.

The data analysis pipeline used to analyse raw data from the GC-IMS in this study generally include pre-processing, data splitting, cross-validation, feature selection, classification, calculation of performance/generating ROC curves, and significant features as shown in Figure 3.1-6.

1. Pre-processing:

The primary goal of pre-processing is to decrease the data dimensionality by eliminating non-informative information from the raw GC-IMS output files, which tend to be complex and high-dimensional. This involves implementing several procedures, such as RIP alignment, cropping, and thresholding [15].

Since the GC-IMS operates at atmospheric pressure, the carrier and drift gas, usually nitrogen or air, contains water, which plays a crucial role in the ionization process. Water molecules interact with the fast electrons emitted by the beta radiation source, leading to the formation of reactant ions known as RIP. These ions are present in all spectra and can be used for sample alignment [16]. The RIP remains consistent across all samples; however, slight shifts in its position may occur due to variations in ambient pressure caused by weather conditions. In order to tackle this issue, the pre-processing stage includes RIP alignment, where a single sample is selected as the reference, and all other RIP samples are adjusted to align with the position of the reference sample. This alignment is crucial for assisting the feature selection process in identifying the most informative features to be used as input for the classification model. The RIP alignment standardizes the data, thereby facilitating accurate feature selection and enhancing the overall performance of the classification model [17-19].

The typical GC-IMS output matrix comprises approximately 11 million data points. However, these high-dimensional data do not align with the actual information content stored in the data. Over 50% of these data points are considered background noise, visually represented as a blue area in Figure 3.1-4. To address this issue, the subsequent pre-processing steps involve cropping out the area of interest and applying a threshold to the dataset. These procedures result in a pre-processed GC-IMS output matrix where background data points are zero-valued, and non-zero data points represent VOCs peaks [20]. Figure 3.1-5 represents a pictorial representation of the process followed for pre-processing the GC-IMS data.

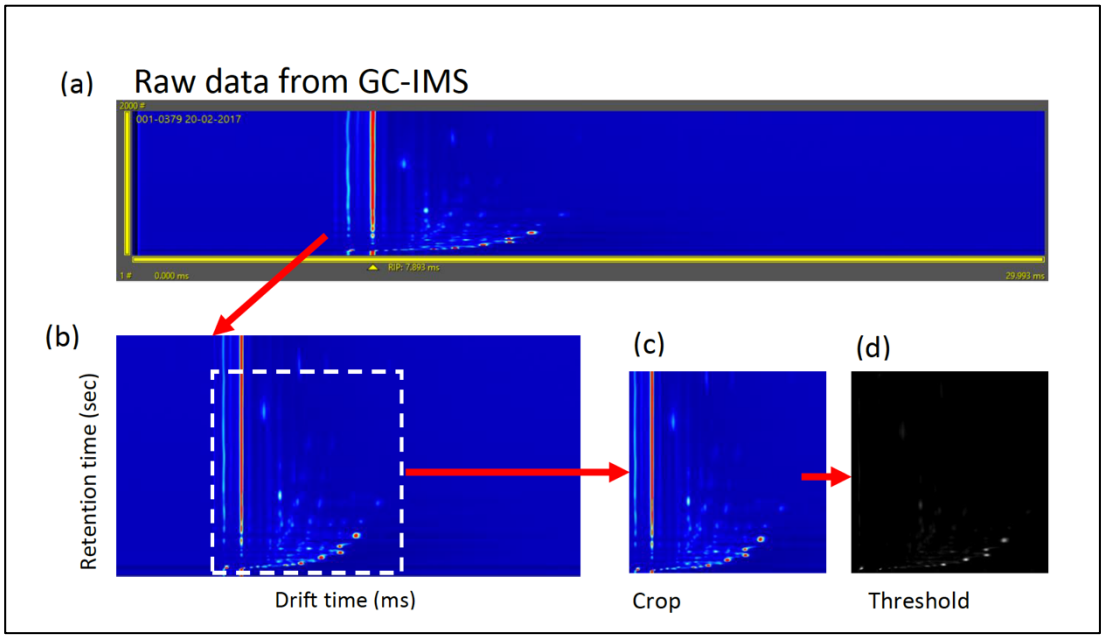


Figure 3.1-5: Pre-processing steps for the GC-IMS instrument (a) Raw data from the GC-IMS device, (b) RIP alignment, (c) Cropping of the important features, and (d) Applying threshold to remove the noise.

2. Data Splitting:

Data splitting constitutes a crucial step in the research workflow, following the pre-processing of the raw data, wherein noise and unwanted variability are effectively removed. This pivotal process takes place prior to the construction of a classification model. Statistical quantification of the data was performed using a custom R program (version 3.6.2) tailored to the specific requirements of the study and can be found at GitHub (https://github.com/JimSkinner/toftools).

A 10-fold Cross-validation approach was adopted in this research to address the delicate balance between computational efficiency and minimizing bias while estimating model performance. The 10-fold cross-validation method involves partitioning the original datasets into ten equivalent subgroups to produce meaningful outcomes. Subsequently, these ten subsets are divided into two distinct groups: one comprising nine subsets for training purposes and the other, constituting the remaining subset, retained for meticulous model assessment. The training subset serves as the foundation for creating feature sets, developing, and meticulously comparing multiple models, all aimed at achieving the final model's optimum configuration [21]. Once the final model is firmly established, it is

subjected to rigorous evaluation using the test subset, enabling a robust quantification of its performance. An exemplary model is one that not only delivers accurate predictions on test data but also showcases its capacity to extrapolate insights beyond the training data [22, 23].

Notwithstanding these advantages, subpar generalization, characterized by overtraining, can pose a potential challenge. Overtraining arises when the model becomes excessively attuned to the patterns within the training data, ultimately impairing its ability to discern patterns in the data beyond the confines of the training set. Ascertaining an appropriate data splitting methodology becomes imperative in the quest to develop a well-generalized model, armed with the capability to proficiently handle new and uncharted data [24].

3. Feature Selection:

Despite applying pre-processing techniques to the GC-IMS data, the dimensionality remains relatively high, encompassing around 1,976,000 data points in the GC-IMS. Consequently, to further mitigate this issue, we incorporated a supervised feature selection technique into the pipeline. Specifically, in this research, we employed the Wilcoxon rank sum test as the supervised feature selection method. By conducting comparisons between groups derived from the pre-processed data, we identified the feature point with the lowest p-value, thereby streamlining the number of features. The p-value is used to determine the significance of the features [25, 26].

In the initial stage, we conduct the classification analysis employing 20, 50, or 100 features with the lowest p-values to determine the optimal number of features that yield the highest classification performance. This approach ensures sufficient information for the algorithms to learn while minimizing the risk of overfitting, while also enabling the identification of significant features at a later stage in the pipeline. Nevertheless, the present limitation of this method lies in the fact that the feature selection process relies solely on statistical criteria and does not incorporate any consideration of biological functions.

4. Classification Training Models:

Following the feature selection with the Wilcoxon rank sum test, the chosen features serve as inputs for machine learning classification algorithms implemented within a 10-fold cross-validation framework. For this study, six sophisticated non-parametric machine learning algorithms were selected, namely Random Forest (RF), Neural Network (NN), Logistic

Regression (LR), Support Vector Machine (SVM), Xtreme Gradient Boost (XGBoost), and Gaussian Process (GP). These algorithms are well-suited for handling high-dimensional data with limited samples, making them suitable choices for the analysis [27].

The Random Forest (RF) method, devised by Breiman in 2001, has emerged as one of the most successful general-purpose algorithms. Comprising numerous decision trees, RF leverages bootstrapping to build each tree, where subsets of the training data are randomly selected with replacement. The individual trees are trained to learn patterns and make class predictions. Aggregating the predictions of each tree results in a final classification outcome. RF is renowned for its accuracy and robustness, especially in scenarios with small sample sizes and high-dimensional feature spaces, justifying its inclusion in our analysis [28].

Neural Network (NN) attempts to emulate the learning process of the human brain, composed of interconnected "neurons" that collaborate to solve problems. The neurons in the NN model are based on non-linear functions that produce outputs based on one or more inputs. Analogous to synapses in the brain, each neuron transmits a signal (a real number) to other neurons, and associated weights determine the strength of input signals relative to other neurons. Typically, NN consists of multiple layers of neurons, with the final layer yielding a single classification result [29, 30].

A Support Vector Machine (SVM) is a supervised machine learning algorithm that seeks to identify an optimal hyperplane with maximum margin distance in an n-dimensional space, effectively distinguishing different classes of data points. The hyperplane serves as a decision surface, aiding the classification of data points based on their respective classes. Support vectors, which are data points nearest to the hyperplane, influence its location and orientation. Maximizing the margin distance reinforces subsequent data point categorization. In handling nonlinear problems, SVM employs a kernel function that transforms low-dimensional data into higher-dimensional data, aiding the search for the optimal hyperplane. SVM is particularly well-suited for binary classification tasks with high-dimensional datasets [31].

A Gaussian Process (GP) is a supervised probabilistic classification model that expresses test predictions as class probabilities. A GP applies a prior probability distribution on a latent function, which is then transformed via a link function. While the latent function facilitates

model construction, its value is neither observable nor pertinent and is thus treated as a nuisance function, eliminated during prediction. For binary classification, GP utilizes the logistic link function [32].

Logistic Regression is a statistical method used for binary classification, where the outcome variable takes only two possible values. It aims to model the relationship between the binary outcome variable (dependent variable) and one or more predictor variables (independent variables) by estimating the probability of the outcome belonging to a specific category. The logistic function converts the linear combination of predictor variables and coefficients into a probability value, determining the classification outcome based on a predefined threshold (usually 0.5) [33].

Xtreme Gradient Boosting (XGBoost) represents a powerful machine learning algorithm suitable for both regression and classification tasks. As an ensemble method, XGBoost combines predictions from multiple weak learners, often decision trees, to construct a robust and highly accurate final model. Its efficiency, scalability, and capacity to handle large datasets render XGBoost a popular choice for various applications [34].

The following machine learning packages were used in this study: RF- random Forest; Neural Network using ‘nnet’ package; Logistic Regression with elastic net regularisation using glmnet; Gaussian Process classifier radial basis kernel (using ‘kernlab’ R package’); extreme gradient boosting XGBoost; and Support Vector Machine (RBF kernel).

5. Performance Parameters:

Following the completion of the training process within the cross-validation, the final model is subsequently evaluated using the testing set. The performance of the models is measured by several key metrics, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), p-value, Receiver Operating Characteristic (ROC) curve, and the Area Under the ROC curve (AUC) [35].

Sensitivity is a measure of the model's ability to correctly identify positive instances from the total actual positive instances in the dataset. High sensitivity indicates that the model is good at identifying positive cases and vice-versa. Specificity is the true negative rate,

representing the model's ability to correctly identify negative instances from the total actual negative instances in the dataset. High specificity indicates that the model is effective in identifying negative cases and vice-versa. Positive Predictive Value (PPV), also known as precision, is the proportion of true positive predictions over the total predicted positive instances. It measures the accuracy of positive predictions made by the model. Negative Predictive Value (NPV) represents the proportion of true negative predictions over the total predicted negative instances. It measures the accuracy of negative predictions made by the model. The p-value is a statistical measure used to determine the significance of a result in hypothesis testing. In the context of model evaluation, it may be used to assess the significance of differences between different models or approaches. A low p-value suggests that the differences are likely not due to random chance and are statistically significant [36].

The Receiver Operating Characteristic (ROC) Curve is a graphical representation of the model's performance at various classification thresholds. It plots the true positive rate (sensitivity) against the false positive rate (1 - specificity). The ROC curve helps visualize how the model performs across different classification thresholds and is particularly useful when dealing with imbalanced datasets. The Area Under the ROC Curve (AUC) is a single scalar value that quantifies the overall performance of a model's ROC curve. It represents the area under the ROC curve, and its value ranges from 0 to 1. A higher AUC indicates better discrimination power of the model, meaning it can distinguish between positive and negative instances more effectively [37].

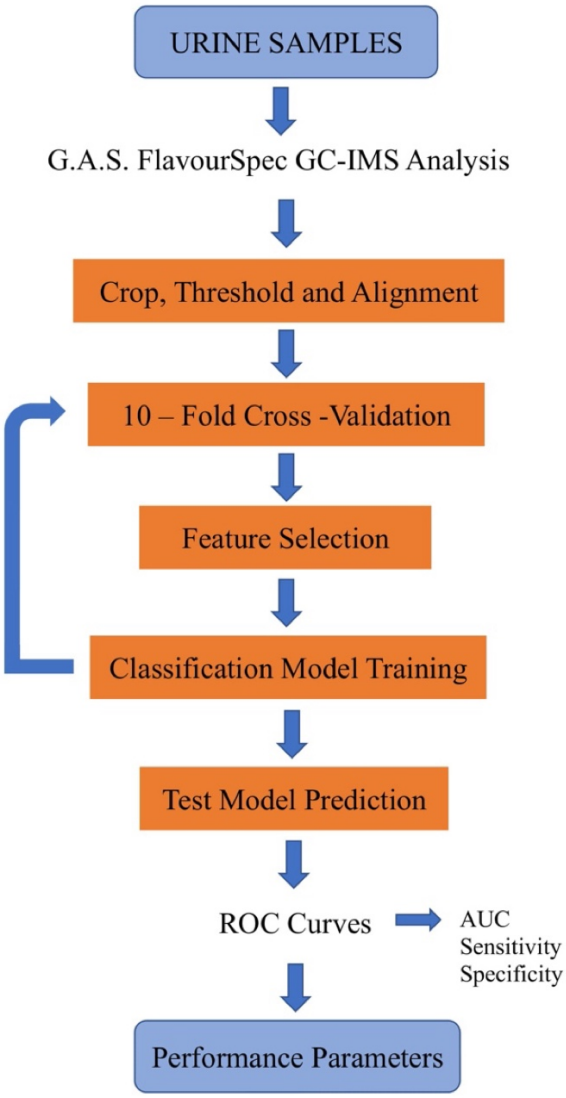


*Figure 3.1-6***:** *Data analysis pipeline for G.A.S. FlavourSpec GC-IMS*

3.2. Gas Chromatography - Time-of-Flight - Mass Spectrometry:

3.2.1. Introduction:

GC-TOF-MS is a combination of GC technology, a separation technique, with TOF- Mass Spectrometry, an identification technique. Time of Flight was first reported by Cameron and Eggers in 1948. In 1958, the first commercial time of flight instrument was manufactured

by Bendix after Wiley and McLaren achieved significant improvements in resolving the power of TOF in 1955 [38, 39]. TOF-MS basically works on the principle of mass separation. In TOF-MS, an ion source is used to generate ions which are accelerated towards a detector by applying high voltage. The ions are then made to travel in a field free region based on their mass to charge ratio. Since the ions traverse in a field-free time of flight tube, the velocity of the ions remains the same. The time taken by the ions to travel to the detector, also called the time of flight, depends on the mass to charge ratio of the ions [40]. This can be seen by the equations below:

The potential energy E_p (Joules, J) of the ion when a high voltage V (Volts, V) is applied to it is given by:

$$E_p = qV \qquad \textit{Equation 2}$$

where q is charge of the ion (Coulomb, C) and V is voltage (Volts, V) applied.

When a voltage V (V) is applied to the ion, it accelerates in a time-of-flight tube and gains kinetic energy E_k (Joules, J) which is expressed by:

$$E_k = 1/2\ mv^2 \qquad \textit{Equation 3}$$

where m is the mass of the ion (Kg), and v is the velocity of the ion (m/s).

Since the potential energy of the ion is converted into kinetic energy, we can equate,

$$E_p = E_k \qquad \textit{Equation 4}$$

$$qV = 1/2\ mv^2 \qquad \textit{Equation 5}$$

Also, we know, velocity can be expressed in terms of distance d cm and time t in sec as:

$$v = d/t \qquad \textit{Equation 6}$$

Substituting these equations, we get,

$$qV = 1/2\ m(d/t)^2 \qquad \text{Equation 7}$$

$$t^2 = 2md^2/qV \qquad \text{Equation 8}$$

Since we know that the distance and voltage applied are constants, we can say,

$$t\ \alpha\ \sqrt{m/q} \qquad \text{Equation 9}$$

where *t* represents time of flight of the ion and *m/q* is the mass to charge ratio.

3.2.2. TRACE 1300 GC - BenchTOF-HD TOF-MS:

The GC-TOF-MS system used is a combination of a TRACE 1300 GC (Thermo Fisher Scientific, Loughborough, UK) and a BenchTOF-HD TOF-MS (Markes Intl., Llantrisant, UK) shown in Figure 3.2-1. GC-TOF-MS provides high acquisition rates and high-resolution capabilities. The high-resolution power allows the user to generate more precise peaks and more accurate mass. The data from GC-TOF-MS analysis were identified using the National Institute of Standards and Technology (NIST) list (2011 and 2020).

TRACE 1300 GC is a user-friendly, fully programmable GC which is highly sensitive and capable of detecting analytes at very low concentrations, even at trace levels. This sensitivity allows for the identification and quantification of compounds present in the sample at extremely low concentrations, making it an ideal tool for trace-level analysis in complex samples. BenchTOF-HD TOF-MS provides a high definition, high sensitivity, high productivity, and good analytical performance. It offers selective elimination of interferences through its high mass resolution and accurate mass measurement capabilities [41].

By resolving closely spaced peaks in the mass spectra, TRACE 1300 GC and BenchTOF-HD TOF-MS can distinguish between different compounds with similar m/z values, reducing the risk of false identifications. Additionally, the accurate mass measurement

allows for the precise determination of the elemental composition of the ions, which aids in the identification of target analytes while discriminating against interferences [42].

Figure 3.2-1: TRACE 1300 GC - BenchTOF-HD TOF-MS with ULTRA-xr and UNITY-xr

a. Principle of Operation: Time – of – Flight – MS (TOF-MS) System:

In TOF-MS, analytes are ionised using Electron Ionisation (EI). The EI source consists of a heated chamber and a heated filament. The heated filament emits electrons which are accelerated to 70 eV. The heated chamber is maintained under a high vacuum where analytes are introduced through a sample hole after passing through the GC. The acceleration of the electrons in the proximity of the analytes results in the collision between them and formation of a charged ion [43]. Figure 3.2-2 illustrates the working principle of EI.

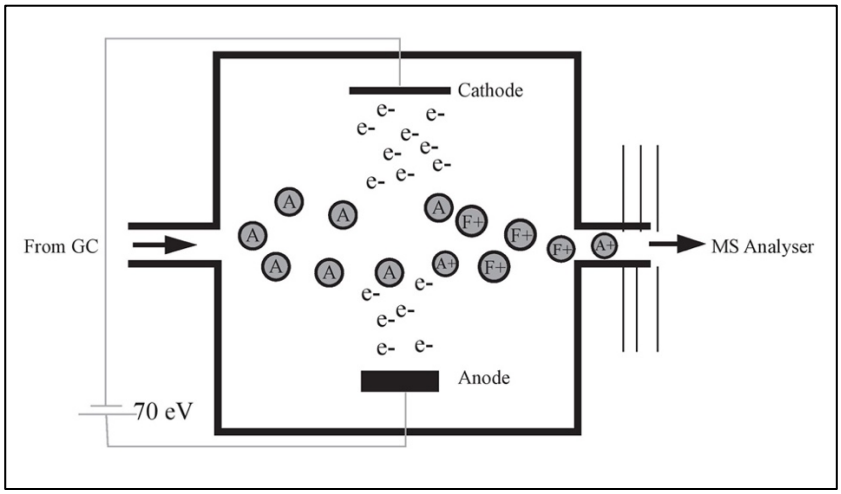


Figure 3.2-2: Electron Ionisation.

These charged ions then pass through a two-stage reflectron as shown in Figure 3.2-3. The reflectron consists of a set of circular shaped electrodes with increasing potential placed outside the field-free area at the end of analyser tube. When the ions reach the electrodes, they decelerate and reflect at a relevant angle due to the electric field applied at the input and end plate of the reflectron [44]. The ions then accelerate again towards the detector. The ions with same mass-to-charge ratio, but higher kinetic energies spend more time in the field free area than the ions with the same m/z values with lower energies.

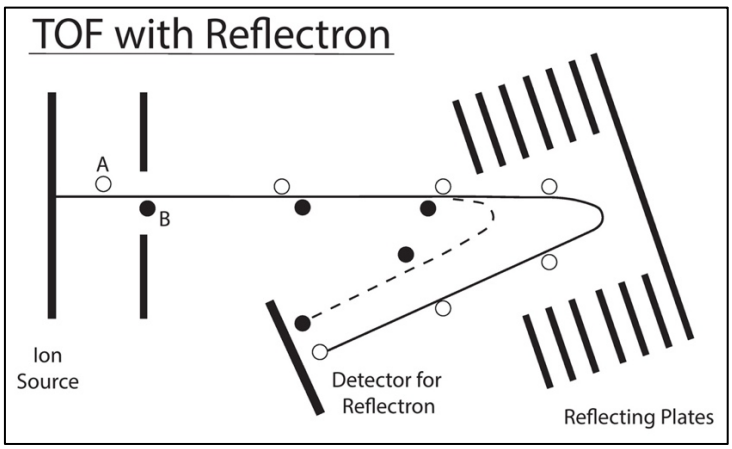


Figure 3.2-3: Reflectron mode for TOF operation. A and B are the two ions with the same m/z value but with different kinetic energy.

These ions are then extracted to the detector. The detector used in the TOF-MS is Micro-channel Plate Detector. A micro-channel plate detector consists of thousands of channel

multipliers in the form of tiny tubes connected to a plate detector. When an ion hits a channel detector, it removes electrons from the outer most shell of the ion resulting in an electron current. Emerging electrons travel and hit multiple channels leading to electron cascade. Every miniaturised channel result in a high signal-to-noise ratio for the detection of ions. The amplified electron current is then collected and converted into a mass spectrum, which provides information about the mass-to-charge ratios of the ions detected in the sample [45, 46]. A basic Micro-channel Plate Detector operation is shown in Figure 3.2-4.

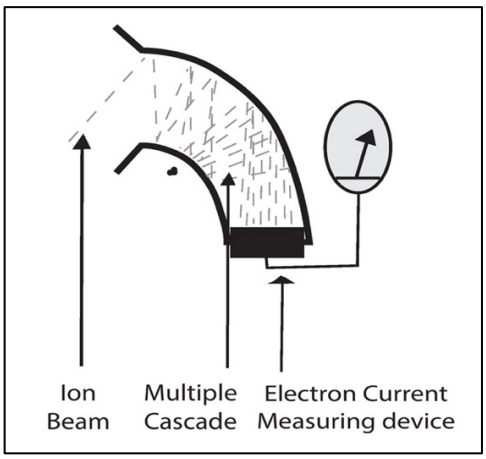


Figure 3.2-4: Micro-channel Plate Detector for BenchTOF-HD TOF-MS.

b. Parameter Settings for TRACE 1300 GC - BenchTOF-HD TOF-MS:

GC-TOF-MS uses a high throughput autosampler and thermal desorption unit ULTRA-xr and UNITY-xr, respectively (both from Markes Intl.). The autosampler allows 100 tubes to be analysed with recollection of both tube and trap desorption repetitively. The injector used for the TRACE 1300 GC is a Split/Splitless Injector (SSL). It consists of an injector body, a heater body, a split and purge vent valve, filters, digital pneumatics for the control of the carrier gas, and fittings for the connection to the analytical column.

The column used in the TRACE 1300 GC was a 20 m, 0.18 mm ID, Rxi-624Sil MS column (Thames Restek, Saunderton, UK). Temperature control and flow control can be set in TRACE 1300 GC by programming the oven temperature, from an initial temperature of

40ºC to a final temperature of 280ºC. The analytical run time for the GC was 25 minutes with a temperature ramp of 20ºC per minute.

c. Sample Preparation:

This study was approved by Coventry and Warwickshire and North-East Yorkshire NHS Ethics Committees (Ref 18717 and Ref 260179). For the analysis of samples by GC-TOF-MS, a 5 mL sample was transferred from the universal sample containers in a 20 mL glass vial from a sample bottle which is sealed with a crimp cap. The headspace of a sample was adsorbed onto a thermal desorption (TD) sorbent tube (C2-AXXX-5149, Markes Intl., Llantrisant, UK). The sorbent biomonitoring tube shown in Figure 3.2-5 was inserted into the septum of a glass vial and heated at 40°C for 20 minutes.

Figure 3.2-5: Biomonitoring Tube.

After this, a pump was attached to the tube whilst still heated to 40°C. The headspace VOCs were then pulled onto the tubes at 20mL/minute for a further 20 minutes. The sorbent tubes were then placed in an ULTRA-xr autosampler for analysis. The analysis began with ULTRA-xr with a stand-by split set to 150°C. Each sample was pre-purged and then desorbed, with the trap purge time set to 1 min. These traps were then cooled at -30°C and the trap was then purged for 3 min at a temperature of 300°C. The data from the GC-TOF-MS analysis were identified using the National Institute of Standards and Technology (NIST) list (2011). The configuration settings are given in Table 3.2-1.

Table 3.2-1: Configuration setting for TRACE 1300 GC - BenchTOF-HD TOF-MS

S. No.	Parameter	Parameter Setting
1	GC run time	25 min
2	Temperature Ramp	40°C to 280°C at 20°C/min
3	Pre-purge time	1 min
4	Desorption temperature & time	250°C for 10 min
5	Temperature of transfer line	250°C
6	Temperature of ion source	250°C

d. Qualitative and Quantitative Analysis:

GC-TOF-MS data are analysed using TOF-DS software (SRA Instruments, Via alla Castellana, Italy). A typical output for a urine sample from GC-TOF-MS is shown in Figure 3.2-6 where the x-axis represents the retention time produced by the GC column (in minutes), and the y-axis represents the abundance of the chemicals with each peak representing the concentration. The TOF-DS software helps in optimising the quantitative and qualitative analysis. TOF-DS supports dynamic baseline correction, which helps to eliminate chromatographic interferences and optimises spectral purity and compound identification [47].

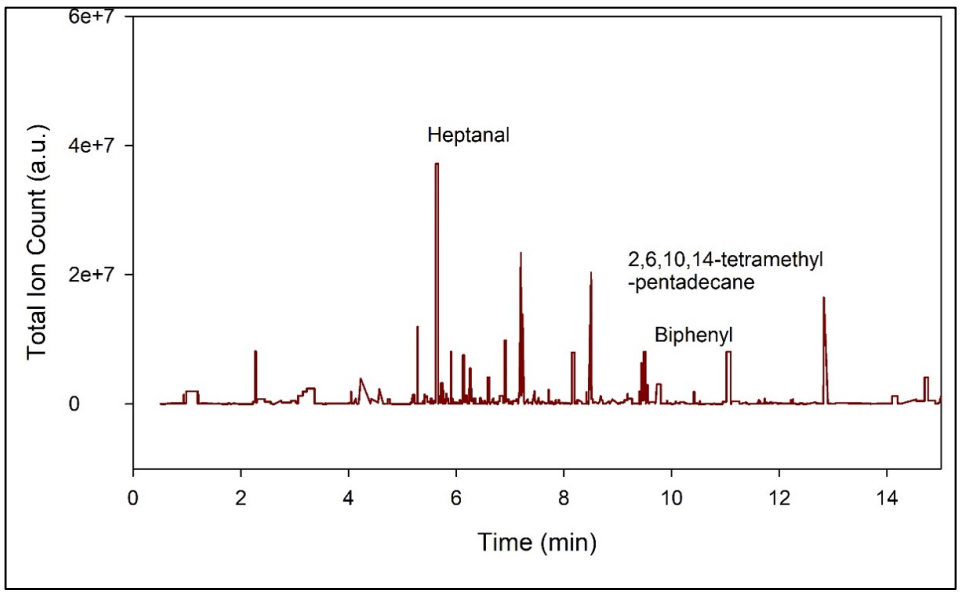


Figure 3.2-6: Output from the GC-TOF-MS from urine sample. Here, the x-axis refers to the retention time, and the y-axis, the total ion count.

The data analysis pipeline utilized in this study to analyse raw data from GC-TOF-MS includes several essential steps, as depicted in Figure 3.2-7. The process commences with pre-processing, data splitting, cross-validation, feature selection, classification, and eventually leads to the calculation of performance metrics and the generation of ROC curves.

The pre-processing stage for GC-TOF-MS data encompasses three distinct phases. Initially, vital information is extracted from the data, such as a list of identified chemicals along with their corresponding peak area and height. Subsequently, a new spreadsheet is generated, where each row represents an individual sample, and the columns correspond to the chemicals detected in each sample. This organized representation facilitates subsequent analyses and data interpretation. In the final pre-processing step, a data cleansing procedure is applied to eliminate chemicals that are present in only a few samples. This selective removal streamlines the dataset, making it more manageable for further analysis. By reducing the number of chemicals from thousands to hundreds, the data become more concise and amenable to subsequent analyses. The removal of rare chemicals allows for a focus on the most relevant and frequently occurring components in the samples, thereby enhancing the efficiency and interpretability of subsequent analyses, and facilitating the identification of meaningful patterns [48].

The subsequent steps in the data analysis pipeline employ the bespoke R program (version 3.6.2) for both qualitative and quantitative analysis in GC-TOF-MS which were identical to those applied in GC-IMS. An R package for automating common tasks can be found at GitHub (https://github.com/JimSkinner/toftools). These steps encompass data processing, feature extraction, statistical analysis, and model construction. In qualitative analysis, the primary objective is to successfully identify and distinguish different classes of sample sets. On the other hand, quantitative analysis aims to quantify the concentration of the identified chemical compounds.

GC-TOF-MS utilizes TOF-DS software to effectively identify the chemical compounds corresponding to the VOCs' signal peaks. TOF-DS software employs a linear Total Ion

Chromatogram (TIC) for 2D acquisitions and performs real-time deconvolution, resulting in a comprehensive representation of peak intensities and capturing reference spectra and essential peak features. This software also enables seamless feature selection by comparing VOCs peaks from the chromatogram with mass spectra [49]. During the feature selection process for GC-TOF-MS data, a rigorous criterion is applied, requiring a p-value of less than 0.05 to retain peaks as significant features. This stringent approach ensures that only the most relevant and statistically significant VOCs peaks are selected, enhancing the accuracy and precision of subsequent analysis [50].

Leveraging TOF-DS software for peak identification and deconvolution, GC-TOF-MS enables the extraction of valuable chemical information from complex chromatograms, facilitating a comprehensive characterization of VOCs present in the samples. The real-time deconvolution process streamlines data processing and analysis, expediting the identification of key compounds and their respective features. The strict p-value-based feature selection ensures that only robust and reliable VOCs' peaks are considered, promoting the generation of high-quality analytical results, and supporting the exploration of meaningful trends and patterns within the data.

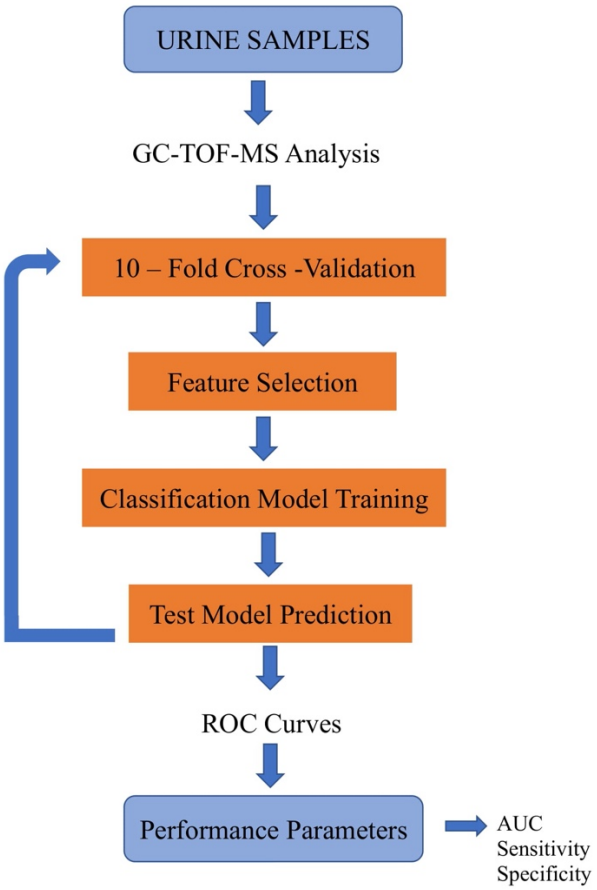


Figure 3.2-7: Data analysis pipeline for the TRACE 1300 GC - BenchTOF-HD TOF-MS

3.3. Electronic Nose:

3.3.1. Introduction:

An Electronic Nose (eNose) is based on the human olfactory system and consists of an array of different sensors combined to identify different organic chemicals. An eNose system typically consists of a sensor array, a signal conditioning and data processing unit and pattern recognition unit. These units combined helps in the recognition of volatile organic compounds present in a sample [51, 52]. When a sample is introduced to an eNose, it induces chemical/physical changes in the sensor array which further causes electrical changes which are detected and used to identify a chemical profile comprising multiple biomarkers. Each sensor in the sensor array responds to different chemicals [53].

3.3.2. AlphaMOS FOX 4000 (Toulouse, France):

The AlphaMOS Fox 4000 is an eNose that uses 18 commercial metal MOS sensors distributed in three temperature-controlled chambers. The FOX 4000 system comprises of a CombiPAL HS100 autosampler, sensor chambers, a mass flow controller, and an acquisition board with a microcontroller. The function of each sensor is shown in Table 3.3-1.

Table 3.3-1: Description of sensors of the AlphaMOS FOX 4000

S. No.	Sensor Name	Description
1	LY2/LG	Oxidising gas
2	LY2/G	Ammonia, Carbon Monoxide
3	LY2/AA	Ethanol
4	LY2/GH	Ammonia/ Organic Amines
5	LY2/gCTL	Hydrogen Sulphide
6	LY2/gCT	Propane/ Butane
7	T30/1	Organic Solvents
8	P10/1	Hydrocarbons
9	P10/2	Methane
10	P40/1	Fluorine
11	T70/2	Aromatic Compounds
12	PA/2	Ethanol, Ammonia/ Organic Amines
13	P30/1	Polar Compounds (Ethanol)
14	P40/2	Heteroatom/Chloride/Aldehydes
15	P30/2	Alcohol
16	T40/2	Aldehydes
17	T40/1	Chlorinated Compounds
18	TA/2	Air quality

The AlphaMOS Fox 4000 is shown in Figure 3.3-1.

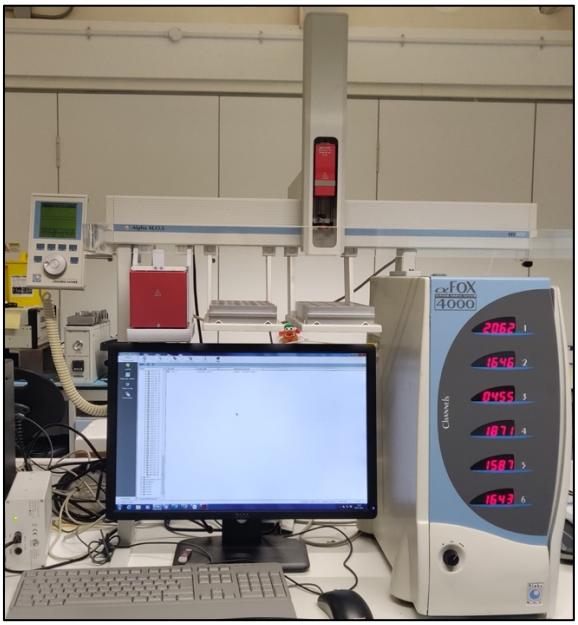


Figure 3.3-1: AlphaMOS FOX 4000 (Toulouse, France) with CombiPAL HS100 autosampler

a. Principle of operation:

The AlphaMOS Fox 4000 uses metal oxide-based sensors out of which there are 6 p-type sensors and 12 n-type sensors. The sensing material in the MOX sensors is typically made of metal oxide nanoparticles that adsorb gas molecules, causing changes in the electrical conductivity of the material [54-57]. However, they may suffer from some limitations, such as poor selectivity and sensitivity to changes in temperature and humidity. N-type sensors are more popular than p-type sensors due to their high sensitivity to oxidizing gases, good stability and reproducibility, lower power consumption, and longer lifetime. However, they suffer from low selectivity. On the other hand, p-type sensors have their own advantages, such as high humidity tolerance, good selectivity, and high temperature tolerance [58].

The output from the MOS sensors is measured in terms of the change in the resistance caused by the presence of a VOCs given by equation below:

$$R = (R_0 - R_T)/ R_0 \qquad \textit{Equation 10}$$

where R (Ω) is the value of the resistance at the end of the measurement, R_0 (Ω) is the initial resistance of the sensor at time 0 and R_T (Ω) is sensor's conductance value.

b. Parameter settings:

AlphaMOS Fox 4000 is combined with CombiPAL HS100 autosampler which is fitted with a 2.5 mL gas syringe, an agitator and two trays with 60 samples capacity (30 per tray). All the samples are placed in the autosampler tray. Every sample is transferred to the agitator where it is heated and agitated. The headspace was then injected into the eNose into a zero-air controlled by mass flow controller. Each sample was analysed for 180 seconds by all of the 18 MOS sensors. The parameter settings for the AlphaMOS FOX 4000 for the analysis of each urine sample are presented in Table 3.3-2.

Table 3.3-2: Configuration setting for the AlphaMOS Fox 4000.

S. No.	Parameter	Parameter Setting
1	Agitator heating time	10 mins
2	Temperature of agitator	40°C
3	Flow rate of sample injection	200 mL/min
4	Flow rate of zero air	150 mL/min
5	Temperature of transfer line	250°C
6	Temperature of ion source	250°C
7	Injection speed	150 ul/s
8	Acquisition time	180 s

c. Quantitative and qualitative analysis for AlphaMOS Fox 4000:

The output from the AlphaMOS Fox 4000 system to a urine sample is shown in Figure 3.3-2. The x-axis in the graph represents the time elapsed in seconds during the measurement, while the y-axis displays the dimensionless response generated by each individual sensor in the eNose upon exposure to the VOCs present in the analysed sample. As each sensor is sensitive to specific chemical compounds, their responses offer valuable insights into the chemical composition of the sample.

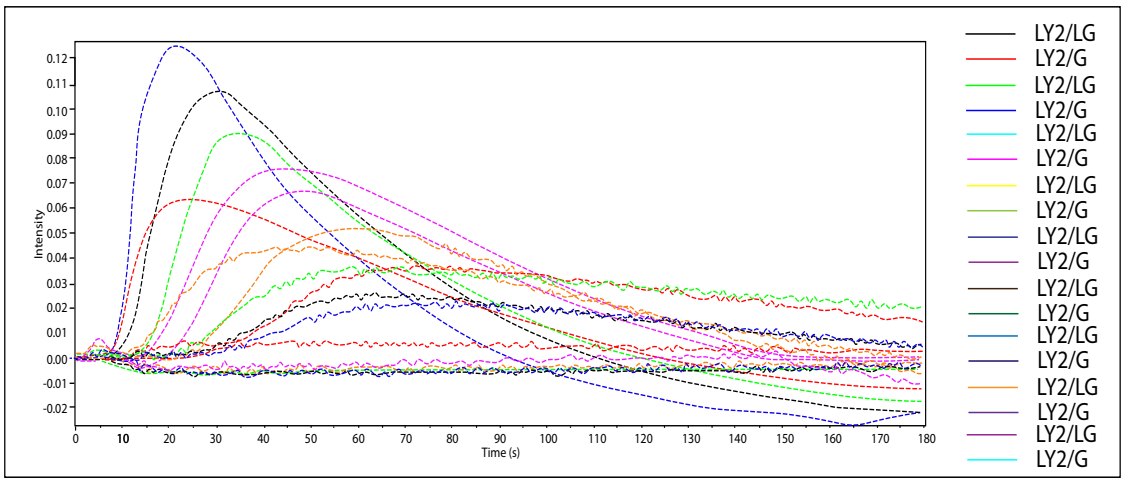


Figure 3.3-2: Output from the FOX 4000 eNose where each curve represents the response of a sensor to a BCa urine sample. Here the sensor response is defined as intensity, which is the change in resistance from the baseline divided by the baseline resistance and therefore, is dimensionless.

The sensor's responses were extracted using AlphaSoft (AlphaMOS v12.36). Once the raw sensor signals are acquired, AlphaSoft performs signal processing to enhance the quality of the data. This includes noise reduction, filtering, and other techniques to improve the accuracy and reliability of the sensor responses. Raw sensor signals can be influenced by noise and unwanted fluctuations originating from various sources, such as electronic interference or environmental factors. To address this, AlphaSoft applies noise reduction algorithms to remove or minimize these unwanted variations, resulting in a clearer representation of the true sensor responses [59].

In addition to noise reduction, filtering techniques are employed to eliminate unwanted frequency components from the sensor signals. Various types of filters, such as low-pass filters or band-pass filters, can be applied to retain the relevant information while removing unwanted artifacts. Furthermore, eNose sensor responses may drift over time due to changes in ambient conditions or instrument calibration. To rectify this issue, AlphaSoft performs baseline correction to adjust the sensor signals and bring them back to a consistent reference level. This step improves the comparability and consistency of the data, ensuring reliable analysis [60].

Another consideration is that different sensors in the eNose may have varying sensitivities or response ranges. To enable meaningful comparison and analysis, AlphaSoft normalizes

the sensor signals, ensuring they are on a consistent scale across different sensors. Moreover, AlphaSoft implements signal averaging techniques to enhance the reliability of the sensor responses. By averaging multiple measurements of the same sample, the software reduces random variations and provides more robust data, enhancing the signal-to-noise ratio [60].

After obtaining enhanced data from AlphaSoft, the next step involves feature extraction. Each measurement data set contains the complete response of all sensors to one sample. However, the complete response curve may contain redundant information, which can be problematic for many pattern recognition algorithms, leading to increased processing time. Therefore, it is desirable to significantly reduce the amount of data to be processed, and this goal was accomplished using a MultiSens Analyzer (JLM Innovation GmbH, Germany) [61].

A MultiSens Analyzer achieves data reduction by extracting specific characteristics (e.g., the maximum of the response curve) from the raw signal. These characteristics are referred to as "features" within the MultiSens Analyzer software. The collection of all these features forms the "feature vector," which consists of a set of values extracted from the sensors' reactions to one sample. The MultiSens Analyzer software employs several different methods for feature extraction, which are listed in Table 3.3-3.

Table 3.3-3: List of the MultiSens Analyzer's Feature Extraction Methods used in this research.

Feature	Description
Area	Calculates the area of the response curve above the baseline (= value of first measurement point).
AreaAt	Calculates the area of the response curve above a baseline. Allows the user to set the start point, end point, and base line determination for the calculation.
Base	Returns the baseline = value of the first measurement point.
Base3	Returns the baseline = average of the first 3 measurement points.
FindSig	Returns the maximum deviation to the baseline that is found between the start point and the end point. Start point, end point and the point where to determine the baseline can be set.

Max	Returns the maximum of the response curve.
Max-Min	Returns the span of the response curve.
MaxDev	Returns the maximum derivative found in the response curve. The derivative is calculated by subtracting the predecessor point for each measurement point.
Min	Returns the minimum of the response curve.
Sig/Base	Returns the Signal divided by the baseline. Signal is determined as the maximum deviation to the baseline in the response curve. Baseline is determined as the value of the first measurement point.
Sig/Base3	Same as Sig/Base, but averages over 3 measurement points to determine signal and baseline.
Sig-Base	Sig-Base is determined as the maximum deviation to the baseline in the response curve. Baseline is determined as the value of the first measurement point.
Sig-Base3	Same as Sig-Base, but averages over 3 measurement points to determine signal and baseline.

The feature vectors, representing measurements in a multi-dimensional feature space, are subjected to evaluation using either PCA or LDA techniques. The PCA (Principal Component Analysis) or LDA (Linear Discriminant Analysis) results are employed for projection, and these outcomes are presented in a scores plot [62]. In a score plot, the different samples are displayed in relation to two principal components (PCs), typically the first and second PCs. The first principal component delineates the direction of maximum variance within the data, while the second principal component captures the second-largest variance orthogonal to the first component, and so forth for subsequent components [63].

The spread of data points along each principal component axis indicates the amount of variance explained by that particular component. Consequently, the greater the spread observed, the higher the contribution of that principal component to the overall variance present in the data. By examining the positioning and distribution of samples, one can gain insights into the underlying structure and variability of the data. This aids in identifying patterns, clusters, or discrimination between groups in the data, thus facilitating data exploration, classification, and other analytical tasks.

The feature matrix was also exported and further processed using a custom analysis pipeline created using a custom R program (version 3.6.2) tailored to the specific requirements of the study and can be found at GitHub (https://github.com/JimSkinner/toftools). This analysis procedure was identical to those applied in GC-IMS and discussed in the section e Qualitative and Quantitative Analysis:. From the resultant probabilities, statistical parameters were calculated including Receiver Operator Characteristic (ROC) curves, area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

3.3.3. PEN3 Electronic Nose (Airsense Analytics GmbH, Schwerin, Germany):

The PEN3 eNose (Airsense Analytics GmbH, Schwerin, Germany) is a portable (92 × 190 × 255 mm) olfactory system used for the identification of chemicals and gases. It is a combination of a gas sampling unit and a sensor array. In our case, the PEN3 eNose is fitted with an autosampler (HT2000H Dynamic Headspace Auto-sampler, Italy), which interfaces directly with the PEN 3 software (WinMuster PEN v 1.6.2.18). The PEN3 system is shown in Figure 3.3-3.

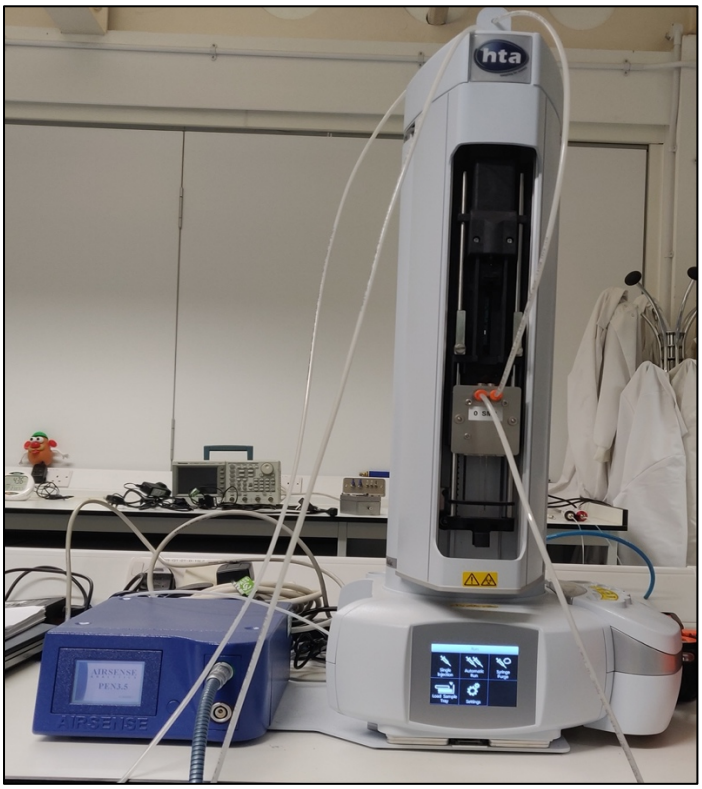


Figure 3.3-3: PEN3 Electronic Nose (Airsense Analytics GmbH, Schwerin, Germany).

The sensor arrays consist of 10 different n-type metal oxide sensors listed in Table 3.3-4. The selectivity of each sensor depends upon several factors, such as the sensing material, the dopant material, the geometry of the sensor and the working temperature. The description of the sensors within PEN3 eNose are described in Table 3.3-4.

Table 3.3-4: Description of the sensors in the PEN3 eNose.

Sensor No.	Sensors Name	Substances for sensing
S1	W1C	Aromatic Compounds
S2	W5S	Broad Range
S3	W3C	Aromatic Compounds
S4	W6S	Hydrogen
S5	W5C	Aromatic and Aliphatic Compounds
S6	W1S	Methane in the Environment, With Broad Range
S7	W1W	Sulphur And Organic Compounds
S8	W2S	Alcohol and Broad Range
S9	W2W	Sulphur Compounds

S10	W3S	Methane and Aliphatic Compounds

a. Principle of operation:

The PEN3 eNose consists of two pumps, one of which is used for pulling the sample gas through the sensor array and the other transfers filtered reference air or zero air into the sensor array. The zero air is also used to clean the system. Before the samples are introduced, the PEN3 system starts with the flushing time. During that period the sensors are rinse with zero gas and their signals move back to the baseline (G/G0=1). This is followed by sample introduction where the air containing the VOCs enters the device, and the sensors in the array come into contact with the gas molecules. When the VOCs interact with the sensor's surface, they cause changes in the sensor's electrical conductivity. The interaction between the gas molecules and the sensor's surface led to the alteration of electrical resistance in the sensors. Zero air is used as a baseline or reference gas, and the sensor responses from the sample gas are measured in comparison to the reference gas.

b. Parameter Settings:

A HT2000H Dynamic Headspace Auto-sampler consists of a sample tray with 42 sample capacity, a 2.5 mL headspace syringe, an oven with 6 sample capacity and orbital shaking capability. The samples are placed in the autosampler tray and transferred one by one from the sample tray to an internal oven and heated to increase concentration above the detection limit of the eNose. After heating up the sample, the headspace was sampled by a syringe. The sample was then analysed for 5 min. The configuration details for the PEN3 eNose are presented in Table 3.3-5.

Table 3.3-5: Configuration setting for the PEN3 eNose.

S. No.	Parameter	Parameter Setting
1	Oven temperature	80°C
2	Shaker on/off time	4 mins/4mins
3	Sample injection volume	2.5 mL
4	Sample injection pressure	5 bars (max)

c. Quantitative and qualitative analysis for PEN3 eNose:

The output from the PEN3 device for a typical urine sample is shown in Figure 3.3-4.

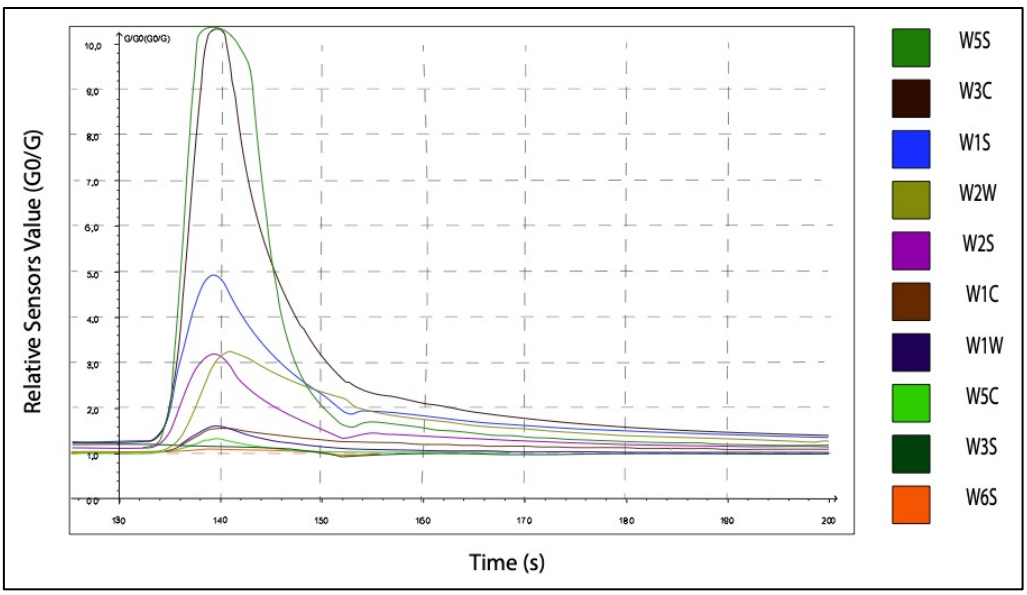


Figure 3.3-4: Output from the PEN3 device. Output from PEN3 device contains sensor output from each of the sensors to a urine sample. Each coloured line in the output represents the response curve of one of the sensors.

The data obtained from PEN3 eNose were analysed using a MultiSens Analyzer (v2.0.0.22, JLM Innovation GmbH, Germany) and an R program (version 3.6.2) which can be found at GitHub (https://github.com/JimSkinner/toftools). This analysis procedure was identical to those applied in the GC-IMS analysis and discussed in the previous section c (Quantitative and qualitative analysis for AlphaMOS Fox 4000:).

The MultiSens Analyzer was used to evaluate measurement data from multi sensor instrument, such as electronic noses. The MultiSens Analyzer classified the data into different groups and then performed feature extraction. The feature used was the maximum deviation of the signal from the baseline to the response which was used to generate a radar plot. The resultant data matrix was then analysed using a 10-fold cross-validation, undertaken using a bespoke R program (version 3.6.2). From the resultant probabilities, final statistical results including a Receiver Operator Characteristic (ROC) curve, sensitivity, selectivity, specificity, a positive predictive value (PPV), and a negative predictive value (NPV) were calculated. An R package for automating the common tasks can be found at GitHub (https://github.com/JimSkinner/toftools).

The AlphaMOS FOX 4000 and PEN3 systems are both electronic nose devices used for gas sensing and analysis. However, there are some key differences between the two devices. Firstly, the AlphaMOS FOX 4000 eNose is equipped with 18 sensors compared to the 10-sensor array in the PEN3 eNose. Secondly, the modular design of the AlphaMOS FOX 4000 gives users the flexibility to choose the sensors that are best suited for their application, making it a highly versatile device.

Thirdly, the AlphaMOS Fox 4000 is known for its high sensitivity and is capable of detecting low concentrations of VOCs in the ppb range whereas the limit of detection for the PEN3 eNose is 0.1 to 5ppm, limiting its sensitivity. However, the PEN3 eNose is portable, easy to use and less expensive in comparison to the bulkier size and higher cost of an AlphaMOS FOX 4000.

3.4. Conclusion:

The detection of volatile organic compounds has been used in many industries such as beverages, environmental, chemical industries etc. There are various different analytical instruments available which are used for this purpose. The GC-IMS and GC-TOF-MS in this study are two analytical tools based on gas chromatography and mass spectrometry (GC-MS). These two instruments are highly sensitive, selective and have enhanced separation abilities. However, these instruments have various disadvantages such as they are not portable, are slow, expensive and require highly trained professionals to operate them. The

other two instruments used in this study were eNoses (AlphaMOS Fox 4000 and PEN3). ENoses overcome the shortcomings of GC-MS based instruments. They are cost-effective, easily operable and can perform real-time data analysis.

3.5. References:

[1] S. Wang, H. Chen, and B. Sun, "Recent progress in food flavor analysis using gas chromatography–ion mobility spectrometry (GC–IMS)," *Food Chemistry,* vol. 315, p. 126158, 2020/06/15/ 2020, doi: https://doi.org/10.1016/j.foodchem.2019.126158.

[2] G. W. Cook, P. T. LaPuma, G. L. Hook, and B. A. Eckenrode, "Using Gas Chromatography with Ion Mobility Spectrometry to Resolve Explosive Compounds in the Presence of Interferents*," *Journal of Forensic Sciences,* https://doi.org/10.1111/j.1556-4029.2010.01522.x vol. 55, no. 6, pp. 1582-1591, 2010/11/01 2010, doi: https://doi.org/10.1111/j.1556-4029.2010.01522.x.

[3] R. L. Grob and E. F. Barry, *Modern practice of gas chromatography*. John Wiley & Sons, 2004.

[4] H. M. McNair, J. M. Miller, and N. H. Snow, *Basic gas chromatography*. John Wiley & Sons, 2019.

[5] W. Jennings, E. Mittlefehldt, and P. Stremple, *Analytical gas chromatography*. Academic Press, 1997.

[6] K. Dettmer-Wilde and W. Engewald, "Practical gas chromatography," in *A Comprehensive Reference*: Springer, 2014, p. 902.

[7] A. B. Kanu and H. H. Hill, "Ion mobility spectrometry detection for gas chromatography," *Journal of Chromatography A,* vol. 1177, no. 1, pp. 12-27, 2008/01/04/ 2008, doi: https://doi.org/10.1016/j.chroma.2007.10.110.

[8] H. Borsdorf and G. A. Eiceman, "Ion Mobility Spectrometry: Principles and Applications," *Applied Spectroscopy Reviews,* vol. 41, no. 4, pp. 323-375, 2006/08/01 2006, doi: 10.1080/05704920600663469.

[9] N. Gerhardt, M. Birkenmeier, D. Sanders, S. Rohn, and P. Weller, "Resolution-optimized headspace gas chromatography-ion mobility spectrometry (HS-GC-IMS) for non-targeted olive oil profiling," *Analytical and Bioanalytical Chemistry,* vol. 409, no. 16, pp. 3933-3942, 2017/06/01 2017, doi: 10.1007/s00216-017-0338-2.

[10] H. H. Hill, W. F. Siems, R. H. S. Louis, and D. G. McMinn, "ION MOBILITY SPECTROMETRY," *Analytical Chemistry,* vol. 62, no. 23, pp. 1201A-1209A, 1990/12/01 1990, doi: 10.1021/ac00222a716.

[11] Owlstone, "3 Types of IMS: An At-a-Glance Guide," ed, 2017.

[12] D. Cavanna, S. Zanardi, C. Dall'Asta, and M. Suman, "Ion mobility spectrometry coupled to gas chromatography: A rapid tool to assess eggs freshness," *Food Chemistry,* vol. 271, pp. 691-696, 2019/01/15/ 2019, doi: https://doi.org/10.1016/j.foodchem.2018.07.204.

[13] H. Borsdorf, T. Mayer, M. Zarejousheghani, and G. A. Eiceman, "Recent Developments in Ion Mobility Spectrometry," *Applied Spectroscopy Reviews,* vol. 46, no. 6, pp. 472-521, 2011/08/01 2011, doi: 10.1080/05704928.2011.582658.

[14] G. A. Eiceman and Z. Karpas, *Ion mobility spectrometry*. CRC press, 2005.

[15] T. Perl, B. Bödeker, M. Jünger, J. Nolte, and W. Vautz, "Alignment of retention time obtained from multicapillary column gas chromatography used for VOC analysis with ion mobility spectrometry," *Analytical and bioanalytical chemistry,* vol. 397, pp. 2385-2394, 2010.

[16] S. Schwolow, N. Gerhardt, S. Rohn, and P. Weller, "Data fusion of GC-IMS data and FT-MIR spectra for the authentication of olive oils and honeys—is it worth to go the extra mile?," *Analytical and Bioanalytical Chemistry,* vol. 411, no. 23, pp. 6005-6019, 2019/09/01 2019, doi: 10.1007/s00216-019-01978-w.

[17] H. Bi, Z. Guo, X. Jia, H. Liu, L. Ma, and L. Xue, "The key points in the pre-analytical procedures of blood and urine samples in metabolomics studies," *Metabolomics,* vol. 16, no. 6, p. 68, 2020/05/25 2020, doi: 10.1007/s11306-020-01666-2.

[18] H. Zheng *et al.*, "Optimal preprocessing of serum and urine metabolomic data fusion for staging prostate cancer through design of experiment," *Analytica Chimica Acta,* vol. 991, pp. 68-75, 2017/10/23/ 2017, doi: https://doi.org/10.1016/j.aca.2017.09.019.

[19] J. Christmann, S. Rohn, and P. Weller, "gc-ims-tools – A new Python package for chemometric analysis of GC–IMS data," *Food Chemistry,* vol. 394, p. 133476, 2022/11/15/ 2022, doi: https://doi.org/10.1016/j.foodchem.2022.133476.

[20] S. A. Alasadi and W. S. Bhaya, "Review of data preprocessing techniques in data mining," *Journal of Engineering and Applied Sciences,* vol. 12, no. 16, pp. 4102-4107, 2017.

[21] R. Brendel, S. Schwolow, S. Rohn, and P. Weller, "Volatilomic Profiling of Citrus Juices by Dual-Detection HS-GC-MS-IMS and Machine Learning—An Alternative Authentication Approach," *Journal of Agricultural and Food Chemistry,* vol. 69, no. 5, pp. 1727-1738, 2021/02/10 2021, doi: 10.1021/acs.jafc.0c07447.

[22] S. Gu, J. Zhang, J. Wang, X. Wang, and D. Du, "Recent development of HS-GC-IMS technology in rapid and non-destructive detection of quality and contamination in agri-food products," *TrAC Trends in Analytical Chemistry,* vol. 144, p. 116435, 2021/11/01/ 2021, doi: https://doi.org/10.1016/j.trac.2021.116435.

[23] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-Validation," in *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu Eds. New York, NY: Springer New York, 2016, pp. 1-7.

[24] D. Berrar, "Cross-Validation," ed, 2019.

[25] J. Christmann, S. Rohn, and P. Weller, "Finding features - variable extraction strategies for dimensionality reduction and marker compounds identification in GC-IMS data," *Food Research International,* vol. 161, p. 111779, 2022/11/01/ 2022, doi: https://doi.org/10.1016/j.foodres.2022.111779.

[26] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine learning research,* vol. 7, pp. 1-30, 2006.

[27] M. Khashei, S. Eftekhari, and J. Parvizian, "Diagnosing diabetes type II using a soft intelligent binary classification model," *Review of Bioinformatics and Biometrics,* vol. 1, no. 1, pp. 9-23, 2012.

[28] Y. Chang, W. Li, and Z. Yang. Network Intrusion Detection Based on Random Forest and Support Vector Machine.

[29] J. B. Butcher, A. V. Rutter, A. J. Wootton, C. R. Day, and J. Sulé-Suso, Cham. Artificial Neural Network Analysis of Volatile Organic Compounds for the Detection of Lung Cancer BT - Advances in Computational Intelligence Systems.

[30] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics,* vol. 35,

no. 5, pp. 352-359, 2002/10/01/ 2002, doi: https://doi.org/10.1016/S1532-0464(03)00034-0.

[31] S. Suthaharan, "Support Vector Machine," in *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, S. Suthaharan Ed. Boston, MA: Springer US, 2016, pp. 207-235.

[32] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell, "Gaussian Processes for Object Categorization," *International Journal of Computer Vision,* vol. 88, no. 2, pp. 169-188, 2010/06/01 2010, doi: 10.1007/s11263-009-0268-3.

[33] J. C. Stoltzfus, "Logistic Regression: A Brief Primer," *Academic Emergency Medicine,* vol. 18, no. 10, pp. 1099-1104, 2011, doi: https://doi.org/10.1111/j.1553-2712.2011.01185.x.

[34] Y.-C. Chang, K.-H. Chang, and G.-J. Wu, "Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions," *Applied Soft Computing,* vol. 73, pp. 914-920, 2018/12/01/ 2018, doi: https://doi.org/10.1016/j.asoc.2018.09.029.

[35] B. Vega-Márquez, I. Nepomuceno-Chamorro, N. Jurado-Campos, and C. Rubio-Escudero, "Deep Learning Techniques to Improve the Performance of Olive Oil Classification," *Frontiers in Chemistry,* Original Research vol. 7, 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fchem.2019.00929.

[36] O. İrsoy, O. T. Yıldız, and E. Alpaydın, "Design and Analysis of Classifier Learning Experiments in Bioinformatics: Survey and Case Studies," *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* vol. 9, no. 6, pp. 1663-1675, 2012, doi: 10.1109/TCBB.2012.117.

[37] M. Hassan, M. A. Butt, and M. Zaman, "An Ensemble random forest algorithm for privacy preserving distributed medical data mining," *International Journal of E-Health and Medical Communications (IJEHMC),* vol. 12, no. 6, pp. 1-23, 2021.

[38] B. A. Mamyrin, "Time-of-flight mass spectrometry (concepts, achievements, and prospects)," *International Journal of Mass Spectrometry,* vol. 206, no. 3, pp. 251-266, 2001/03/22/ 2001, doi: https://doi.org/10.1016/S1387-3806(00)00392-4.

[39] U. Boesl, "Time-of-flight mass spectrometry: Introduction to the basics," *Mass Spectrometry Reviews,* https://doi.org/10.1002/mas.21520 vol. 36, no. 1, pp. 86-109, 2017/01/01 2017, doi: https://doi.org/10.1002/mas.21520.

[40] R. J. Cotter, "Time-of-flight mass spectrometry for the structural analysis of biological molecules," *Analytical chemistry,* vol. 64, no. 21, pp. 1027A-1039A, 1992.

[41] H. Tyagi, E. Daulton, A. S. Bannaga, R. P. Arasaradnam, and J. A. Covington, "Urinary volatiles and chemical characterisation for the non-invasive detection of prostate and bladder cancers," *Biosensors,* vol. 11, no. 11, p. 437, 2021.

[42] E. Daulton *et al.*, "Volatile organic compounds (VOC) for the non-invasive detection of pancreatic cancer from urine," *Talanta,* vol. 221, p. 121604, 2021.

[43] M. Guilhaus, "Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts," *Journal of mass spectrometry,* vol. 30, no. 11, pp. 1519-1532, 1995.

[44] B. A. Mamyrin, "Laser assisted reflectron time-of-flight mass spectrometry," *International Journal of Mass Spectrometry and Ion Processes,* vol. 131, pp. 1-19, 1994/02/24/ 1994, doi: https://doi.org/10.1016/0168-1176(93)03891-O.

[45] G. Montagnoli *et al.*, "The large-area micro-channel plate entrance detector of the heavy-ion magnetic spectrometer PRISMA," *Nuclear Instruments and Methods in Physics*

Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 547, no. 2, pp. 455-463, 2005/08/01/ 2005, doi: https://doi.org/10.1016/j.nima.2005.03.158.

[46] T. Gys, "Micro-channel plates and vacuum detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 787, pp. 254-260, 2015/07/01/ 2015, doi: https://doi.org/10.1016/j.nima.2014.12.044.

[47] N. M. Frerichs *et al.*, "Fecal Volatile Metabolomics Predict Gram-Negative Late-Onset Sepsis in Preterm Infants: A Nationwide Case-Control Study," *Microorganisms*, vol. 11, no. 3, doi: 10.3390/microorganisms11030572.

[48] W. Jiang, Y. Qiu, Y. Ni, M. Su, W. Jia, and X. Du, "An Automated Data Analysis Pipeline for GC−TOF−MS Metabonomics Studies," *Journal of Proteome Research,* vol. 9, no. 11, pp. 5974-5981, 2010/11/05 2010, doi: 10.1021/pr1007703.

[49] Y. Ni *et al.*, "ADAP-GC 2.0: Deconvolution of Coeluting Metabolites from GC/TOF-MS Data for Metabolomics Studies," *Analytical Chemistry,* vol. 84, no. 15, pp. 6619-6629, 2012/08/07 2012, doi: 10.1021/ac300898h.

[50] A. Peralbo-Molina, M. Calderón-Santiago, B. Jurado-Gámez, M. D. Luque de Castro, and F. Priego-Capote, "Exhaled breath condensate to discriminate individuals with different smoking habits by GC–TOF/MS," *Scientific Reports,* vol. 7, no. 1, p. 1421, 2017/05/03 2017, doi: 10.1038/s41598-017-01564-z.

[51] S. M. Scott, D. James, and Z. Ali, "Data analysis for electronic nose systems," *Microchimica Acta,* vol. 156, no. 3, pp. 183-207, 2006/12/01 2006, doi: 10.1007/s00604-006-0623-9.

[52] D. Karakaya, O. Ulucan, and M. Turkan, "Electronic Nose and Its Applications: A Survey," *International Journal of Automation and Computing,* vol. 17, no. 2, pp. 179-209, 2020/04/01 2020, doi: 10.1007/s11633-019-1212-9.

[53] J.-E. Haugen and K. Kvaal, "Electronic nose and artificial neural network," *Meat Science,* vol. 49, pp. S273-S286, 1998/01/01/ 1998, doi: https://doi.org/10.1016/S0309-1740(98)90054-7.

[54] S. Bai, D. Li, D. Han, R. Luo, A. Chen, and C. L. Chung, "Preparation, characterization of WO3–SnO2 nanocomposites and their sensing properties for NO2," *Sensors and Actuators B: Chemical,* vol. 150, no. 2, pp. 749-755, 2010/10/28/ 2010, doi: https://doi.org/10.1016/j.snb.2010.08.007.

[55] J. Zhang, Z. Qin, D. Zeng, and C. Xie, "Metal-oxide-semiconductor based gas sensors: screening, preparation, and integration," *Physical Chemistry Chemical Physics,* 10.1039/C6CP07799D vol. 19, no. 9, pp. 6313-6329, 2017, doi: 10.1039/C6CP07799D.

[56] S. D. Senturia, "The role of the MOS structure in integrated sensors," *Sensors and Actuators,* vol. 4, pp. 507-526, 1983/01/01/ 1983, doi: https://doi.org/10.1016/0250-6874(83)85063-X.

[57] N. Yamazoe, G. Sakai, and K. Shimanoe, "Oxide Semiconductor Gas Sensors," *Catalysis Surveys from Asia,* vol. 7, no. 1, pp. 63-75, 2003/04/01 2003, doi: 10.1023/A:1023436725457.

[58] A. Dey, "Semiconductor metal oxide gas sensors: A review," *Materials Science and Engineering: B,* vol. 229, pp. 206-217, 2018/03/01/ 2018, doi: https://doi.org/10.1016/j.mseb.2017.12.036.

[59] S. Mildner-Szkudlarz, H. H. Jeleń, and R. Zawirska-Wojtasiak, "The use of electronic and human nose for monitoring rapeseed oil autoxidation," *European Journal of Lipid Science and Technology,* https://doi.org/10.1002/ejlt.200700009 vol. 110, no. 1, pp. 61-72, 2008/01/01 2008, doi: https://doi.org/10.1002/ejlt.200700009.

[60] A. MOS. "AlphaSoft, Sensory Analysis Software." https://www.alpha-mos.com/smell-analysis-heracles-electronic-nose#alphasoft (accessed.

[61] J. A. Covington *et al.*, "The detection of patients at risk of gastrointestinal toxicity during pelvic radiotherapy by electronic nose and FAIMS: a pilot study," *Sensors,* vol. 12, no. 10, pp. 13002-13018, 2012.

[62] M. F. Rutolo, D. Iliescu, J. P. Clarkson, and J. A. Covington, "Early identification of potato storage disease using an array of metal-oxide based gas sensors," *Postharvest Biology and Technology,* vol. 116, pp. 50-58, 2016/06/01/ 2016, doi: https://doi.org/10.1016/j.postharvbio.2015.12.028.

[63] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence,* vol. 97, no. 1, pp. 245-271, 1997/12/01/ 1997, doi: https://doi.org/10.1016/S0004-3702(97)00063-5.

Chapter 4. Analysis of Cancer Samples using GC-IMS and GC-TOF-MS

This chapter presents the experimental outcomes that have been achieved from the analytical instruments, namely, GC-IMS and GC-TOF-MS, used during the course of this research. The experiments were carried out utilizing urine samples of four different types of cancers with the objective of obtaining the biomarker fingerprint and understanding the ability of these instruments to be able to distinguish cancer samples from healthy controls with high sensitivity and specificity. Comparative analyses were conducted between the samples obtained from cancerous and healthy individuals, as well as between distinct categories of cancers. Additionally, the statistical performances of several analytical instruments were evaluated and compared.

4.1. Introduction:

As discussed in chapter 2, cancer remains a leading cause of death worldwide, with approximately 19.3 million new cases and 10 million deaths in 2020 [1]. However, the rate of survival depends highly on the detection of cancer in its early stages. Most of the current diagnostic methods do not provide high sensitivity or specificity towards the detection of early stages of the cancer [2]. As stated earlier, one of the potential methods that could support cancer diagnosis is through the measurement of Volatile Organic Compounds (VOCs) that reflect the biological process of disease. These bodily VOCs are the reflection of the physiological effects and metabolism of the individual and the environment surrounding them. Cancer causes changes in these biological pathways leading to the emission or omission of specific VOCs [3-5].

The cancers included in this study were Colorectal cancer (CRC), Bladder cancer (BCa), Prostate cancer (PCa) and Hepatocellular carcinoma (HCC). Urine is a common biological source of VOCs, as the components present are either the intermediate products or end products of metabolic activities occurring inside the human body [6]. In this research, we

aimed to evaluate the use of GC-IMS and GC-TOF-MS as potential analytical instruments for the detection and discrimination between cancer groups and healthy controls [7, 8].

4.2. Materials and methods:

4.2.1. Urine samples:

A total of 186 patients were recruited after providing informed consent at University Hospital Coventry & Warwickshire NHS Trust, UK, between July 2013 and November 2019. Patients were recruited prior to anti-cancer treatment. This study was approved by Coventry and Warwickshire and North-East Yorkshire NHS Ethics Committees (Ref 18717 and Ref 260179). Urine samples were collected in standard universal sterile specimen containers and frozen within 2 hours at -80˚C for subsequent batch analysis and according to standard operating procedures compliant with tissue bank requirements under Human Tissue Act 2004. No chemicals were added to the urine before freezing, as we have previously shown that urine samples remain stable for extended periods of time at this temperature [9]. Prior to analysis the samples were transferred to the University of Warwick and briefly stored at -20˚C. The samples were defrosted in a laboratory fridge at 4°C and aliquoted into 20mL glass sample vials with a crimp cap. 5mL of each urine sample was used for the analysis. Of the 186 urine samples collected, 15 patients had confirmed BCa, 55 were confirmed PCa, 20 patients had confirmed HCC, 58 patients had confirmed CRC and 38 patients were non-cancerous controls. The mean age of the BCa patients was 70 years, for PCa patients was 72 years, for CRC patients was 74 years and the mean age for HCC patients as 72 years. The demographic data of the subjects are illustrated in Table 4-4.2-1.

Table 4-4.2-1: Demographic data for subject groups.

N= 186	BCa	CRC (Early & Late stage)	PCa	HCC	Non-Cancer (Fibrosis & Non-Fibrosis)
Number	15	58 (24-34)	55	6	38 (7-31)
Mean Age (max-min) years	70.0(95-51)	74.3(92-46)	71.9(90-47)	71.9(89-48)	62.5(90-32)
Sex: M/F	12:3	2.33:1	All Male	4.83:1	2:1
Avg. BMI	24.4	28.6	27.5	25.8	30.9
Current Smoker (% of whole population)	1 (6.7%)	3 (5.2%)	6 (10.9%)	0 (0%)	3 (8.3%)

Controls were recruited from two sources to decrease bias. The first source was healthy individuals without liver disease. The second source was patients with different stages of NAFLD, the advantage here is that these patients represent those at risk of becoming HCC cases in the future. Controls were then further divided into 31 non-fibrotic and 7 fibrotic/cirrhotic cases. Exclusion criteria were pregnancy and age <18 years. All participants were recruited prior to any anticancer treatment. The study did not take into consideration the various covariates mentioned in Table 4-4.2-1 such as age, smoking, sex, and BMI (Body Mass Index), during the research process.

58 CRC samples were further distributed into 24 early-stage CRC samples and 34 late-stage CRC samples based on TNM (tumour/node/metastasis) staging. We assigned T1 and T2 stages as early-stage and T3 and T4 as late-stage samples.

4.2.2. Analytical Devices and Setup:

a. G.A.S. FlavourSpec GC-IMS:

The G.A.S FlavourSpec (Germany) uses a GC-IMS measurement technique to analyse VOCs. The operation and working principle of GC-IMS is discussed in detail in chapter 3. GC-IMS is a method used in various applications, such as detection of explosives and chemicals [10, 11], air quality [12], health and disease detection [7, 13] and food [14, 15]. The sample preparation methods that were utilized for the analysis have been explained in detail in Chapter 3, section d (Sample Preparation:, pg.:61).

b. Markes GC-TOF-MS:

GC-TOF-MS operates by analysing the time of flight of ions and analyse them according to their mass-to-charge ratio. The operation and working principle of GC-TOF-MS is discussed in detail in chapter 3. The GC-TOF-MS system used is a combination of TRACE 1300 GC (Thermo Fisher Scientific, Loughborough, UK) and BenchTOF-HD TOF-MS (Markes Intl., Llantrisant, UK). The sample preparation methods that were utilized for the analysis have been explained in detail in Chapter 3, section c (Sample Preparation:, pg.: 76).

4.3. Results:

4.3.1. G.A.S. FlavourSpec GC-IMS:

The results of the statistical analysis of the GC-IMS represents results between different cancer groups and the non-cancerous group are given in Table 4.3-1, Table 4.3-2, Table 4.3-3, and Table 4.3-4. Figure 4.3-1 shows the chromatogram obtained from the GC-IMS FlavourSpec output, which displays the raw data. The chromatogram illustrates the output generated by GC-IMS for bladder cancer samples and control samples.

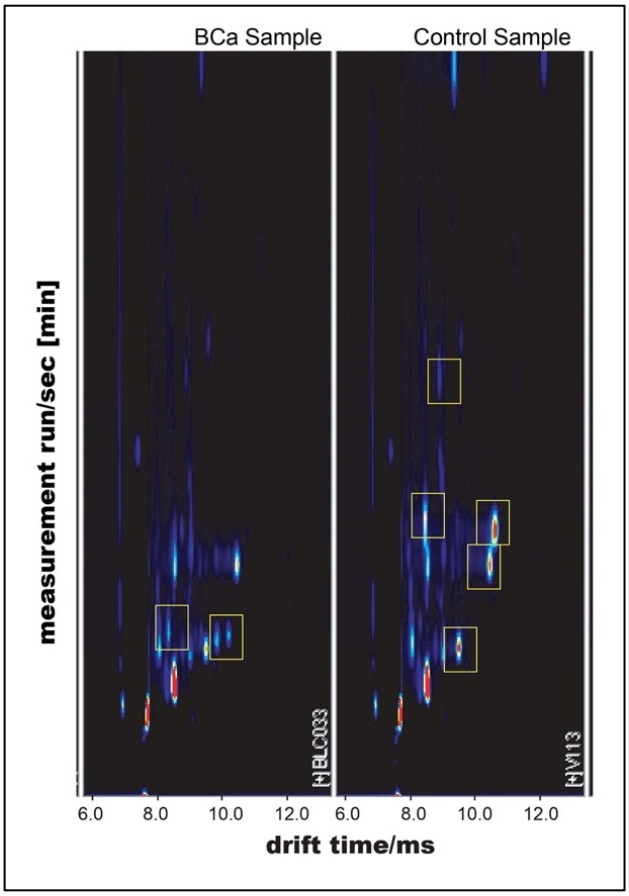


Figure 4.3-1: Chromatogram generated by GC-IMS.

The results demonstrate high sensitivity and specificity, indicating that there are significant differences between the VOCs profiles of the different groups.

Table 4.3-1, Table 4.3-2, Table 4.3-3 and Table 4.3-4 show that output generated by a bespoke R-program. The tables present the classifiers that achieved the highest level of separation among the six classifiers introduced in the Chapter 3 in section e (Qualitative and Quantitative Analysis:, pg.: 62) that generated the highest level of separations to distinguish between two classes for each comparison. Each of these classifiers was carefully selected to distinguish between two classes for each comparison. The outputs in the tables highlight the specific classifiers that demonstrated the maximum separation capabilities, indicating their effectiveness in accurately differentiating between the classes compared. Similarly, sensitivity and specificity obtained from the classifiers show the proportion of cancer groups that were correctly classified and the proportion of the non-cancerous groups that were

correctly classified. The detailed analysis process is discussed in Chapter 3, section e (Qualitative and Quantitative Analysis:, pg.: 62).

The underlined entries in the tables represent the outputs that stood out, either showing remarkably high separation or, conversely, extremely poor separation. These specific entries signify exceptional performance or challenges faced by the classifiers in distinguishing between the classes being compared. The underlining draws attention to these critical points of interest, indicating instances of high success or areas where further improvement may be required. Further discussion regarding the underlined outputs in the tables can be found at the end in the "Discussion" section.

GC-IMS was able to distinguish between bladder cancer and hepatocellular cancer and bladder cancer and fibrosis disease with 100% sensitivity and 100% specificity. This signifies that FlavourSpec was able to identify all bladder cancer cases, hepatocellular cancer cases and fibrosis samples correctly. For BCa in comparison with CRC, logistic Regression reported 14 true positive cases out of 15 and 56 true negative cases out of 58. BCa vs PCa reported 9 true positive cases out of 15 and 53 true negative cases out of 55.

Table 4.3-1: GC-IMS Output for the comparison of BCa with different cancers and non-cancerous urine samples (presenting the most significant results)

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
BCa vs non-Cancerous	Logistic Regression	0.95 (0.9-0.99)	0.87 (0.70-1.00)	0.92 (0.84-0.98)	0.81 (0.64-0.95)	0.95 (0.88-1)
BCa Vs CRC	Logistic Regression	0.99 (0.99-1)	0.93 (0.81-1)	0.98 (0.95-1)	0.93 (0.81-1)	0.98 (0.95-1)
BCa vs PCa	Logistic Regression	0.94 (0.88-0.98)	0.60 (0.38-0.81)	0.96 (0.92-1)	0.81 (0.6-1)	0.9 (0.8-0.97)
BCa Vs HCC	Extreme Gradient Boosting	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)
BCa Vs Fibrosis	Extreme Gradient Boosting	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)
BCa Vs Non-Fibrosis	Logistic Regression	0.86 (0.76-0.94)	0.87 (0.71-1)	0.77 (0.65-0.89)	0.65 (0.48-0.82)	0.92 (0.83-1)

Table 4.3-2: GC-IMS Output for the comparison of CRC with different cancers and non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
CRC Vs non-cancerous	Logistic Regression	0.83 (0.75-0.89)	0.84 (0.76-0.92)	0.68 (0.56-0.81)	0.80 (0.72-0.89)	0.74 (0.62-0.86)
CRC Vs PCa	Logistic Regression	0.99 (0.98-0.99)	0.91 (0.84-0.97)	0.95 (0.89-0.98)	0.95 (0.89-0.98)	0.92 (0.85-0.97)
CRC Vs HCC	Extreme Gradient Boosting	0.94 (0.87-0.99)	0.97 (0.92-1)	0.75 (0.57-0.91)	0.92 (0.85-0.97)	0.88 (0.73-1)
CRC Vs Fibrosis	Extreme Gradient Boosting	0.71 (0.49-0.89)	0.98 (0.95-1)	0 (0-0)	0.89 (0.82-0.95)	0 (0-0)
CRC Vs Non-Fibrosis	Logistic Regression	0.87 (0.79-0.93)	0.93 (0.87-0.98)	0.74 (0.61-0.87)	0.87 (0.79-0.94)	0.85 (0.73-0.96)

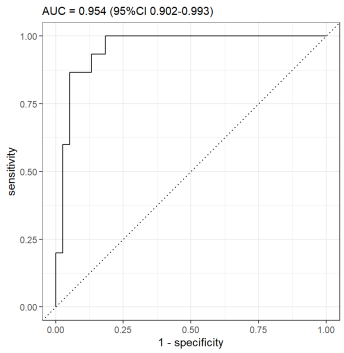
Table 4.3-3: GC-IMS Output for the comparison of PCa with different cancers and non-cancerous urine samples.

Comparisons	Classifiers	AUC	Sensitivity	Specificity	PPV	NPV
PCa vs non-Cancerous	Extreme Gradient Boosting	0.89 (0.83-0.94)	0.76 (0.64-0.88)	0.88 (0.80-0.95)	0.81 (0.69-0.91)	0.85 (0.76-0.92)
PCa Vs HCC	Extreme Gradient Boosting	0.89 (0.80-0.97)	0.6 (0.4-0.78)	0.96 (0.92-1)	0.86 (0.69-1)	0.87 (0.80-0.94)
PCa Vs Fibrosis	Logistic Regression	0.84 (0.65-0.99)	0.43 (0-0.75)	0.98 (0.95-1)	0.75 (0.32-1)	0.93 (0.88-0.98)
PCa Vs Non-Fibrosis	Extreme Gradient Boosting	0.86 (0.78-0.93)	0.65 (0.5-0.79)	0.93 (0.87-0.98)	0.83 (0.69-0.95)	0.83 (0.75-0.91)

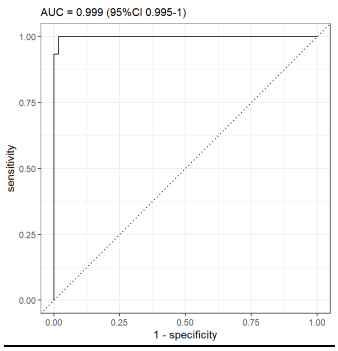
Table 4.3-4: GC-IMS Output for the comparison of HCC with non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
HCC Vs non-Cancerous	Extreme Gradient Boosting	0.78 (0.67-0.88)	0.65 (0.47-0.82)	0.84 (0.74-0.93)	0.68 (0.5-0.86)	0.82 (0.72-0.91)
HCC Vs Fibrosis	Extreme Gradient Boosting	0.97 (0.91-1)	0.43 (0.13-0.74)	0.95 (0.86-1)	0.75 (0.33-1)	0.83 (0.68-0.95)
HCC Vs NON-Fibrosis	Extreme Gradient Boosting	0.62 (0.48-0.76)	0.6 (0.41-0.78)	0.74 (0.61-0.87)	0.6 (0.42-0.78)	0.74 (0.61-0.88)
Fibrosis Vs Non-Fibrosis	Logistic Regression	0.63 (0.36-0.89)	0.29 (0-0.6)	0.9 (0.81-0.97)	0.4 (0-0.83)	0.85 (0.74-0.94)

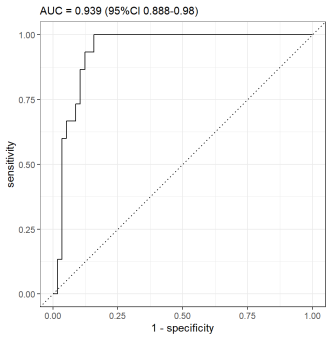
The ROC curves obtained from GC-IMS data comparing different groups are shown in Figure 4.3-2. The results indicate that among BCa patients and HCC patients, the AUC (area under the curve) was 1 (1-1) with sensitivity and specificity of 1 (1-1) and 1 (1-1) respectively. Also, the separation between BCa and Fibrosis samples was high with a sensitivity 1 (1-1), specificity 1 (1-1) and 1 (1-1). Similarly, for CRC and BCa sample comparison using GC-IMS, the separation was significant with sensitivity 0.93 (0.81-1), specificity 0.98 (0.95-1) and AUC 0.99 (0.99-1).



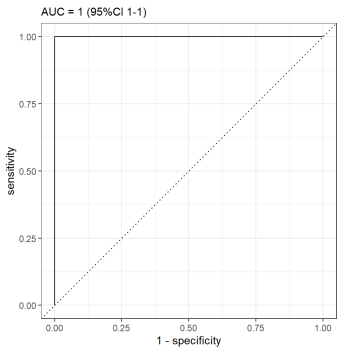
(a)



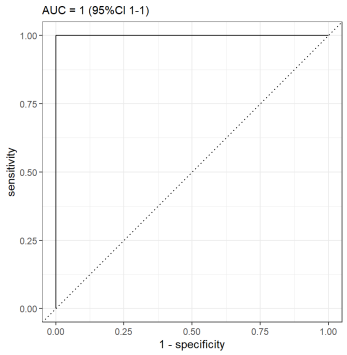
(b)



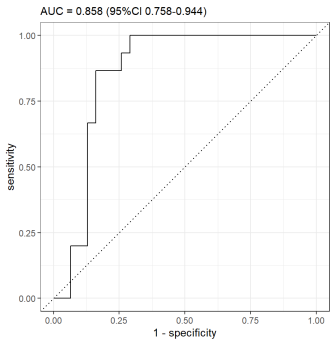
(c)



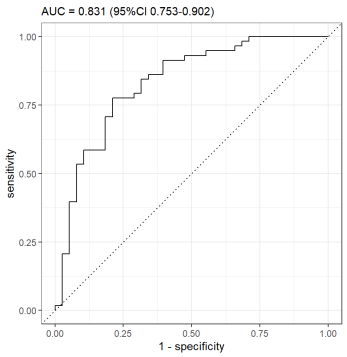
(d)



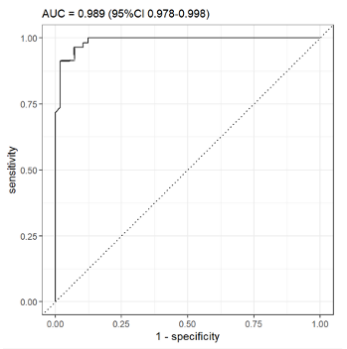
(e)



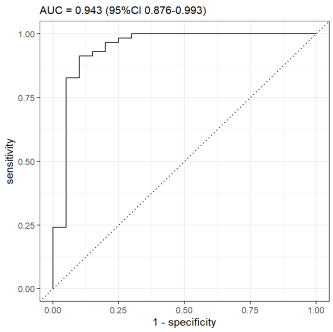
(f)



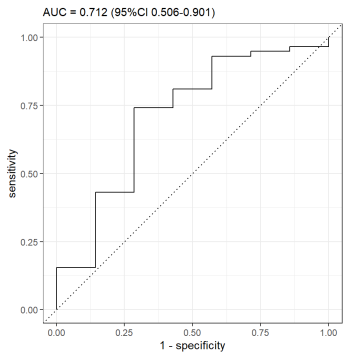
(g)



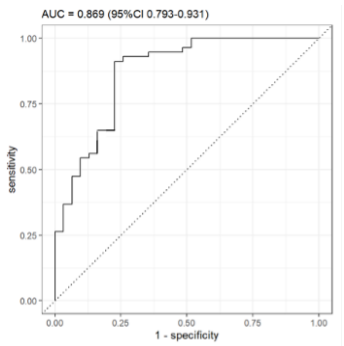
(h)



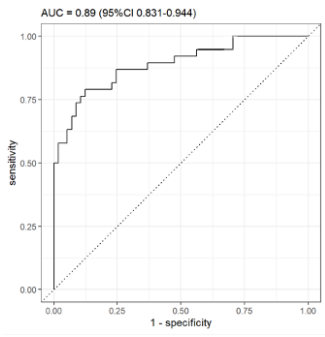
(i)



(j)



(k)



(l)

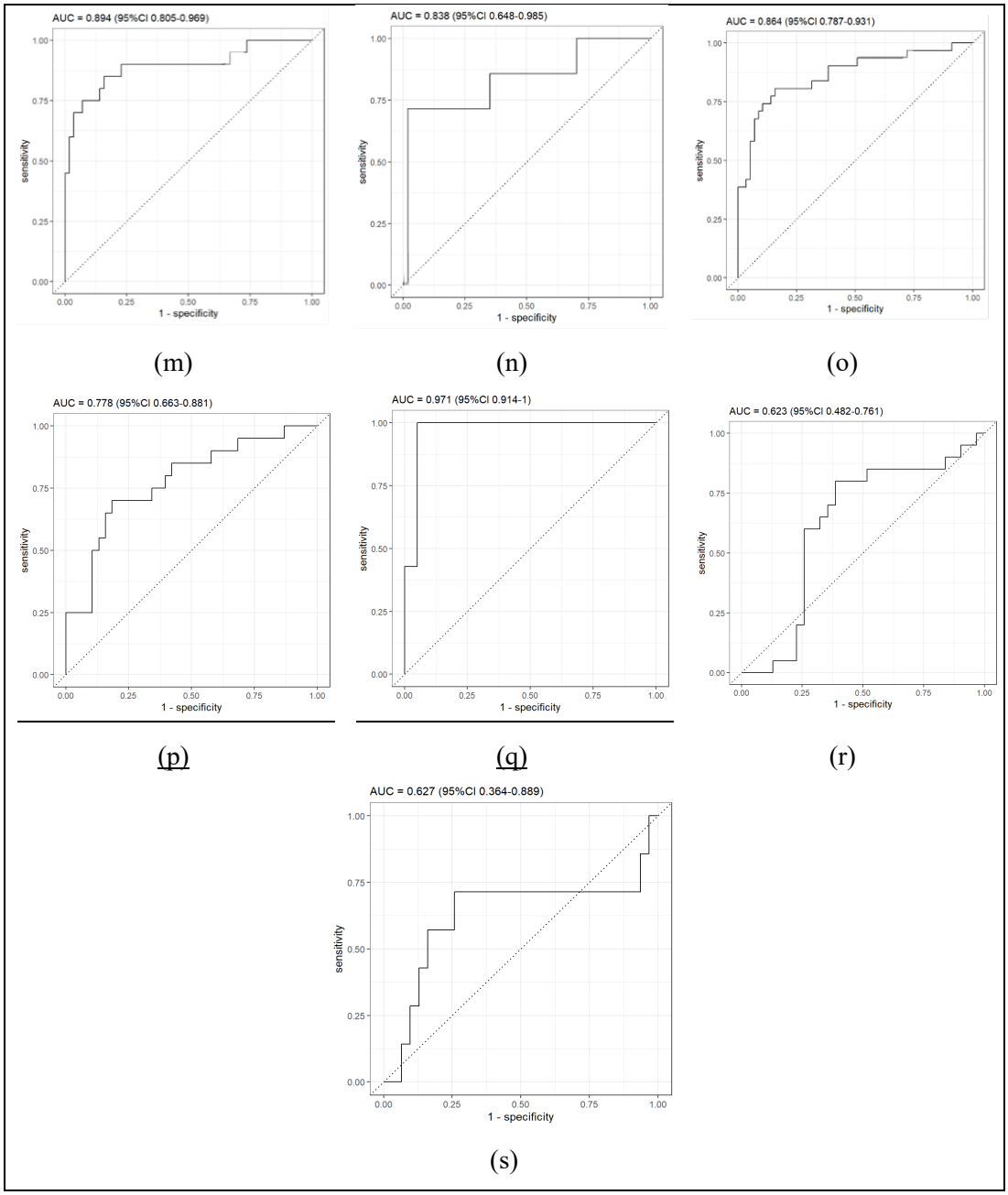


Figure 4.3-2: ROC curves obtained from the R-Program using GC-IMS for different cancer groups where (a) Bladder cancer Vs non-cancerous; (b) Bladder cancer Vs colorectal cancer; (c) Bladder cancer Vs prostate cancer; (d) Bladder cancer Vs hepatocellular cancer; (e) Bladder cancer Vs fibrosis; (f) Bladder cancer Vs non-fibrosis; (g)Colorectal cancer Vs non-cancerous group; (h) Colorectal cancer Vs prostate cancer; (i) Colorectal cancer Vs hepatocellular cancer; (j) Colorectal cancer Vs fibrosis; (k) Colorectal cancer Vs non-fibrosis; (l) Prostate cancer Vs non-cancerous; (m) Prostate cancer Vs hepatocellular cancer; (n) Prostate cancer Vs fibrosis; (o) Prostate cancer Vs non-fibrosis;(p) Hepatocellular cancer Vs non-cancerous; (q) Hepatocellular cancer Vs fibrosis; (r) Hepatocellular cancer Vs non-fibrosis; and (s) Fibrosis Vs non-fibrosis.

4.3.2. Markes GC-TOF-MS:

The results of the statistical analysis between different cancer groups for GC-TOF-IMS are given in Table 4.3-5, Table 4.3-6, Table 4.3-7 and Table 4.3-8. The results demonstrate high sensitivity and specificity. GC-TOF-MS was successfully able to distinguish between different cancer groups and diseases. The highest sensitivity and specificity were reported with CRC and PCa group with 100% sensitivity and specificity. Also, the separation between bladder cancer and CRC samples were high with GC-TOF-MS analysis. It correctly reported 11 out of 15 true positive cases and 58 out of 58 negative cases.

Table 4.3-5: GC-TOF-MS Output the comparison of BCa with different cancers and non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
BCa vs non-Cancerous	Random Forest	0.82 (0.69-0.89)	0.07 (0-0.19)	0.94 (0.88-1)	0.33 (0-1)	0.71 (0.6-0.81)
BCa Vs CRC	Logistic Regression	0.99 (0.98-1)	0.73 (0.53-0.9)	1 (1-1)	1 (1-1)	0.94 (0.88-0.98)
BCa vs PCa	Random Forest	0.87 (0.79-0.93)	0.33 (0.13-0.55)	0.91 (0.84-0.97)	0.5 (0.22-0.75)	0.83 (0.75-0.91)
BCa Vs HCC	Logistic Regression	0.97 (0.89-1)	0.93 (0.81-1)	0.67 (0.33-1)	0.88 (0.72-1)	0.8 (0.5-1)
BCa Vs Fibrosis	Random Forest	0.96 (0.88-1)	0.93 (0.81-1)	0.43 (0.13-0.8)	0.78 (0.6-0.94)	0.75 (0.33-1)
BCa Vs Non-Fibrosis	Random Forest	0.81 (0.69-0.91)	0.4 (0.19-0.6)	0.77 (0.64-0.89)	0.46 (0.23-0.7)	0.72 (0.58-0.85)

Table 4.3-6: GC-TOF-MS Output the comparison of CRC with different cancers and non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
CRC Vs non-cancerous	Neural Network	0.95 (0.91-0.98)	0.88 (0.81-0.95)	0.89 (0.79-0.97)	0.93 (0.86-0.98)	0.82 (0.72-0.92)
CRC Vs PCa	Random Forest	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)

CRC Vs HCC	Neural Network	0.70 (0.46-0.91)	0.9 (0.93-1)	0.29 (0-0.6)	0.92 (0.85-0.97)	0.5 (0-1)
CRC Vs Fibrosis	Neural Network	0.81 (0.66-0.94)	0.91 (0.85-0.97)	0.43 (0.11-0.78)	0.93 (0.87-0.98)	0.38 (0.1-0.67)
CRC Vs Non-Fibrosis	Neural Network	0.95 (0.9-0.98)	0.88 (0.8-0.95)	0.87 (0.76-0.96)	0.93 (0.86-0.98)	0.79 (0.67-0.9)

Table 4.3-7: GC-TOF-MS Output the comparison of PCa with different cancers and non-cancerous urine samples.

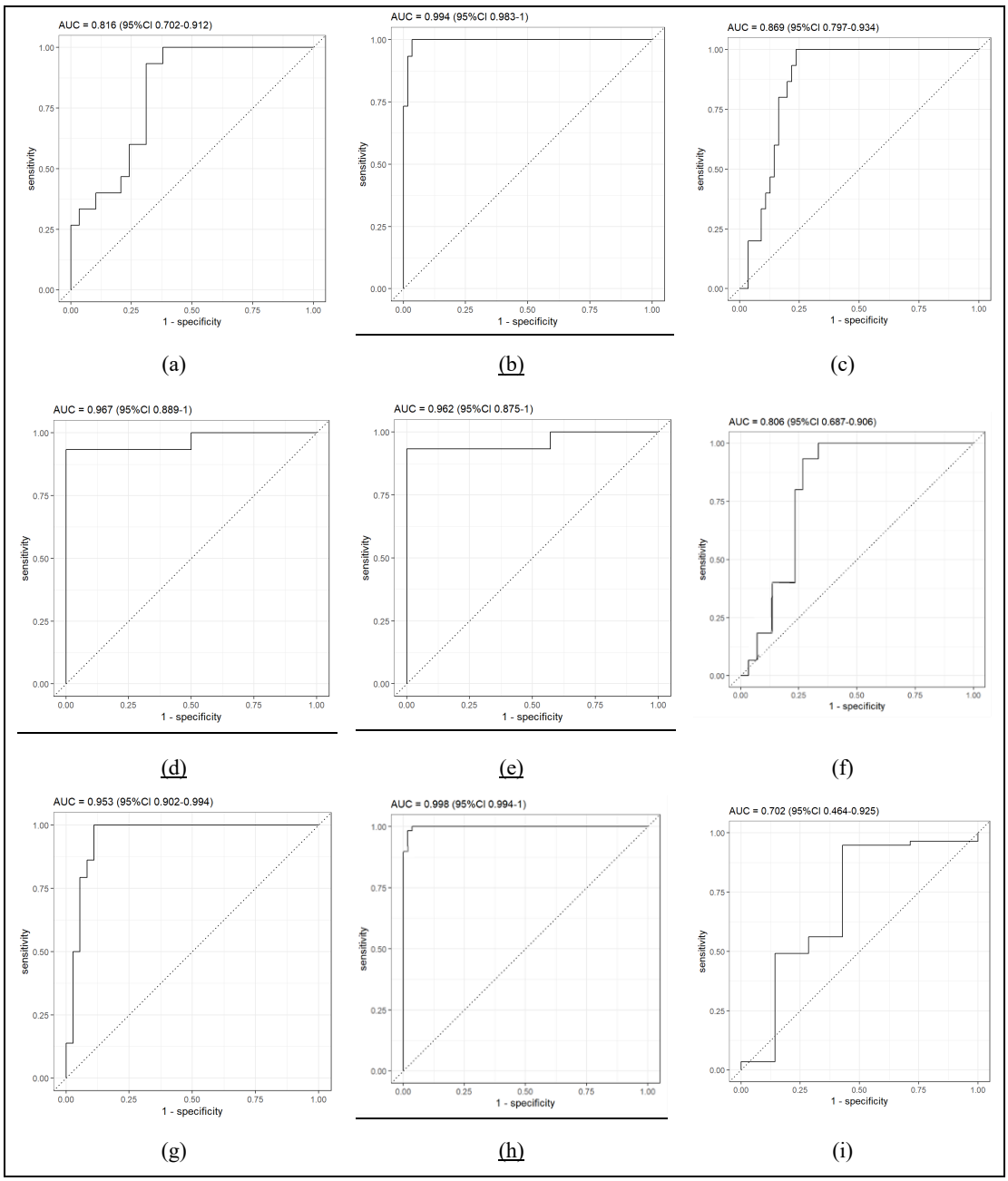
Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
PCa vs non-Cancerous	Random Forest	0.93 (0.89-0.97)	0.72 (0.59-0.84)	0.88 (0.80-0.95)	0.81 (0.69-0.91)	0.82 (0.73-0.91)
PCa Vs HCC	Neural Network	0.87 (0.75-0.97)	0.5 (0-0.86)	1 (1-1)	1 (1-1)	0.94 (0.89-0.98)
PCa Vs Fibrosis	Neural Network	0.96 (0.91-0.99)	0.29 (0-0.6)	1 (1-1)	1 (1-1)	0.92 (0.85-0.97)
PCa Vs Non-Fibrosis	Random Forest	0.93 (0.89-0.97)	0.73 (0.59-0.86)	0.91 (0.83-0.96)	0.82 (0.69-0.93)	0.85 (0.77-0.93)

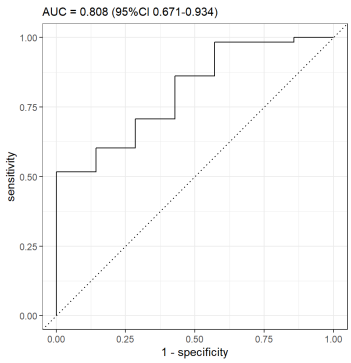
Table 4.3-8: GC-TOF-MS Output the comparison of HCC with non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
HCC Vs non-Cancerous	Neural Network	0.52 (0.32-0.73)	0 (0-0)	0.94 (0.88-1)	0 (0-0)	0.85 (0.76-0.93)
HCC Vs Fibrosis	Neural Network	0.67 (0.36-0.91)	0.57 (0.25-0.88)	0.5 (0.14-0.83)	0.57 (0.22-0.86)	0.5 (0.14-0.83)
HCC Vs Non-Fibrosis	Gaussian Process classifier	0.48 (0.21-0.73)	0 (0-0)	0.93 (0.85-1)	0 (0-0)	0.85 (0.74-0.94)
Fibrosis Vs Non-Fibrosis	Logistic Regression	0.65 (0.41-0.87)	0.43 (0.11-0.78)	1 (1-1)	1 (1-1)	0.88 (0.79-0.97)

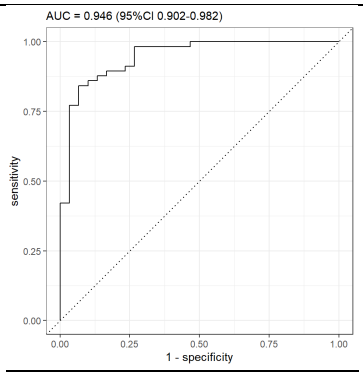
The ROC curves obtained from GC-TOF-MS data comparing different groups are shown in Figure 4.3-3. The results indicate CRC and PCa urine samples had the highest sensitivity and specificity of 1 (1-1) and 1 (1-1) respectively. Also, for CRC and BCa samples comparison using GC-TOF-MS, the separation was high with sensitivity 0.73 (0.53-0.9),

specificity 1 (1-1) and AUC 0.99 (0.98-1). However, some of the groups showed poor sensitivity for example, both HCC and the non-fibrosis group with HCC and non-cancerous group showed zero sensitivity.

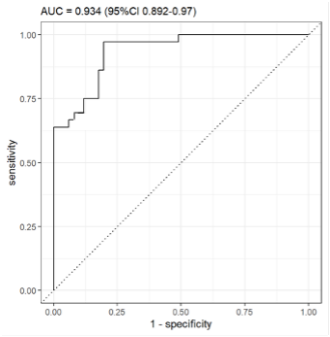



AUC = 0.808 (95%CI 0.671-0.934)
sensitivity
1 - specificity

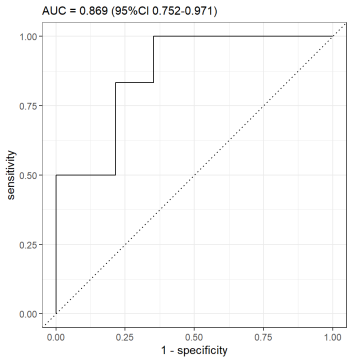
(j)


AUC = 0.946 (95%CI 0.902-0.982)
sensitivity
1 - specificity

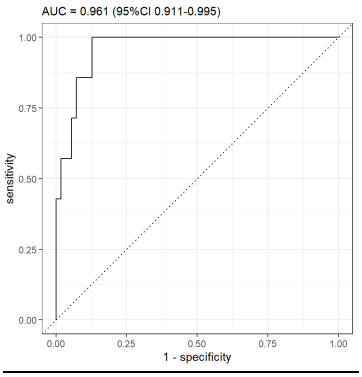
(k)


AUC = 0.934 (95%CI 0.892-0.97)
sensitivity
1 - specificity

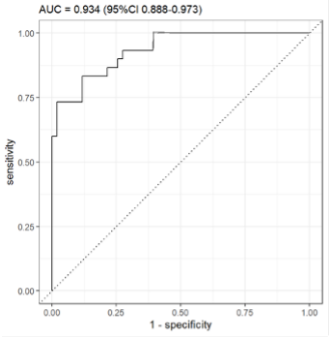
(l)


AUC = 0.869 (95%CI 0.752-0.971)
sensitivity
1 - specificity

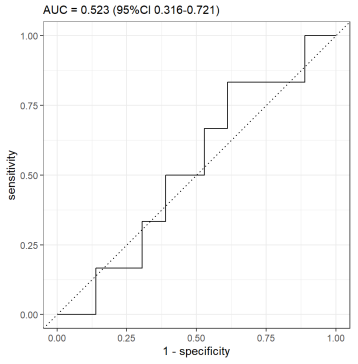
(m)


AUC = 0.961 (95%CI 0.911-0.995)
sensitivity
1 - specificity

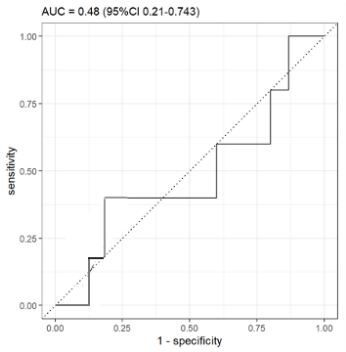
(n)


AUC = 0.934 (95%CI 0.888-0.973)
sensitivity
1 - specificity

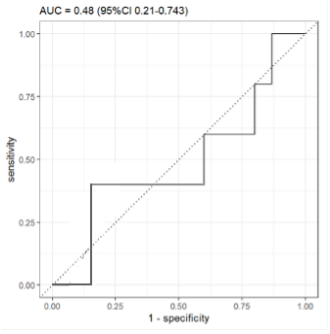
(o)


AUC = 0.523 (95%CI 0.316-0.721)
sensitivity
1 - specificity

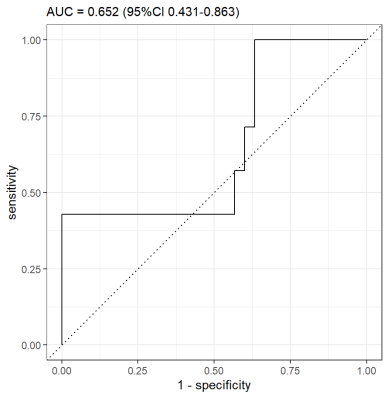
(p)


AUC = 0.48 (95%CI 0.21-0.743)
sensitivity
1 - specificity

(q)


AUC = 0.48 (95%CI 0.21-0.743)
sensitivity
1 - specificity

(r)


AUC = 0.652 (95%CI 0.431-0.863)
sensitivity
1 - specificity

(s)

Figure 4.3-3: ROC curves obtained from the R-Program using GC-TOF-MS for different cancer groups where (a) Bladder cancer Vs non-cancerous; (b) Bladder cancer Vs colorectal cancer; (c) Bladder cancer Vs prostate cancer; (d) Bladder cancer Vs hepatocellular cancer; (e) Bladder cancer Vs fibrosis; (f) Bladder cancer Vs non-fibrosis; (g)Colorectal cancer Vs non-cancerous group; (h) Colorectal cancer Vs prostate cancer; (i) Colorectal cancer Vs hepatocellular cancer; (j) Colorectal cancer Vs fibrosis; (k) Colorectal cancer Vs non-fibrosis; (l) Prostate cancer Vs non-cancerous; (m) Prostate cancer Vs hepatocellular cancer; (n) Prostate cancer Vs fibrosis; (o) Prostate cancer Vs non-fibrosis;(p) Hepatocellular cancer Vs non-cancerous; (q) Hepatocellular cancer Vs fibrosis; (r) Hepatocellular cancer Vs non-fibrosis; and (s) Fibrosis Vs non-fibrosis.

In addition, data obtained from the GC-TOF-MS were used to identify the unknown VOCs in urine samples for each group. TOF-DS software identified the chemicals based on a NIST list using a criterion of p-value<0.05. We were able to identify 46 VOCs significant for all the cancer groups based on selection criteria, different studies, and PubChem, we found 13 VOCs which were relevant to BCa as shown in Table 4.3-9. For colorectal cancer, 17 VOCs were identified as significant VOCs leading to CRC as shown in Table 4.3-10. GC-TOF-MS was able to identify 9 VOCs which may be biomarkers for Prostate cancer as shown in Table 4.3-11. There were 7 chemicals reported in this study found relevant to HCC shown in Table 4.3-12.

Table 4.3-9: A list of possible biomarkers from the analysis of urine samples by GC-TOF-MS identified using PubChem, NIST and publications significant to Bladder Cancer.

S. No.	Chemicals	p-value	Molecular weight (g/mol)
1	Biphenyl*	<0.01	154.21
2	Nonanal*	<0.01	142.24
3	Tetradecane*	<0.01	198.39
4	Pentadecane, 2,6,10,14-tetramethyl-	0.012	268.5
5	2-Pentanone	0.012	86.13
6	Undecane	0.014	156.31
7	4-Heptanone	0.018	114.19
8	Dodecane	0.025	170.33
9	Hexadecane	0.026	226.44
10	Heptanal	0.026	114.19
11	Methyl Isobutyl Ketone	0.045	100.16

12	Naphthalene	0.046	128.169
13	Benzoic acid	0.049	122.12
***Indicates the biomarkers that did not overlap with any previous study and may be indicators of Bladder cancer**			

Table 4.3-10: A list of possible biomarkers from the analysis of urine samples by GC-TOF-MS identified using PubChem, NIST and publications significant to Colorectal Cancer.

S. No.	Chemicals	p-value	Molecular weight (g/mol)
1	Octanal	<0.001	128.21
2	Nonanal	<0.001	142.24
3	Decanal	<0.001	156.26
4	Heptanal	<0.001	114.19
5	Hexanal	<0.001	100.16
6	Acetone	<0.001	58.08
7	Acetic acid	<0.001	60.05
8	2-Pentanone	0.001	86.13
9	Toluene	0.00219	92.14
10	2-Heptanone	0.00429	114.19
11	4-Heptanone	0.00307	114.19
12	Ethylbenzene	0.00499	106.16
13	Nonane	0.01102	128.25
14	p-Xylene	0.01478	106.16
15	Methyl Isobutyl Ketone	0.01602	100.16
16	Undecane	0.02246	156.31
17	Naphthalene	0.02613	128.17

Table 4.3-11: A list of possible biomarkers from the analysis of urine samples by GC-TOF-MS identified using PubChem, NIST and publications significant to Prostate Cancer.

S. No.	Chemicals	p-value	Molecular weight (g/mol)
1	Toluene	<0.01	92.14

2	Phenol	<0.01	325.4
3	Acetic acid	<0.01	60.05
4	Ethylbenzene	<0.01	106.16
5	1-Hexanol, 2-ethyl-	0.011	130.229
6	Disulfide, dimethyl	0.012	94.2
7	Cyclopentanone, 2-methyl-	0.017	98.14
8	Naphthalene	0.029	128.17
9	Pyrrole	0.033	67.09

Table 4.3-12: A list of possible biomarkers from the analysis of urine samples by GC-TOF-MS identified using PubChem, NIST and publications significant to Hepatocellular Cancer.

S. No.	Chemicals	p-value	Molecular weight (g/mol)
1	4-Methyl-2,4-bis(p-hydroxyphenyl) pent-1-ene, 2TMS derivative	<0.01	412.7
2	2-Butanone	0.03637	72.11
3	2-Hexanone	0.04309	100.16
4	Benzene, 1-ethyl-2-methyl-	0.04183	120.19
5	3-Butene-1,2-diol, 1-(2-furanyl)-	0.03247	154.16
6	Bicycloheptane, 3,7,7-trimethyl	0.03553	138.25
7	Sulpiride	0.04369	341.4

4.4. Discussion:

The use of volatile organic compounds (VOCs) for the detection of cancer has garnered considerable attention due to its non-invasive nature and potential cost-effectiveness. In this research, we employed two different methods, namely GC-IMS and GC-TOF-MS to distinguish various cancer and disease groups from one another, as well as from the non-cancerous group. By analysing the VOCs profiles of urine samples through GC-MS-TOF data, we were able to identify the chemical compounds involved in the process.

The study aimed to assess the effectiveness of several methods in distinguishing between cancer and non-cancerous groups in urine samples. Results obtained from GC-IMS are presented in Table 4.3 1, Table 4.3 2, Table 4.3 3 and Table 4.3 4, with the underlined entries representing the data that stood out. For instance, Table 4.3-1 shows a comparison between Bladder Cancer (BCa) and Colorectal Cancer (CRC) using the Logistic Regression classifier, achieving an AUC of 0.99, indicating excellent discrimination power. The sensitivity for the Logistic Regression model is 0.93, with a 95% confidence interval ranging from 0.81 to 1. This means that the model correctly identified 93% of the BCa cases in the dataset. Similarly, the specificity for the Logistic Regression model is 0.98, with a 95% confidence interval ranging from 0.95 to 1. This means that the model correctly identified 98% of the CRC cases in the dataset.

In Table 4.3-2, another comparison is presented, involving colorectal cancer and fibrosis disease, using the Extreme Gradient Boosting (XGBoost) classifier. The XGBoost model demonstrated relatively good discrimination ability, with an AUC value of 0.71. However, it showed a high sensitivity of 0.98 but a low specificity of 0, indicating that it had a higher tendency to correctly identify CRC cases but struggled to correctly identify Fibrosis cases. The observed high variation in the data may be attributed to the imbalanced distribution of samples between the CRC class (consisting of 58 samples) and the Fibrosis class (consisting of only 7 samples). When the dataset is imbalanced, the classifier may become biased towards the majority class (in this case, CRC) and may not perform as well on the minority class (Fibrosis). Another reason for the high variation in the data could be the technique used for data splitting, which was 10-fold cross-validation. In this technique, the dataset was divided into 10 equal-sized subsets, and the classifier was trained and evaluated 10 times, with each fold serving as the test set once, while the remaining 9 folds served as the training set. With an imbalanced dataset and 10-fold cross-validation, it is crucial to ensure that each fold contains a proportional representation of both classes to prevent any bias. However, due to the small number of Fibrosis samples, maintaining this balance in every fold was challenging, which could further affect the performance estimates of the classifier [16, 17].

Table 4.3-8 presents similar findings, focusing on the comparison between HCC and Non-Fibrosis samples. For this comparison, we utilized the Gaussian Process classifier, and the values in parentheses indicate the 95% confidence interval for each metric. The AUC value,

with a confidence interval of (0.21-0.73), reveals some uncertainty about the classifier's performance. Moreover, the sensitivity and PPV values have confidence intervals of (0-0), indicating that the classifier failed to correctly identify any HCC samples or make any accurate positive predictions. On the other hand, the specificity and NPV values have confidence intervals of (0.85-1), suggesting some uncertainty in the classifier's ability to correctly identify non-fibrosis samples. Overall, the GP classifier exhibited limited capability to distinguish between HCC and Non-Fibrosis samples, but it demonstrated better performance in accurately identifying non-fibrosis samples.

Conversely, we observed a perfect score of 1 for all metrics achieved using the XGBoost classifier when distinguishing between BCa and HCC, as well as between BCa and fibrosis. Perfect classification indicates that the classifier makes no errors and successfully separates all samples into their respective classes. Nevertheless, it is essential to recognize that the GC-IMS data may contain inherent noise, measurement errors, and variations resulting from various factors. Additionally, the complexity and heterogeneity of biological samples can present challenges in achieving perfect classification. Moreover, GC-IMS data often involve high-dimensional feature spaces, and some samples may exhibit overlapping characteristics, making it difficult for any classifier to achieve perfect separation. Several potential reasons could account for obtaining a perfect score. For instance, if the dataset is small or heavily imbalanced, with one class significantly outnumbering the others, a classifier may achieve high accuracy by merely predicting the majority class for all samples [18]. Another factor to consider is overfitting, where the classifier becomes overly specialized to the specific data points in the training set, leading to exceptional performance on the training data but poor generalization to new, unseen data. Overfitting can inflate performance metrics artificially and may not accurately reflect the classifier's true performance on real-world data. Similar patterns were observed in GC-TOF-MS results when differentiating between CRC samples and PCa samples [19].

Figure 4.3-2 and Figure 4.3-3 displays the ROC curves obtained from GC-IMS and GC-TOF-MS data, respectively. These ROC curves illustrate the performance of the classifiers for various comparison groups. Among BCa patients and PCa patients, the AUC (area under the curve) was found to be 0.97 (with a confidence interval of 0.93 to 1.00). This suggests that the classifier exhibits a moderate level of discrimination power for distinguishing

between these two groups. The ROC curve graphically represents the performance of the classifier across various classification thresholds. The AUC serves as a comprehensive metric, ranging from 0 to 1, with higher values indicating better discrimination ability. A value of 1 represents a perfect classifier, while 0.5 signifies a random classifier. For the BCa and PCa comparison, the AUC of 0.97 indicates that the classifier has a high probability of accurately classifying positive instances (sensitivity) and negative instances (specificity). The 95% confidence interval (0.93 to 1.00) suggests a relatively high likelihood that the true AUC value falls within this range. On the other hand, the comparison between HCC and non-fibrosis shows an AUC of 0.62, indicating that the classifier's ability to distinguish between HCC samples and non-fibrosis samples is only slightly better than random chance. This suggests that the classifier's performance in correctly classifying data into their appropriate categories may not be very reliable.

It is important to interpret these results carefully, considering the classifier's performance in differentiating between distinct groups. While the AUC of 0.97 for BCa and PCa indicates a reasonably good discrimination power, the AUC of 0.62 for HCC and non-fibrosis indicates that the classifier's performance in this specific comparison is modest. These findings highlight the importance of assessing the classifier's performance across various datasets and conditions to obtain a comprehensive understanding of its discriminatory capabilities [20].

Comparing our results with previous studies, we found that GC-IMS can be used for detection of different disease groups. A study conducted by Gasparri *et al* [21] analysed urinary VOCs for the early diagnosis of lung cancer with a gas chromatography-ion mobility spectrometer (GC-IMS) and an electronic nose (e-nose) made by a matrix of twelve quartz microbalances complemented by a photoionization detector. They showed that GC-IMS identified lung cancer with respect to the control group with a diagnostic AUC of 91%, sensitivity of 85%, and specificity of 90. Another study conducted by Liu *et al.* [22] showed that GC-IMS was able to detect VOCs in urine samples from 66 patients with PCa and to comparatively analyses samples from 87 patients with non-cancerous controls (NCs) with an AUC for the RF and SVM model were 0.955 and 0.981, respectively. Similarly, we found that Gao *et al* [23] calculated an AUC of 0.92 (0.96 sensitivity and 0.80 specificity) for urinary VOCs analysis of PCa while Lima *et al* [24] reported an AUC of 0.83 (84%

sensitivity and 80% specificity) using PLS-DA for PCa detection. GC-TOF-MS was also highly effective in separating colorectal cancer (CRC) and PCa groups with an accuracy of 0.99 and sensitivity and specificity of 0.97 and 0.98. The highest accuracy obtained by GC-TOF-MS for CRC and the non-cancerous group was 0.93 and 0.94 for PCa and the non-cancerous groups. These high values indicate that GC-TOF-MS was able to accurately recognize and separate CRC, PCa, and non-cancerous urine samples.

In this study, we developed urinary VOCs profiles for different cancer groups using GC-TOF-MS data. Table 4.3-9 consists of the chemicals that have been identified in our study and have been cross verified using PubChem, NIST and previously published research, which may have relevance to BCa diagnosis. Similarly, Table 4.3-10 consists of chemicals that may indicate the presence of Colorectal cancer. Chemicals present in Table 4.3-11 represent the biomarkers which may represent the possibility of Prostate cancer. Lastly Table 4.3-12 contains the chemicals that may have relevance to HCC diagnosis.

Out of 13 VOCs found noteworthy to BCa, Biphenyl, Heptanal, and 2, 6, 10, 14-tetramethyl-Pentadecane were the three distinct biomarkers found in our study that did not overlap with other studies. Biphenyl has been identified as the most significant biomarker in our study. Biphenyl has been linked to various diseases, including carcinoma. It has been proven that Biphenyl is a promoter of BCa in rats [25]. Nonanal, Tetradecane, Dodecane, Hexadecane, Naphthalene, and Methyl Isobutyl Ketone have been suggested by Rodrigues *et al.* [26] in their study using GC-MS on BCa cell lines, whereas 2-pentanone and 4-Heptanone overlap with the findings of Cauchi *et al.* [27]. Benzoic acid was another chemical found in our study that overlapped in both Rodrigues *et al.* [26] and Cauchi *et al.* [27] studies.

Table 4.3-10 illustrates 17 chemicals relevant for CRC detection. Octanal is a human metabolite present in the membrane and generally reported in saliva or faeces inside the human body [28]. It was reported by Batty *et al.* [29] on analysis of faecal samples of CRC with PLS-DA following feature selection with Wilcoxon T test. Nonanal is another VOC found in our study. It is a toxic compound and has been found related with several diseases [28]. Nonanal has frequently been reported as a breath biomarker for CRC in different studies [30-32]. Decanal is another important biomarker observed in our study, which has been reported as a CRC biomarker in several studies as a breath biomarker [30, 32] and

VOCs in cell culture studies [33-35]. Both 2-Pentanone and 4-Heptanone have been reported as CRC biomarkers by Arasaradnam *et al.* [36] in their study of colorectal cancer using urinary samples. 2-Heptanone exists at the membrane level inside living species including humans, and outside, it can be found in milk, corns, and peppermints. It has been identified as a significant CRC biomarker using cell culture in two different studies [33, 35]. It is reported as a CRC VOCs in three different studies [30, 32, 37]. Nonane can be found in breath, saliva and faeces and is in membrane inside the human body [28, 38]. It has been found as a CRC biomarker in plasma samples using GC x GC-MS by Kim *et al*. [39]. Methyl Isobutyl Ketone is a human metabolite and found in urine, saliva, and faeces. It is found in several diseases and causes weak effects on the central nervous system and causes memory impairment [40]. It has been reported as a breath biomarker of CRC [30, 32]. Ethylbenzene is a human metabolite present in subcellular level (membrane) in the human body. It is present in tobacco smoke and is a water and air pollutant [28]. De Vietro *et al.* [31] found that ethylbenzene was present in 4 out of the 7 and toluene in 6 out of 7 CRC patient's breath samples and tissue samples. Also, a study conducted by Altomare *et al.* [41] showed that ethylbenzene and toluene were associated with a chemical fingerprint of CRC.

Table 4.3-11 consists of the list of chemicals that were reported by our study as PCa biomarkers. In our study, we found toluene as the most significant chemical for PCa. Toluene has been published previously as a significant biomarker for PCa [42]. In addition, it has been reported that toluene has been found to be associated with testicular diseases [43, 44]. Toluene was suggested as a PCa biomarker by Peng *et al.* [42] in their study of detection of different cancers using sensor arrays. Phenol is a toxic compound and is used as a disinfectant and can cause dizziness, uraemia, hypotension, confusion, headache, diarrhoea etc. Struck-Lewicka *et al.* [45] found Phenol and acetic acid as a significant biomarker for prostate cancer in their study of urine metabolic fingerprinting using LC-MS and GC-MS. Ethylbenzene was also demonstrated as PCa biomarker in a study conducted with urine samples using GC-MS (Gas Chromatography-Mass Spectrometry) [46]. Lima *et al.* [47] demonstrated in their work that 2-methyl-Cyclopentanone does not naturally occur in the human body. It is found to be present in urine as a potential biomarker for prostate cancer. They also suggested Naphthalene as a PCa biomarker.

In this research, two analytical instruments, GC-IMS, and GC-TOF-MS, were utilized to assess their respective abilities to distinguish between various cancer groups and a non-cancerous group. The results obtained from the GC-IMS demonstrated superior statistical performance, indicating its effectiveness in differentiating between the cancerous and non-cancerous samples. On the other hand, GC-TOF-MS was employed specifically for VOC pattern generation.

The reason why GC-IMS generated better statistical results can be attributed to its capability to provide more precise and sensitive measurements of volatile organic compounds (VOCs). GC-IMS is known for its high sensitivity, rapid response, and excellent resolution, enabling it to detect and quantify subtle differences in VOC profiles, which are crucial for accurate classification. The advanced ion mobility technique utilized by GC-IMS facilitates the separation and identification of various VOCs present in the samples, enhancing the discriminatory power of the instrument.

Conversely, GC-TOF-MS was chosen for VOC pattern generation due to its ability to capture the comprehensive profile of VOCs in the samples. This analytical technique excels in detecting a wide range of VOCs, thereby allowing researchers to create detailed VOC patterns that can provide valuable insights into the molecular composition and variations among different cancer groups. GC-TOF-MS, with its high-resolution mass spectrometry, is capable of identifying numerous VOCs simultaneously, enabling the construction of comprehensive VOC profiles for in-depth analysis.

By leveraging the strengths of both GC-IMS and GC-TOF-MS, this research aimed to achieve a comprehensive understanding of the VOC profiles associated with various cancer groups. While GC-IMS excelled in generating statistically significant results for classification, GC-TOF-MS complemented the study by providing detailed VOC patterns, contributing to a more comprehensive characterization of the volatile compounds present in the cancer samples. This combination of analytical instruments allowed for a holistic approach in studying the VOC signatures of different cancer types, thereby advancing our understanding of their unique metabolic profiles and potential diagnostic applications.

The use of urinary VOCs analysis is an attractive option due to the non-invasive nature. It also has the potential to be used in early cancer and disease diagnosis with further validation

studies. This approach may also prove to be efficient, whilst lowering the cost per patient, and increasing patient compliance due to its non-invasive nature. We believe that using VOCs to analyse human waste will be an important diagnostic tool for the future. Cancer may well be one area of focus and be used as part of the UK two-week wait screening programme to help reduce the number of un-needed procedures. However, the success of this approach depends on conducting larger studies targeting various cancers and having tools that are CE marked or equivalent for cancer diagnosis.

This thesis chapter highlights several limitations of the study results. Firstly, the study did not account for the various contributory factors that can lead to abnormal metabolism and subsequent excretion of various concentrations of these chemicals in urine, including stress, alcohol, smoking, certain food products, medicines, and environmental factors. Several studies have reported the effects of smoking on VOCs [48, 49]. A study conducted by A. McWilliams *et al*. [50] showed that active smoking had an impact on urinary VOCs profiles associated with current smokers and ex-smokers. Secondly, the study lacked a healthy control group, as it compared different cancer groups with a non-cancerous group consisting of patients with a history of different diseases and was conducted at a single site. Thirdly, urine was used as a sample, which can be easily influenced by exogenous sources, potentially leading to biased results. Finally, the study did not undertake chemical identification with calibration standards, nor did it attempt to quantify the identified chemicals, although many of these chemicals were previously reported in other studies, which indicates that these VOCs can be potential indicators of these diseases. Future research should aim to address these limitations by utilizing a larger sample set, including a healthy control group, and accounting for the various contributory factors that may impact on the excretion of VOCs in urine.

4.5. Conclusion:

In this research, GC-IMS and GC-TOF-MS were used to identify VOCs fingerprints using urine headspace and establish an interdependence between different cancer groups, disease groups and non-cancerous samples. It was found that all these analytical devices have the potential to differentiate between different cancer groups. A total of 46 VOCs were found to be relevant for identifying these cancer groups, with several VOCs distinct to each cancer.

These VOCs profiles for each cancer group may support the use of VOCs for the screening of cancer and confirm clinical diagnostic assessments. This will help in avoiding the inefficient analytical methods currently used for screening and give better, cheaper, and non-invasive approaches for cancer and disease diagnosis and detection. In the future, these VOCs profiles obtained from these analytical devices could be used as a reference for developing low-cost devices. It is plausible that VOCs profiles can be used as an adjunct to diagnosis perhaps selecting only high-risk groups to undergo cystoscopy examinations which will be widely beneficial considering their limited capacity and cost.

4.6. References:

[1] H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians,* vol. 71, no. 3, pp. 209-249, 2021.

[2] V. S. A. Jayanthi, A. B. Das, and U. Saxena, "Recent advances in biosensor development for the detection of cancer biomarkers," *Biosensors and Bioelectronics,* vol. 91, pp. 15-23, 2017.

[3] Q. Gao and W.-Y. Lee, "Urinary metabolites for urological cancer detection: A review on the application of volatile organic compounds for cancers," *American journal of clinical and experimental urology,* vol. 7, no. 4, p. 232, 2019.

[4] B. de Lacy Costello, N. M. Ratcliffe, A. Amann, and D. Smith, *Volatile organic compounds (VOCs) found in urine and stool*. Elsevier, Amsterdam, The Netherlands, 2013.

[5] R. Arasaradnam, J. Covington, C. Harmston, and C. Nwokolo, "Next generation diagnostic modalities in gastroenterology–gas phase volatile compound biomarker detection," *Alimentary pharmacology & therapeutics,* vol. 39, no. 8, pp. 780-789, 2014.

[6] R. Becker, "Non-invasive cancer detection using volatile biomarkers: Is urine superior to breath?," in *Medical Hypotheses* vol. 143, ed, 2020, p. 110060.

[7] 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," in *Biosensors* vol. 9, ed, 2019.

[8] S. Bosch *et al.*, "Early detection and follow-up of colorectal neoplasia based on faecal volatile organic compounds," in *Colorectal Disease* vol. 22, ed: John Wiley & Sons, Ltd, 2020, pp. 1119-1129.

[9] S. Esfahani *et al.*, "Variation in Gas and Volatile Compound Emissions from Human Urine as It Ages, Measured by an Electronic Nose," in *Biosensors* vol. 6, ed, 2016.

[10] L. V. Haley and J. M. Romeskie. GC-IMS: a technology for many applications.

[11] G. W. Cook, P. T. LaPuma, G. L. Hook, and B. A. Eckenrode, "Using Gas Chromatography with Ion Mobility Spectrometry to Resolve Explosive Compounds in the Presence of Interferents*," in *Journal of Forensic Sciences* vol. 55, ed: John Wiley & Sons, Ltd, 2010, pp. 1582-1591.

[12] P. C. Moura, V. Vassilenko, J. M. Fernandes, and P. H. Santos, Cham. Indoor and Outdoor Air Profiling with GC-IMS BT - Technological Innovation for Life Improvement.

[13] M. Allers *et al.*, "Measurement of exhaled volatile organic compounds from patients with chronic obstructive pulmonary disease (COPD) using closed gas loop GC-IMS and GC-APCI-MS," in *Journal of Breath Research* vol. 10, ed: IOP Publishing, 2016, p. 26004.

[14] E. Valli *et al.*, "An HS-GC-IMS Method for the Quality Classification of Virgin Olive Oils as Screening Support for the Panel Test," in *Foods* vol. 9, ed, 2020.

[15] S. Wang, H. Chen, and B. Sun, "Recent progress in food flavor analysis using gas chromatography–ion mobility spectrometry (GC–IMS)," in *Food Chemistry* vol. 315, ed, 2020, p. 126158.

[16] R. Das and A. Sengur, "Evaluation of ensemble methods for diagnosing of valvular heart disease," *Expert Systems with Applications,* vol. 37, no. 7, pp. 5110-5115, 2010/07/01/ 2010, doi: https://doi.org/10.1016/j.eswa.2009.12.085.

[17] N. Japkowicz, "Learning from imbalanced data sets: a comparison of various strategies," in *AAAI workshop on learning from imbalanced data sets*, 2000, vol. 68: Citeseer, pp. 10-15.

[18] Y. Zhang *et al.*, "Using ensemble methods to deal with imbalanced data in predicting protein–protein interactions," *Computational Biology and Chemistry,* vol. 36, pp. 36-41, 2012/02/01/ 2012, doi: https://doi.org/10.1016/j.compbiolchem.2011.12.003.

[19] M. S. Santos, J. P. Soares, P. H. Abreu, H. Araujo, and J. Santos, "Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier]," *ieee ComputatioNal iNtelligeNCe magaziNe,* vol. 13, no. 4, pp. 59-76, 2018.

[20] F. He *et al.*, "Exploration of key aroma active compounds in strong flavor Baijiu during the distillation by modern instrument detection technology combined with multivariate statistical analysis methods," *Journal of Food Composition and Analysis,* vol. 110, p. 104577, 2022/07/01/ 2022, doi: https://doi.org/10.1016/j.jfca.2022.104577.

[21] R. Gasparri *et al.*, "Volatolomic urinary profile analysis for diagnosis of the early stage of lung cancer," *Journal of Breath Research,* vol. 16, no. 4, p. 046008, 2022/09/02 2022, doi: 10.1088/1752-7163/ac88ec.

[22] Q. Liu *et al.*, "Volatile organic compounds for early detection of prostate cancer from urine," *Heliyon,* vol. 9, no. 6, 2023.

[23] Q. Gao and W.-Y. Lee, "Urinary metabolites for urological cancer detection: a review on the application of volatile organic compounds for cancers," in *American journal of clinical and experimental urology* vol. 7, ed: e-Century Publishing Corporation, 2019, pp. 232-248.

[24] A. R. Lima *et al.*, "A Panel of Urinary Volatile Biomarkers for Differential Diagnosis of Prostate Cancer from Other Urological Cancers," in *Cancers* vol. 12, ed, 2020.

[25] Y. Kurata, M. Asamoto, A. Hagiwara, T. Masui, and S. Fukushima, "Promoting effects of various agents in rat urinary bladder carcinogenesis initiated by N-butyl-N-(4-hydroxybutyl)nitrosamine," in *Cancer Letters* vol. 32, ed, 1986, pp. 125-135.

[26] D. Rodrigues *et al.*, "Volatile metabolomic signature of bladder cancer cell lines based on gas chromatography–mass spectrometry," in *Metabolomics* vol. 14, ed, 2018, p. 62.

[27] M. Cauchi *et al.*, "Evaluation of gas chromatography mass spectrometry and pattern recognition for the identification of bladder cancer from urine headspace," in *Analytical Methods* vol. 8, ed: Royal Society of Chemistry, 2016, pp. 4037-4046.

[28] D. S. Wishart *et al.*, "HMDB 4.0: the human metabolome database for 2018," in *Nucleic Acids Research* vol. 46, ed, 2018, pp. D608-D617.

[29] C. A. Batty, M. Cauchi, C. Lourenço, J. O. Hunter, and C. Turner, "Use of the Analysis of the Volatile Faecal Metabolome in Screening for Colorectal Cancer," in *PLOS ONE* vol. 10, ed: Public Library of Science, 2015, p. e0130301.

[30] D. Altomare *et al.*, "Exhaled volatile organic compounds identify patients with colorectal cancer," in *The British journal of surgery* vol. 100, ed, 2013, pp. 144-150.

[31] N. De Vietro *et al.*, "Relationship between cancer tissue derived and exhaled volatile organic compound from colorectal cancer patients. Preliminary results," in *Journal of Pharmaceutical and Biomedical Analysis* vol. 180, ed, 2020, p. 113055.

[32] D. F. Altomare *et al.*, "Effects of Curative Colorectal Cancer Surgery on Exhaled Volatile Organic Compounds and Potential Implications in Clinical Follow-up," in *Annals of Surgery* vol. 262, ed, 2015.

[33] G. Wang *et al.*, "Determination of volatile organic compounds in SW620 colorectal cancer cells and tumor-bearing mice," in *Journal of Pharmaceutical and Biomedical Analysis* vol. 167, ed, 2019, pp. 30-37.

[34] G. Zonta *et al.*, "Detection of colorectal cancer biomarkers in the presence of interfering gases," in *Sensors and Actuators B: Chemical* vol. 218, ed, 2015, pp. 289-295.

[35] M. Liu *et al.*, "Release of volatile organic compounds (VOCs) from colorectal cancer cell line LS174T," in *Analytical Biochemistry* vol. 581, ed, 2019, p. 113340.

[36] R. P. Arasaradnam *et al.*, "Detection of Colorectal Cancer (CRC) by Urinary Volatile Organic Compound Analysis," in *PLOS ONE* vol. 9, ed: Public Library of Science, 2014, p. e108750.

[37] A. Bond *et al.*, "Volatile organic compounds emitted from faeces as a biomarker for colorectal cancer," in *Alimentary Pharmacology & Therapeutics* vol. 49, ed: John Wiley & Sons, Ltd, 2019, pp. 1005-1012.

[38] B. de Lacy Costello *et al.*, "A review of the volatiles from the healthy human body," in *Journal of Breath Research* vol. 8, ed: IOP Publishing, 2014, p. 14001.

[39] S. Kim, X. Yin, M. A. I. Prodhan, X. Zhang, Z. Zhong, and I. Kato, "Global Plasma Profiling for Colorectal Cancer-Associated Volatile Organic Compounds: a Proof-of-Principle Study," in *Journal of Chromatographic Science* vol. 57, ed, 2019, pp. 385-396.

[40] E. Grober and H. H. Schaumburg, "Occupational exposure to methyl isobutyl ketone causes lasting impairment in working memory," in *Neurology* vol. 54, ed, 2000, pp. 1853 LP - 1855.

[41] D. Altomare *et al.*, "Chemical signature of colorectal cancer: case-control study for profiling the breath print," in *BJS Open*, ed, 2020.

[42] G. Peng *et al.*, "Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors," in *British Journal of Cancer* vol. 103, ed, 2010, pp. 542-551.

[43] W.-G. Chung, I.-J. Yu, C.-S. Park, K.-H. Lee, H.-K. Roh, and Y.-N. Cha, "Decreased formation of ethoxyacetic acid from ethylene glycol monoethyl ether and reduced atrophy of testes in male rats upon combined administration with toluene and xylene," in *Toxicology Letters* vol. 104, ed, 1999, pp. 143-150.

[44] I.-J. Yu *et al.*, "Co-administration of toluene and xylene antagonized the testicular toxicity but not the hematopoietic toxicity caused by ethylene glycol monoethyl ether in Sprague–Dawley rats," in *Toxicology Letters* vol. 109, ed, 1999, pp. 11-20.

[45] W. Struck-Lewicka *et al.*, "Urine metabolic fingerprinting using LC–MS and GC–MS reveals metabolite changes in prostate cancer: A pilot study," in *Journal of Pharmaceutical and Biomedical Analysis* vol. 111, ed, 2015, pp. 351-361.

[46] A. Jiménez-Pacheco, M. Salinero-Bachiller, M. C. Iribar, A. López-Luque, J. L. Miján-Ortiz, and J. M. Peinado, "Furan and p-xylene as candidate biomarkers for prostate cancer," in *Urologic Oncology: Seminars and Original Investigations* vol. 36, ed, 2018, pp. 243.e21-243.e27.

[47] A. R. Lima *et al.*, "Identification of a biomarker panel for improvement of prostate cancer diagnosis by volatile metabolic profiling of urine," in *British Journal of Cancer* vol. 121, ed, 2019, pp. 857-868.

[48] Y.-H. Kim and K.-H. Kim, "A novel method to quantify the emission and conversion of VOCs in the smoking of electronic cigarettes," *Scientific reports,* vol. 5, no. 1, pp. 1-9, 2015.

[49] S. M. Charles, S. Batterman, and C. Jia, "Composition and emissions of VOCs in main-and side-stream smoke of research cigarettes," *Atmospheric Environment,* vol. 41, no. 26, pp. 5371-5384, 2007.

[50] A. McWilliams, P. Beigi, A. Srinidhi, S. Lam, and C. E. MacAulay, "Sex and smoking status effects on the early detection of early lung cancer in high-risk smokers using an electronic nose," *IEEE Transactions on Biomedical Engineering,* vol. 62, no. 8, pp. 2044-2054, 2015.

Chapter 5. Analysis of Cancer Samples using eNoses

This chapter focuses on the application of two eNose systems, namely, the AlphaMOS FOX 4000 and PEN3, in the analysis of urinary cancer samples. eNoses were used to analyse urine samples obtained from cancer patients as well as healthy controls. The chapter discusses the analysis as well as the results obtained from the eNose analysis. It also explores the potential of eNose technology for cancer diagnosis and its possible implications for distinguishing cancer samples from healthy controls.

5.1. Introduction:

As discussed in the chapters before, the eNose is capable of detecting VOCs present in biological samples, which potentially can differ between cancerous and healthy individuals. Studies have reported that the eNose can distinguish between cancer and healthy samples with high sensitivity and specificity. For instance, in a study using urine samples, the eNose could differentiate between bladder cancer patients and healthy controls with an accuracy of 94% [1]. Similarly, in another study using breath samples, the eNose showed a high accuracy of 86% in distinguishing lung cancer patients from healthy individuals [2]. Therefore, eNose has demonstrated a promising ability to differentiate cancer samples from healthy controls, highlighting its potential as a non-invasive diagnostic tool for cancer.

In this research, we aimed to evaluate the use of the AlphaMOS FOX 4000 eNose and PEN3 eNose as potential analytical instruments for the detection and discrimination between cancer groups and healthy controls.

5.2. Materials and Methods:

5.2.1. Urine Samples:

The details about the urine sample collection, storage, analysis, and demography have been discussed in chapter 4.

5.2.2. Analytical devices and setup:

a. The Alpha-MOS electric Nose - FOX 4000 with HS100 Autosampler:

The AlphaMOS Fox 4000 is an electronic nose that comprises 18 commercial metal oxide sensors (MOS), 6 p-type sensors and 12 n-type sensors, distributed in three temperature-controlled chambers. The output of the sensors is measured in terms of change in resistance.

The change of resistance measured refers to the quantitative electrical resistance values obtained from the gas sensors during exposure to samples. These values are subsequently used to generate the VOC profiles and facilitate the classification of different samples. The resistance measurements are fundamental for the AlphaMOS Fox 4000's functionality and its ability to differentiate between various cancer samples and non-cancerous samples. A comprehensive explanation of the working principles of the AlphaMOS FOX 4000 can be found in Chapter 3 in section 3.3.2 (AlphaMOS FOX 4000 (Toulouse, France):, pg.: 81).

The data analysis methods that were utilized in this research have been explained in detail in Chapter 3 in section c (Quantitative and qualitative analysis for AlphaMOS Fox 4000:, pg.: 83).

b. The PEN3 Electronic Nose (Airsense Analytics GmbH, Schwerin, Germany):

The PEN3 eNose (Airsense Analytics GmbH, Schwerin, Germany) is a portable (92 × 190 × 255 mm) olfactory system used for the identification of chemicals and gases. It is a combination of a gas sampling unit and a sensor array. In our case the PEN3 eNose is fitted with an autosampler (HT2000H Dynamic Headspace Auto-sampler, Brescia (BS), Italy), which interfaces directly with the PEN 3 software (WinMuster PEN v 1.6.2.18). A comprehensive explanation of the working principles of the AlphaMOS FOX 4000 can be found in Chapter 3 in section 3.3.3 (PEN3 Electronic Nose (Airsense Analytics GmbH, Schwerin, Germany):, pg.: 87).

The data analysis methods that were utilized in this research have been explained in detail in Chapter 3 in section c (Quantitative and qualitative analysis for AlphaMOS Fox 4000:, pg.: 83).

5.3. Results:

5.3.1. AlphaMOS electronic Nose- FOX 4000 with HS100 Autosampler:

The AlphaMOS FOX 4000 used 18 sensors to evaluate the urine samples and data were analysed using the MultiSens Analyzer and RStudio program. The initial analysis performed on the features vectors extracted by the MultiSens Analyzer was PCA (Principal Component Analysis). An example of the output from the PCA is shown in Figure 5.3-1.

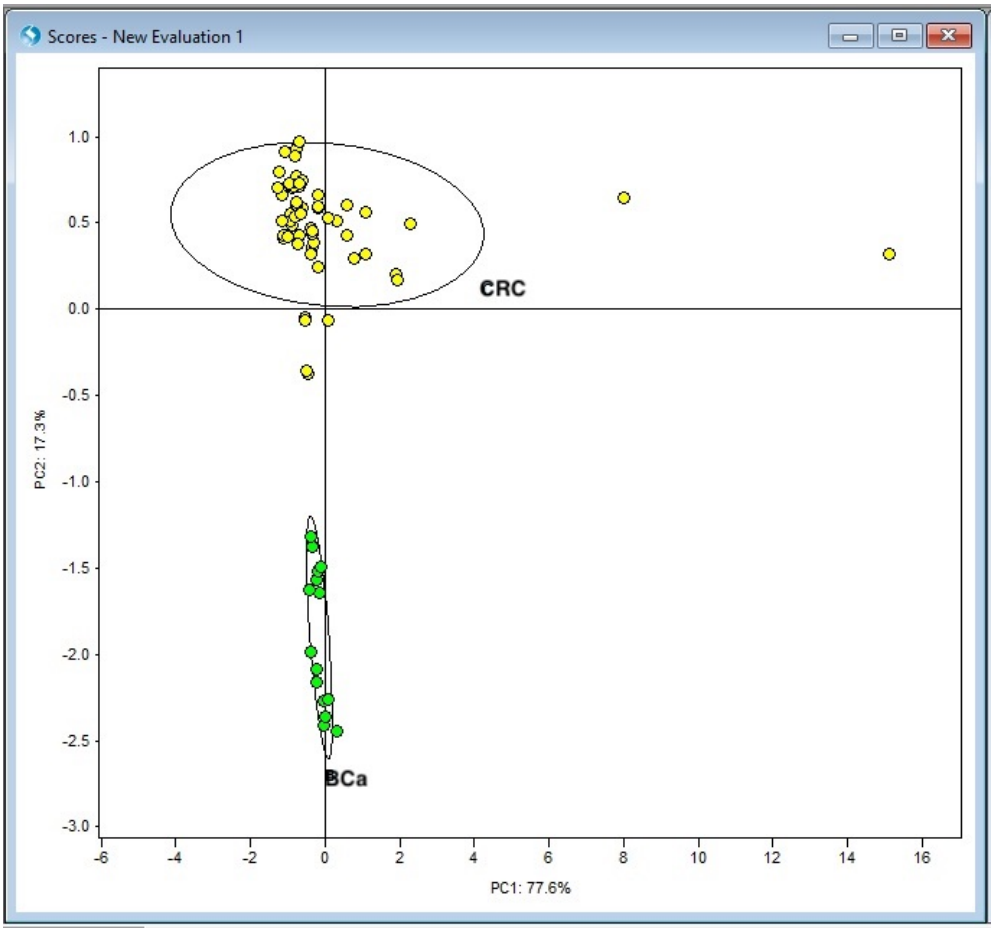


Figure 5.3-1: Principal Component Analysis of AlphaMOS Fox 4000 results where green dots represent the features extracted from the bladder cancer samples and yellow dots represent the features extracted from the colorectal cancer samples.

The results from the statistical analysis are shown in Table 5.3 1, Table 5.3 2, Table 5.3-3 and Table 5.3-4. The tables present the top-performing classifiers among the six introduced in Chapter 3, section e (Qualitative and Quantitative Analysis:, pg.: 62). These classifiers were carefully selected to effectively distinguish between two classes for each comparison. A detailed analysis of this process is discussed in Chapter 3, section c (Quantitative and qualitative analysis for AlphaMOS Fox 4000:, pg.: 83). The underlined entries in the tables highlight the output that stand out, representing outputs with either remarkably high separation or, on the contrary, extremely poor separation. These specific entries highlight exceptional performance or challenges faced by the classifiers in distinguishing between the compared classes. The underlining emphasizes these crucial points of interest, showcasing instances of remarkable success or areas where further improvement may be needed. For

more in-depth discussions about the underlined outputs in the tables, refer to the "Discussion" section at the end.

Table 5.3-1: AlphaMOS FOX 4000 Output for the comparison of BCa with different cancers and non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
BCa vs non-Cancerous	Random Forest	0.86 (0.76-0.95)	0.67 (0.44-0.86)	0.85 (0.76-0.94)	0.63 (0.41-0.83)	0.88 (0.78-0.95)
BCa Vs CRC	Neural Network	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)
BCa Vs PCa	Random Forest	0.88 (0.81-0.94)	0.47 (0.25-0.69)	0.90 (0.83-0.96)	0.54 (0.29-0.78)	0.87 (0.79-0.94)
BCa Vs HCC	Logistic Regression	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)
BCa Vs Fibrosis	Random Forest	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)
BCa Vs Non-Fibrosis	Random Forest	0.85 (0.75-0.93)	0.6 (0.38-0.8)	0.84 (0.73-0.94)	0.64 (0.41-0.86)	0.82 (0.71-0.93)

Table 5.3-2: AlphaMOS FOX 4000 Output for the comparison of CRC with different cancers and non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
CRC Vs non-cancerous	Neural Network	0.67 (0.57-0.75)	0.78 (0.69-0.86)	0.34 (0.22-0.47)	0.63 (0.53-0.72)	0.52 (0.36-0.67)
CRC Vs PCa	Neural Network	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)	1 (1-1)
CRC Vs HCC	Neural Network	0.99 (0.98-1)	0.8 (0.67-0.92)	1 (1-1)	1 (1-1)	0.92 (0.86-0.97)
CRC Vs Fibrosis	Support Vector Machine	0.9 (0.79-0.99)	0.98 (0.95-1)	0.33 (0-0.63)	0.9 (0.84-0.97)	0.75 (0.33-1)
CRC Vs Non-Fibrosis	Neural Network	0.93 (0.86-0.98)	0.93 (0.88-0.98)	0.88 (0.77-0.97)	0.93 (0.87-0.98)	0.88 (0.77-0.97)

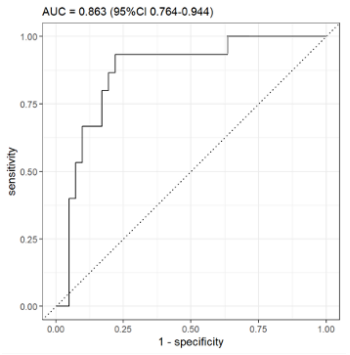
Table 5.3-3: AlphaMOS FOX 4000 Output for the comparison of PCa with different cancers and non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
PCa vs non-Cancerous	Random Forest	0.93 (0.88-0.97)	0.78 (0.67-0.88)	0.89 (0.81-0.96)	0.84 (0.73-0.94)	0.85 (0.76-0.92)
PCa Vs HCC	Neural network	0.97 (0.94-0.99)	0.84 (0.71-0.96)	0.98 (0.95-1)	0.95 (0.87-1)	0.94 (0.88-0.98)
PCa Vs Fibrosis	Neural network	0.99 (0.99-1)	0.78 (0.5-1)	1 (1-1)	1 (1-1)	0.97 (0.92-1)
PCa Vs Non-Fibrosis	Random Forest	0.91 (0.84-0.96)	0.81 (0.7-0.91)	0.88 (0.81-0.95)	0.79 (0.67-0.91)	0.89 (0.83-0.96)

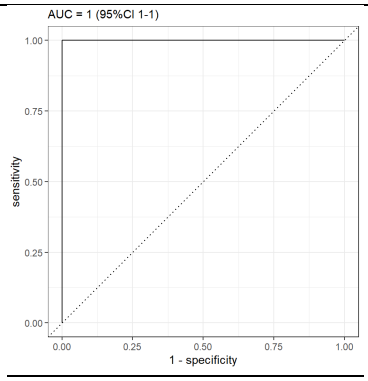
Table 5.3-4: AlphaMOS FOX 4000 Output for the comparison of HCC with different non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
HCC Vs non-Cancerous	Random Forest	0.69 (0.58-0.8)	0.52 (0.35-0.69)	0.66 (0.52-0.78)	0.48 (0.31-0.64)	0.69 (0.57-0.82)
HCC Vs Fibrosis	Gaussian Process	0.65 (0.48-0.8)	0.72 (0.57-0.86)	0.44 (0.14-0.75)	0.78 (0.64-0.92)	0.36 (0.12-0.6)
HCC Vs NON-Fibrosis	Neural Network	0.78 (0.67-0.88)	0.64 (0.48-0.79)	0.75 (0.62-0.88)	0.67 (0.5-0.83)	0.73 (0.59-0.85)
Fibrosis Vs Non-Fibrosis	Random Forest	0.7 (0.48-0.89)	0.44 (0.17-0.73)	0.97 (0.91-1)	0.8 (0.5-1)	0.86 (0.76-0.95)

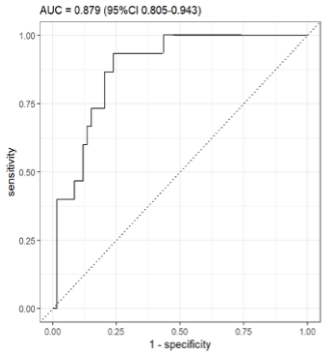
The ROC curves for these comparisons are shown in Figure 5.3-2. These ROC curves show that AlphaMOS FOX was able to discriminate between cancer and fibrosis patients from the non-cancerous group with high sensitivity and specificity. The sensitivity, specificity and AUC for colorectal cancer and hepatocellular cancer were 0.8 (0.67-0.92), 1 (1-1) and 0.99 (0.98-1). The separation between prostate cancer and fibrosis groups were high AUC at 0.99 (0.99-1), sensitivity 0.78 (0.5-1) and high specificity 1 (1-1). Similarly, the ROC curves for BCa and CRC, BCa and HCC, BCa and Fibrosis, CRC and PCa showed that AlphaMOS FOX was able to identify all these cancer groups correctly. These comparisons reported a sensitivity of 1 (1-1), specificity of 1 (1-1) and AUC of 1 (1-1).



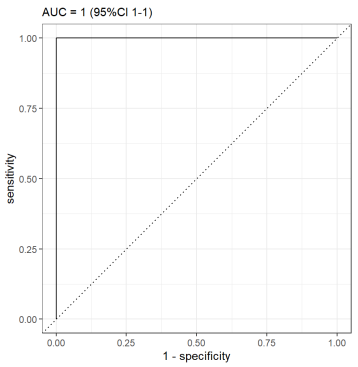
(a)



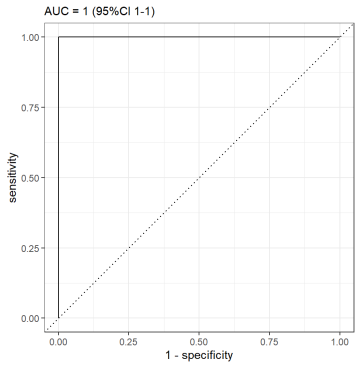
(b)



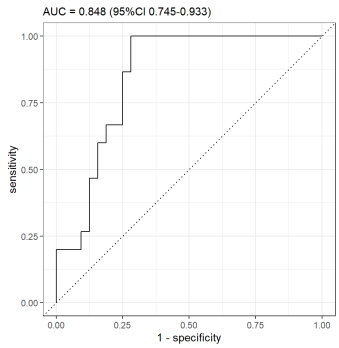
(c)



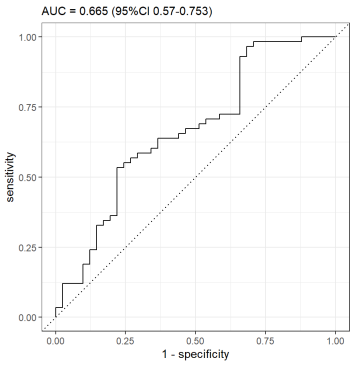
(d)



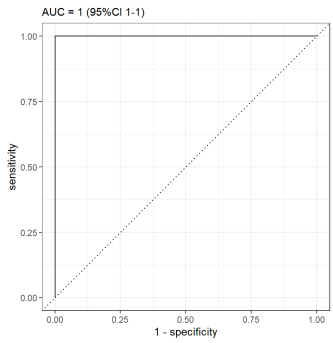
(e)



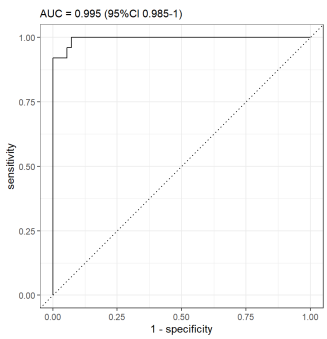
(f)



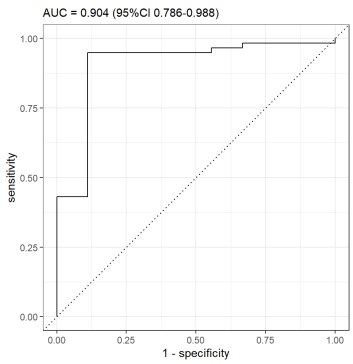
(g)



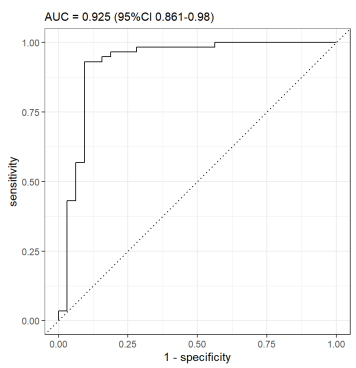
(h)



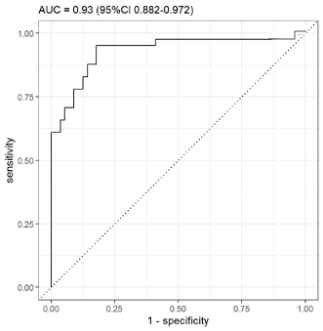
(i)



(j)



(k)



(l)

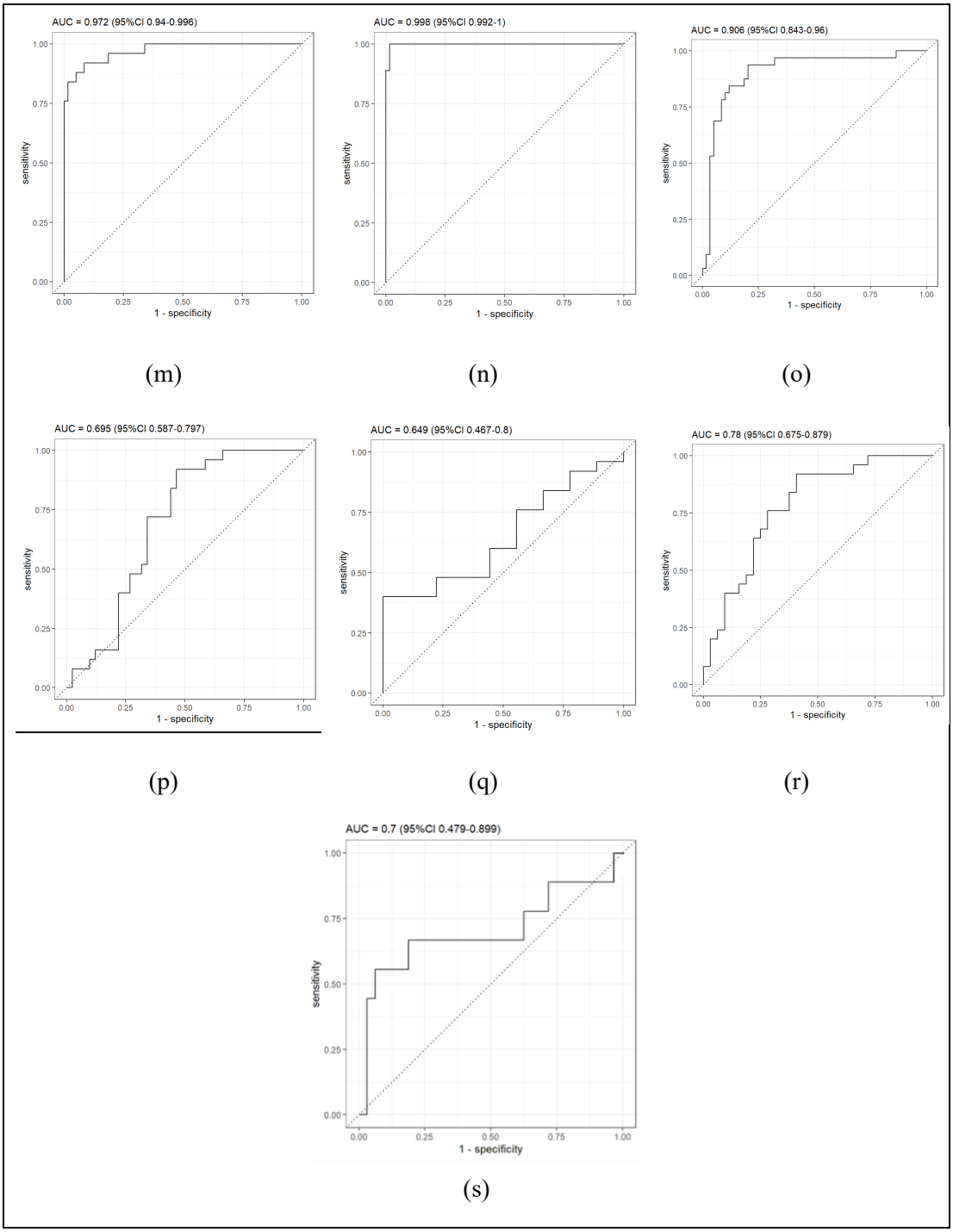


Figure 5.3-2: ROC curves obtained from the R-Program using GC-TOF-MS for different cancer groups where (a) Bladder cancer Vs non-cancerous; (b) Bladder cancer Vs colorectal cancer; (c) Bladder cancer Vs prostate cancer; (d) Bladder cancer Vs hepatocellular cancer; (e) Bladder cancer Vs fibrosis; (f) Bladder cancer Vs non-fibrosis; (g) Colorectal cancer Vs non-cancerous group; (h) Colorectal cancer Vs prostate cancer; (i) Colorectal cancer Vs hepatocellular cancer; (j) Colorectal cancer Vs fibrosis; (k) Colorectal cancer Vs non-fibrosis; (l) Prostate cancer Vs non-cancerous; (m) Prostate cancer Vs hepatocellular cancer; (n) Prostate cancer Vs fibrosis; (o)

Prostate cancer Vs non-fibrosis; (p) Hepatocellular cancer Vs non-cancerous; (q) Hepatocellular cancer Vs fibrosis; (r) Hepatocellular cancer Vs non-fibrosis and (s) Fibrosis Vs non-fibrosis.

The responses generated by all the sensors for different urine samples can be seen in Figure 5.3-3.

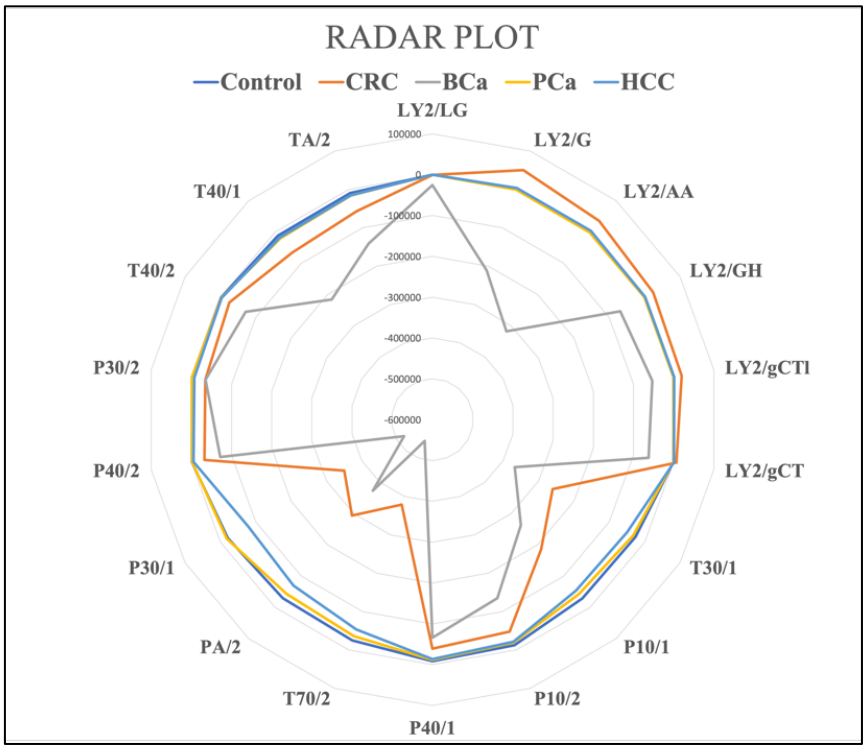


Figure 5.3-3: Response of AlphaMOS Fox 4000 sensors for different urine samples.

The radar plot show that each sensor in the Fox 4000 system responded to specific VOCs that were present in the different cancer urine samples. Analysis of the plot revealed that the sensors in the Fox 4000 exhibited the highest responses to the bladder cancer and the colorectal cancer samples, when compared to the other types of cancer samples. Among the sensors, LY2/GH, T30/1, P10/1, T70/2, PA/2, P30/1, and T40/1 demonstrated the most significant responses to the VOCs present in these cancer urine samples.

5.3.2. PEN3 Electronic Nose (Airsense Analytics GmbH, Schwerin, Germany):

The output of the statistical analysis for the data obtained from PEN3 system was analysed using MultiSens Analyzer and RStudio. The PCA plot obtained is shown in Figure 5.3-4.

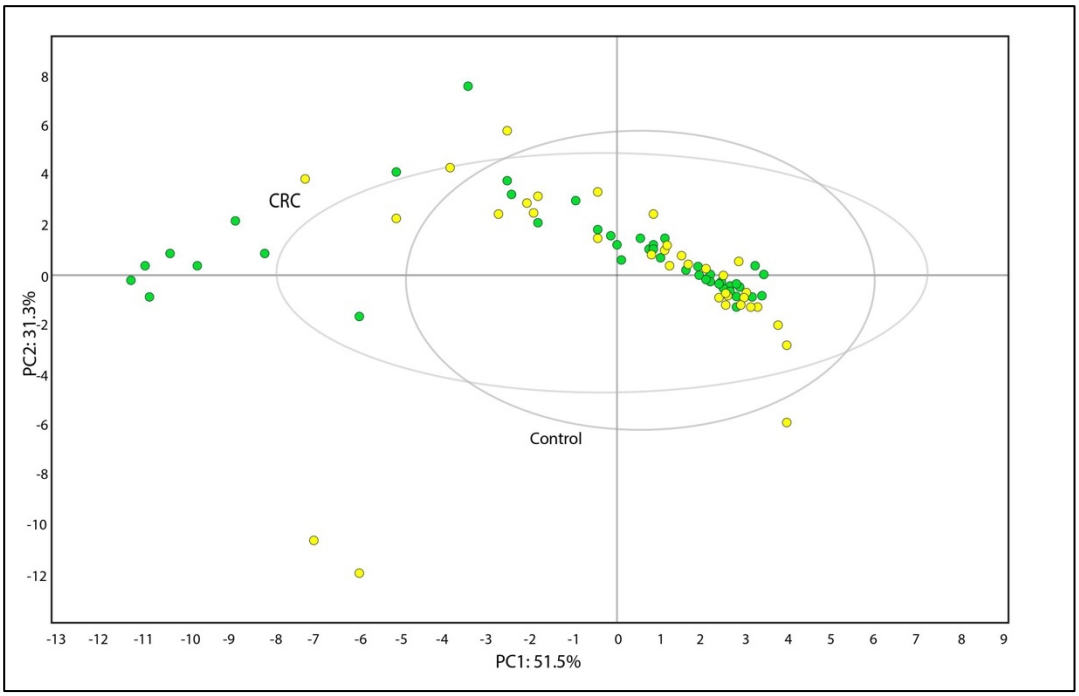


Figure 5.3-4: Principal Component Analysis of PEN3 eNose results where the green dots represent the features extracted from the prostate cancer samples and the yellow dots represent the features extracted from the control samples.

The output show that the PEN3 system was successfully able to distinguish different cancer and disease groups from each other. The statistical results are represented in Table 5.3-5, Table 5.3-6, Table 5.3-7 and Table 5.3-8. These tables display the top-performing classifiers out of the six introduced in Chapter 3, section e (Qualitative and Quantitative Analysis:, pg.: 62). Each of these classifiers was meticulously chosen to effectively distinguish between the classes for every comparison. To better understand the analysis pipeline, a thorough description is provided in Chapter 3, section c (Quantitative and qualitative analysis for AlphaMOS Fox 4000:, pg.: 83). The underlined entries in the tables draw attention to outputs that are particularly noteworthy, either exhibiting remarkably high separation or, conversely,

extremely poor separation. These specific entries highlight instances of exceptional classifier performance or challenges faced in differentiating between the compared classes. The underlining serves to emphasize these critical points of interest, indicating both remarkable successes and areas where further improvement may be necessary. For more comprehensive discussions about the underlined outputs in the tables, please refer to the "Discussion" section at the end of the chapter.

Table 5.3-5: PEN3 eNose Output for the comparison of BCa with different cancers and non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
BCa vs non-Cancerous	Random Forest	0.92 (0.83-0.99)	0.83 (0.64-1)	0.92 (0.84-0.98)	0.77 (0.56-0.94)	0.95 (0.88-1)
BCa Vs CRC	Random Forest	0.96 (0.91-0.99)	0.67 (0.43-0.9)	0.98 (0.95-1)	0.89 (0.69-1)	0.99 (0.88-0.98)
BCa vs PCa	Random Forest	0.75 (0.63-0.86)	0.17 (0-0.38)	0.89 (0.81-0.96)	0.25 (0-0.5)	0.83 (0.75-0.91)
BCa Vs HCC	Random Forest	0.82 (0.68-0.93)	0.67 (0.43-0.89)	0.79 (0.62-0.94)	0.67 (0.43-0.89)	0.79 (0.63-0.98)
BCa Vs Fibrosis	Gaussian Process classifier	0.76 (0.56-0.98)	0.75 (0.55-0.93)	0.38 (0.11-0.67)	0.64 (0.43-0.86)	0.5 (0.17-0.86)
BCa Vs Non-Fibrosis	Random Forest	0.89 (0.76-0.98)	0.83 (0.64-1)	0.93 (0.85-1)	0.83 (0.64-1)	0.93 (0.85-1)

Table 5.3-6: PEN3 eNose Output for the comparison of CRC with different cancers and non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
CRC Vs non-cancerous	Random Forest	0.75 (0.65-0.83)	0.81 (0.72-0.89)	0.58 (0.44-0.71)	0.75 (0.66-0.83)	0.67 (0.52-0.81)
CRC Vs PCa	Random Forest	0.97 (0.94-0.99)	0.95 (0.89-0.99)	0.94 (0.88-1)	0.95 (0.89-1)	0.94 (0.89-0.98)
CRC Vs HCC	Random Forest	0.87 (0.78-0.95)	0.95 (0.89-0.98)	0.58 (0.39-0.75)	0.87 (0.80-0.94)	0.79 (0.58-0.94)
CRC Vs Fibrosis	Logistic Regression	0.87 (0.78-0.95)	0.95 (0.89-0.98)	0.58 (0.39-0.75)	0.87 (0.8-0.94)	0.79 (0.58-0.94)

CRC Vs Non-Fibrosis	Random Forest	0.74 (0.64-0.83)	0.88 (0.8-0.95)	0.43 (0.29-0.59)	0.75 (0.66-0.83)	0.65 (0.47-0.82)

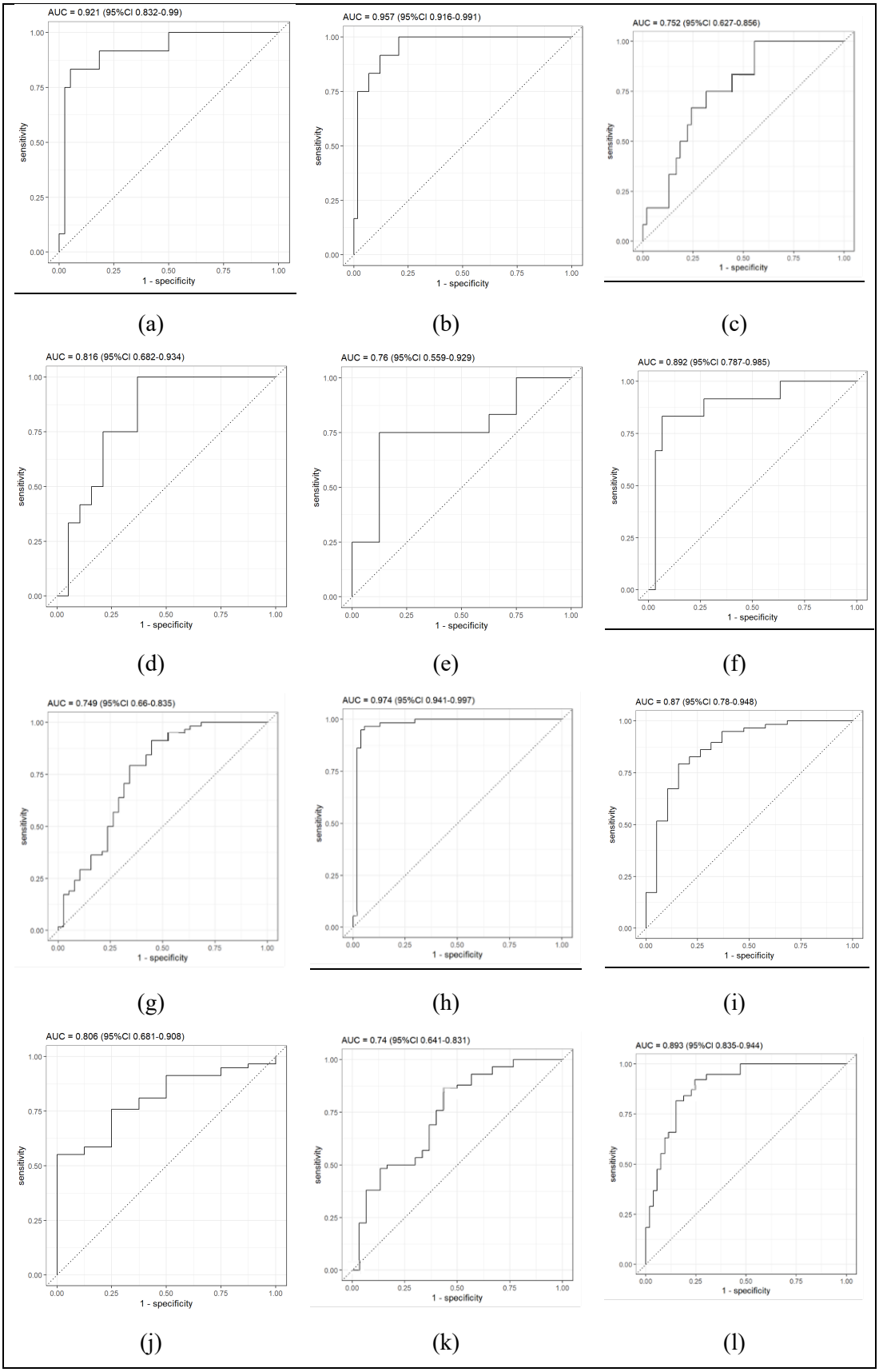
Table 5.3-7: PEN3 eNose Output for the comparison of PCa with different cancers and non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
PCa vs non-Cancerous	Random Forest	0.89 (0.83-0.94)	0.66 (0.53-0.79)	0.87 (0.79-0.94)	0.78 (0.66-0.89)	0.78 (0.69-0.87)
PCa Vs HCC	Random Forest	0.89 (0.82-0.96)	0.37 (0.19-0.56)	0.98 (0.95-1)	0.88 (0.67-1)	0.82 (0.74-0.89)
PCa Vs Fibrosis	Random Forest	0.86 (0.74-0.97)	0.13 (0-0.33)	0.98 (0.95-1)	0.5 (0-1)	0.88 (0.81-0.95)
PCa Vs Non-Fibrosis	Random Forest	0.87 (0.80-0.93)	0.57 (0.42-0.71)	0.87 (0.79-0.94)	0.71 (0.55-0.85)	0.78 (0.69-0.87)

Table 5.3-8: PEN3 eNose Output for the comparison of CRC with non-cancerous urine samples.

Comparisons	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
HCC Vs non-Cancerous	Logistic Regression	0.65 (0.51-0.77)	0.0 (0-0)	0.95 (0.88-1)	0.0 (0-0)	0.65 (0.55-0.75)
HCC Vs Fibrosis	Gaussian Process classifier	0.4 (0.18-0.65)	0.13 (0-0.33)	0.95 (0.85-1)	0.5 (0-1)	0.72 (0.58-0.87)
HCC Vs NON-Fibrosis	Gaussian Process classifier	0.71 (0.59-0.83)	0.68 (0.5-0.86)	0.63 (0.48-0.78)	0.54 (0.38-0.71)	0.76 (0.61-0.89)
Fibrosis Vs Non-Fibrosis	Neural Network	0.68 (0.51-0.84)	0 (0-0)	0.97 (0.91-1)	0 (0-0)	0.78 (0.66-0.89)

The PEN3 system shows promising results in discriminating the cancer groups, disease groups and non-cancerous groups. The ROC curves for the eNose are shown in Figure 5.3-5. The results indicate that BCa, and non-cancerous group had a sensitivity and specificity of 0.83 (0.64-1) and 0.92 (0.84-0.98) respectively indicating that the PEN3 system recognised 13 BCa samples correctly out of 15 and 35 non-cancerous samples out of 38. Similarly, for the comparison of the PCa and the non-fibrosis group, the PEN3 system was able to recognise 48 out of 55 PCa samples and 13 out of 38 non-fibrosis samples.


AUC = 0.921 (95%CI 0.832-0.99)
AUC = 0.957 (95%CI 0.916-0.991)
AUC = 0.752 (95%CI 0.627-0.856)
AUC = 0.816 (95%CI 0.682-0.934)
AUC = 0.76 (95%CI 0.559-0.929)
AUC = 0.892 (95%CI 0.787-0.985)
AUC = 0.749 (95%CI 0.66-0.835)
AUC = 0.974 (95%CI 0.941-0.997)
AUC = 0.87 (95%CI 0.78-0.948)
AUC = 0.806 (95%CI 0.681-0.908)
AUC = 0.74 (95%CI 0.641-0.831)
AUC = 0.893 (95%CI 0.835-0.944)
sensitivity
1 - specificity
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)

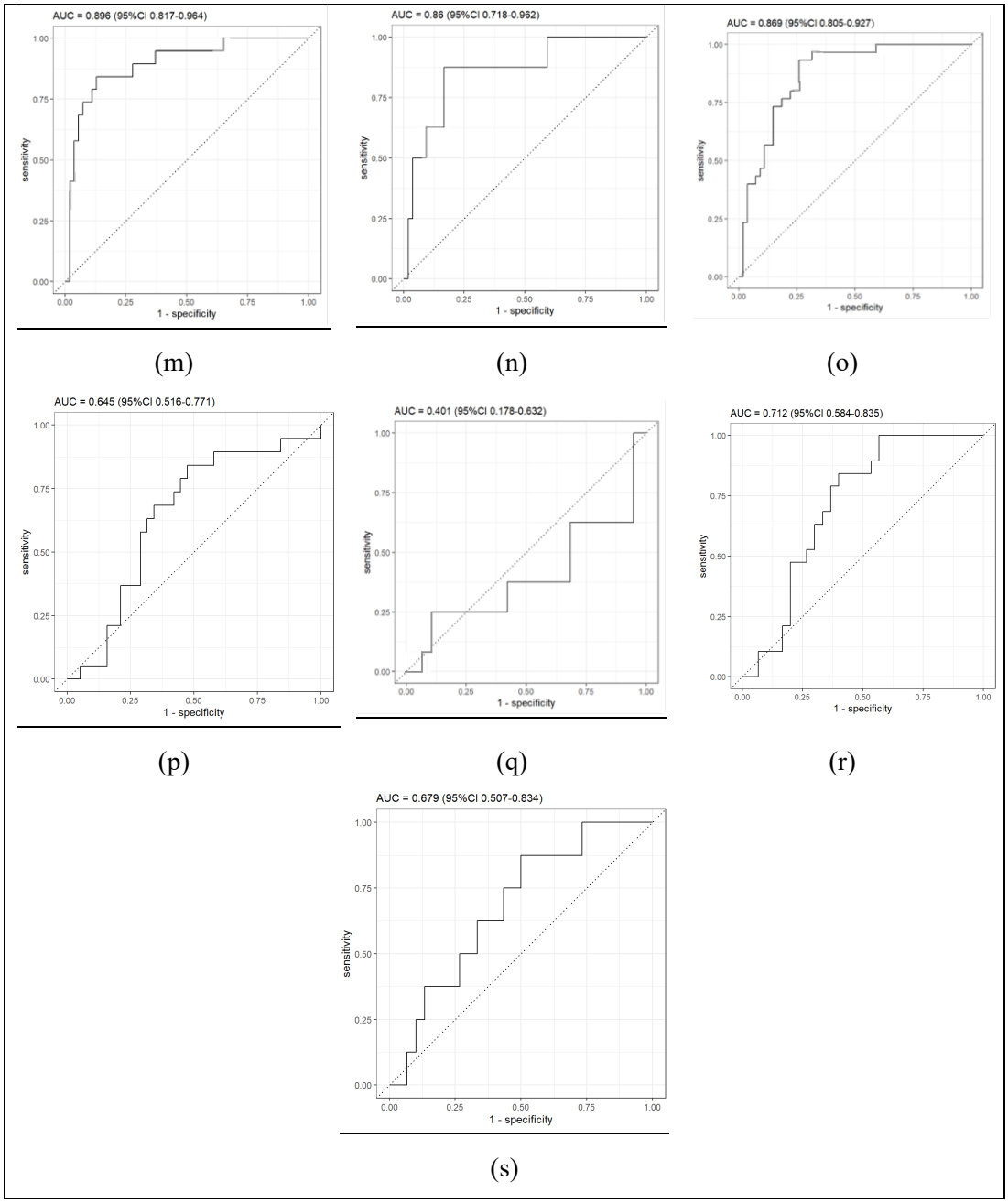


Figure 5.3-5: ROC curves obtained from the R-Program using the PEN3 system for different cancer groups where (a) Bladder cancer Vs non-cancerous; (b) Bladder cancer Vs colorectal cancer; (c) Bladder cancer Vs non-fibrosis; (d) Colorectal cancer Vs non-cancerous group; (e) Colorectal cancer Vs prostate cancer; (f) Colorectal cancer Vs non-fibrosis; (g) Prostate cancer Vs non-cancerous; (h) Prostate cancer Vs non-fibrosis; (i) Hepatocellular cancer Vs non-cancerous; (j) Hepatocellular cancer Vs fibrosis; (k) Fibrosis Vs non-fibrosis.

The response generated by the senor array of the PEN3 system is represented in Figure 5.3-6 as radar plot.

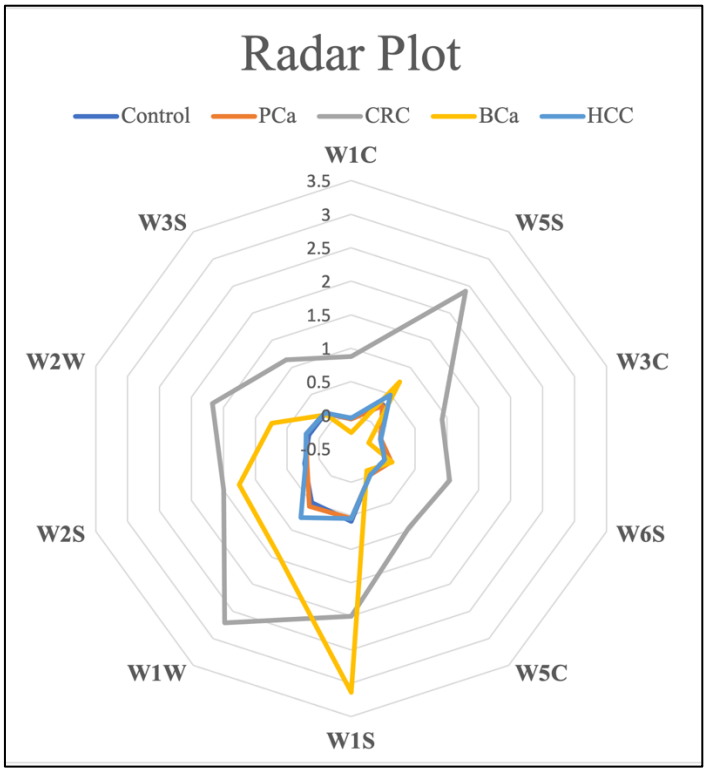


Figure 5.3-6: Radar plot for the response of PEN3 eNose sensors for different urine samples.

The output generated by the PEN3 eNose in response to urine samples from patients with bladder, prostate, colorectal and hepatocellular cancer was visualized using the radar plot shown above. The plot showed that the PEN3 eNose sensors exhibited the highest reactivity to BCa and CRC urine samples. Specifically, the sensors with the greatest intensity readings were W5S, W1S, W1W, and W2S. These results suggest that the PEN3 eNose has the potential to detect differences in the chemical composition of urine samples from patients with different types of cancer, which may be useful for diagnostic or screening purposes.

5.4. Discussion:

The study aimed to evaluate the diagnostic potential of eNoses namely, the AlphaMOS FOX 4000 and PEN3 system, in discriminating between the cancer patients and healthy controls based on the VOCs detected in the urine samples. The results suggest that the eNose could be a promising tool for non-invasive cancer diagnosis, although further studies are needed to validate these findings. The limitations of the study include a small sample size and potential confounding factors such as age and gender.

In the comprehensive study undertaken that focused on distinguishing between cancerous and non-cancerous conditions, it was observed that the Random Forest classifier consistently yielded the most favourable results. This outcome can be attributed to several intrinsic characteristics of the Random Forest algorithm that contribute to its superior performance in this context. Random Forest operates by constructing an ensemble of decision trees, where each tree is trained on a random subset of the dataset and makes its predictions independently. This approach addresses the issue of overfitting, which can be prevalent when dealing with complex and high-dimensional medical datasets. The ensemble nature of Random Forest reduces the risk of capturing noise or idiosyncrasies present in individual samples, thereby enhancing its generalization capability to new, unseen data [28].

The application of PCA on the AlphaMOS FOX 4000 eNose data, shown in Figure 5.3-1, has provided valuable insights into the variance and patterns present in the sensor responses. The PCA analysis results obtained from the AlphaMOS FOX 4000 eNose, shown in Figure 5.3-1, reveal that PC1 accounted for 77.6% of the total variance in the data, while PC2 explained 17.3% of the variance. PCA is a powerful statistical technique used to reduce the dimensionality of high-dimensional data while retaining as much variability as possible. In our study, we utilized the AlphaMOS FOX 4000 eNose to differentiate between BCa samples and CRC samples.

The high PC1 value of 77.6% indicates that the first principal component captured the most significant discriminatory features in the VOC profiles of the samples. This suggests that the FOX 4000 eNose successfully distinguished between BCa and CRC samples based on the dominant patterns in the VOC data, contributing significantly to the separation of these two groups. Additionally, the PC2 value of 17.3% reveals that the second principal component identified additional discriminatory information in the VOC profiles, further aiding in the differentiation of BCa and CRC samples.

When considering the cumulative variance explained by PC1 and PC2 (94.9%), it becomes evident that these two principal components contain the majority of the discriminatory power in the VOC data. This indicates that the FOX 4000 eNose demonstrated high accuracy in distinguishing between BCa and CRC samples based on the VOC profiles of the urine samples.

The radar plot shown in Figure 5.3-3 provides valuable insights into the sensitivity of each sensor to specific VOCs associated with different cancer types. Upon analysing the radar plot, we observed that the sensors in the FOX 4000 system exhibited varying responses to the VOCs present in the different cancer urine samples. Notably, the sensors showed the highest responses to the Bladder Cancer (BCa) and Colorectal Cancer (CRC) samples compared to other types of cancer samples and the non-cancerous group. Specifically, the sensors LY2/GH, T30/1, P10/1, T70/2, PA/2, P30/1, and T40/1 demonstrated the most significant responses to the VOCs present in the BCa and CRC urine samples. These sensors appeared to be highly sensitive to the specific chemical compounds or VOCs profiles associated with these cancer types, leading to distinguishable and robust responses in the radar plot. The sensitivity of the identified sensors to BCa and CRC urine samples underscores the importance of sensor selection in eNose analysis. The modular design of the FOX 4000 system allows for flexibility in choosing the most appropriate sensors, which proved beneficial in capturing the unique VOCs patterns associated with specific cancer types. A study conducted by Saary *et al.* showed the importance of visualising the multi-variate data and delivering its meaning using radar plots [3].

Table 5.3-1 displays the outcomes of various classifiers employed to assess the capability of VOCs in distinguishing between different cancer types and healthy controls. The performance of the classifiers was evaluated based on different parameters such as AUC, sensitivity, specificity, PPV, and NPV.

Interestingly, some comparisons achieved perfect results as 100%, such as the Neural Network classifier for the comparison between BCa and CRC, BCa and HCC, and the comparison between BCa and fibrosis that meant that the classifier achieved a perfect performance. When both sensitivity and specificity were reported as 100%, this means that the classifier had correctly identified all true positives and true negatives, respectively.

However, perfect classifier performance can be influenced by the distribution of samples across different classes. When the dataset is imbalanced, as observed in the BCa and CRC sample comparison, one class may significantly outweigh the others. Consequently, the classifier might become biased towards the majority class, leading to high accuracy, specifically for that class [4, 5]. Additionally, when dealing with a relatively small dataset,

there is a higher risk of overfitting. If the dataset is limited in size and contains only a few samples for each class, then the classifier may memorize the training data and exhibit exceptional performance on it. Nevertheless, it might struggle to generalize effectively to new, unseen data [6, 7]. This situation might potentially account for the 100% results achieved in the comparison of BCa and HCC samples, as well as in the comparison between BCa and fibrosis samples.

Similarly, in the case of CRC vs non-cancerous, the neural network classifier achieved an AUC of 0.67, indicating moderate performance, while sensitivity was relatively high at 0.78, suggesting that the classifier correctly identified a significant proportion of positive cases as shown in Table 5.3-2. However, specificity was low at 0.34, meaning that the classifier had a relatively high false positive rate. In contrast, for CRC vs PCa, the neural network classifier achieved perfect performance across all parameters, including AUC, sensitivity, specificity, PPV, and NPV, indicating excellent discrimination between the two conditions. Similarly, in CRC vs HCC, the neural network classifier achieved an AUC of 0.99, and a sensitivity of 0.8, but with perfect specificity and PPV, indicating that it could distinguish between these two conditions with high accuracy. For PCa vs HCC and Fibrosis as shown in Table 5.3-3, the Neural Network achieved an AUC of 0.97 and 0.99, respectively, indicating excellent diagnostic ability. The sensitivity, specificity, PPV, and NPV values were all high for these comparisons, suggesting that the model was effective in identifying PCa from HCC and Fibrosis samples. Finally, as shown in Table 5.3-4, HCC and non-fibrosis comparison resulted in an AUC of 0.78, along with a sensitivity of 0.64 and specificity of 0.75. A similar study conducted by Esfahani *et al.* [8] found that the AlphaMOS FOX 4000 eNose had an accuracy of 86.4% in discriminating between urine samples from 69 diabetes patients and 67 healthy controls.

Similarly, Figure 5.3-4 shows the PCA plot obtained for identifying the most significant patterns and variations in the data obtained from the PEN3 eNose. The PCA analysis of the PEN3 eNose with PC1 explaining 51.5% of the variance and PC2 explaining 31.3% of the variance provides valuable insights into the ability of the PEN3 eNose instrument to distinguish between PCa samples and non-cancerous samples based on their VOCs patterns.

The PC1 value of 51.5% indicates that the first principal component captured 51.5% of the total variance in the data. PC1 represents the direction in the data with the highest variance. Therefore, this means that the PEN3 eNose was able to extract key information from the VOC profiles of the samples, allowing it to differentiate between different PCa and non-cancerous samples, as reflected by the variance captured by PC1. Similarly, the PC2 value of 31.3% indicates that the second principal component captured 31.3% of the total variance in the data. PC2 represents a direction orthogonal to PC1 that explains the second-highest amount of variance in the data. The fact that PC2 captures a substantial portion of the variance suggests that the PEN3 eNose was able to further discriminate between different PCa samples and non-cancerous samples using additional features from the VOC profiles.

The cumulative variance explained by PC1 and PC2 (51.5% + 31.3% = 82.8%) suggests that these two principal components contain the majority of the discriminatory information present in the VOC data. This implies that the PEN3 eNose has the capability to differentiate between PCa samples and non-cancerous samples with a significant level of confidence, as evidenced by the variance captured by the first two principal components. The remaining 17.2% of the variance may encompass less influential variations or noise that are not as critical in distinguishing between PCa samples and non-cancerous samples. While not as influential as PC1 and PC2, still contributes to the overall characterization of the data and might contain some subtler patterns or variations. This remaining variance represents information that is orthogonal to (i.e., not aligned with) the PC1-PC2 plane.

Overall, the PCA analysis results showed that the AlphaMOS FOX 4000 eNose exhibited higher variance in the first and second principal components (PC1: 77.6% and PC2: 17.3%) in comparison to the PEN3 eNose (PC1: 51.5% and PC2: 31.3%). This suggests that the AlphaMOS eNose was able to capture more significant discriminatory features in the VOC data compared to the PEN3 eNose, which might have lower variance in its PCA analysis.

The results for the statistical analysis shown in Table 5.3-5, Table 5.3-6, Table 5.3-7, and Table 5.3-8 represent output obtained by the PEN3 system from distinguishing between various types of cancer and non-cancerous samples. In the case of distinguishing between BCa and non-cancerous conditions, the Random Forest classifier achieved an AUC of 0.92, with a sensitivity of 0.83 and specificity of 0.92. However, in the comparison of BCa vs PCa

samples, the Random Forest classifier demonstrated a moderate level of discrimination ability in distinguishing between BCa and PCa samples, with an AUC value of 0.75. The sensitivity value of 0.17 indicates that the Random Forest classifier had a relatively low ability to correctly identify BCa samples among all the BCa samples present in the dataset, while the specificity value of 0.89 suggests that the classifier had a high ability to correctly identify PCa samples among all the PCa samples in the dataset. This discrepancy in performance could be attributed to the imbalanced distribution of samples between the two classes (BCa and PCa). Such an imbalance can lead to biased model training, causing the classifier to become more specialized in the majority class (PCa) and perform poorly on the minority class (BCa). Another possible reason is that the BCa and PCa samples in the dataset might have overlapping distributions in the feature space, presenting a challenge for the classifier to establish clear boundaries between the two classes. This overlap can hinder the classifier's sensitivity, making it difficult to correctly identify samples from the minority class. Therefore, these factors could collectively contribute to the observed limitations in the Random Forest classifier's performance in the BCa vs PCa comparison.

Meanwhile, the PEN3 system showed strong performances in distinguishing between CRC samples and other cancers as well as non-cancerous samples, as evidenced by high sensitivity, specificity, PPV, NPV, and AUC values shown in Table 5.3-6. However, we can see in Table 5.3-7, that when distinguishing between PCa vs HCC samples and PCa vs Fibrosis samples, the Random Forest classifier demonstrated a relatively moderate discrimination ability. Although the high values of AUC of 0.89 and 0.86 for both PCa vs HCC samples, and PCa vs Fibrosis samples, respectively, suggested relatively good discrimination ability, the low sensitivity values of 0.37 and 0.13, respectively, indicated that that the Random Forest classifier had only a moderate ability to correctly identify Prostate Cancer (PCa) samples among all the PCa samples present in the dataset.

The performance metrics for two different comparisons: HCC vs. Non-Cancerous samples, and Fibrosis vs. Non-Fibrosis samples presented in Table 5.3-8 indicate that both the Logistic Regression and Neural Network classifiers exhibited poor performance in distinguishing between HCC and Non-Cancerous samples and between Fibrosis and Non-Fibrosis samples. The classifiers had low sensitivity, which means they failed to detect positive cases correctly. The high specificity indicates that they performed well in

identifying true negatives. The low PPV and NPV values indicate a high number of false positives and false negatives, respectively, leading to the classifiers' limited ability to accurately classify samples into their respective categories. Additionally, the 95% confidence intervals provide a range of likely values for each metric, considering the uncertainty in the estimates due to the limited sample size. In a similar study, the PEN3 eNose was used to analyse breath samples from patients with colorectal cancer and healthy controls, and achieved a high sensitivity of 93.33%, which means that it correctly identified most of the CRC patients. However, the specificity was very low at only 10%, which means that the PEN3 eNose had a high false-positive rate, misidentifying many healthy individuals as CRC patients [9].

The findings of the research demonstrate the ability of eNose system to distinguish between different cancer types and healthy controls using urinary VOCs. The perfect performance achieved for certain comparisons is particularly significant, indicating that it correctly identified all true positives and true negatives. While the classifier's performance was only moderate or low for the other comparisons. There are several factors that can influence the performance of an eNose system for distinguishing between cancer samples and healthy controls. One possible factor is the variability in the composition of the VOCs in different cancer types, which can affect the ability of the eNose to accurately detect and differentiate between them. Additionally, the size and diversity of the sample population used to train and test the classifiers can also impact their performance. If the sample size is too small, the classifiers may not have enough data to accurately learn the patterns in the VOCs detected by an eNose, resulting in lower performance [10, 11]. Overall, the performance of the eNose depends on a range of factors and may vary depending on the specific cancer type and the characteristics of the sample population.

However, the AlphaMOS FOX 4000 and PEN3 eNose systems, both employed for gas sensing and analysis, exhibit distinctive features that set them apart. Notably, the AlphaMOS FOX 4000 features a significantly larger sensor array comprising 18 sensors, as opposed to the PEN3's 10-sensor array. This augmented sensor array confers a greater capacity for detecting and analysing a diverse range of volatile organic compounds (VOCs), thereby enhancing its potential for comprehensive gas sensing and analysis. Moreover, the modular design inherent to the AlphaMOS FOX 4000 endows users with the flexibility to selectively

choose and customize sensors tailored to their specific application requirements. This adaptability amplifies the eNose's versatility, enabling it to cater for a broad spectrum of analytical needs with precision.

Another notable distinction is the substantial difference in sensitivity levels between the two devices. The AlphaMOS FOX 4000 exhibits exceptional sensitivity, capable of detecting minute VOC concentrations within the parts per billion (ppb) range. In contrast, the PEN3 eNose displays limited sensitivity, with a lower limit of detection ranging from 0.1 to 5 parts per million (ppm). This discrepancy in sensitivity confers a considerable advantage to the AlphaMOS FOX 4000, rendering it better suited for applications necessitating high sensitivity and accurate detection of low-concentration VOCs.

In summary, the comprehensive data analysis supports the contention that the AlphaMOS FOX 4000 eNose surpasses the PEN3 eNose in various critical aspects, including sensor array size, versatility, and sensitivity. Consequently, the AlphaMOS FOX 4000 emerges as the superior choice for applications wherein extensive gas sensing and analysis capabilities are essential for precise and reliable results.

Overall, these results indicate that the AlphaMOS FOX 4000 and PEN3 systems have high potential for use for the identification of cancer samples and offer several advantages for cancer detection, including non-invasiveness, rapid detection, portability, high sensitivity and specificity, and the ability to detect cancer at an early stage. While the high specificity and sensitivity methods achieved in some comparisons are promising, it is important to continue to refine and optimize the technology for more accurate and reliable results. These findings suggest that VOCs analysis combined with machine learning algorithms, particularly neural networks, may have a promising potential for the diagnosis and discrimination of different cancer types from healthy controls. Overall, the study highlights the potential of eNose technology in cancer diagnosis and warrants further research in this area.

5.5. Conclusion:

In conclusion, our findings suggest that these methods have the potential to be developed as non-invasive and cost-effective tools for the detection of various cancers and diseases. It should be noted that this is the first study to utilize the AlphaMOS FOX 4000 to compare bladder cancer urine samples with non-cancerous samples and the first to employ a PEN3 eNose for testing urine samples from patients with CRC and PC.

5.6. References:

[1] S. Dragonieri *et al.*, "An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD," *Lung Cancer,* vol. 64, no. 2, pp. 166-170, 2009/05/01/ 2009, doi: https://doi.org/10.1016/j.lungcan.2008.08.008.

[2] G. Peng *et al.*, "Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors," in *British Journal of Cancer* vol. 103, ed, 2010, pp. 542-551.

[3] M. J. Saary, "Radar plots: a useful way for presenting multivariate health care data," *Journal of Clinical Epidemiology,* vol. 61, no. 4, pp. 311-317, 2008/04/01/ 2008, doi: https://doi.org/10.1016/j.jclinepi.2007.04.021.

[4] F. Thabtah, "An accessible and efficient autism screening method for behavioural data and predictive analyses," *Health informatics journal,* vol. 25, no. 4, pp. 1739-1755, 2019.

[5] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou, "On the Class Imbalance Problem," in *2008 Fourth International Conference on Natural Computation*, 18-20 Oct. 2008 2008, vol. 4, pp. 192-201, doi: 10.1109/ICNC.2008.871.

[6] G. D. Magoulas and A. Prentza, "Machine learning in medical applications," in *Advanced course on artificial intelligence*: Springer, 1999, pp. 300-307.

[7] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Computers in Biology and Medicine,* vol. 112, p. 103375, 2019/09/01/ 2019, doi: https://doi.org/10.1016/j.compbiomed.2019.103375.

[8] 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, doi: 10.3390/bios8040121.

[9] D. F. Altomare *et al.*, "The use of the PEN3 e-nose in the screening of colorectal cancer and polyps," *Techniques in Coloproctology,* vol. 20, no. 6, pp. 405-409, 2016/06/01 2016, doi: 10.1007/s10151-016-1457-z.

[10] M. Tirzïte, M. Bukovskis, G. Strazda, N. Jurka, and I. Taivans, "Detection of lung cancer with electronic nose and logistic regression analysis," *Journal of Breath Research,* vol. 13, no. 1, p. 016006, 2018/11/20 2019, doi: 10.1088/1752-7163/aae1b8.

[11] A. D. Wilson, "Review of Electronic-nose Technologies and Algorithms to Detect Hazardous Chemicals in the Environment," *Procedia Technology,* vol. 1, pp. 453-463, 2012/01/01/ 2012, doi: https://doi.org/10.1016/j.protcy.2012.02.101.

Chapter 6. Analysis of Infectious Disease VOCs:

This chapter includes the analysis of urine samples of infectious disease subjects, namely Urinary Tract Infections (UTIs). The chapter gives detailed information on the background of UTI diseases and the effects of UTI on the population, especially women. The present chapter explains the methodology of the process, as well as the analytical instruments utilized in the course of this study. Specifically, the investigation employed two GC-IMS instruments - the G.A.S. FlavourSpec GC-IMS and the Silox GC-IMS (IMSPEX, U.K.) - as well as two electronic noses (eNose), namely the AlphaMOS FOX 4000 and PEN3. Moreover, this chapter incorporates the outcomes of statistical analyses based on data gleaned from the aforementioned instruments, culminating in a comprehensive discussion.

6.1. Introduction:

Urinary Tract Infections (UTIs) are bacterial infections of the urinary tract and are considered to be amongst the most prevalent infections. It has been observed that the incidence of UTIs is higher in women as compared to men, with an estimated 50% to 60% of adult women experiencing a UTI at least once during their lifetime [1, 2]. UTI infections can be divided into two categories: uncomplicated and complicated UTIs. Uncomplicated UTIs occur in healthy patients with properly functioning urinary tracts whereas complicated UTIs result from abnormal functioning of urinary tract [3-5]. UTI treatment is known to be financially demanding and can impose significant social and personal burdens [6]. At present, commonly employed diagnostic techniques for UTI detection include Dipstick urinalysis, urine culture, among others. Despite its low cost, Dipstick urinalysis has been found to exhibit a low level of specificity[7]. On the other hand, urine cultures demonstrate high specificity. However, the analysis time can be up to 24 to 48 hours [8, 9]. The analysis of VOCs can be used for the identification and diagnosis of the disease but due to its complicated and complex nature, it is important that the analysis is precise and efficient.

In this chapter, we aim to analyse the UTI samples with two different GC-IMS instruments and two different electronic nose devices and investigate the use of VOCs for the diagnosis and detection of UTIs in comparison to healthy controls.

6.2. Materials and methods:

6.2.1. Sample design:

The samples for this study were collected and cultured at University Hospital Coventry and Warwickshire (UHCW) to confirm the disease. The samples were stored at -80°C in universal sample containers according to standard operating procedures compliant with tissue bank requirements under Human Tissue Act 2004 and were transferred to the University of Warwick in three batches where they were stored at -20°C. The samples were then defrosted at room temperature in a fume cupboard before testing. 5 ml of each sample was then transferred into 20 ml glass vials which were sealed with Blue BiMetal Seal, 3mm Thick PTFE/Blue Silicone, Ultra Low Bleed crimp caps (Thames Restek, FI5150BMUL-20B) using an appropriate crimp tool. These samples were then analysed by different analytical instruments as discussed in this study. The demographic data of the subjects are illustrated in Table 6.2-1.

Table 6.2-1: Demographic data for subject groups.

N= 81	**Positive Samples**	**Negative Samples**	**Samples with <10 CFU/ml**	**Mixed Growth**
No. of samples	9	56	13	3

6.2.2. Analytical instruments:

The analytical instruments used in this study were a G.A.S. FlavourSpec GC-IMS, Silox GC-IMS, an AlphaMOS FOX 4000 and a PEN3 system. The working and the data analysis methods for the FlavourSpec, AlphaMOS FOX 4000 and PEN3 device are presented in chapter 3. The details for the Silox GC-IMS are discussed below.

a. Silox GC-IMS (IMSPEX, UK):

The operating principle of a Silox GC-IMS device is similar to the FlavourSpec. The Silox GC-IMS contained a 30 m long column which provides better selectivity for light molecules. The Silox GC-IMS consist of an integrated temperature and flow controller which generates better separation of highly volatile compounds and a shorter run time.

For analysis, the samples were collected and analysed at the hospital within hours after collection. After collecting the samples from the patients, they were stored in a fridge at 4^{o}C. The samples were then transferred to 20 ml glass vials each containing a 5 ml sample. The sample was heated for 10 minutes at 40°C using a dry-block heater. The urine samples for analysis were introduced to the Silox device using a 21g needle which was attached to the input port of the Silox. The needle was then inserted into the headspace of a glass vial containing the urine sample at least 1 cm above the urine. The needle was held in place for 6 seconds. This allowed the VOCs present in the headspace of the urine sample to enter and be analysed by the GC-IMS. The carrier gas used for the Silox was nitrogen, the carrier gas flowrate was set to 150 ml/min and the sample flowrate was 20 ml/min. The Silox was set to run for 5 min for every sample with a heating temperature set to 80^{o}C.

6.2.3. Result:

a. GC-IMS Results:

The urine samples were characterised as negative and positive UTI samples based on the confirmation obtained after culturing them. These samples were analysed using two GC-IMS instruments and the output from the two GC-IMS instruments used in this study. Figure 6.2-1 shows the chromatogram obtained from the GC-IMS FlavourSpec, which displays the raw data. The chromatogram acquired from a negative UTI sample exhibits a relatively flat baseline, with minimal or no significant peaks that suggest the absence of UTI-causing VOCs. In contrast, a chromatogram obtained from a positive UTI sample reveals several prominent peaks that indicate the presence of UTI-causing VOCs. These peaks are believed to correspond to the specific VOCs produced by the bacteria responsible for the infection.

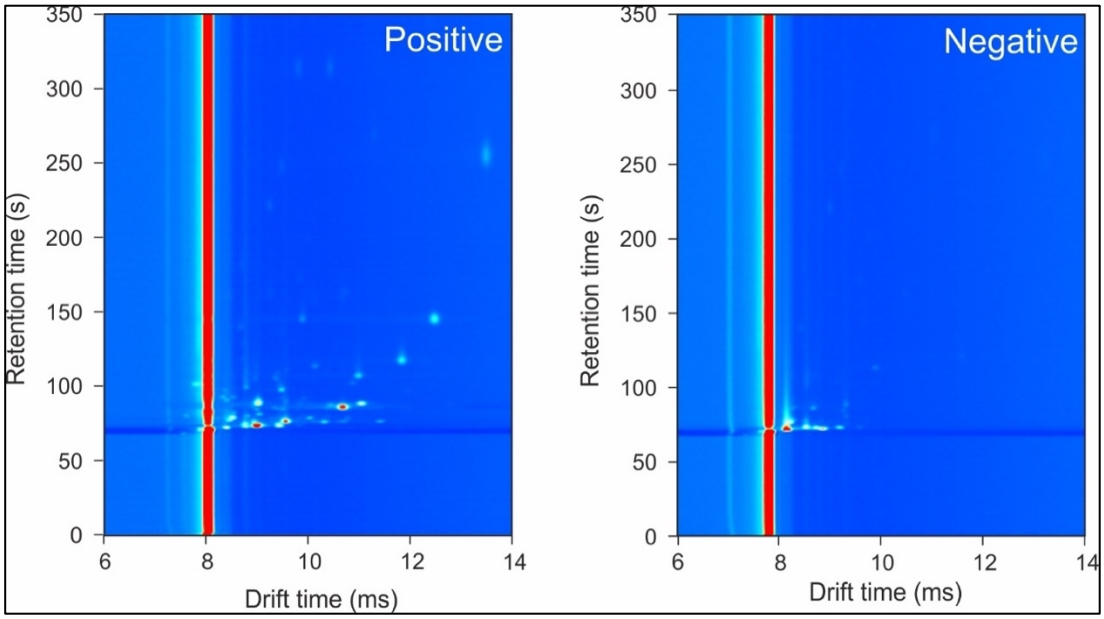


Figure 6.2-1: represents the output of the GC-IMS and its ability to differentiate between negative UTI samples and positive UTI samples.

The output was further quantified using the data analysis process as described in chapter 3 (Qualitative and Quantitative Analysis:) and this is shown in Table 6.2-2 for negative UTI sample and positive UTI samples.

Table 6.2-2: Statistical output from the FlavourSpec GC-IMS and the Silox GC-IMS for the comparison between negative UTI samples and positive UTI samples.

Instruments	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
FlavourSpec GC-IMS	Extreme Gradient Boosting	0.571 (0.459-0.681)	0.947 (0.907-0.981)	0.076 (0-0.179)	0.789 (0.726-0.851)	0.286 (0-0.615)
Silox GC-IMS	Extreme Gradient Boosting	0.885 (0.819-0.945)	0.968 (0.935-0.99)	0.5 (0.345-0.667)	0.876 (0.821-0.927)	0.813 (0.643-0.952)

The statistical analysis showed that the Silox GC-IMS was able to differentiate between negative and positive samples with much better accuracy in comparison to the FlavourSpec GC-IMS. Though the sensitivity achieved from FlavourSpec GC-IMS was much higher with 94.7%, it failed identify positive urine samples and resulted in poor specificity of only 7.6%.

On the other hand, the Silox GC-IMS showed remarkably high sensitivity with 96.8% and it successfully identified 50% of the positive samples. Overall, the results obtained from the Silox showed an AUC of 88.5% whereas the FlavourSpec resulted in 57.1% for the AUC value.

The ROC curves generated by the bespoke R program from the two instruments are shown in Figure 6.2-2.

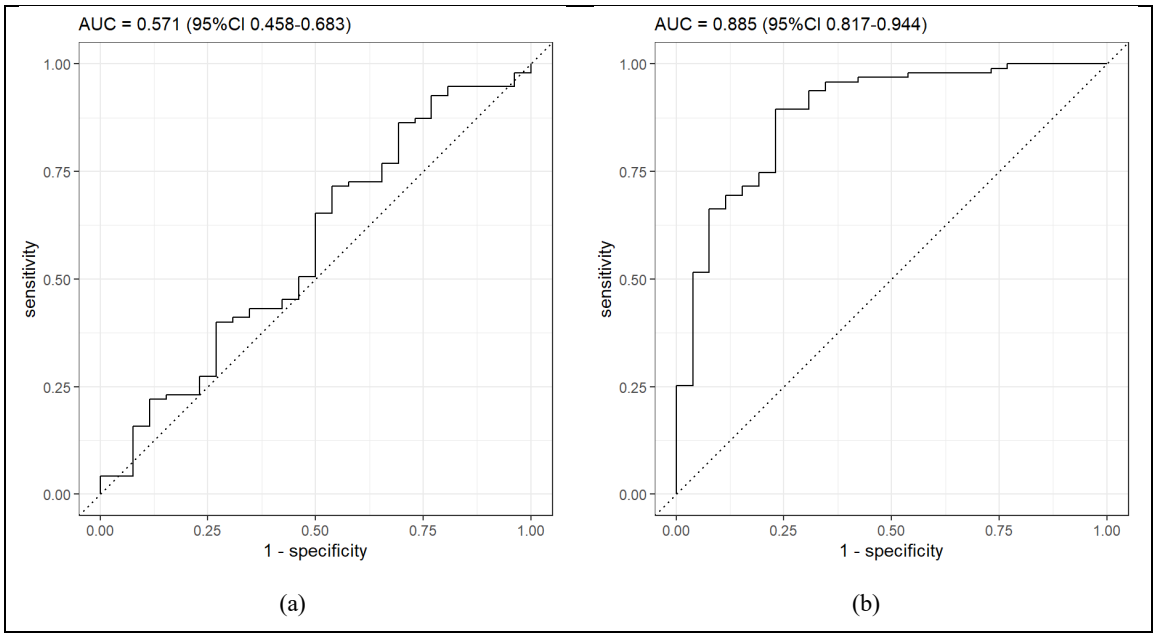


Figure 6.2-2: represents the output of the GC-IMS to differentiate negative UTI samples and positive UTI samples. ***(a)*** *represents the output generated by the FlavourSpec GC-IMS and,* ***(b)*** *represents the output generated by the Silox GC-IMS.*

b. eNose Results:

There were two eNoses used in this study and the outputs obtained from the two eNoses, the AlphaMOS FOX 4000 and PEN3, were analysed using the data analysis process described in chapter 3 (Quantitative and qualitative analysis for AlphaMOS Fox 4000: and Quantitative and qualitative analysis for PEN3 eNose:). The statistical output generated is shown in Table 6.2-3 for the comparison of negative and positive UTI samples.

Table 6.2-3: Statistical output from the AlphaMOS FOX 4000 eNose and the PEN3 eNose for the comparison between negative UTI samples and positive UTI samples.

Instruments	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
AlphaMOS FOX 4000	Random Forest	0.885 (0.769-0.978)	0.982 (0.946-1)	0.667 (0.375-0.909)	0.948 (0.896-0.983)	0.857 (0.6-1)
PEN3	Random Forest	0.675 (0.433-0.905)	1 (1-1)	0.444 (0.167-0.727)	0.911 (0.842-0.965)	1 (1-1)

Out of the two eNoses, the results show that FOX 4000 was able to distinguish between the two UTI groups much better than PEN3 eNose. The results obtained from FOX 4000 showed remarkably high sensitivity of 98.2% and high specificity of 66.7%, whereas the PEN3 eNose showed 100% sensitivity but low specificity of 44.4%. Overall, the FOX 4000 demonstrated a high AUC of 88.5% and PEN3 eNose demonstrated a moderate AUC of 67.5%. Further discussion regarding the comparison between different outputs in the tables can be found at the end in the "Discussion" section.

The ROC curves for the two eNose statistical output are shown in Figure 6.2-3.

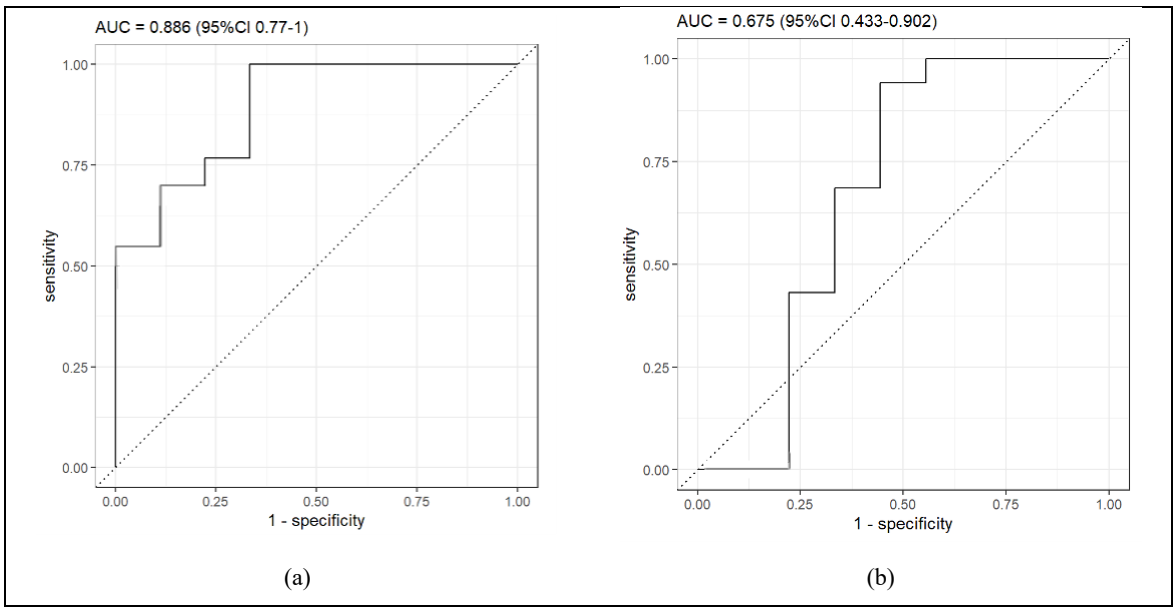


Figure 6.2-3: represents the output of the two Electronic Noses to differentiate negative UTI samples and positive UTI samples. ***(a)*** *represents the output generated by the AlphaMOS FOX 4000 device and,* ***(b)*** *represents the output generated by the PEN3 device.*

c. AlphaMOS FOX 4000 Results:

The AlphaMOS FOX 4000 consists of 18 MOS gas sensors which represent the presence of VOCs in terms of change in resistance [10]. The typical output of the FOX 4000 eNose is shown in Figure 6.2-4. Each colour in the figure represents the response of the different MOS sensors to a positive UTI sample.

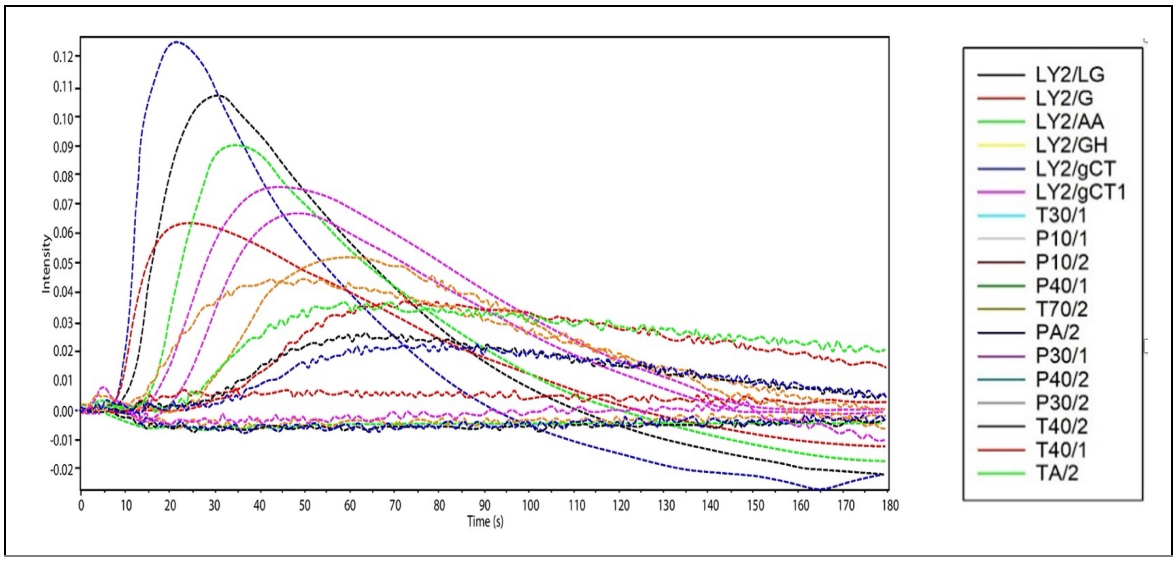


Figure 6.2-4: Output of the FOX 4000 for positive UTI sample.

Figure 6.2-5 is a bar graph generated by the FOX 4000 device for positive UTI samples where each graph represents the response of the sensors to the sample. The graph shows that the most responsive sensors were P30/1, P10/1, P10/2, PA/2, T30/1, and TA/2.

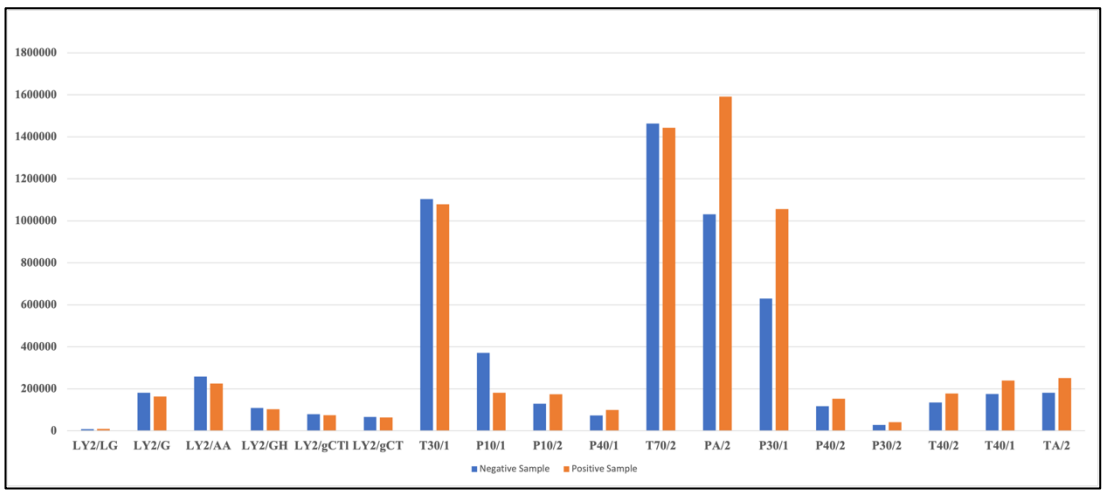


Figure 6.2-5: Bar graph representation of the output generated by the FOX 4000 in response to a positive UTI sample (Orange) and a Negative UTI sample (Blue).

The data generated from the FOX 4000 was then analysed. The UTI samples were divided into positive UTIs, negative UTIs, samples with less than 10 CFU (colony forming units) and mixed growth samples. These groups were then analysed using an R program and the MultiSens Analyzer. The output generated from the R program is shown in Table 6.2-4. The table presents the classifiers that achieved the highest level of separation among the six classifiers outlined in Chapter 3 in section e (Qualitative and Quantitative Analysis:, pg.: 62) that generate the highest level of separations to distinguish between two classes for each comparison.

Table 6.2-4: Statistical Output of UTI urine Samples using the AlphaMOS FOX 4000.

Groups	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
Negative Vs Positive	Random Forest	0.886 (0.769-0.978)	0.982 (0.946-1)	0.667 (0.375-0.909)	0.948 (0.896-0.984)	0.857 (0.6-1)
Negative Vs Mixed Growth	Neural Network	0.643 (0.477-0.778)	1 (1-1)	0 (0-0)	0.949 (0.898-0.983)	NA (NA)
Negative Vs <10 CFU	Random Forest	0.563 (0.437-0.685)	0.077 (0.0-0.22)	0.911 (0.845-0.965)	0.167 (0.0-0.5)	0.809 (0.725-0.889)
Negative Vs Positive and Mixed Growth	Random Forest	0.859 (0.747-0.958)	0.964 (0.914-1)	0.5 (0.25-0.733)	0.9 (0.836-0.962)	0.75 (0.454-1)
Negative Vs Positive and <10 CFU	Random Forest	0.743 (0.629-0.841)	0.892 (0.82-0.96)	0.454 (0.275-0.631)	0.806 (0.721-0.885)	0.625 (0.417-0.818)
Negative Vs Positive and mixed growth and <10CFU	Random Forest	0.734 (0.632-0.831)	0.857 (0.79-0.92)	0.44 (0.28-0.608)	0.774 (0.688-0.861)	0.578 (0.388-0.762)

The ROC curves for the data analysis for the AlphaMOS FOX 4000 eNose output can be seen in Figure 6.2-6.

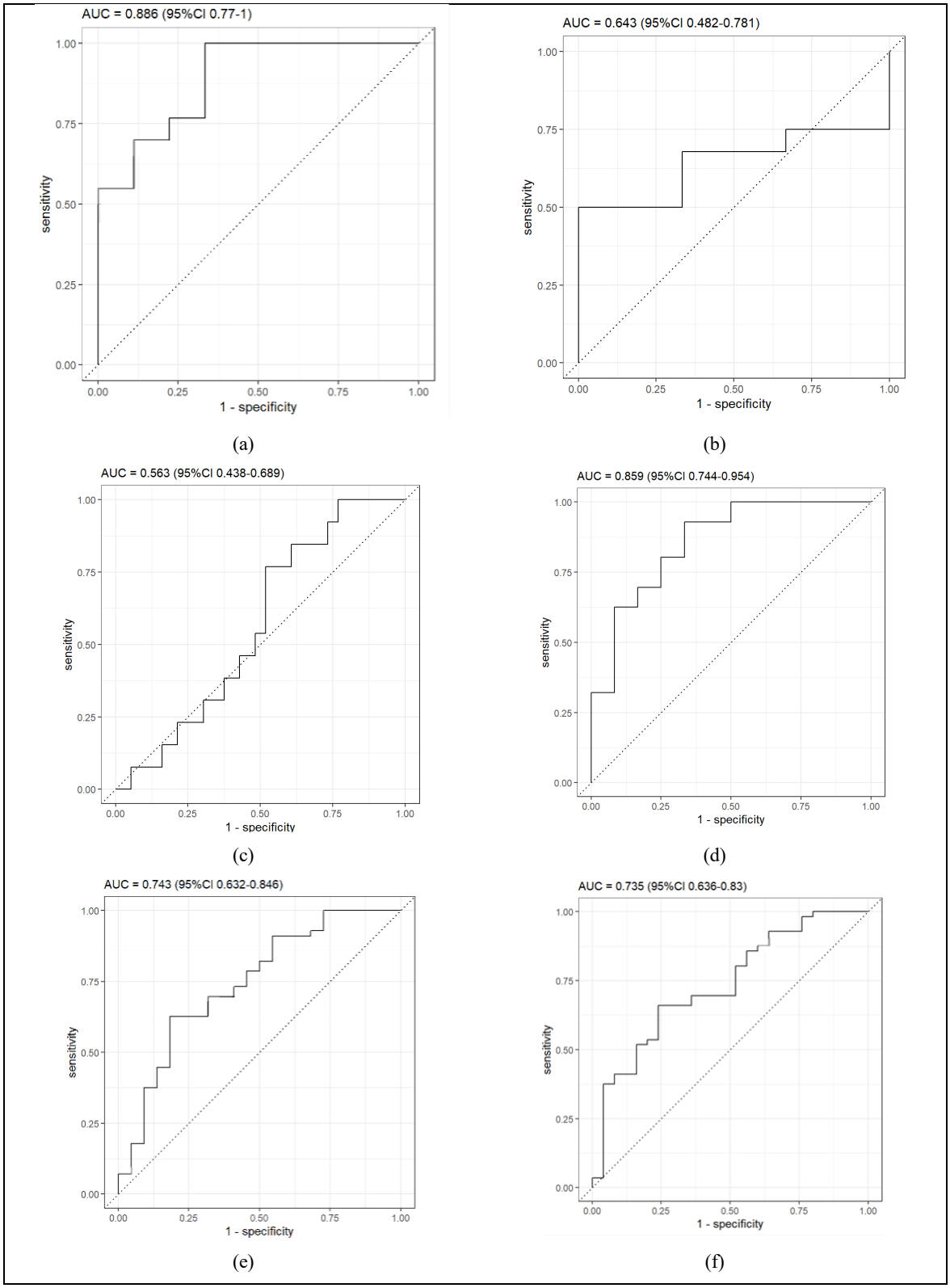


Figure 6.2-6: ROC curves for UTI urine samples using the AlphaMOS FOX 4000 eNose where ***(a)*** *ROC curve for Negative Vs Positive UTI samples;* ***(b)*** *ROC curve for Negative Vs Mixed growth UTI samples;* ***(c)*** *ROC curve for Negative Vs <10 CFU growth UTI samples;* ***(d)*** *ROC curve for Negative Vs Positive and Mixed growth UTI samples;* ***(e)*** *ROC curve for Negative Vs Positive and*

<10 CFU growth UTI samples; and ***(f)*** *ROC curve for Negative Vs Positive, <10 CFU growth and mixed growth UTI samples.*

Figure 6.2-6 shows the different ROC curves for the comparison of the different categories. The R program analysis showed that the best result was generated by the random forest classification algorithm for comparison between negative and positive UTI samples. Out of the 56 negative samples, the FOX 4000 eNose was able to identify 55 samples correctly whereas out of the 9 positive UTI samples, it was able to identify 6 samples correctly. However, for the comparison of negative UTI samples with mixed growth samples, the neural network showed the best results where it showed that the FOX 4000 eNose identified all samples as negative samples. The reason for this could be the small set size of mixed growth samples. Similarly, for the comparison between negative samples and samples with infection less than 10 cfu/ml of urine, the neural Network classification algorithm showed that the FOX 4000 eNose was able to identify 51 negative samples out of 56 samples correctly and just one sample for the samples with infection less than 10 cfu/ml of urine.

d. PEN3 eNose Results:

Figure 6.2-7 illustrates the graphical representation of the measurement of relative sensor values on the y-axis versus time on the x-axis for the gas sensors within the PEN3 eNose. Each coloured curve corresponds to the response of different sensors in the PEN3 device to a positive UTI sample.

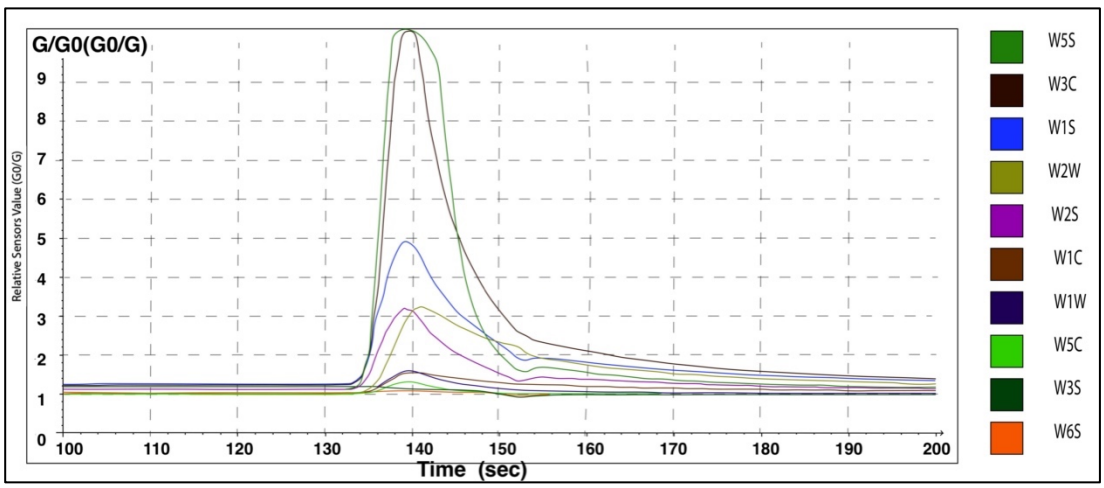


Figure 6.2-7: represents the output of the PEN3 device to UTI urine sample.

A radar plot was created to represent the response of each sensor towards the positive and negative UTI sample as shown in Figure 6.2-8. The radar plot compared the sensor's response for a positive UTI sample (blue line) and a negative UTI sample (orange line). It was observed that the sensors W2S, W1W and W1S showed the most promising response to the two UTI samples.

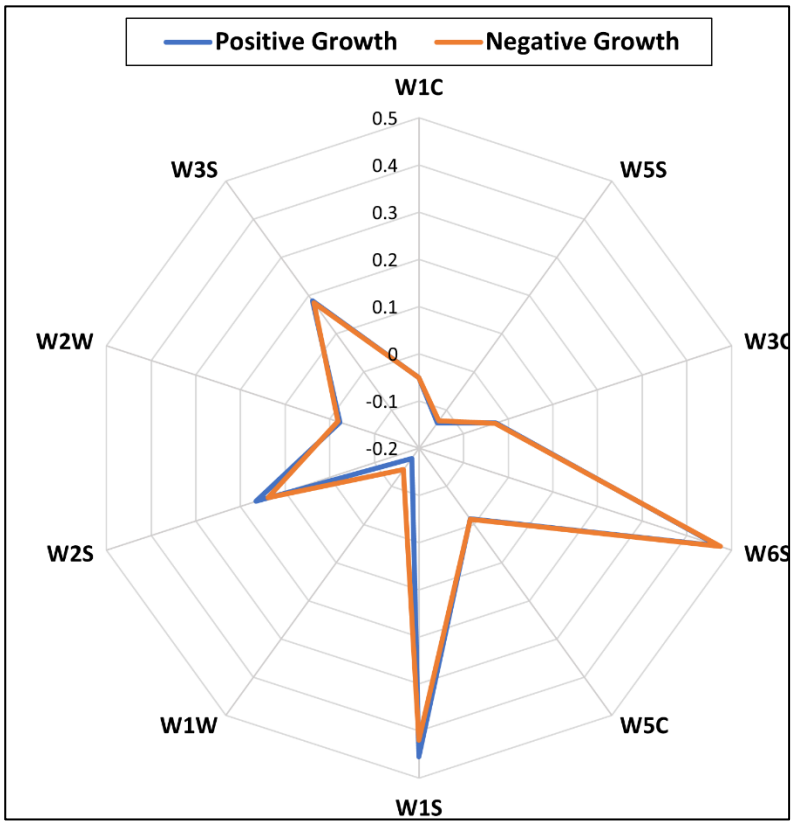


Figure 6.2-8: The output from the sensors of the PEN3 eNose for positive UTI urine samples and Negative UTI urine samples.

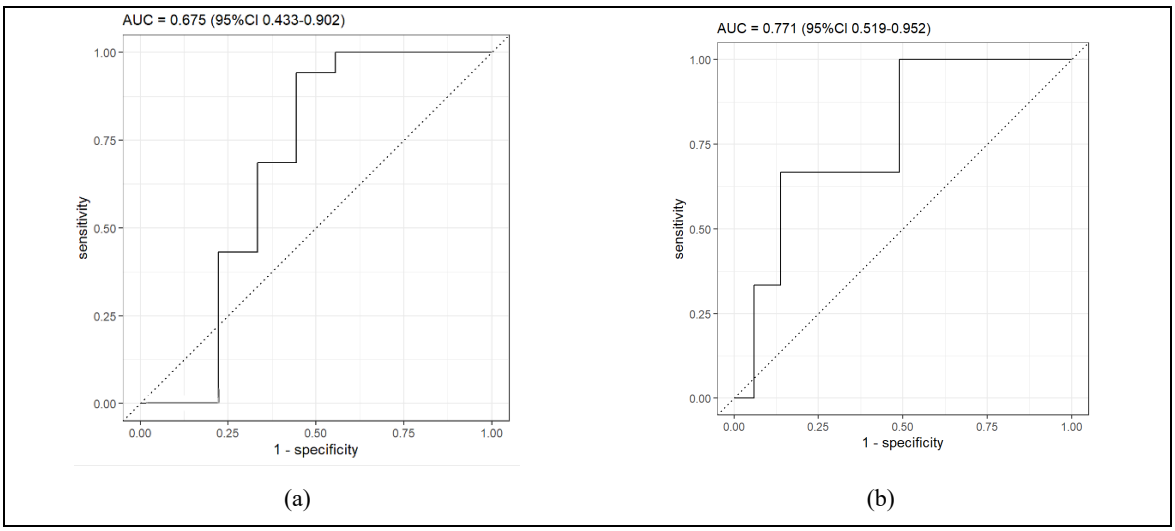
The R program was used to analyse the different UTI sample groups. The 81 UTI samples were divided into four groups, namely, Positive UTI samples, samples with less than 10 CFU (colony forming units), mixed growth samples and negative samples which were used as controls for this study. The results obtained are shown in Table 6.2-5. The statistical analysis between the different UTI groups and the controls showed that the PEN3 device was able to discriminate between UTI samples and controls. The table presents the classifiers that achieved the highest level of separation among the six classifiers mentioned

in the Chapter 3 in section e (Qualitative and Quantitative Analysis:, pg.: 62) that generate the highest level of separations to distinguish between two classes for each comparison.

Table 6.2-5: Statistical Output of UTI urine Samples using PEN3eNose.

Groups	Training Strategies	AUC	Sensitivity	Specificity	PPV	NPV
Negative Vs Positive	Random Forest	0.66 (0.43-0.91)	1 (1-1)	0.44 (0.17-0.73)	0.91 (0.84-0.97)	1 (1-1)
Negative Vs Mixed Growth	Neural Network	0.77 (0.54-0.94)	0.00 (0.00.00)	0.98 (0.94-1)	0.00 (0.00.00)	0.94 (0.89-0.98)
Negative Vs <10 CFU	Random Forest	0.48 (0.32-0.65)	0.00 (0.00.00)	0.94 (0.88-0.98)	0.00 (0.00.00)	0.879 (0.69-0.87)
Negative Vs Positive and Mixed Growth	Random Forest	0.69 (0.51-0.87)	0.98 (0.94-1.00)	0.42 (0.18-0.67)	0.88 (0.80-0.95)	0.83 (0.50-1)
Negative Vs Positive and <10 CFU	Random Forest	0.66 (0.54-0.78)	0.86 (0.78-0.94)	0.41 (0.24-0.59)	0.77 (0.68-0.86)	0.56 (0.35-0.77)
Negative Vs Positive and mixed growth and <10CFU	Random Forest	0.73 (0.63-0.83)	0.86 (0.78-0.93)	0.44 (0.28-0.61)	0.77 (0.69-0.86)	0.58 (0.39-0.76)

The ROC curves generated by the R program is shown in Figure 6.2-9.



(a) (b)

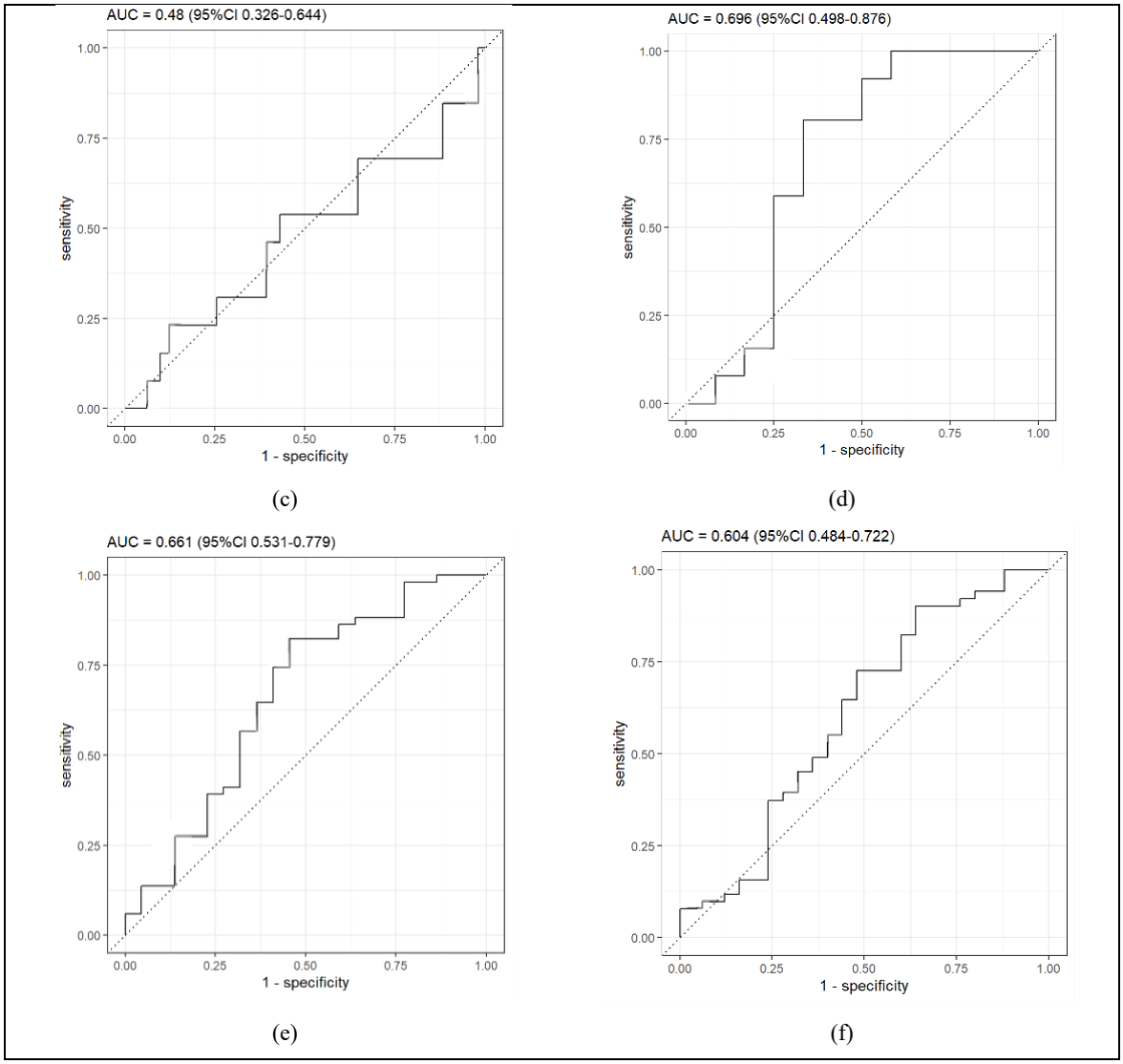


*Figure 6.2-9: ROC curves for the UTI urine samples using the PEN3 eNose where **(a)** ROC curve for Negative Vs Positive UTI samples; **(b)** ROC curve for Negative Vs Mixed growth UTI samples; **(c)** ROC curve for Negative Vs <10 CFU growth UTI samples; **(d)** ROC curve for Negative Vs Positive and Mixed growth UTI samples; **(e)** ROC curve for Negative Vs Positive and <10 CFU growth UTI samples; and **(f)** ROC curve for Negative Vs Positive, <10 CFU growth and mixed growth UTI samples.*

These curves represent the accuracy of the classification models to correctly classify the different UTI samples and the non-UTI samples. The ROC curves for the negative UTI samples in comparison to mixed growth resulted in 0.77 AUC, whereas the AUC for positive UTI samples and negative UTI samples was 0.68. The lowest AUC obtained was for the negative UTI samples and samples with <10 CFU and these showed that the PEN3 eNose was not able to accurately distinguish between these groups.

6.3. Discussion:

Urinary tract infections remain the cause of high medical expenditure and morbidity with infrequent occurrence among 2- to 13-year-old girls and leading to multiple episodes for young and adult females [11, 12]. UTIs have been regarded as a consistent disease that affects 1% to 3% of all consultations [13]. The most challenging aspect of UTIs is diagnostic delays. Any delays in the diagnosis or detection of UTIs can lead to loss of renal function and damaged kidneys. The currently available diagnostic methods include urinalysis and urine cultures. These methods are considered gold standard for the detection of UTIs and can detect the presence of pathogens in the urine [14]. Urinalysis is a quick and non-invasive method, but it has low sensitivity and selectivity whereas urine culture is more specific, but the diagnosis through culture takes a long time. Though these are non-invasive methods, they have their own limitations [15, 16]. Therefore, it would be beneficial to have a fast, cheap, and non-invasive diagnostic method.

As observed in Chapter 5, consistent outcomes were evident in this study, aligning with the Random Forest classifier yielding the most favourable results. The Random Forest algorithm is an ensemble learning technique that merges multiple decision trees to render robust predictions. The algorithm's proficiency in capturing non-linear correlations, accommodating feature interactions, and curbing overfitting through its ensemble approach collectively contribute to its heightened performance. The ensemble nature of the algorithm guarantees stability, diminishing the likelihood of model bias or variance, thereby engendering a well-balanced predictive efficacy. The observed supremacy of Random Forest in distinguishing between disease samples and healthy controls instances underscores its relevance and potential for enhancing medical diagnostic endeavours.

Table 6.2-2 and Table 6.2-3 represent the results obtained from the comparison between negative UTI samples and positive UTI samples using four different analytical instruments, namely, the AlphaMOS FOX 4000, the PEN3 eNose, the FlavourSpec GC-IMS, and the Silox GC-IMS. Comparing these instruments, we observed that the AlphaMOS FOX 4000 and the Silox GC-IMS exhibited higher Area Under the Curve (AUC) values of 0.885, indicating their superior overall ability to discriminate between negative and positive UTI samples. These instruments demonstrated better separation power by capturing the essential

discriminatory features present in the VOC profiles of the samples. The AlphaMOS FOX 4000 exhibited excellent sensitivity (0.982) and NPV (0.857), implying its ability to correctly identify positive and negative UTI samples, respectively. On the other hand, the Silox GC-IMS showed high sensitivity (0.968) and PPV (0.876), signifying its effectiveness in distinguishing between the two groups by correctly detecting positive UTI samples and maintaining a high level of confidence in its positive predictions.

Conversely, the FlavourSpec GC-IMS exhibited relatively low specificity (0.076) and NPV (0.286), indicating its challenges in correctly identifying negative UTI samples. This suggests that the instrument may have faced difficulties in achieving a strong separation between the two groups, particularly with respect to correctly identifying true negatives and maintaining confidence in its negative predictions. The PEN3, although achieving an AUC of 0.675, demonstrated perfect sensitivity (1.0) and NPV (1.0) along with low specificity (0.444) and PPV (0.911). This implies that the instrument was successful in correctly identifying all positive UTI cases while exhibiting some misclassifications in negative UTI samples, leading to a compromised specificity and PPV. While achieving a sensitivity of 1.0 (100%) and a negative predictive value (NPV) of 1.0 (100%) might seem impressive, it could potentially indicate overfitting of the model to the training data, data imbalance, generalizability, validation, noise, and clinical relevance.

In conclusion, the AlphaMOS FOX 4000 and the Silox GC-IMS instrument demonstrated superior performance in distinguishing between negative and positive UTI samples. However, the evaluation of these results should be considered in the context of the imbalanced dataset to ensure that the classifier's performance is not biased towards the majority class, and further validation on more balanced datasets would be beneficial to assess the robustness and generalizability of these findings.

Detection of UTIs using volatile organic compounds have been investigated over the years. The analysis of urine for the detection of bacteria was used by Hayward *et al.* [17] in his study of diagnosis of UTIs using gas-liquid headspace. They found that HS-GC-MS was able to detect all the infected urine samples correctly and reported chemical biomarkers such as ethanol, methyl mercaptan or dimethyl sulphide for the detection of UTIs. In this study, we found that the GC-IMS was able to distinguish between different UTI sample categories

with high success. The FlavourSpec GC-IMS showed very high sensitivity, i.e., 94.7% whereas the specificity achieved was very low, i.e., 7.6%. In comparison to the FlavourSpec GC-IMS, the Silox GC-IMS resulted in high AUC value. The Silox GC-IMS was able to differentiate between negative samples to the infected samples at high rates. The sensitivity obtained by the Silox was 96.8% and specificity was 50%. The result generated through an R program showed that the FlavourSpec was able to correctly identify 51 negative UTI samples out of 56 negative samples, whereas it was only able to identify 2 samples with positive UTIs. It was observed that there was a wide difference between the sensitivity and specificity of the Silox GC-IMS and the FlavourSpec GC-IMS devices. These discrepancies can be attributed to the fact that the samples for the Silox GC-IMS device were collected and used directly for analysis within a few hours of collection, while it was reported that the urine samples used in the analysis of the FlavourSpec GC-IMS were mishandled and not refrigerated properly. This mishandling resulted in the degradation of the chemical biomarkers present in the samples. A number of studies have shown that the mishandling of samples can lead to the potential degradation of the quality of the sample. The study conducted by Esfahani *et al.* [18] showed that the urinary metabolite degradation caused by the storage and mishandling of samples led to a loss of chemical information from the urine headspace for gas analysis at −80 °C. The disadvantages of the GC-IMS technologies are that they cannot be used in real time, they require trained personal and are very expensive. Therefore, the focus has moved to an eNose device. eNoses analyse data in real-time and are an inexpensive technology.

Figure 6.2-5 is a bar graph obtained from the AlphaMOS FOX 4000 eNose, representing the response of the sensors to positive UTI samples and negative UTI samples. Each bar in the graph corresponds to a specific sensor in the eNose, and the height of the bar indicates the magnitude of the response exhibited by the sensor to the respective sample. The graph reveals that certain sensors demonstrated significant differences in their responses between positive and negative UTI samples. These sensors, namely P30/1, P10/1, P10/2, PA/2, T30/1, and TA/2, exhibited notably higher responses to positive UTI samples compared to negative UTI samples. This disparity in response of these specific sensors to positive UTI samples indicates their potential significance in discriminating between positive and negative UTI cases. The ability of the eNose to detect such discriminatory patterns offers valuable insights into the VOC profiles associated with UTI infections, which can aid in the

development of diagnostic tools and contribute to better disease classification and management.

There is limited research on the use of eNoses for UTI diagnosis and detection. Roine *et al.* [19] conducted a study on 101 samples of different UTI pathogens, namely, *Escherichia coli, Staphylococcus saprophyticus, Klebsiella species, Enterococcus faecalis,* and healthy controls. They used a commercially available eNose (ChemPro 100i, Environics Inc., Mikkeli, Finland) and analysed the samples using Linear discriminant analysis (LDA) and logistic regression (LR) classification models. In their study, they found out that eNose was able to achieve a sensitivity and specificity of 95% and 97% respectively, using logistic regression and a sensitivity of 90% and specificity of 96% using linear discriminant analysis.

In this study, we compared the performance of two electronic nose devices, the AlphaMOS FOX 4000 eNose and the PEN3 eNose, for their ability to discriminate between different sample groups based on volatile organic compounds (VOCs). The statistical outputs obtained from the two eNoses were analysed to evaluate their separation capabilities and discriminatory power. For the comparison between negative and positive urinary tract infection (UTI) samples, the AlphaMOS FOX 4000 eNose exhibited an AUC of 0.886 with a sensitivity of 0.982 and a specificity of 0.667. On the other hand, the PEN3 eNose achieved an AUC of 0.66 with a sensitivity of 1 and a specificity of 0.44. These results indicate that both eNoses showed some level of discriminatory ability, but the AlphaMOS FOX 4000 eNose demonstrated a slightly better performance in distinguishing between positive and negative UTI samples.

Similarly, when comparing negative UTI samples with samples showing mixed growth, the AlphaMOS FOX 4000 eNose achieved an AUC of 0.643 with a sensitivity of 1 and a specificity of 0, whereas the PEN3 eNose obtained an AUC of 0.77 with a sensitivity of 0 and a specificity of 0.98. In this case, the PEN3 eNose showed higher specificity, while the AlphaMOS FOX 4000 eNose had perfect sensitivity. Furthermore, for the comparison between negative UTI samples and samples with a bacterial count of less than 10 colony-forming units (CFU), the AlphaMOS FOX 4000 eNose obtained an AUC of 0.563 with a very low sensitivity of 0.077 and a specificity of 0.911. The PEN3 eNose achieved an AUC of 0.48 with a sensitivity of 0 and a specificity of 0.94. Here, the AlphaMOS FOX 4000

eNose showed higher specificity, while the PEN3 eNose had lower sensitivity. The results suggest that the classifiers had no ability to differentiate negative samples from samples with mixed growth, resulting in a specificity value of 0. This could be due to the imbalanced distribution of samples, with a larger number of negative samples (56 samples) and a smaller number of mixed growth samples (3 samples) as well as samples with a bacterial count of less than 10 CFU (13 samples), and this could also contribute to the skewed performance of the classifier.

When considering the comparisons involving combinations of different sample groups, both eNoses generally demonstrated reasonable discriminatory capabilities, but their performances varied depending on the specific sample groups being compared. The AlphaMOS FOX 4000 eNose often exhibited higher AUC values and sensitivity in comparison to the PEN3 eNose, while the PEN3 eNose showed higher specificity in some cases. In conclusion, the AlphaMOS FOX 4000 eNose and the PEN3 eNose showed varying levels of separation capabilities in distinguishing between different sample groups based on VOC profiles. The AlphaMOS FOX 4000 eNose generally demonstrated slightly better discriminatory performance in comparison to the PEN3 eNose in most cases.

The foremost limitation of this study was that the urinary samples used during this study specifically for the analysis using the FlavourSpec GC-IMS, the AlphaMOS FOX 4000 eNose and the PEN3 eNose were reported to have been mishandled during storage which may have resulted in the decay of the metabolic characteristics of the samples. The mishandling of the samples may have led to substantial contamination, degradation, or cross-contamination of the samples which could have significantly impacted on the integrity of the data and compromised the data to the extent that the results may be unreliable for making any meaningful conclusions. In such cases, the analysis may be deemed null and void, and the data should not be used for making any scientific or clinical decisions [20]. Another limitation of this study was the unbalanced sample set sizes for the different categories of the samples. Due to this, the classification algorithm may be unable to factor in the unbalanced classes and this invariance may affect the results [21].

In this research, we have explored the statistical analysis of several GC-IMS devices, such as, the FlavourSpec GC-IMS and the Silox GC-IMS, and eNose devices, particularly the

AlphaMOS FOX 4000 and PEN3, for distinguishing between various sample groups based on VOCs. The success of the statistical analysis discussed relies on several key criteria. First and foremost, the analysis should exhibit a high discriminatory power, effectively distinguishing between different groups or classes based on the VOC profiles detected by the eNose devices. For instance, in one of the comparisons, the AlphaMOS FOX 4000 eNose demonstrated an AUC of 0.886 (with a 95% confidence interval of 0.769 to 0.978) when differentiating between negative and positive UTI samples. This suggests that the eNose has a relatively high ability to correctly identify positive UTI samples among all the UTI samples present in the dataset [22].

Another essential criterion is the specificity of the analysis, ensuring that the classifier can accurately identify negative instances, thereby minimizing false positives. In the same UTI example, the specificity of the AlphaMOS FOX 4000 eNose was reported as 0.667 (with a 95% confidence interval of 0.375 to 0.909), indicating that it had a moderate ability to correctly identify negative UTI samples among all the negative samples in the dataset. Confidence intervals also play a critical role in providing a measure of uncertainty around the estimated performance metrics. Wider confidence intervals, such as in the case of the FlavourSpec GC-IMS, with an AUC of 0.571 (with a 95% confidence interval of 0.459 to 0.681), signify higher uncertainty regarding the classifier's performance.

Moreover, consistency and reproducibility of such results are crucial for reliable analysis. For instance, in the comparison between negative and mixed growth samples using the Neural Network classifier, the AUC of the PEN3 eNose was 0.77 (with a 95% confidence interval of 0.54 to 0.94). This suggests that the classifier consistently performed well in distinguishing between these two groups in repeated runs. The consideration of data imbalance is also critical. For example, in the comparison between Negative and Positive UTI samples, the dataset was imbalanced with 56 negative samples and 9 positive samples. The analysis should address such imbalanced data to prevent biases [23].

By meeting these key criteria, the statistical analysis can provide meaningful insights into the performance of the devices used in this research. These insights can thus potentially aid in clinical diagnosis, disease classification, and monitoring, with potential applications in various healthcare settings [24].

The use of an eNose for the detection of bacteria looks promising and a large amount of research has been done for the detection of bacteria specially using urinary headspace. The early detection of UTIs using an eNose has shown inadequate results due to the low sensitivity and specificity of the sensors. Improvements are required in the sensors' sensitivity and pattern recognition in order to offer an effective clinical implementation.

6.4. Conclusion:

The use of VOCs to understand the chemical fingerprint of diseases and use them for the early diagnosis and prognosis has revealed significant potential for a very long time. UTI infections remain the reason for higher medical costs and high mortality with low sensitivity and time on using current diagnostic methods. It is important to develop a highly sensitive, cheaper, and fast solution to this problem. In conclusion, the results found in our study showed the potential of urinary VOCs for the discrimination of UTI samples from healthy controls. Both the GC-IMS technology and the eNose technologies were able to differentiate between negative UTI samples and positive UTI samples but the sensitivities obtained from these instruments were not high enough to confirm that they are reliable enough for early UTI diagnosis. However, an eNose can deliver a solution to these limitations over the GC-IMS technologies as eNoses are fast, cheap, and portable devices. They provide a considerable cost reduction solution required currently for the detection and diagnosis of UTIs as well as possibly helping in the early detection of UTIs which in turn would result in further cost reductions. However, more research is required for the development of eNose sensors to improve in the sensitivity for the early detection of UTIs.

6.5. References:

[1] G. Schmiemann, E. Kniehl, K. Gebhardt, M. M. Matejczyk, and E. Hummers-Pradier, "The diagnosis of urinary tract infection: a systematic review," (in eng), *Deutsches Arzteblatt international,* vol. 107, no. 21, pp. 361-367, 2010, doi: 10.3238/arztebl.2010.0361.

[2] L. K. McLellan and D. A. Hunstad, "Urinary Tract Infection: Pathogenesis and Outlook," *Trends in Molecular Medicine,* vol. 22, no. 11, pp. 946-957, 2016/11/01/ 2016, doi: https://doi.org/10.1016/j.molmed.2016.09.003.

[3] N. S. Sheerin, "Urinary tract infection," *Medicine,* vol. 39, no. 7, pp. 384-389, 2011/07/01/ 2011, doi: https://doi.org/10.1016/j.mpmed.2011.04.003.

[4] T. M. Hooton, "Uncomplicated urinary tract infection," *New England Journal of Medicine,* vol. 366, no. 11, pp. 1028-1037, 2012.

[5] L. Nicolle, "Complicated urinary tract infection in adults," *Canadian Journal of Infectious Diseases and Medical Microbiology,* vol. 16, no. 6, pp. 349-360, 2005.

[6] M. Medina and E. Castillo-Pino, "An introduction to the epidemiology and burden of urinary tract infections," (in eng), *Ther Adv Urol,* vol. 11, pp. 1756287219832172-1756287219832172, 2019, doi: 10.1177/1756287219832172.

[7] J. W. Bartges, "Diagnosis of urinary tract infections," *Veterinary Clinics: Small Animal Practice,* vol. 34, no. 4, pp. 923-933, 2004.

[8] C. M. Chu and J. L. Lowder, "Diagnosis and treatment of urinary tract infections across age groups," *American Journal of Obstetrics and Gynecology,* vol. 219, no. 1, pp. 40-51, 2018/07/01/ 2018, doi: https://doi.org/10.1016/j.ajog.2017.12.231.

[9] S. Najeeb, T. Munir, S. Rehman, A. Hafiz, M. Gilani, and M. Latif, "Comparison of urine dipstick test with conventional urine culture in diagnosis of urinary tract infection," *J Coll Physicians Surg Pak,* vol. 25, no. 2, pp. 108-10, 2015.

[10] J. A. Covington *et al.*, "Application of a Novel Tool for Diagnosing Bile Acid Diarrhoea," *Sensors,* vol. 13, no. 9, pp. 11899-11912, 2013. [Online]. Available: https://www.mdpi.com/1424-8220/13/9/11899.

[11] W. E. Stamm and S. R. Norrby, "Urinary Tract Infections: Disease Panorama and Challenges," *The Journal of Infectious Diseases,* vol. 183, no. Supplement_1, pp. S1-S4, 2001, doi: 10.1086/318850.

[12] J. B. Lee and G. H. Neild, "Urinary tract infection," *Medicine,* vol. 35, no. 8, pp. 423-428, 2007.

[13] K. Gupta, L. Grigoryan, and B. Trautner, "Urinary tract infection," *Annals of internal medicine,* vol. 167, no. 7, pp. ITC49-ITC64, 2017.

[14] M. Bitsori and E. Galanakis, "Pediatric urinary tract infections: diagnosis and treatment," *Expert review of anti-infective therapy,* vol. 10, no. 10, pp. 1153-1164, 2012.

[15] S. M. E. Finnell, A. E. Carroll, S. M. Downs, and t. S. o. U. T. Infection, "Diagnosis and Management of an Initial UTI in Febrile Infants and Young Children," *Pediatrics,* vol. 128, no. 3, pp. e749-e770, 2011, doi: 10.1542/peds.2011-1332.

[16] T. K. Price, E. E. Hilt, T. J. Dune, E. R. Mueller, A. J. Wolfe, and L. Brubaker, "Urine trouble: should we think differently about UTI?," *International Urogynecology Journal,* vol. 29, no. 2, pp. 205-210, 2018/02/01 2018, doi: 10.1007/s00192-017-3528-8.

[17] N. J. Hayward, "Head-space gas—liquid chromatography for the rapid laboratory diagnosis of urinary tract infections caused by enterobacteria," *Journal of Chromatography B: Biomedical Sciences and Applications,* vol. 274, pp. 27-35, 1983/01/01/ 1983, doi: https://doi.org/10.1016/S0378-4347(00)84405-9.

[18] S. Esfahani *et al.*, "Variation in Gas and Volatile Compound Emissions from Human Urine as It Ages, Measured by an Electronic Nose," in *Biosensors* vol. 6, ed, 2016.

[19] A. Roine *et al.*, "Rapid and Accurate Detection of Urinary Pathogens by Mobile IMS-Based Electronic Nose: A Proof-of-Principle Study," *PLOS ONE,* vol. 9, no. 12, p. e114279, 2014, doi: 10.1371/journal.pone.0114279.

[20] M. Rotter *et al.*, "Stability of targeted metabolite profiles of urine samples under different storage conditions," *Metabolomics,* vol. 13, no. 1, p. 4, 2016/11/28 2016, doi: 10.1007/s11306-016-1137-z.

[21] M. Yoon and M. H. C. Lai, "Testing Factorial Invariance With Unbalanced Samples," *Structural Equation Modeling: A Multidisciplinary Journal,* vol. 25, no. 2, pp. 201-213, 2018/03/04 2018, doi: 10.1080/10705511.2017.1387859.

[22] G. Noctor, A. Mhamdi, and C. H. Foyer, "Oxidative stress and antioxidative systems: recipes for successful data collection and interpretation," *Plant, Cell & Environment,* vol. 39, no. 5, pp. 1140-1160, 2016/05/01 2016, doi: https://doi.org/10.1111/pce.12726.

[23] R. Nisbet, J. Elder, and G. D. Miner, *Handbook of statistical analysis and data mining applications*. Academic press, 2009.

[24] A. O. Kirdar, K. D. Green, and A. S. Rathore, "Application of Multivariate Data Analysis for Identification and Successful Resolution of a Root Cause for a Bioprocessing Application," *Biotechnology Progress,* vol. 24, no. 3, pp. 720-726, 2008/05/01 2008, doi: https://doi.org/10.1021/bp0704384.

Chapter 7. Development of enhanced Photoionisation Detector

Chapter 7 of this thesis presents a comprehensive overview of the Photoionization Detector (PID) technology and outlines the design approaches taken to develop three in-house prototypes: PID+ version 1, version 2, and version 3. The chapter commences with a brief introduction to PID technology, followed by a detailed description of the principle of ionization operation. Subsequently, the chapter delves into the approaches adopted for prototype development, including a comprehensive account of the faults and errors encountered. Furthermore, this chapter and the following chapter present detailed information on the results obtained during device development. The comprehensive and detailed information provided in this section facilitates the reproducibility of the study by other researchers.

7.1. Introduction:

The presence of Volatile Organic Compounds (VOCs) beyond a threshold value has an adverse effect on human health and on other living organisms. VOCs are toxic gases that may cause disease or eventually death [1]. VOCs originate everywhere around us in the form of alkanes, alkenes, aromatic hydrocarbons, ketones, aldehydes, nitrogenous compounds, etc. [2]. The analysis of VOCs has attracted attention in recent years and with the development in the scientific research technologies, the monitoring of the VOCs associated with various issues such as medical diagnosis [3], indoor/outdoor air pollution [4], climate changes [5], changes in soil microbiology [6] etc.

The most successful analytical tools available for the analysis of VOCs are GC-MS, IMS, PID, eNose, chemical sensors such as Conducting Polymer sensors, SAW sensor arrays, MOS sensors, etc. and these have been discussed in detail in chapter 3.

The PID is comprised of an ionization chamber and a high voltage detecting unit, with a UV lamp situated inside the ionization chamber to generate UV light used to ionize gaseous

compounds. When the ionization energy produced by the lamp is more than the ionization energy of the compounds, they ionise into charged molecules. These ionised analytes are then detected by the high voltage detectors [7]. With developments in technology, the demands for accurate, improved, simple and sensitive devices have increased in recent years. Conventional methods for the detection of the VOCs are costly, bulkier, and heavier, time-consuming and involve limited sampling. Technologies such as GC-IMS and IMS have high sensitivity and high selectivity [8]. GC and MS have become especially useful analytical methods within the last few years. However, the main disadvantages of GC–MS are that they have a long analysis time, a demand for qualified technicians, excessive cost, and bulkier size, whereas eNoses are cost-effective, easily portable and have shorter analysis time. The major drawbacks of eNose systems are that since they use an array of sensors, every sensor has its own advantages and drawbacks. eNose sensors require regular calibration to maintain accuracy and may experience drift over time, affecting their reliability. Environmental factors, such as temperature, humidity, and the presence of other odours, can interfere with eNose performance, compromising the accuracy of VOCs' detection. Moreover, different types of eNose sensors may exhibit varying sensitivities and response ranges, making it challenging to achieve consistent and comparable results across different sensor types [9-11].

The PID was initially reported in 1960 as a means of detecting gases and vapours. At the time, it was already known that ionization of gases and vapours was an effective method for measuring their concentrations. However, the use of UV radiation for ionization proved to be a more stable process than previous methods, as it did not require the use of inert gases and eliminated problems associated with coatings and column bleed that were observed with PID using flowing inert gases. The introduction of UV radiation in PID technology allowed for the creation of more reliable and efficient devices for the detection of VOCs. The use of UV radiation in PID also made the devices more cost-effective, portable, and easier to use [12]. Since the development of the first PID, there have been many advancements in the technology, including the development of in-house versions, which have improved upon the original design to make them more efficient, accurate, and sensitive. The first PID with a sealed UV source was developed at HNU Systems, Inc. (Newton Upper Falls, Massachusetts). The PID developed at HNU consisted of two parts, a power supply to provide high voltages and a detector module. The detector consisted of a sealed UV chamber

which enclosed an UV bulb with potential energy of 10.2 eV emitting UV radiations through a magnesium fluoride window into the chamber resulting in the photoionization of all the molecules with potential energy equal or less than 10.2 eV. The ionised molecules were then collected at the detecting electrodes and measured using the FID electrometer in the Tracer Model 550 gas chromatograph. They found that the PID showed sensitivity towards the organic compounds and some inorganic compounds with potential energy less than 10.2 eV. They also found that the PID showed a 30-fold increase in sensitivity compared to an FID [13]. After this, several studies were conducted to evaluate the strength of PID using different chemicals. The study conducted by Marsha *et al.* [14, 15] used an HNU systems PID using a 10.2 eV lamp with different standards. They found that the PID was more sensitive to the chemicals with higher carbon number, functional groups, and bonding types of the chemicals [14, 15]. Studies done on the PID show the capabilities of the use of a PID for the detection of gases/chemicals.

7.2. Theory:

The PID operates on the basic principle of photoionization, which involves the use of photons to generate ions. However, it is crucial to note that not all photons can produce ionization, as the energy of the photon must be higher than that of the gas molecules or atoms for ionization to occur. The energy required to displace an electron from VOCs molecules is referred to as the ionization energy, and the larger the molecule, the less energy is required for the detection. The PID utilizes UV radiation to ionize gas samples into charged molecules that can be detected by electrodes. UV radiation has a short wavelength and higher frequencies, which implies higher energy. Therefore, the PID's detection capability is based on the energy of the photons, making it selective and sensitive to low concentrations of VOCs. The ionization process transpires within an ionization chamber, which is visually demonstrated in Figure 7.2-1.

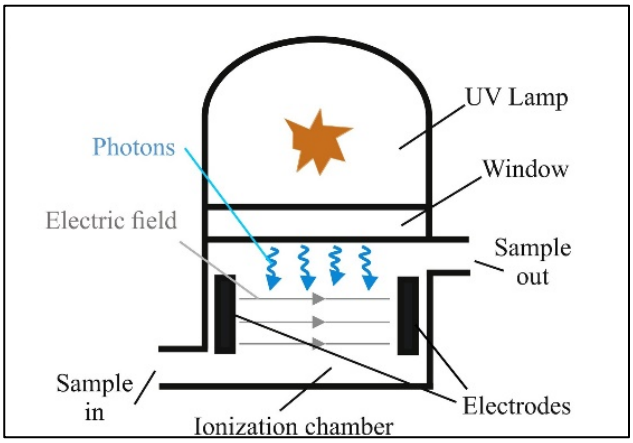


Figure 7.2-1: Ionization Process

Photons emitted by the UV bulb are absorbed by the sample molecules resulting in the formation of a charged molecule. When the photoionization energy emitted by the lamp is more than the photoionization energy of the sample molecules, it results in the generation of charged molecules and free electrons.

hv + Molecule -> Molecule (charged molecule) + e-* *Equation 11*

The ions generated are collected by the electrodes by applying a bias voltage of a few hundred Volts. The design of the chamber is one of the crucial factors affecting the sensitivity, selectivity, response time and linearity of the device.

The core components of a PID design are the main circuitry, the ionisation chamber and the source, and electrodes.

7.2.1. Ionisation source:

The ion source or ionisation source used in a PID is typically a UV lamp. The UV lamp is comprised of a glass lamp body and a crystal window, where the crystal window serves as the exit window for the UV photons essential for the photoionization process. The level of selectivity exhibited by the ionization source is reliant on the photoionization energy of the photons, which can vary depending on the gas and window crystal utilized. For instance, krypton emits radiation at 123.9 nm and 116.9 nm. Commercial ionization sources currently available on the market are capable of producing ionization energies ranging from 8.3 eV to

11.8 eV [16, 17]. The ionization process that takes place in the UV lamp is a function of the energy produced by the gas discharge, which excites the inert gas molecules and leads to the emission of photons. These photons then enter the crystal window and interact with the gas sample in the ionization chamber, ionizing it into charged molecules that can be detected by electrodes. The selection of the appropriate gas and crystal window combination plays a significant role in enhancing the sensitivity and selectivity of the ionization source, which is essential for effective gas detection. Table 7.2-1 contains the list of material used for the fill gas and window crystal. Transmission windows combined with gases filled in the lamps also determine the life of the lamp and the selectivity. The average life of UV lamps can vary depending on the specific manufacturer, model, and usage conditions. Some lamps may last around 500 to 1000 hours, while others designed for more heavy-duty or industrial applications may last up to 2000 to 4000 hours. The most common lamp is the 10.6 eV lamp as it is the strongest and longest-lived lamp. It uses krypton as fill gas and a magnesium fluoride window [18]. The UV lamp used is in this research is a 10.6 eV Heraeus UV lamp shown in Figure 7.2-2.

Figure 7.2-2: Heraeus UV lamp

Table 7.2-1: Transmission Window and fill gas

Nominal lamp Photon Energies (eV)	Fill Gas	Window Crystal	Crystal Transmittance wavelength range (nm)
11.7-11.8	Ar	LiF	105-5000
10.6	Kr	MgF_2	115-7000
10.2	H_2	MgF_2	115-7000
9.8-10.0	Kr	CaF_2	125-8000
9.5-9.6	Xe	BaF_2	135-9900
9.5	O_2	CaF_2	135-9900

8.4	Xe	Al_2O_2	145-4500
8.4	Xe	SiO_2	145-2300

The lamps can be driven by either a DC or RF mode, with the RF mode being particularly advantageous for miniaturisation. In the RF mode, the lamp is excited by the voltage applied to the conductors present external to the lamp. The RF excited UV lamps consist of a cylindrical conductor placed concentric to the lamp cavity, which is used for RF coupling to the external conductors. The lamp does not require internal electrodes, and hence can be miniaturised. A high voltage is applied to the external conductors, which excites the gas inside the lamp. The excited gas absorbs all the energy corresponding to the frequency of the UV light and transmits the light out through the transmission window. These transmitted photons enter the UV chamber where they ionise the gas molecules.

The RF mode has a higher energy efficiency and a longer lamp lifetime than the DC mode. The increased lamp life is primarily attributed to the more efficient and gentler ionization process that occurs in the RF mode.

In RF excitation mode, the UV lamp operates at a higher frequency, typically in the radio frequency range (e.g., 13.56 MHz). The RF power is applied to the lamp using a high-frequency oscillator, which creates an oscillating electric field within the lamp. This oscillating electric field induces rapid ionization and recombination of the gas molecules inside the lamp, producing UV photons. On the other hand, in DC mode, the lamp operates at a constant direct current, which does not provide the same level of efficient ionization and recombination as RF excitation. The lower efficiency in DC mode can result in a lower UV light output for the same electrical power input.

The rapid and efficient ionization and recombination processes in RF mode cause less wear and tear on the lamp components compared to the continuous ionization and recombination processes in DC mode. As a result, the RF mode can significantly extend the lifespan of the UV lamp. In some cases, the lamp life can be extended to thousands of hours in RF mode, while in DC mode, it might only last for hundreds of hours [19, 20].

7.2.2. Ionisation chamber:

The ionisation chamber is a crucial component in the operation of the PID and serves as the location for the detection and measurement of ionised chemicals. It is equipped with two inlets for the gas to pass through, a cavity where the gas is contained, a hole for the placement of the UV bulb, and two electrodes (anode and cathode) for the measurement of ionised chemicals.

The gas enters the chamber through one of the inlets, where it comes into contact with UV radiation emitted by the lamp. When the UV radiation strikes a gaseous molecule, it results in the emission of a free electron, which is then collected at the detector. This generates an ionisation current, which is measured by the electrodes.

The size of the ionisation chamber is an important factor that affects the performance of the PID. The size of the chamber plays a role in determining the sensitivity, response time, and signal-to-noise ratio of the device. Several studies have shown that a larger ionisation chamber leads to a lower response time, whereas an excessively small chamber can result in lower sensitivity. Therefore, the size of the ionisation chamber must be chosen carefully to optimize the performance of the PID [21-23].

7.2.3. Detector:

When UV light with sufficient energy strikes a molecule, it ionizes it by knocking off an electron, leaving behind a positively charged ion and a free electron. These free electrons are collected by two parallel conducting plates, which are maintained at a high voltage, creating an electric field that attracts the electrons towards the detector plates and repels the positively charged ions. This movement of electrons towards the plates leads to the generation of a current, which is directly proportional to the concentration of the gas sample.

7.3. Materials and Methodology:

7.3.1. Design of first-generation prototype of PID+:

The main elements of the first-generation prototype of PID+ are the power supply module for the UV lamp, the detection and amplification circuit, and an embedded signal processing circuit. Figure 7.3-1 illustrates the fundamental block diagram of the first generation PID+ prototype. The power circuit or power supply unit takes power from an external DC supply and regulates it and powers the device. A microcontroller is used to control, operate, and convert the various signals used for the operation of the PID+ and to convert the signal produced by the PID+ in to a readable format. The bulb circuit powers up the UV lamp used for the ionization. The output from the lamp photo-ionises gaseous molecules. The bulb used is a 10.6 eV Heraeus Photoionization lamp. The circuit uses an H-bridge and a step-up transformer. The transformer is used to produce a high voltage to turn on the UV bulb. The detector unit is used to detect the ionised gases and convert them into current. This generated current is amplified using an amplifier unit and read using a data logger.

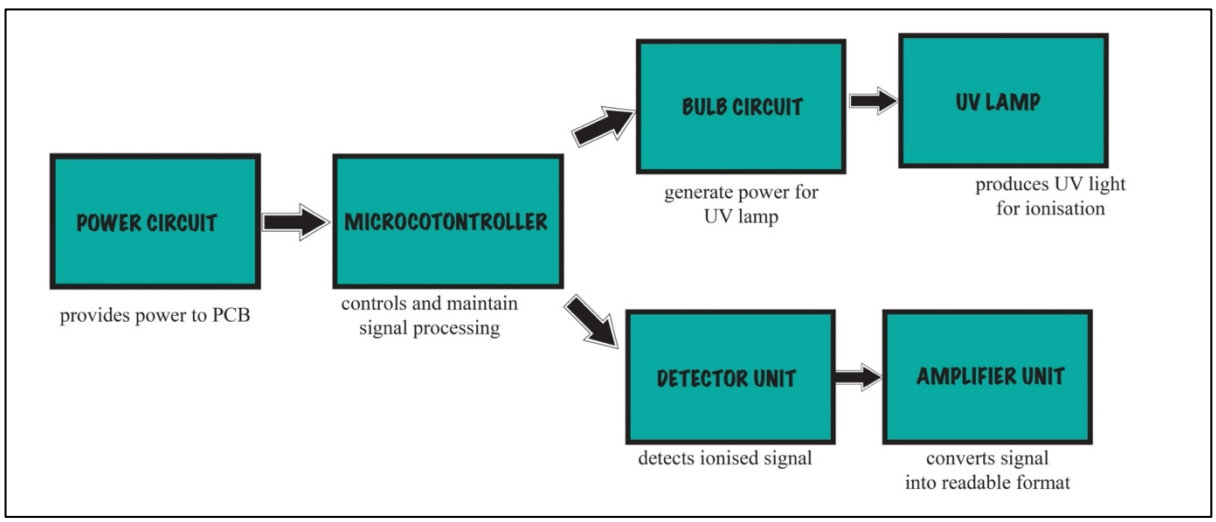


Figure 7.3-1: illustrates the fundamental block diagram of the first generation PID+ prototype.

a. Power Supply and microcontroller unit:

The main function of the power supply unit was to deliver the power to the different components/ ICs on the device. Figure 7.3-2 shows a schematic diagram of the power supply and microcontroller unit of the first generation PID+ prototype. The main components of the power supply unit were voltage regulators. The voltage regulators were used to generate a stable and constant +/- 5V DC and +/- 3.3V supply from an external 12V DC power supply and deliver it to the different components populated on the board. The microcontroller used in this board was an Arduino M0. The Arduino M0 is powered by an Atmel SAMD21 MCU, featuring a 32-bit ARM Cortex® M0 core. The microcontroller was used to control the operation of the sensor and interpret the output of the sensor.

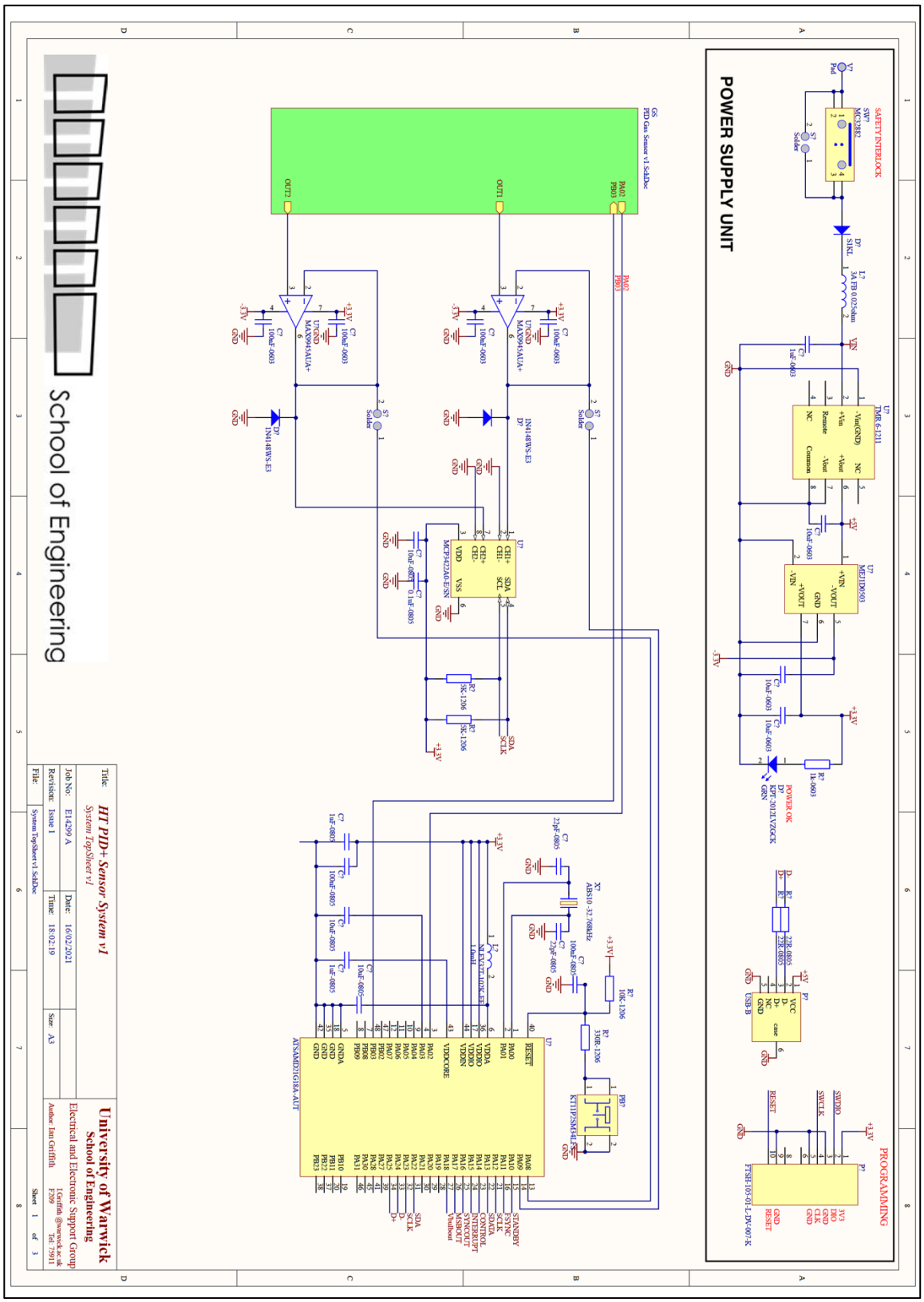


Figure 7.3-2: Power Supply and Micro-controller Unit

b. Bulb circuit and UV lamp/UV chamber:

The bulb circuit consisted of the power supply required for the excitation of UV lamp. The power supply required for the UV lamp was high frequency (around 100 KHz) and high voltage (around 1000V). It was required to have a stable frequency and stable voltage to maintain and prolong the life cycle of the UV lamp. The bulb circuit comprised of an H-bridge and a step-up transformer. The H-bridge was used to take the DC voltage supply from the power supply unit and use internal flip-flops to convert the DC voltage into 50% duty cycle supply which was further fed to the primary of a step-up transformer. The output signal generated by the bulb circuit is used for the excitation of the UV lamp which was placed in a resin printed UV chamber which was sealed so that there was no leakage of UV light as well as no contamination of the gaseous samples.

The bulb circuit powers up the UV lamp used for the ionization. The output from the lamp photo-ionises the gaseous molecules. The bulb used is a 10.6 eV Heraeus Photoionization lamp. The circuit uses an H-bridge and a step-up transformer. The transformer was used to produce a high voltage to turn on the UV bulb. Figure 7.3-3 shows the ionisation chamber with two inlets for the passage of the sample and represents the pictures and CAD drawings of the first version of the UV chamber.

The ionisation chamber used for this research was 3D printed and measured 43.81*17.75*8.07 mm. The dimensions of the ionization chamber were carefully selected to achieve a compact and lightweight design. A smaller size was chosen to minimize the overall footprint of the PID+ device, making it more portable and convenient for various applications. Two $1/8^{th}$ tube fittings were screwed into threads on the sides to serve as inlets and outlets. This chamber extends to cover the electrode plates, guiding the ions from inlet to the outlet. The chamber was designed so that it assured leakage free flow of the test gas. The UV bulb is placed inside the photoionization chamber in such a way that the high energy photons travelled perpendicular to the flow of test gas. The placement of the UV bulb inside the photoionization chamber was strategically chosen to optimize the ionization process. By positioning the UV bulb in such a way that high-energy photons travelled perpendicular to the flow of the test gas, maximum interaction between the gas molecules and UV photons

was ensured. This configuration enhanced the efficiency of ionization, leading to better sensitivity and detection capabilities of the PID+ device.

The rationale behind these design choices was to create an efficient, reliable, and compact ionization chamber for the PID+ device. The combination of 3D printing technology, precise measurements, and thoughtful design allowed for a leak-free gas flow and optimized interaction between the UV bulb and the test gas. These features were crucial in enhancing the overall performance and sensitivity of the PID+ device, making it a valuable tool for gas detection and analysis in various applications.

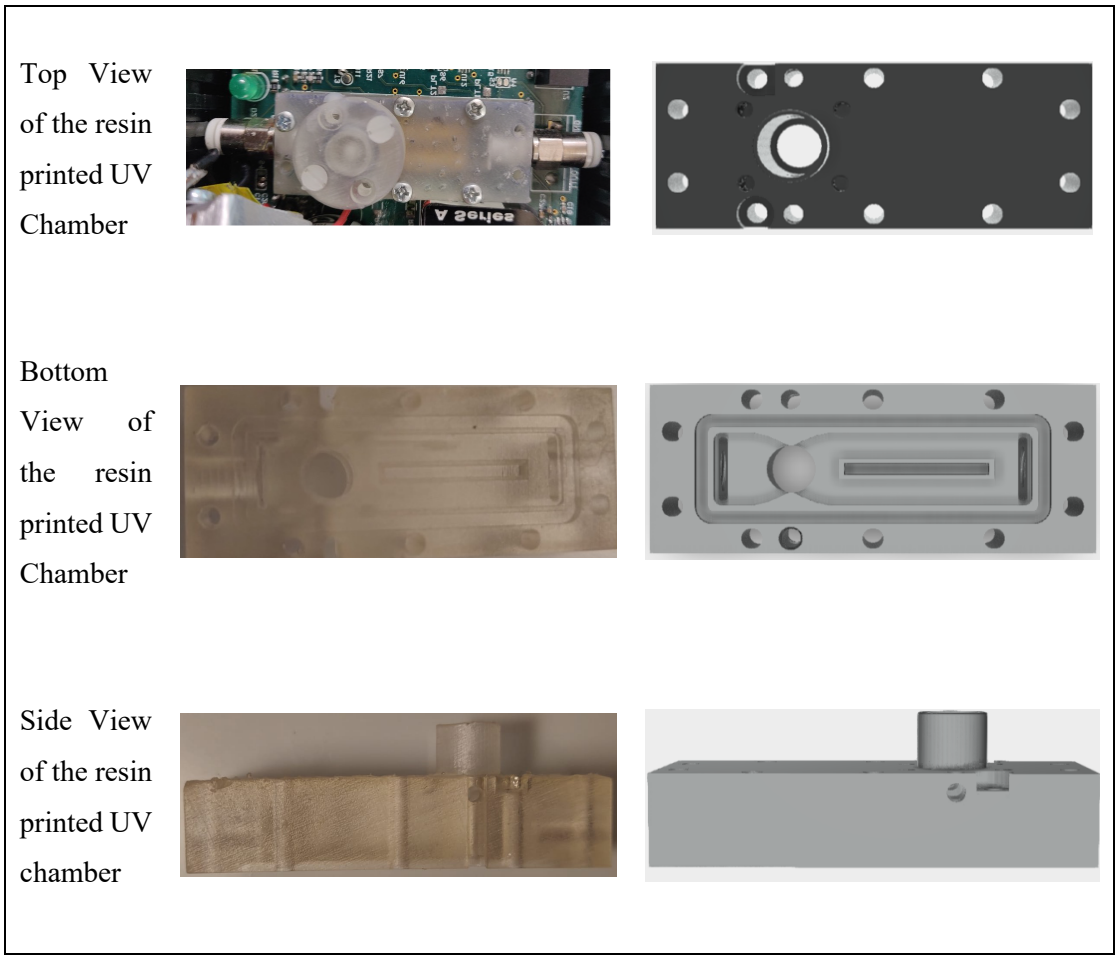


Figure 7.3-3: first version Ionisation chamber

c. Detector unit and Amplifier stage:

The detectors used were two parallel plates of conductor placed perpendicular to the UV lamp inside the UV chamber. The main purpose of these conducting plates was to collect

the ionised molecules/free electrons generated after becoming ionised by the UV radiation. An extremely high voltage was applied at the detectors to generate a high electric field. The high electric field was used to attract the free electrons or ions towards the detector plates. Figure 7.3-4 shows a picture of the detector plates.

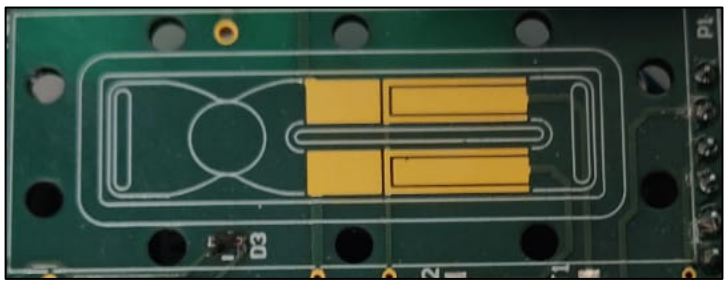


Figure 7.3-4: pictorial representation of the detector plates.

The detection unit consisted of a DC voltage (5V) to a High Voltage DC converter (500V). This high electric field was used to deflect the ions towards the electrode detectors where one of the electrode plates is supplied with the HV DC and the other one was grounded. The HV DC was altered by programming the microcontroller to study the effect of changing the bias voltage to the sensitivity of the electrodes. The program used to change the bias voltage can be found in Appendix A. The signal generated by the electrodes was fed to a two-stage amplifier to amplify the ion current obtained from both the electrodes.

In the first stage, the transimpedance amplifier was utilized to convert the current output from the electrodes into a corresponding voltage signal and the second stage was used for amplification of the signal to meaningful voltage levels. This approach was advantageous in providing high sensitivity to detect even the slightest changes in gas concentrations. By accurately converting the weak current signals to voltage, the PID+ could achieve precise measurements and respond effectively to low-level gas concentrations, making it a reliable gas sensor.

One of the primary considerations in selecting the amplifier components was to ensure low-noise performance. Low-Noise Op Amps were employed in the circuit design to minimize electrical noise and interference. This resulted in an improved signal-to-noise ratio, ensuring the accuracy and reliability of the PID+ output. The reduction of unwanted electrical noise further enhanced the PID+'s ability to detect gas concentrations accurately, especially in

environments with potential electromagnetic interferences. Another critical aspect of the amplifier strategy was the utilization of MOS-Input Op Amps. The MOS technology provided high input impedance, reducing the current drawn from the electrodes and minimizing signal loading effects. This characteristic was instrumental in preserving the sensor's electrical properties and maintaining the stability of the sensor's response.

Additionally, considering the PID+ device's intended use in various settings, the choice of Low-Power Op Amps was essential to optimize power consumption. These op amps were designed to be energy-efficient, allowing the amplifier stage to operate with minimal power consumption. As a result, the PID+ device could be used in portable and battery-operated applications, maximizing its operational time, and reducing the need for frequent battery replacements.

The outputs generated by the two electrodes were measured and illustrated as output 1 and output 2 during measurement. Figure 7.3-5 shows a schematic of the bulb circuit and detector circuit.

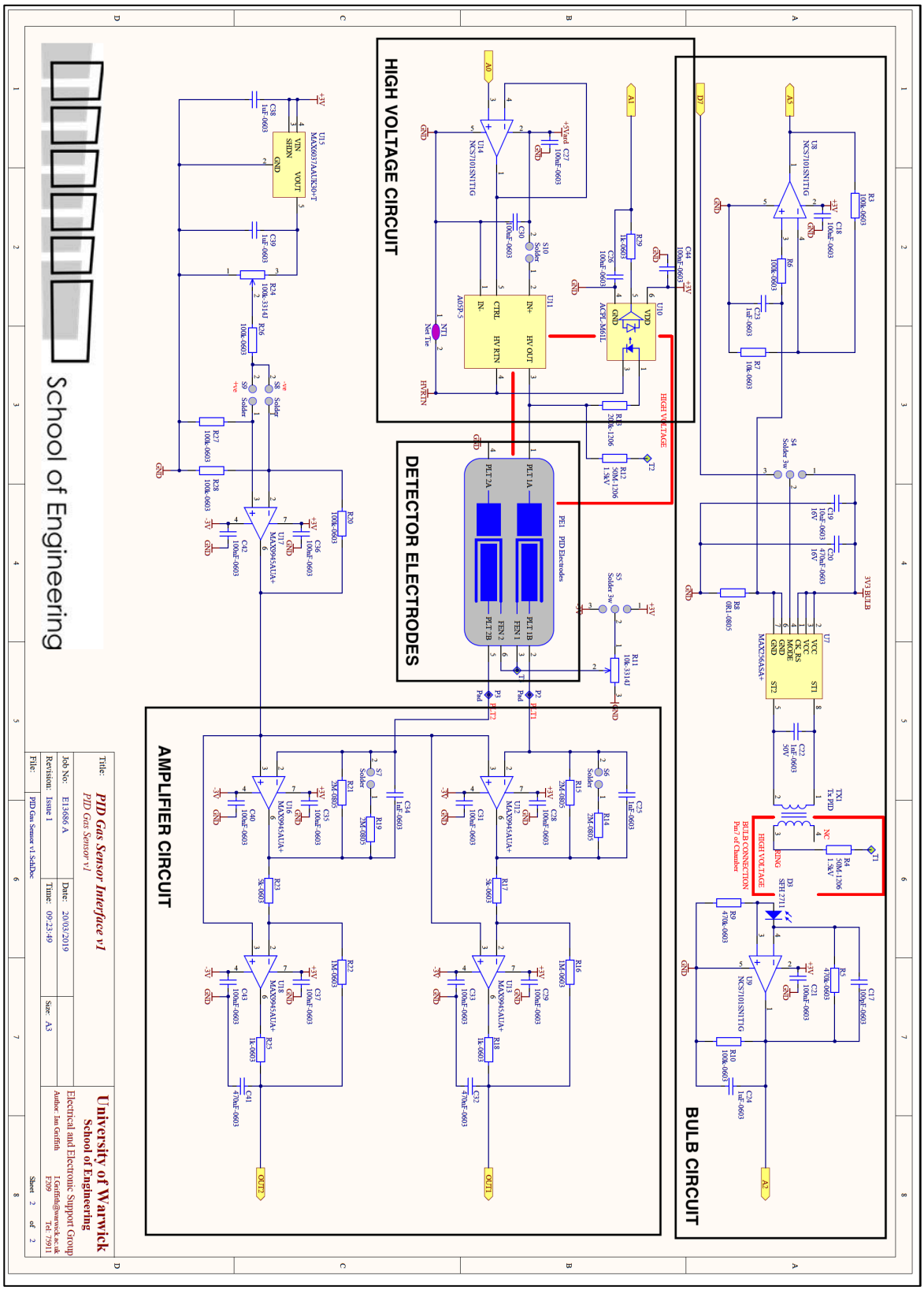


Figure 7.3-5: Schematic of the main circuitry

d. Hardware setup:

Figure 7.3-6 shows the hardware setup clearly illustrating all the main elements of the first generation PID+ prototype.

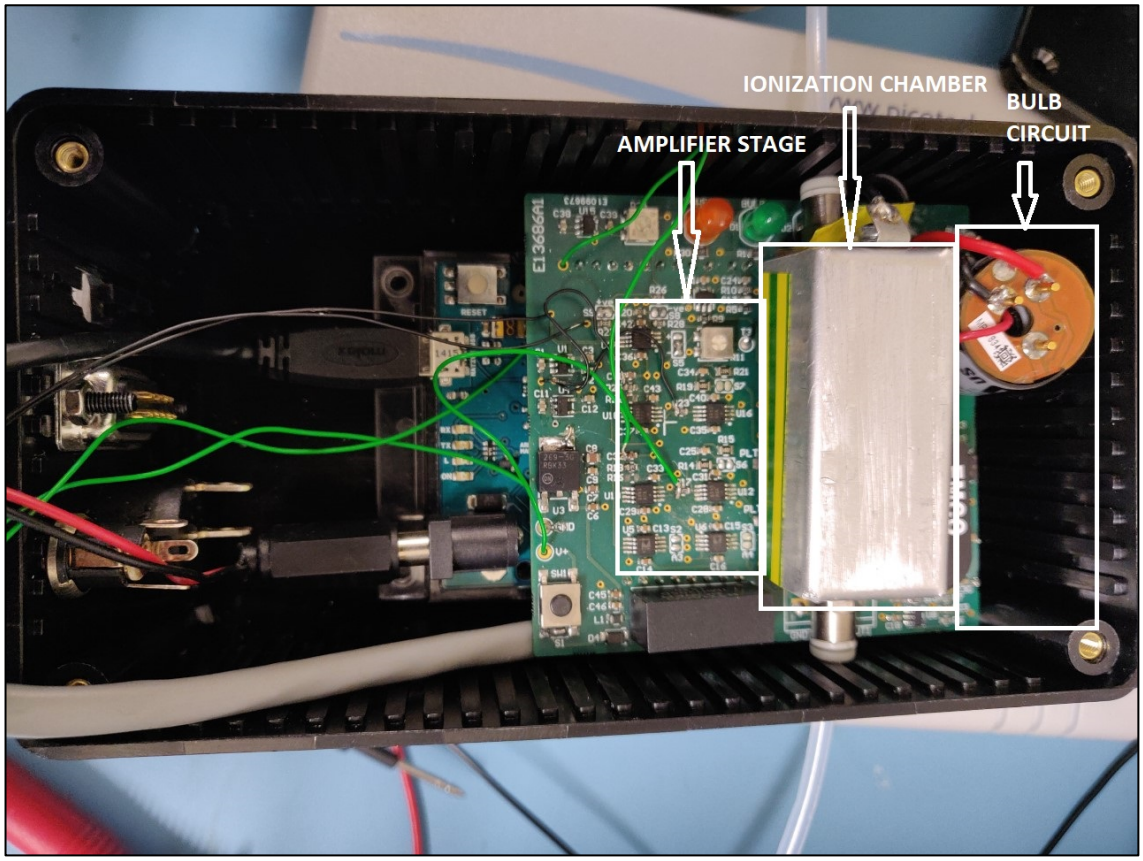


Figure 7.3-6: Hardware Setup

e. Experimental setup

The first generation PID+ prototype was tested using isobutylene (ISB). Isobutylene (ISB) had been used as the primary VOCs to test the performance of the first generation PID+ prototype. ISB is a common gas used for testing because it is inexpensive, readily available, has average sensitivity and is non-toxic [30].

To understand the sensitivity of the prototype, it was tested under various conditions. The testing conditions included changing the flow rates of the ISB, the concentrations of ISB

and the biasing of the electrodes. An API Dilution calibrator- Model 700 was used to generate the desired concentrations and flow rates of the gas and regulate them into the PID. The circuit was powered up with a power supply of 12 Volts and consumed 64 mA current. The output was visualised using a Pico ADC24 Data logger.

The results obtained from the various tests are described in chapter 8.

7.3.2. Design of second generation PID+ prototype:

The second generation PID+ prototype followed the same building blocks as the first generation PID+ prototype. The main fundamental changes were made to increase the sensitivity of the prototype which included integrating the microcontroller within the design, a change in the bulb circuit for the excitation of the UV lamp and some updates to the detector circuit.

a. Power Supply and microcontroller unit:

Figure 7.3-7 shows the schematic of the power supply and microcontroller unit. The second generation PID+ prototype was powered using either a USB or a DC power supply. USB-B connectors were connected to the PCB as well as to interface the PCB with a PC. 5V voltage applied to the board which was regulated to +/- 3.3 V for the different components on the PCB. Instead of an Arduino M0 board, a low-power, high performance ATSAMD21G18 ARM® Cortex®-M0+ based flash microcontroller was used. This helped in decreasing the size of the board and to simplify the board design. The main functions of the ATSAMD21G18 microcontroller were to control the functionality of the bulb circuit, control the bias voltage of the detector electrodes and stream the data generated by PID+ to the PC. A 10 position SMD header was populated on the board to boot load the microcontroller. The microcontroller was flashed using a SEGGER J-Link BASE JTAG/ SWD Debugger. Another important change made in this prototype was that an additional ADC (Analog-to-Digital Converter) was added to the outputs of the detecting electrodes.

The ADC was used to convert the analogue signal detected by the electrodes and convert it into digital form and process it using the microcontroller.

The idea of integrating the microcontroller directly onto the PCB in the second version of the PID+ device was to bring several benefits and improvements to the system. By eliminating the need for an external Arduino board, the overall size and footprint of the PID+ device were reduced, making it more compact and lightweight. This miniaturization proved to be advantageous for applications where space was limited, such as portable or handheld gas detectors. The direct integration of the microcontroller minimized the risk of loose connections or signal interruptions that could have occurred with an external board. Additionally, integrating the microcontroller on the PCB contributed to cost savings in the manufacturing process, as it reduced the number of external components required and streamlined the assembly steps. The ADC added to the circuit enabled the PID+ to record and store the digitized measurement data, facilitating data logging capabilities. This was especially valuable for monitoring gas concentrations over time and for capturing data during specific events or conditions.

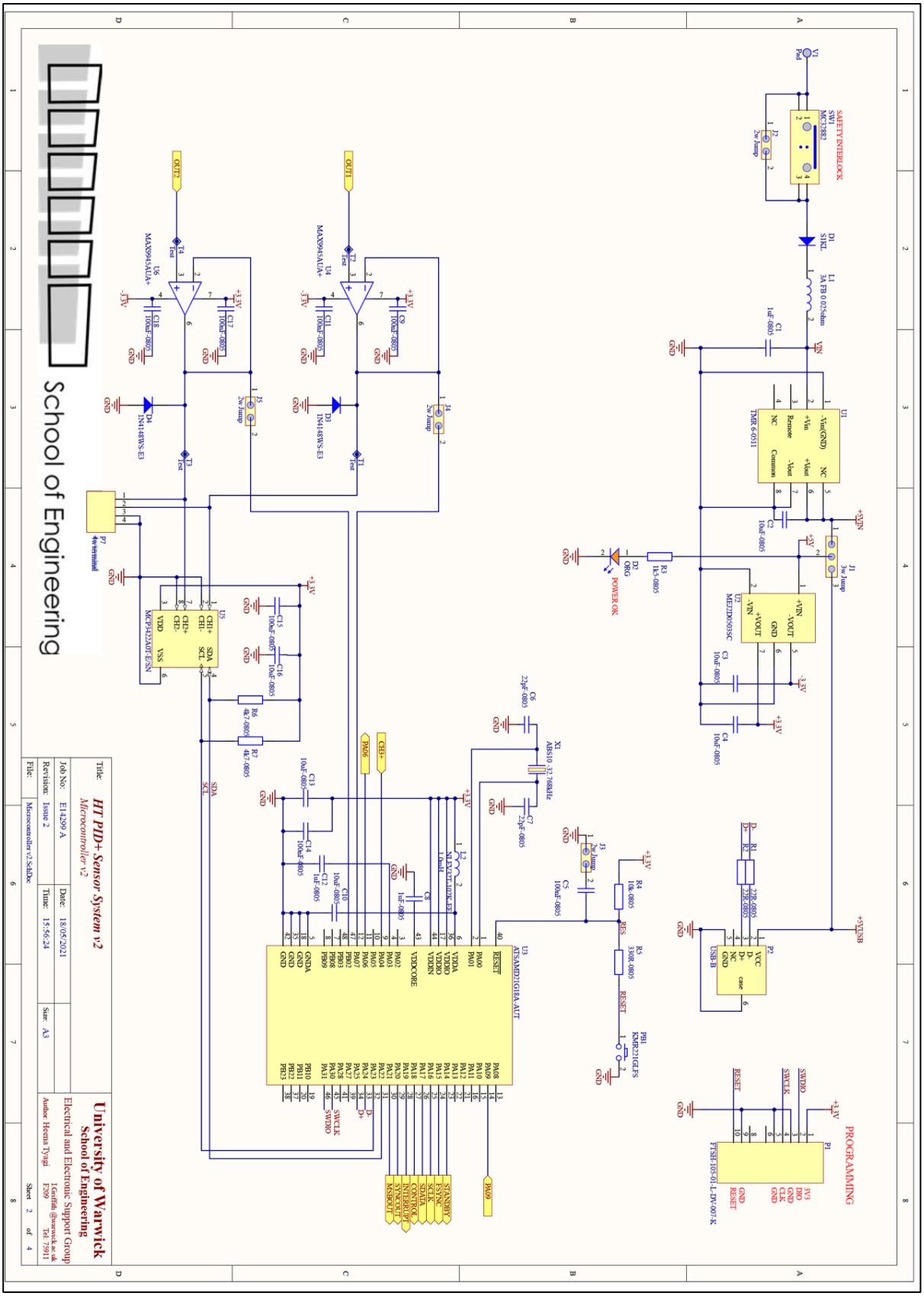


Figure 7.3-7: Schematics for Power Supply Unit and Microcontroller unit.

b. Bulb circuit and UV lamp/UV chamber:

The main changes to the bulb circuit were made in the circuit design for the excitation of the UV lamp. The main aim for the development of the bulb circuit was to generate a stable high frequency sine wave to drive the primary side of the transformer. Different concepts were applied for the sine-wave circuit, some of which are explained below:

1. Pulse Width Modulation-Low pass filter:

   The main concept behind the circuit shown in Figure 7.3-8 is to generate a Pulse Width Modulation (PWM) wave using the microcontroller and use an external RC (Resistor-Capacitor) low-Pass filter to generate the sine wave. When a PWM signal is passed through a low pass filter, it removes the high frequency component and converts the digital PWM signal into an analogue signal. A passive filter was used for this circuit. The passive filter has an advantage over an active filter as it offers low cost and reduced complexity. However, they can suffer from impedance loading issues and high distortions as can be seen in Figure 7.3-9.

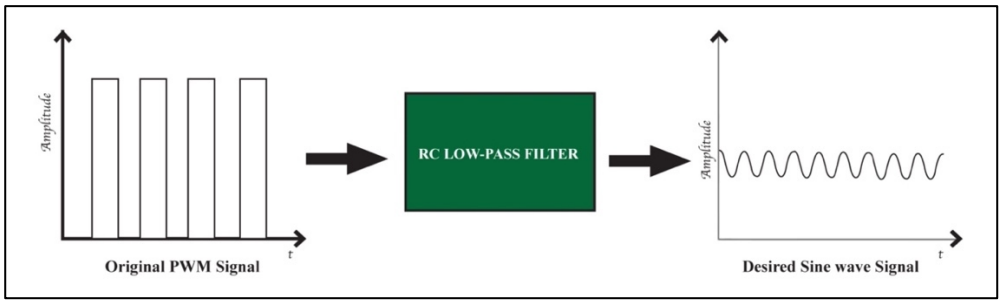


Figure 7.3-8: RC Filter for sine wave generation

Figure 7.3-9: output of the Pulse Width Modulation-Low pass filter circuit

2. H-Bridge using mosfet ICs:

An H-bridge with an RC filter is a circuit which consists of four switches with a load in the centre and a DC supply connected at the top and ground at the bottom. These switches were individually controlled in such a way that the voltage across the load changed polarity after each half cycle of the control signal which was being used to sequentially drive the switches. This control signal was a PWM signal generated by an Arduino. To generate a sine wave, two pins of the Arduino were used, one for the positive cycle and the other for the negative half cycle.

Figure 7.3-10 shows the circuit diagram for the H-Bridge circuit. The PWM generated by one pin (A) of the Arduino was applied to one p-channel mosfet (IRF4905) using a driver IC TC4421A which outputs an inverted signal and one n-channel mosfet (IRLZ34N) using a driver IC TC4422A and PWM generated by another pin (B) of the Arduino was applied to another p-channel mosfet (IRF4905) using a driver IC TC4421A and another n-channel mosfet (IRLZ34N) using driver IC TC4422A. When a high PWM signal was applied, the mosfet M1 (pmos) and M3 (nmos) turned on and when a low PWM signal was applied, the mosfet M2 (pmos) and M4 (nmos) turned on. The output generated by the H-Bridge was made to pass through a low pass filter which generated a sine wave. The output generated

through the circuit was not stable and could not produce a smooth sine wave. The output generated using this circuit is illustrated in Figure 7.3-11.

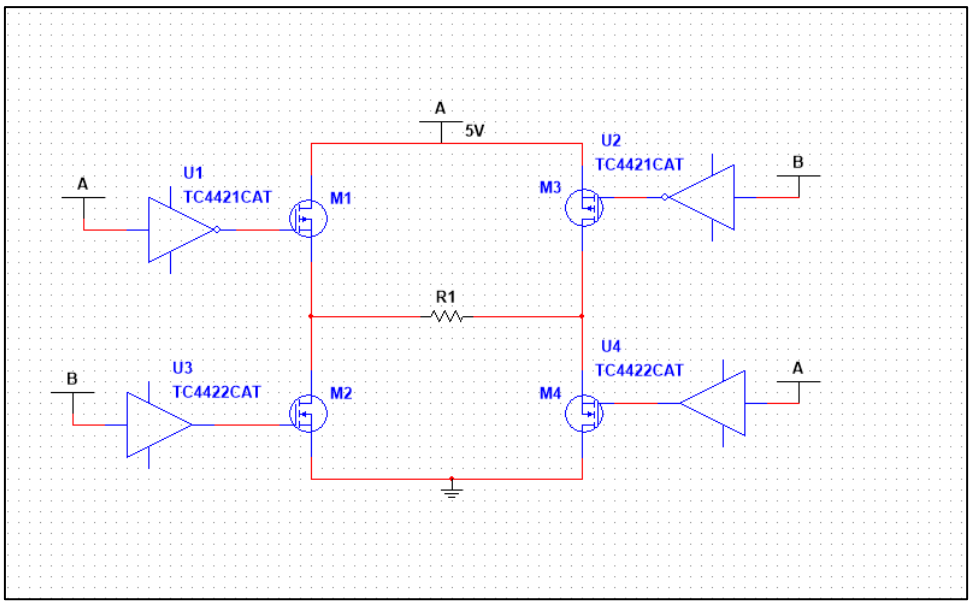


Figure 7.3-10: H-Bridge Schematics.

(a) (b)

Figure 7.3-11: (a) output from the H-bridge before the Filter, and (b) output of the H-bridge circuit after the filter

3. Sine wave generation using Common Cathode Florescent Lamp (CCFL) inverter:

CCFL inverters are devices that transform a DC voltage into an AC voltage using a resonant mode power supply called the Royer circuit, which comprises of two transistors and a centre trapped transformer. When an input voltage is applied to the base of one of the transistors, it enters the saturation region and causes the other transistor to enter the cut-off region as it is connected to the collector of the first transistor. The first transistor then enters the cut-off

region, while the second transistor turns on, generating a continuous output that is applied to the primary side of the transformer. Figure 7.3-12 shows the schematic for the CCFL inverter used in this approach, while the experimental setup is shown in Figure 7.3-13, where a neon bulb was used instead of a UV bulb to prove the concept. However, the drawback of the CCFL inverter is that, in order to extend the lifetime of the UV bulb, it is necessary to generate a stable voltage and resonance frequency, which changes during the bulb's lifetime, making the CCFL inverter non-adjustable.

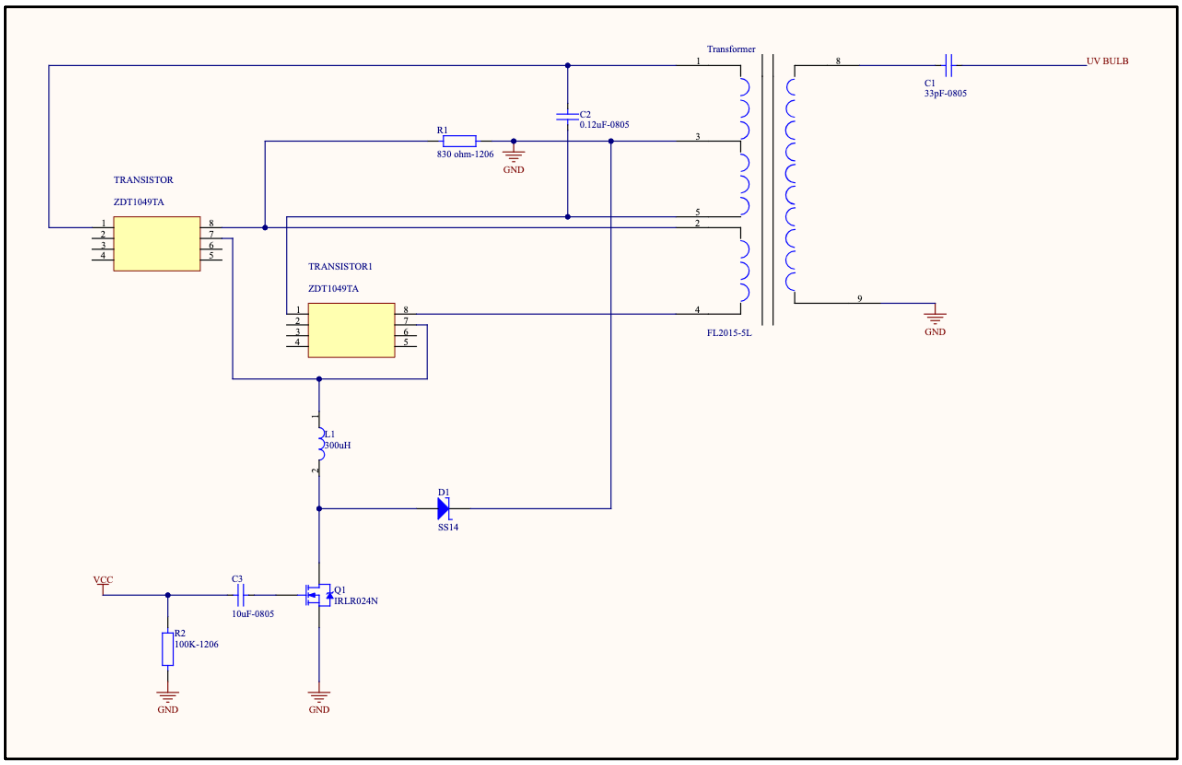


Figure 7.3-12: Schematic for the CCFL inverter

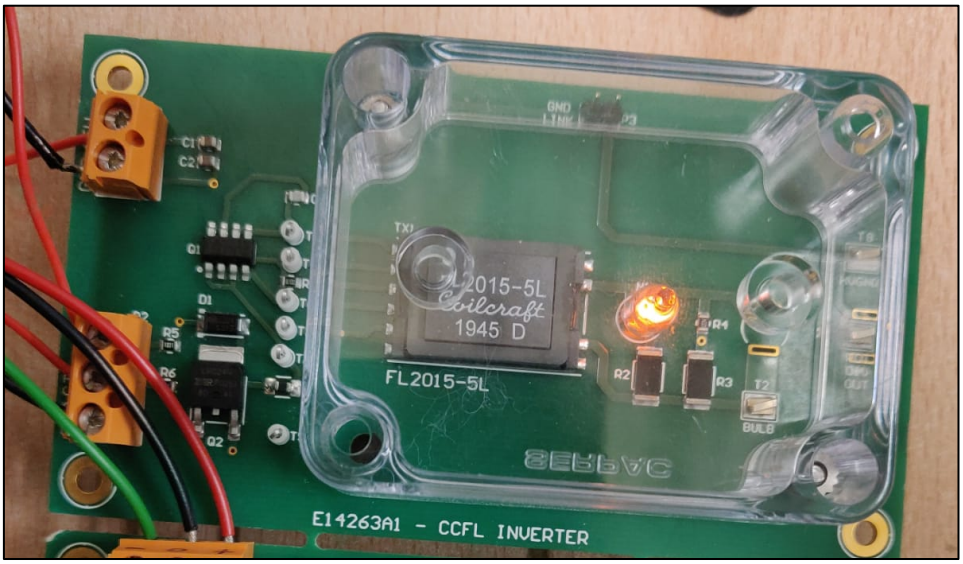


Figure 7.3-13: Experimental setup for the CCFL inverter

4. Sine wave generation using an AD5932:

The AD5932 is a programmable frequency generator, which can be used to generate sinusoidal wave, triangular wave, or square wave. The AD5932 was powdered by a 3.3V external power supply which was separated into digital power and analogue power. Some inductors and capacitors were used at the supplies in order to remove the noise and ripples. The AD5932 was programmed using an Arduino M0. The code for the AD5932 can be found in Appendix A. The AD5932 was programmed for a range of the frequencies from 90KHz to 130KHz. The schematic for the circuit and the experimental setup is shown in Figure 7.3-14 and Figure 7.3-15. The output generated was a sine wave as shown in Figure 7.3-16.

The key innovation of this circuit was to leverage the inherent property of the UV bulb, where it generates a frequency equal to its resonance frequency when powered up. As the UV bulb degrades over time, its resonance frequency shifts. Exploiting this phenomenon, an AD5932 integrated circuit was incorporated into the circuit design. The AD5932 was programmed to conduct a frequency scan from 90 KHz to 130 KHz at every power-up sequence of the PID. At each 1 KHz interval, the AD5932 measured the voltage produced by the UV bulb. By correlating the resonance frequency of the UV bulb with the corresponding voltage measurements, the AD5932 effectively determined the lowest voltage value during the frequency scan. This approach allowed the circuit to automatically

set the appropriate operating frequency for the UV bulb, ensuring optimal performance. A bespoke firmware was written to control the circuit's operation effectively.

The aforementioned measures were implemented with the primary objective of prolonging the operational lifespan of the UV bulb while concurrently optimizing its efficiency, thereby effectively reducing the recurrent costs associated with frequent bulb replacements. In summary, the developed circuit and firmware offered a novel and automated method to monitor and adjust the operating frequency of the UV bulb based on its resonance characteristics. This adaptive approach aimed to maximize the efficiency and longevity of the UV bulb, optimizing its performance as a vital component of the in-house PID.

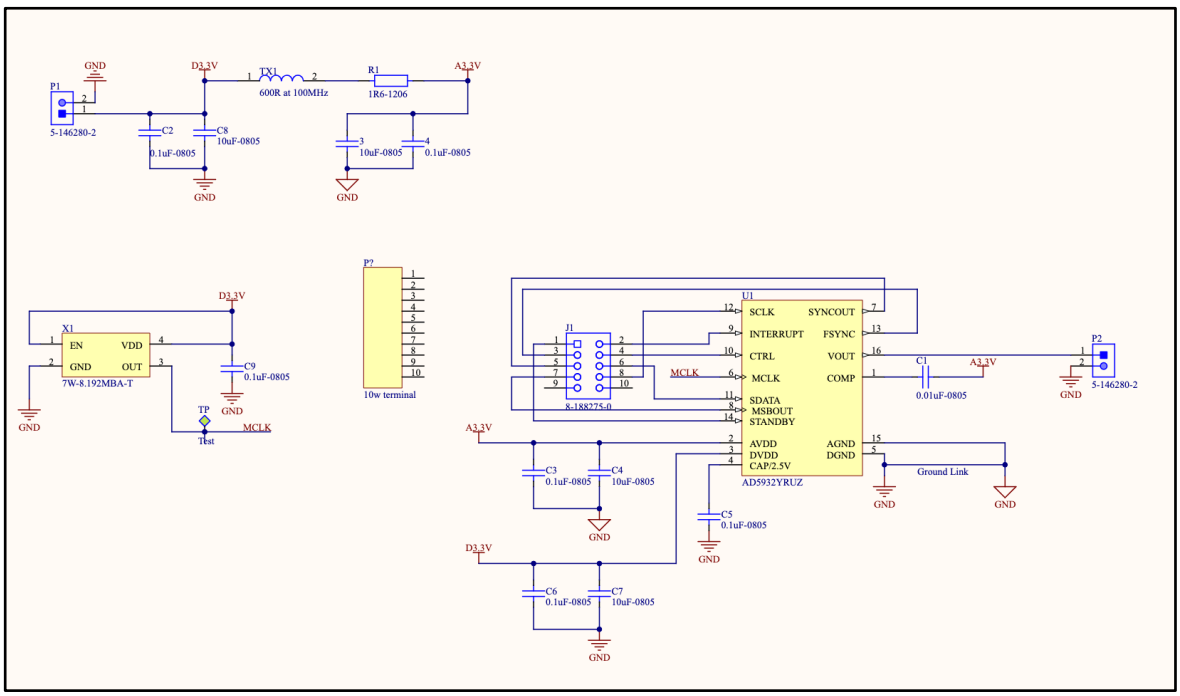
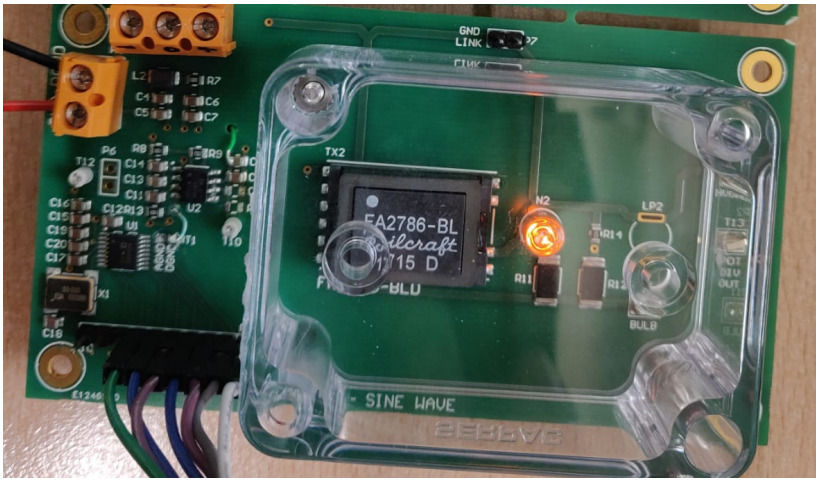
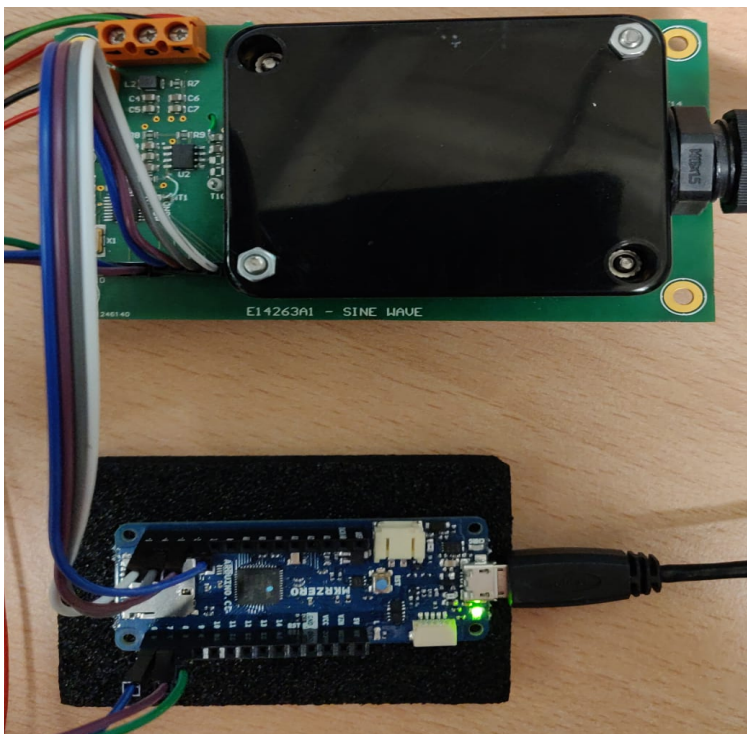


Figure 7.3-14: Schematics for the AD5932 sine wave generation circuit.



Experimental setup for the AD5932



Setup to test the AD5932 with a UV lamp

Visualisation of the UV bulb using an inspection camera

Figure 7.3-15: Experimental setup from AD5932 sine wave generation.

Figure 7.3-16: Output from the AD5932 sine wave generation.

Figure 7.3-17 shows the schematic for the bulb circuit. As discussed above, the AD5932 was used to generate a sine wave which was programmed using the microcontroller present on the PCB. The signal generated by the AD5932 was amplified using an op-amp and fed to the primary of a transformer. The output generated by the transformer was used to excite the UV bulb. The signal from the primary of the transformer was fed to an amplifier and a filter circuit to monitor the stability of the circuit.

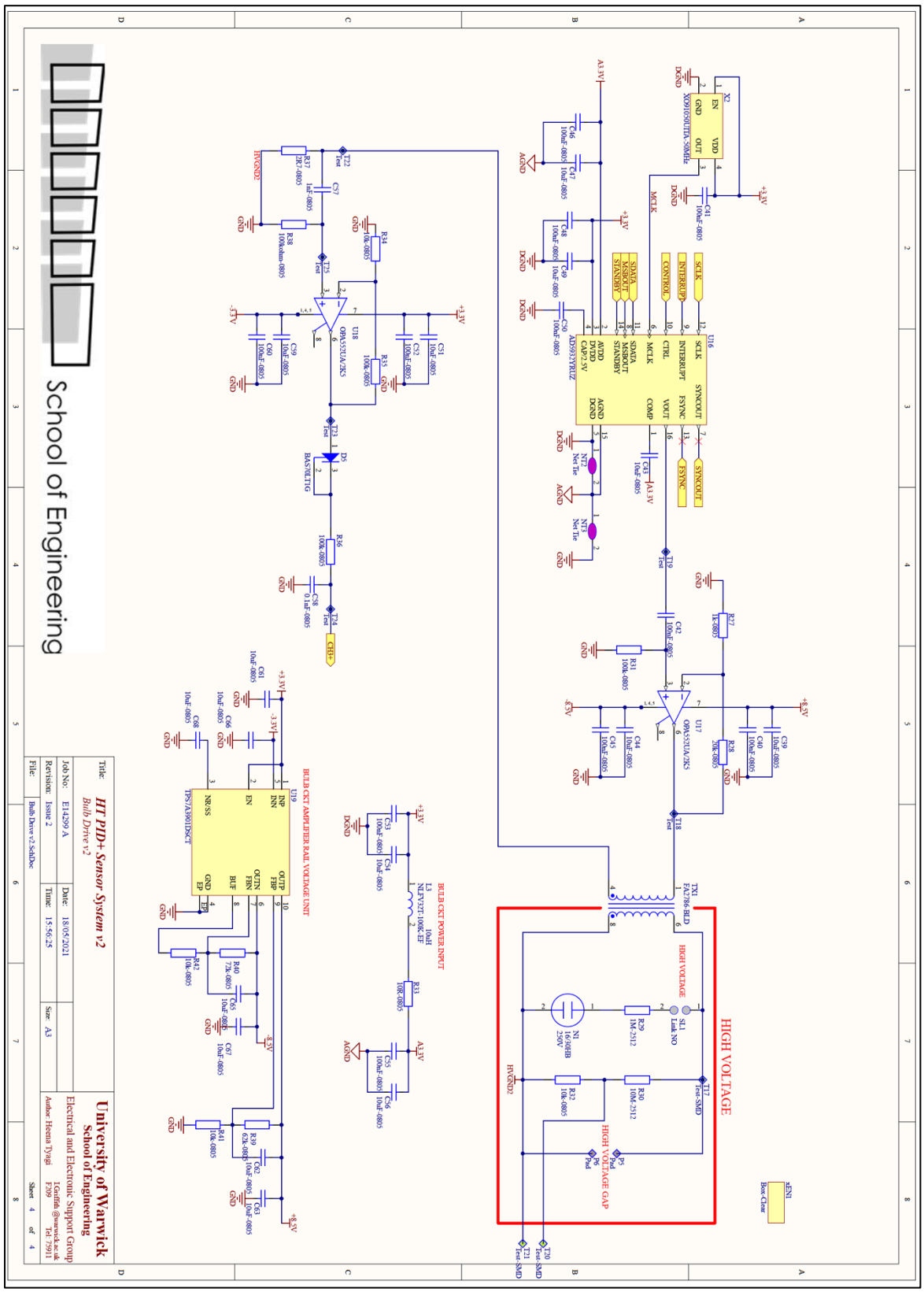


Figure 7.3-17: Schematic for the bulb circuit

c. Detector unit and Amplifier stage:

There were few changes to the detector circuit. The main changes were that the high voltage regulator IC was changed to generate 1000V DC instead of 500V. Also, the feedback resistors for the amplifier stage were changed to potentiometers to experiment with the gain for the amplifiers to generate better outputs. In the physical design of the board, the detector plates and the fencing around the detector plates were increased to avoid electric arc generation due to the increase in the bias voltage from 500V DC to 1000V DC. The schematic for the detector and amplifier stage are shown in Figure 7.3-18 respectively.

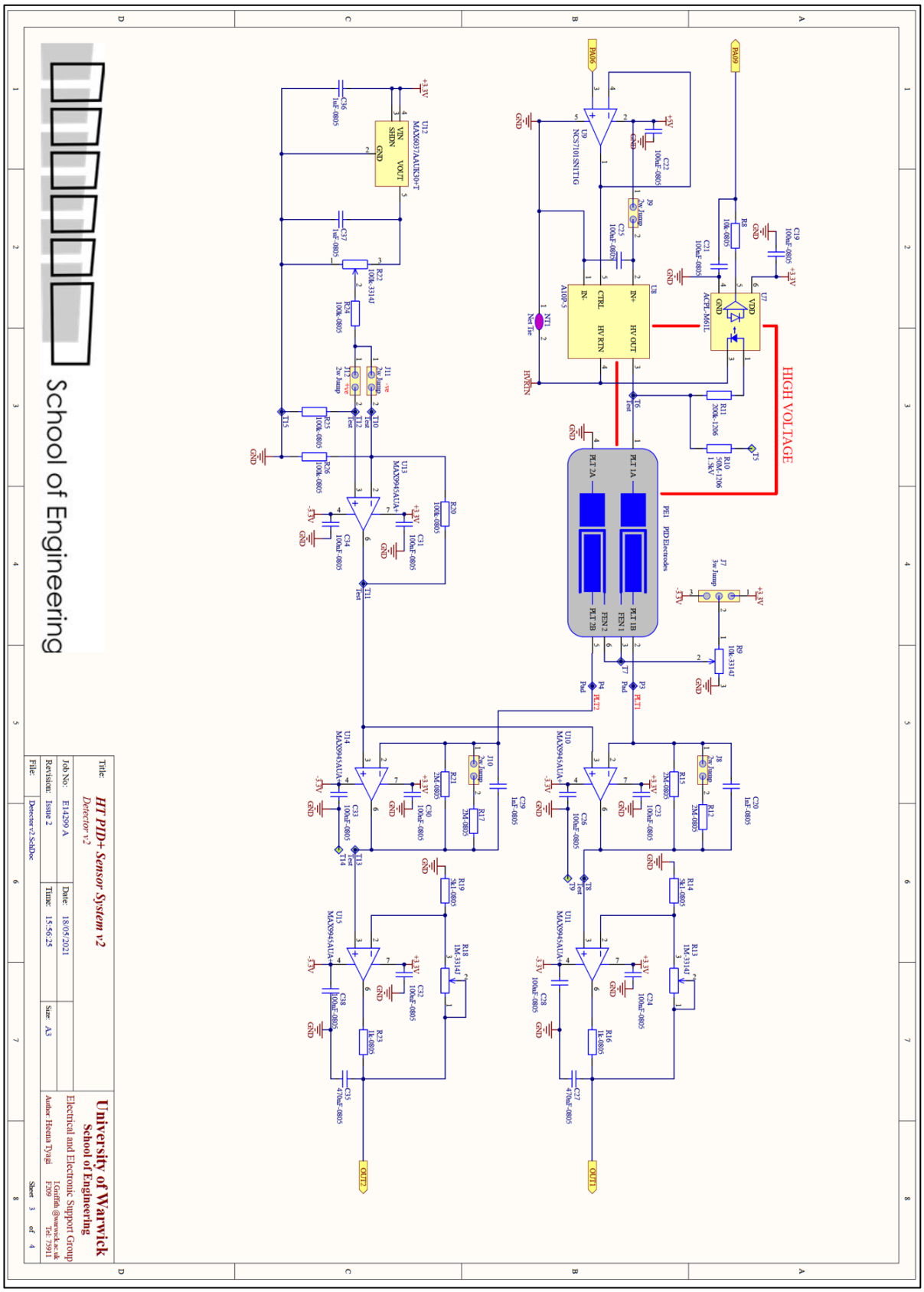


Figure 7.3-18: Schematic for detector and amplifier stage for second version PID+ prototype

d. Hardware setup:

Figure 7.3-19 shows the hardware setup for the second generation PID+ PCB. The high voltage circuitry was concealed under a 3D printed enclosure. An inspection camera was fitted on the side of the UV chamber to visualise the UV lamp. The opening for the camera was concealed using rubber stoppers to stop any leakage of UV light.

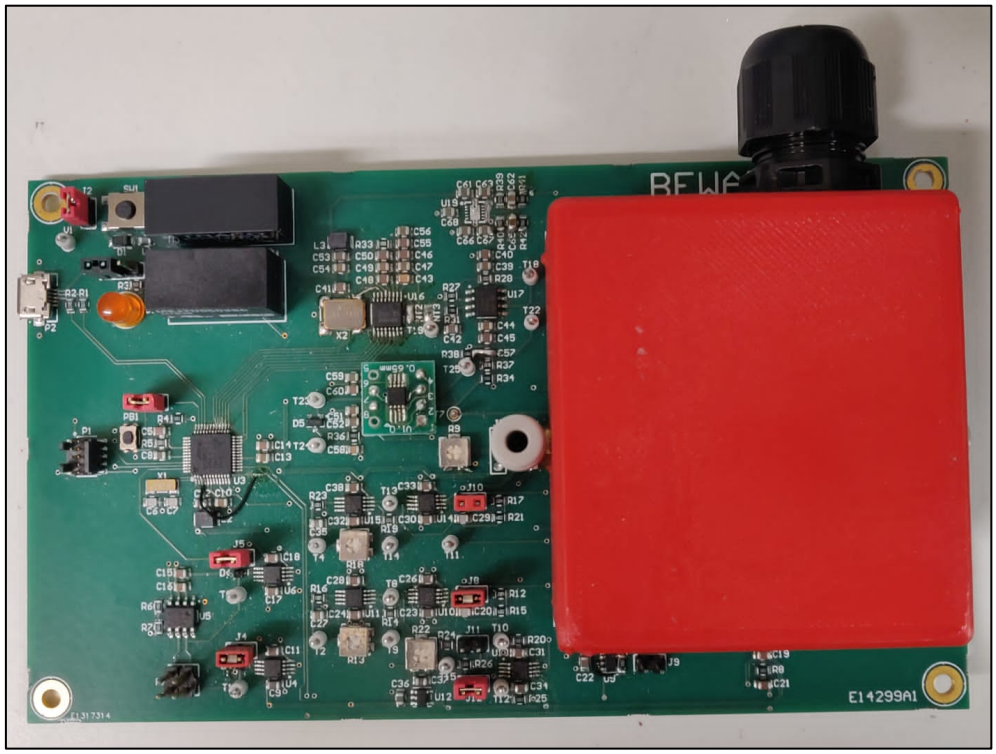


Figure 7.3-19: Hardware setup for the second generation PID+ prototype

7.3.3. Design of third generation PID+ prototype:

a. Power Supply and microcontroller unit:

The main idea behind the changes made in third version PID+ prototype was to make the device handheld. Therefore, a rechargeable battery and a battery charger were added to the design. The battery charger circuit involved a load sharer circuit using an MCP73871. In the

presence of the rechargeable battery, the PID+ prototype was powered through the battery and when the battery was discharged and the USB power was connected, the MCP73871 would stop powering the board with the battery and would power directly through the USB while allowing the battery to charge normally. This input power was then used to power the whole board through different voltage regulators. Figure 7.3-20 shows the schematic for the battery charger circuit and Figure 7.3-21 shows the schematic for the power supply and microcontroller unit for the third-generation prototype.

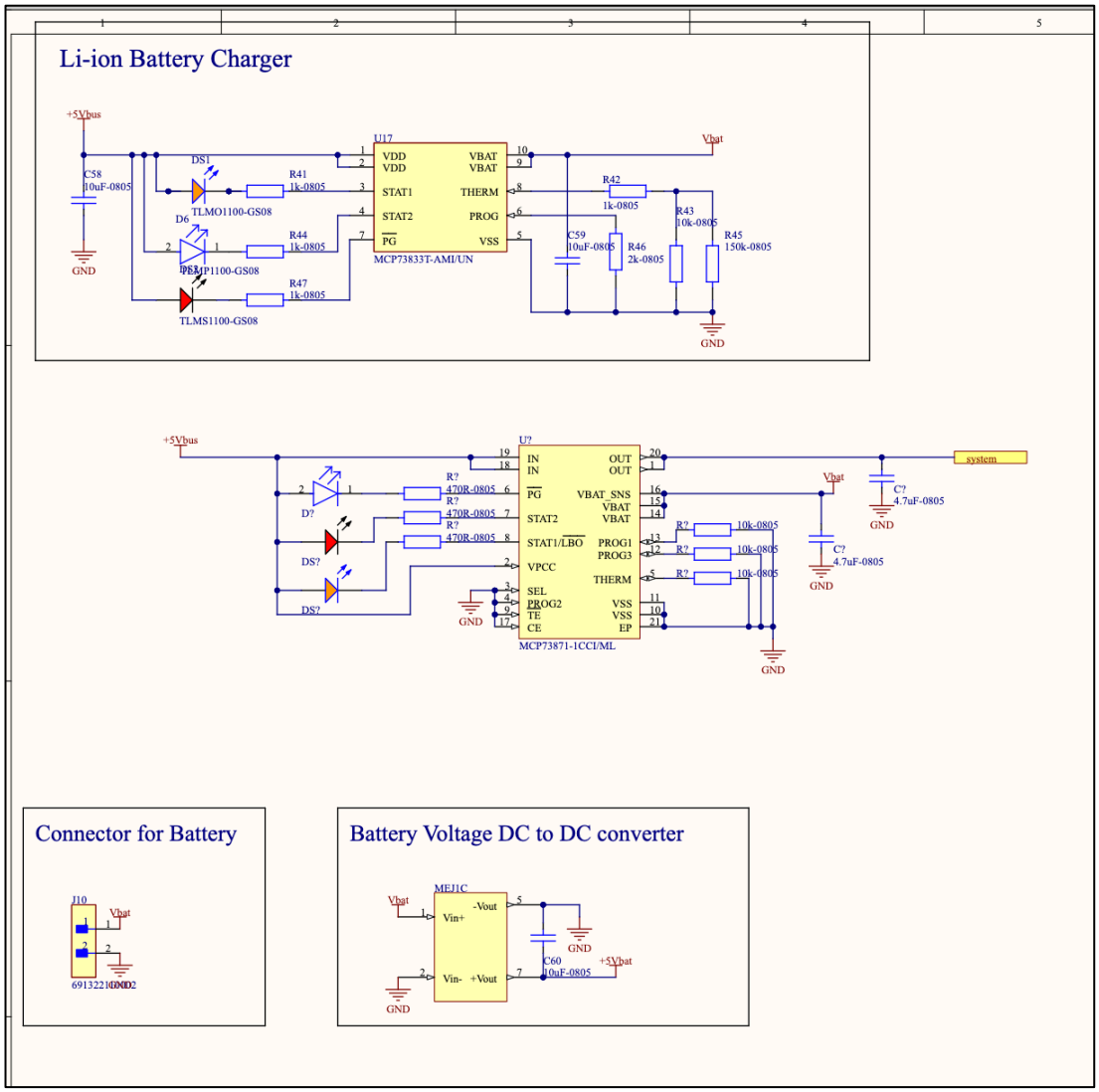


Figure 7.3-20: Schematic for the battery charger circuit

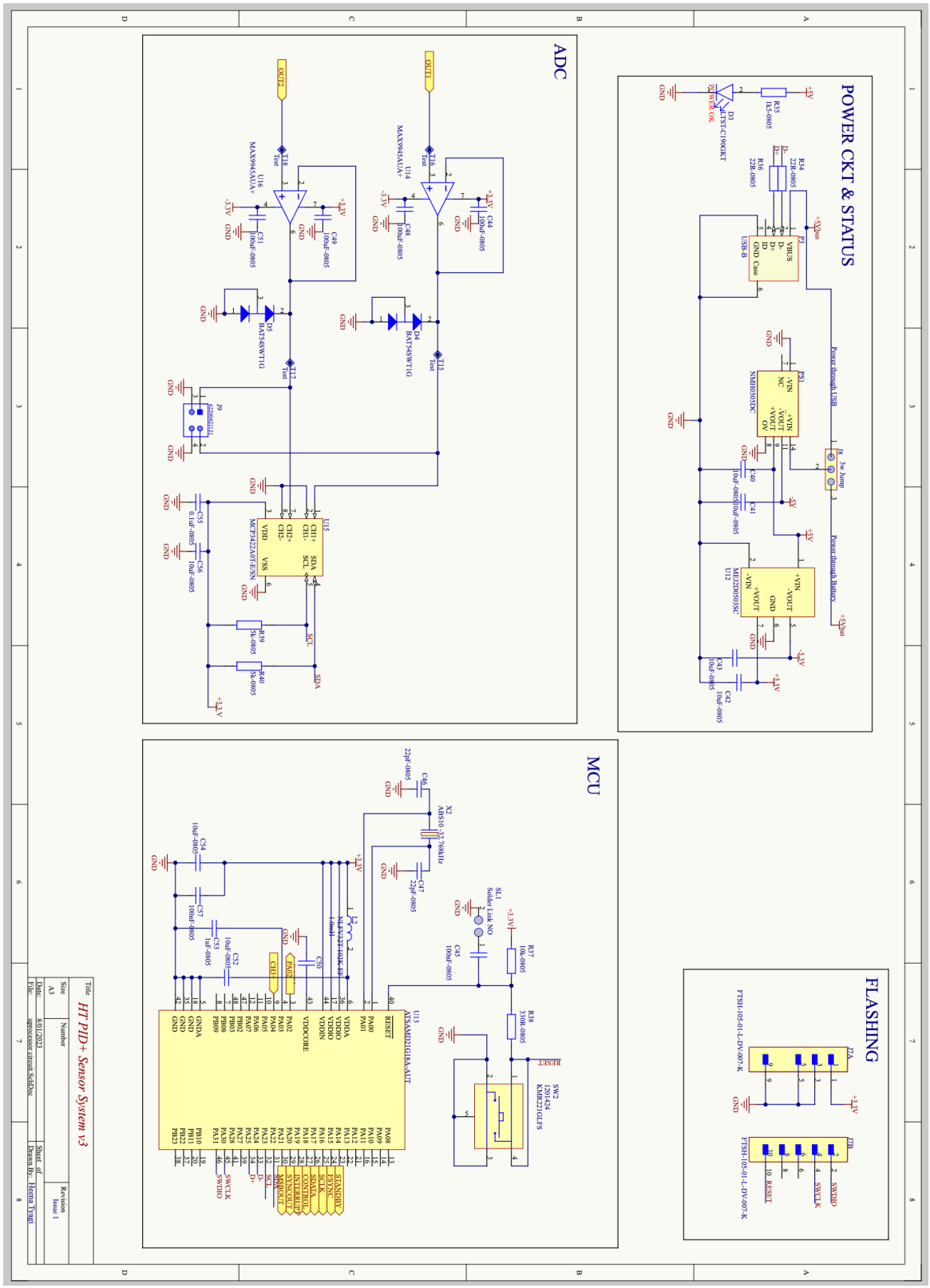


Figure 7.3-21: Schematic for the power supply and the microcontroller unit.

b. Bulb circuit and UV lamp/UV chamber:

The bulb circuit design consisted of a bulb circuit power supply, which was used to generate and separate the analogue and digital power supplied required for the AD5932. An integrated code was written to generate the signal using an AD5932, scan resonant frequency for the bulb circuit, read the ADC converter and for controlling the bias voltage added to Appendix A. The other changes made to the bulb circuit design included changing the amplifier and changing the feedback circuit. Figure 7.3-22 shows the schematic for the bulb circuit.

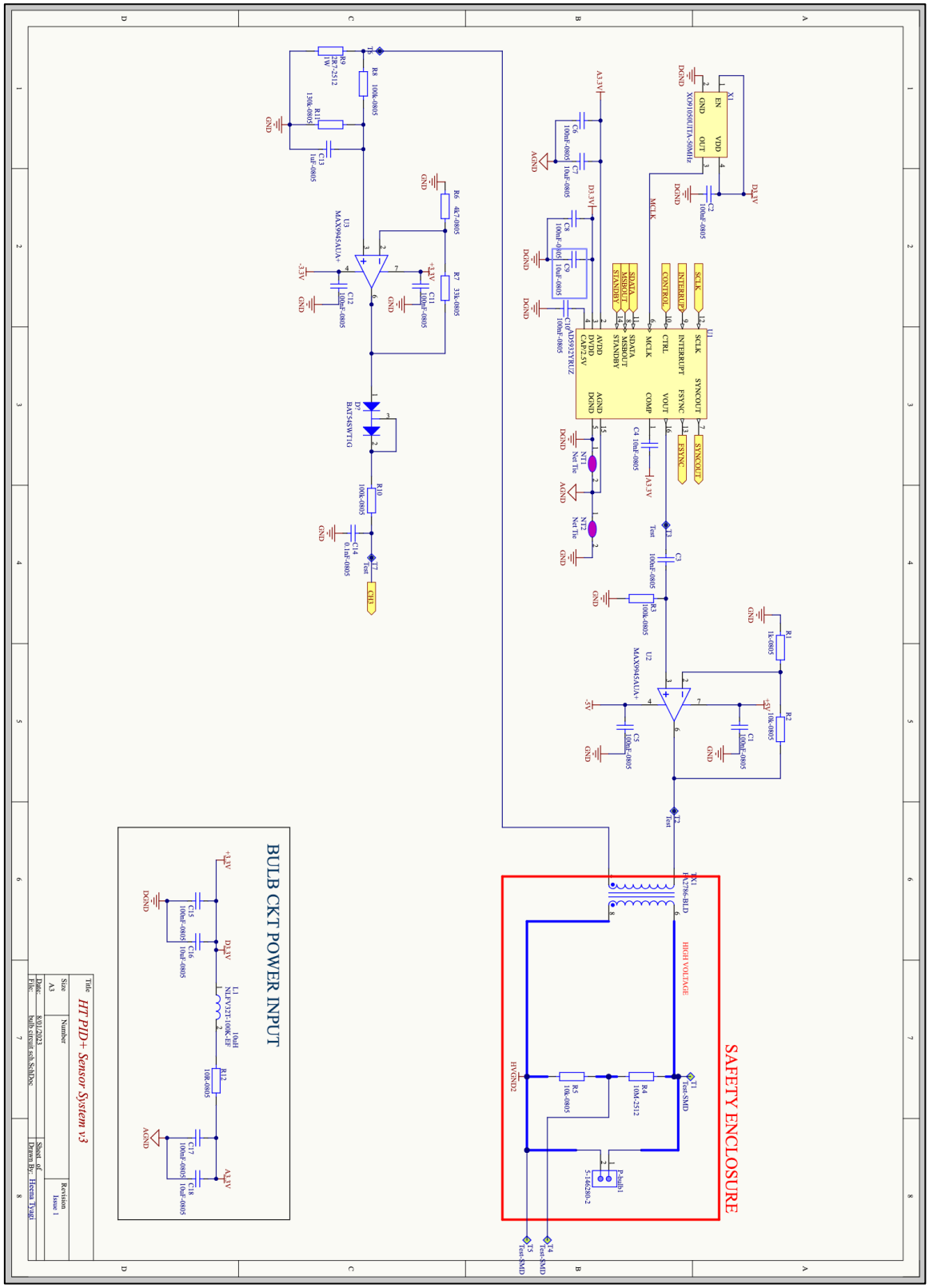


Figure 7.3-22: schematic for the bulb circuit for third generation PID+ prototype

c. Detector unit and Amplifier stage:

The main changes made to the detector unit and amplification changes included removing the optocoupler from the circuit design. A Schottky diode was added to prevent reverse current flow from the detector plates to the high voltage IC. The potentiometers were removed from the circuit and instead a few fixed resistors were added to the feedback circuit of the amplifier with an option to increase or decrease the gain according to the application. Figure 7.3-23 shows the schematic for the detector and amplifier stage.

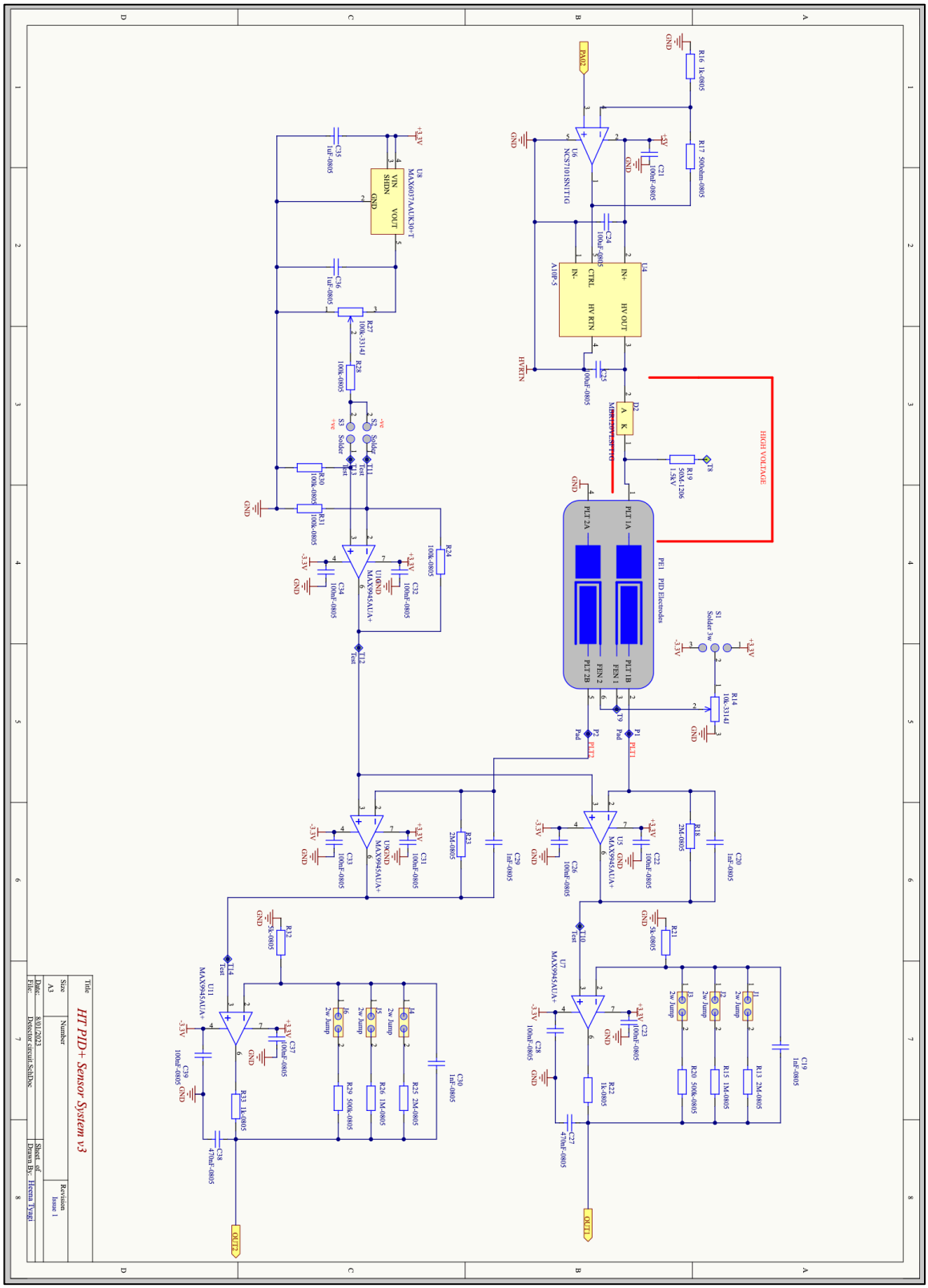


Figure 7.3-23: Schematic for Detector and Amplifier stage for third generation PID+ prototype.

7.4. Conclusion:

In conclusion, this chapter has presented a detailed account of the development of three in-house Photoionization Detector (PID) prototypes. The chapter commenced with a theoretical explanation of a PID and the various components that make up the device. The three prototypes were developed using different methodologies, with each version building upon the previous one to improve the device's performance. The first version was based on an Arduino M0 microcontroller, while the second version integrated a high performance ATSAMD21G18 ARM® Cortex®-M0+ based flash microcontroller on a PCB. Different techniques were employed to improve the bulb circuit, with the AD5932 generating the most stable sine wave to ignite the UV bulb. In the third version, the power system for the circuit was improved by including a battery power, making the device portable. An integrated code was written to control the bias voltage, scan the resonant frequency for the bulb circuit, and read the ADC converter. This chapter provides a comprehensive account of the development process, which can serve as a reference for other researchers in this field. The results obtained from the three prototypes will be presented in the following chapter.

7.5. References:

[1] S.-O. Baek, L. Suvarapu, and Y.-K. Seo, "Occurrence and Concentrations of Toxic VOC in the Ambient Air of Gumi, an Electronics-Industrial City in Korea," in *Sensors* vol. 15, ed: Multidisciplinary Digital Publishing Institute, 2015, pp. 19102-19123.

[2] S. Cheng *et al.*, "Global Research Trends in Health Effects of Volatile Organic Compounds during the Last 16 Years: A Bibliometric Analysis," in *Aerosol and Air Quality Research* vol. 19, ed: 台灣氣膠研究學會, 2019, pp. 1834-1843.

[3] Z. Jia, A. Patra, V. K. Kutty, and T. Venkatesan, "Critical review of volatile organic compound analysis in breath and in vitro cell culture for detection of lung cancer," in *Metabolites* vol. 9, ed: Multidisciplinary Digital Publishing Institute, 2019, p. 52.

[4] B. Berglund *et al.*, "Effects of Indoor Air Pollution on Human Health," in *Indoor Air* vol. 2, ed: John Wiley & Sons, Ltd (10.1111), 1992, pp. 2-25.

[5] J. S. Yuan, S. J. Himanen, J. K. Holopainen, F. Chen, and C. N. Stewart, "Smelling global climate change: mitigation of function for plant volatile organic compounds," in *Trends in Ecology & Evolution* vol. 24, ed, 2009, pp. 323-331.

[6] H. Insam and M. S. A. Seewald, "Volatile organic compounds (VOC) in soils," in *Biology and Fertility of Soils* vol. 46, ed, 2010, pp. 199-213.

[7] P. Hsi, W. Yeh, D. Zhang, X. U. Yuhui, and K. Lu, "Low power photoionization detector (pid)," ed: Google Patents, 2019.

[8] H. H. Hill and G. Simpson, "Capabilities and limitations of ion mobility spectrometry for field screening applications," in *Field Analytical Chemistry & Technology* vol. 1, ed: John Wiley & Sons, Ltd, 1997, pp. 119-134.

[9] H. K. Patel and M. J. Kunpara, "Electronic nose sensor response and qualitative review of e-nose sensors," in *2011 Nirma University International Conference on Engineering*, 2011: IEEE, pp. 1-6.

[10] A. D. Wilson, "Review of Electronic-nose Technologies and Algorithms to Detect Hazardous Chemicals in the Environment," in *Procedia Technology* vol. 1, ed, 2012, pp. 453-463.

[11] P. Boeker, "On 'Electronic Nose' methodology," in *Sensors and Actuators B: Chemical* vol. 204, ed, 2014, pp. 2-17.

[12] J. Lovelock, "A photoionization detector for gases and vapours," *Nature,* vol. 188, no. 4748, pp. 401-401, 1960.

[13] J. N. Driscoll, "Evaluation of a new photoionization detector for organic compounds," *Journal of Chromatography A,* vol. 134, no. 1, pp. 49-55, 1977/04/01/ 1977, doi: https://doi.org/10.1016/S0021-9673(00)82568-6.

[14] M. L. Langhorst, "Photoionization Detector Sensitivity of Organic Compounds*," *Journal of Chromatographic Science,* vol. 19, no. 2, pp. 98-103, 1981, doi: 10.1093/chromsci/19.2.98.

[15] A. N. Freedman, "Photoionization detector response," *Journal of Chromatography A,* vol. 236, no. 1, pp. 11-15, 1982/02/05/ 1982, doi: https://doi.org/10.1016/S0021-9673(00)82494-2.

[16] G. Coelho Rezende, S. Le Calvé, J. J. Brandner, and D. Newport, "Micro photoionization detectors," in *Sensors and Actuators B: Chemical* vol. 287, ed, 2019, pp. 86-94.

[17] J. N. Driscoll, "Evaluation of a new photoionization detector for organic compounds," in *Journal of Chromatography A* vol. 134, ed, 1977, pp. 49-55.

[18] S. O. Agbroko and J. Covington, "A novel, low-cost, portable PID sensor for the detection of volatile organic compounds," in *Sensors and Actuators B: Chemical* vol. 275, ed, 2018, pp. 10-15.

[19] Y. Morimoto, T. Sumitomo, M. Yoshioka, and T. Takemura, "Recent progress on UV lamps for industries," in *Conference Record of the 2004 IEEE Industry Applications Conference, 2004. 39th IAS Annual Meeting.*, 3-7 Oct. 2004 2004, vol. 2, pp. 1008-1015 vol.2, doi: 10.1109/IAS.2004.1348536.

[20] C. Ferrari, I. Longo, L. Socci, and M. Cavagnaro, "Coaxially driven microwave electrodeless UV lamp," *Journal of Electromagnetic Waves and Applications,* vol. 28, no. 6, pp. 669-684, 2014/04/13 2014, doi: 10.1080/09205071.2014.883944.

[21] J. Sun, F. Guan, D. Cui, X. Chen, L. Zhang, and J. Chen, "An improved photoionization detector with a micro gas chromatography column for portable rapid gas chromatography system," *Sensors and Actuators B: Chemical,* vol. 188, pp. 513-518, 2013/11/01/ 2013, doi: https://doi.org/10.1016/j.snb.2013.07.066.

[22] H. Zhu *et al.*, "Flow-through microfluidic photoionization detectors for rapid and highly sensitive vapor detection," *Lab on a Chip,* 10.1039/C5LC00328H vol. 15, no. 14, pp. 3021-3029, 2015, doi: 10.1039/C5LC00328H.

[23] J. P. HENNES, R. SCOLNIK, and A. K. STOBER, "Characteristics of a Photoionization Chamber for the Vacuum Ultraviolet," *Publications of Goddard Space Flight Center,* vol. 2, no. 1461, p. 374, 1959.

Chapter 8. PID+ Experimental Results and Discussion

Chapter 8 consists of the details about the experiment setup, results, and discussion on the performance of the different generations of the PID+ prototype. Chapter 8 starts with a detailed description of the experimental procedure followed to analyse and test the performance of the different PID+ prototypes. Since the procedure followed for testing the different prototypes varied little, it has been described in one section. This is followed by the results generated by the different testing methods divided according to each prototype. Finally, the chapter is concluded by providing a brief discussion on the results obtained from each generation of PID+ prototype and the challenges faced with each prototype which led to the development of the next generation of PID+ prototype.

8.1. Experimental Setup for the first version of the PID+:

The testing and analyses of the PID+ prototypes were divided into two main parts, first was a functionality test and performance test. The functionality test was performed on all the PCBs to test the operation of the various parts and components of the circuit design and the structural integrity of the prototype boards. The first step of the functionality test was visual inspection in which the board was visually examined to check for shorts, wrong placement of the components, broken traces, or faulty components. Once the board passed the visual checks, the board was powered, and the different voltage rails were checked at different parts of the board with a multi-meter. Once the board passed these tests, the board was flashed with the bootloader using a JTAG/SWD debugger (SEGGER J-Link BASE - JTAG/SWD Debugger) and Atmel Studio 7.0 software. When the flashing proved successful, the board was flashed with an Adafruit feather M0 bootloader.

After completing the functionality test, the PID+ was tested for performance. The performance test was carried out by testing the responses of the device. To determine the

responses of the first prototype, it was tested under various conditions using isobutylene (ISB) gas. Isobutylene (ISB) had been used as the primary VOC as it is the gas used for calibration and testing of commercial PID sensors. ISB is a common gas used for testing because it is inexpensive, readily available, has average sensitivity and low toxicity. The testing conditions included changing the flow rates of ISB, concentrations of ISB and biasing of the electrodes. An API Dilution calibrator- Model 700 was used to generate the desired concentrations and flow rates of the gas and regulate it into the PID. The circuit was powered up with a power supply of 12 Volts and consumed 64 mA current. The output was visualised using a Pico ADC24 Data logger.

The methods for testing the performance of the first generation PID+ prototype was executed as follows:

1. The first test included supplying ISB through a 50 ppm ISB cylinder to the dilution calibrator which was used to generate a constant flow rate of 1 LPM of zero air with varying concentrations of ISB. The process included first subjecting clean air to the PID+ for 20 minutes followed by continuous flow for 20 minutes of 50 ppb (parts per billion), 20 minutes of 100 ppb, 20 minutes of 500 ppb, 20 minutes of 1000 ppb and 20 minutes of 5000 ppb of ISB respectively. Finally, the test was concluded with another 20 mins of flow of clean air in order to flush out the UV chamber completely.

2. The second test was conducted in the same manner as the first test except that during the second test the flow rate was changed from 1 LPM (Litre Per Minute) to 1.5 LPM.

3. The third test was also conducted in the same manner as the tests above except that the flow rate was changed to 0.5 LPM.

4. During the fourth test, the bias voltage applied to electrode 1 of the two detector plates was changed. The bias voltage was configured to change after every 20 mins via firmware and the output generated by the first prototype was observed. A constant flow rate of 1 LPM and constant concentration of ISB of 5 ppm (parts per million) was applied to the device during this test.

5. After observing the output generated by test 4, the same experiment was conducted except that instead of applying a continuous bias voltage, a pulse voltage was applied. During this test, each bias voltage was applied for 10 mins followed by 10 mins of 0 voltage.

6. The tests conducted up to that stage indicated that the first generation PID+ prototype was most sensitive at a concentration of 5 ppm of ISB at a flowrate of 1 LPM. Therefore, the last test was conducted in order to verify this. This test was conducted in three steps. During first step, the PID+ was subjected to a constant concentration of 500 ppb of ISB but the flow rate of the gas was varied. Similarly, during second step, the PID+ was subjected to a constant concentration of 1000 ppb of ISB and finally, the PID+ was subjected to a constant concentration of 5 ppm of ISB.

8.2. Experimental Results:

The first tests conducted to check the efficiency of the first PID+ prototype included keeping the flow rate of ISB constant at a rate of 1 LPM and changing the concentration of ISB. Table 8.2-1 illustrates the results obtained from the experiment and Figure 8.2-1 shows a pictorial representation of the outcome from the test. The two outputs reported in these measurements are the output generated by the detector unit as explained in chapter 7, section 7.3.1.

In Table 8.2-1, Output 1 and Output 2 signify the current resulting from the ionized molecules of ISB that have been gathered at the electrode plates of the PID+. After optimizing the circuit, a bias of 0V was chosen and fixed for the output plate 2. When the PID+ was exposed to air, both electrodes produced an output of 4.464 mV and 35.638 mV. Conversely, when ISB was passed through the PID+, the signal strength at both electrodes increased (indicated by a negative current), with the highest gas concentration being 1 ppm. The electrode output generated at a flow rate of 1 LPM and a concentration of 5000 ppb (5 ppm) demonstrated the most significant voltage level shift, according to the observations.

Table 8.2-1: PID+ first prototype result at a flow rate of 1 LPM

Concentration (ppb)	Output 1 (mV)	Output 2 (mV)
Air	4.464	35.638
158(*50)	4.455	35.549
158(*100)	4.449	35.329
500	4.441	35.272
1000	4.434	34.692
5000	4.432	32.382
Air	4.448	35.419

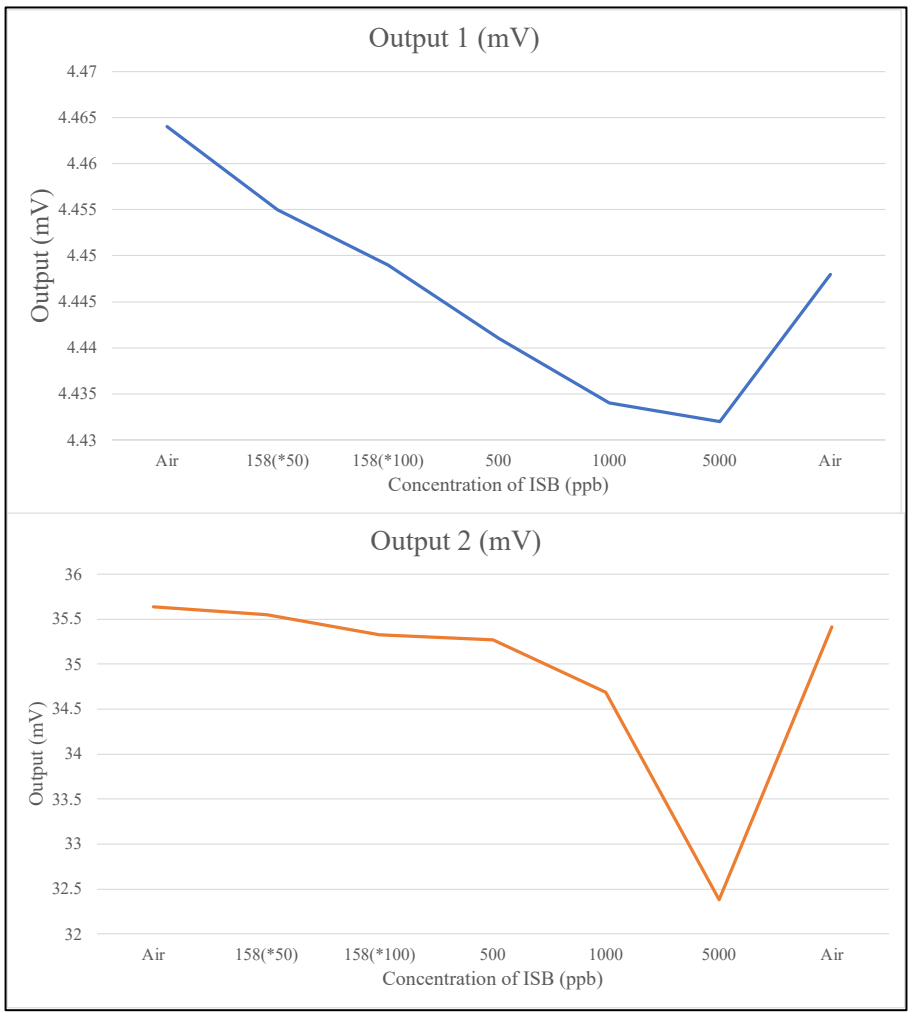


Figure 8.2-1: shows the output of the first generation PID+ prototype at a flow rate of 1 LPM with different concentrations of ISB.

After obtaining the results from the first experiment, the second test was conducted during which ISB was applied to the PID+ with a flow rate of 1.5 LPM. The outputs generated from

test 2 are represented in Table 8.2-2. Figure 8.2-2 shows a pictorial representation of the outcome of second test.

Table 8.2-2: PID+ first prototype result at a flow rate 1.5 LPM

Concentration (ppb)	Output 1 (mV)	Output 2 (mV)
Air	4.454	35.642
105(*50)	4.446	35.576
105(*100)	4.445	35.454
500	4.444	35.434
1000	4.44	34.909
5000	4.438	32.992
Air	4.45	35.933

The output obtained from the test 2 showed that the output generated by the PID+ for air was 4.454 mV and 35.642 mV for output 1 and output 2 respectively, whereas the outputs generated during the presence of ISB varied from 4.446 to 4.438 mV for output 1 and 35.576 to 32.992 mV for output 2. The output generated by the PID+ at a concentration of 5 ppm with a 1.5 LPM flow rate generated the most response.

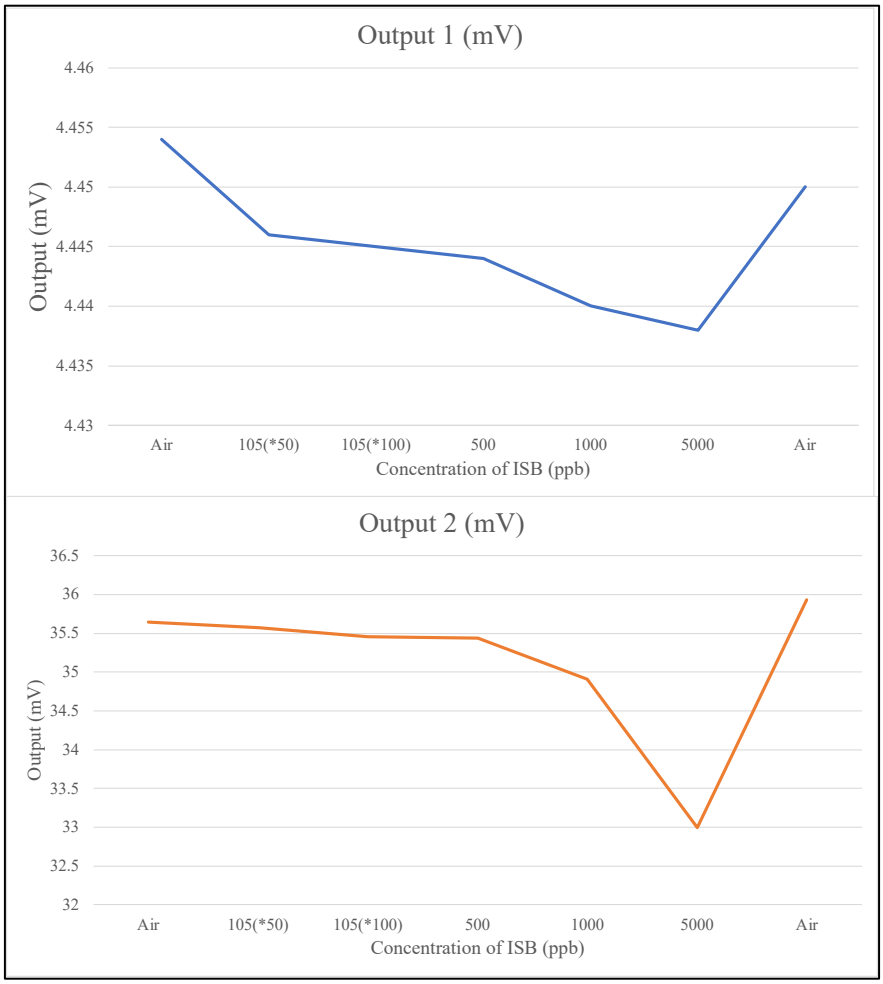


Figure 8.2-2: shows the output of the first generation PID+ prototype at a flow rate of 1.5 LPM with the different concentration of ISB.

The results generated by the third test are shown in Table 8.2-3. The third test constituted the same experimental procedure as the two tests above except that during the third test, flow rate was dropped to 0.5 LPM. Figure 8.2-3 illustrates the outcome of the third test.

Table 8.2-3: PID+ first prototype result at a flow rate 0.5 LPM

Concentration (ppb)	Output 1 (mV)	Output 2 (mV)
Air	4.464	35.732
357(*50)	4.451	35.584
357*100)	4.448	35.555
500	4.445	35.516
1000	4.441	35.431
5000	4.439	33.262
Air	4.461	35.652

The output obtained from the third set of tests showed a similar pattern as the other two tests. The output generated by the PID+ when it was subjected to clean air was 4.464 mV and 37.732 mV respectively. It was shown that the response generated by electrode 1 was highest at 5000 ppb.

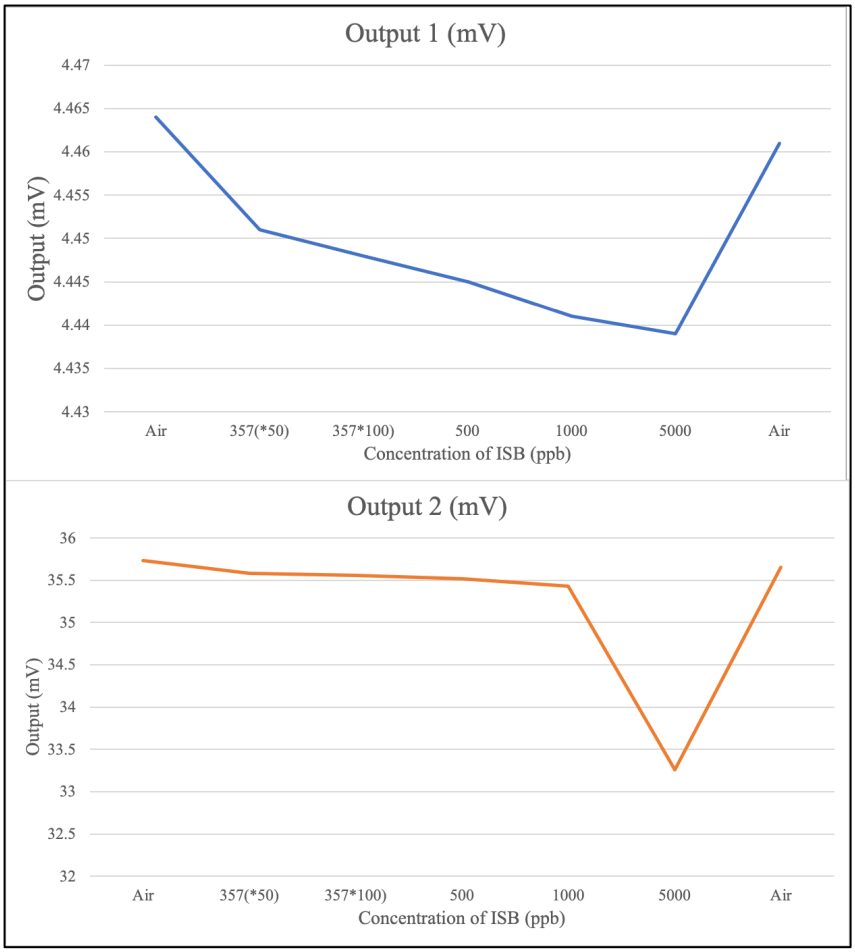


Figure 8.2-3: shows the output of the first generation PID+ prototype at a flow rate of 0.5 LPM with the different concentrations of ISB.

The results obtained so far demonstrated that the PID+ was moderately capable of sensing the gas at concentrations of 500 ppb to 5 ppm at different flow rates. Figure 8.2-4 illustrates the response of output 1 for the first generation PID+ prototype at different concentrations of ISB at different flow rates generated by a dilution calibrator. Similarly, Figure 8.2-5 illustrates the response of output 2 of the first generation PID+ prototype. It was observed that there was a minimal response from the prototype for 500 ppb. At a concentration of 500 ppb, the PID+ represented a very low response at different flow rates. The best response was

seen at a flow rate of 1 LPM. With an increase in concentration, the response was increasing. At a flow rate of 1.5 LPM, the performance of the PID+ was again moderate.

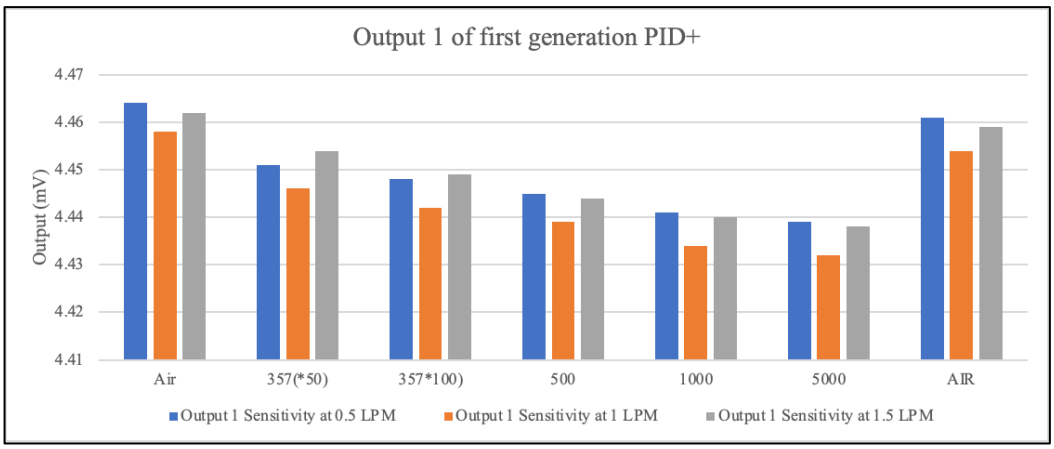


Figure 8.2-4: shows the response generated by electrode 1for the first generation PID+ prototype for different concentrations of ISB at flow rates of 0.5 LPM, 1 LPM and 1.5 LPM.

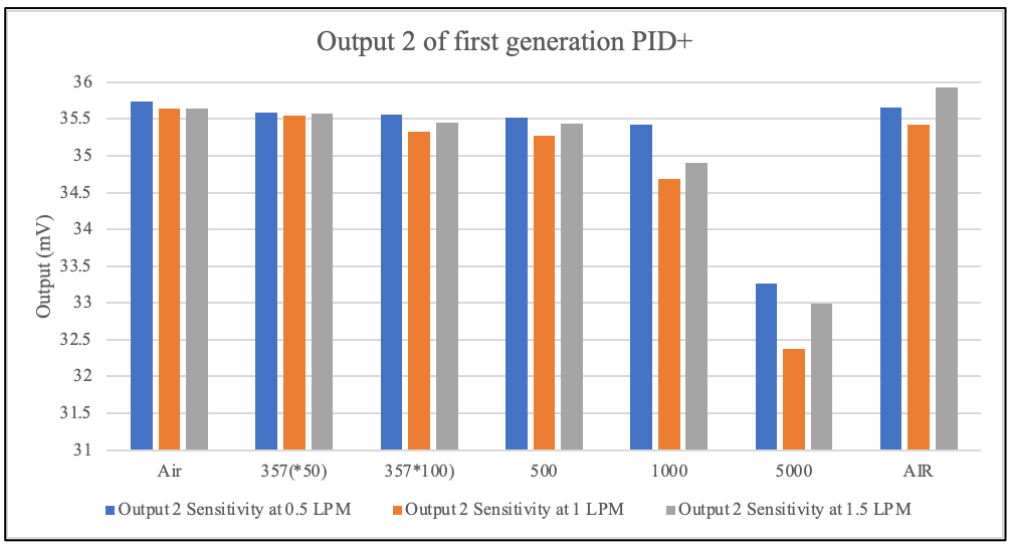


Figure 8.2-5: shows the response generated by electrode 2 for the first generation PID+ prototype for different concentrations of ISB at flow rates of 0.5 LPM, 1 LPM and 1.5 LPM.

As mentioned in chapter 6, out of the two electrodes of the detector, the bias voltage of one of the electrodes was controlled via microcontroller and the other electrode was grounded. During the first three tests, the bias voltage of the electrode on the detector was set to 0V. After observing the response of the PID+ with different flow rates at 0 volts, the bias voltage was varied and the responses of the PID+ to different bias voltages were observed. Table 8.2-4 contains the output generated by the two electrodes of the first generation PID+

prototype when the bias voltage applied to the electrode was changed using the microcontroller. The output was noted when the device was at standby followed by increasing the bias voltage and finally applying 0 Volts to the device.

Table 8.2-4: Effect of bias voltage on the PID+ first prototype outputs.

Bias voltage (V)	Output 1 (mV)	Output 2 (mV)
AIR	7.613	33.072
150	7.502	31.683
300	7.495	31.863
450	7.461	32.208
600	7.432	31.923
750	7.428	31.519
900	7.467	31.568
0	7.552	32.856
AIR	7.586	33.332

It was seen that, when there was no flow of gas, the outputs generated by the two electrodes were 7.613 mV and 33.072 mV respectively. When the bias voltage across the electrode was changed, it was observed there was a decrease in the voltage representing the sensing of the presence of the ISB gas. It was observed that the maximum response signal was seen when the bias voltage was 750 V. Figure 8.2-6 illustrates the output from the PID+ during test four.

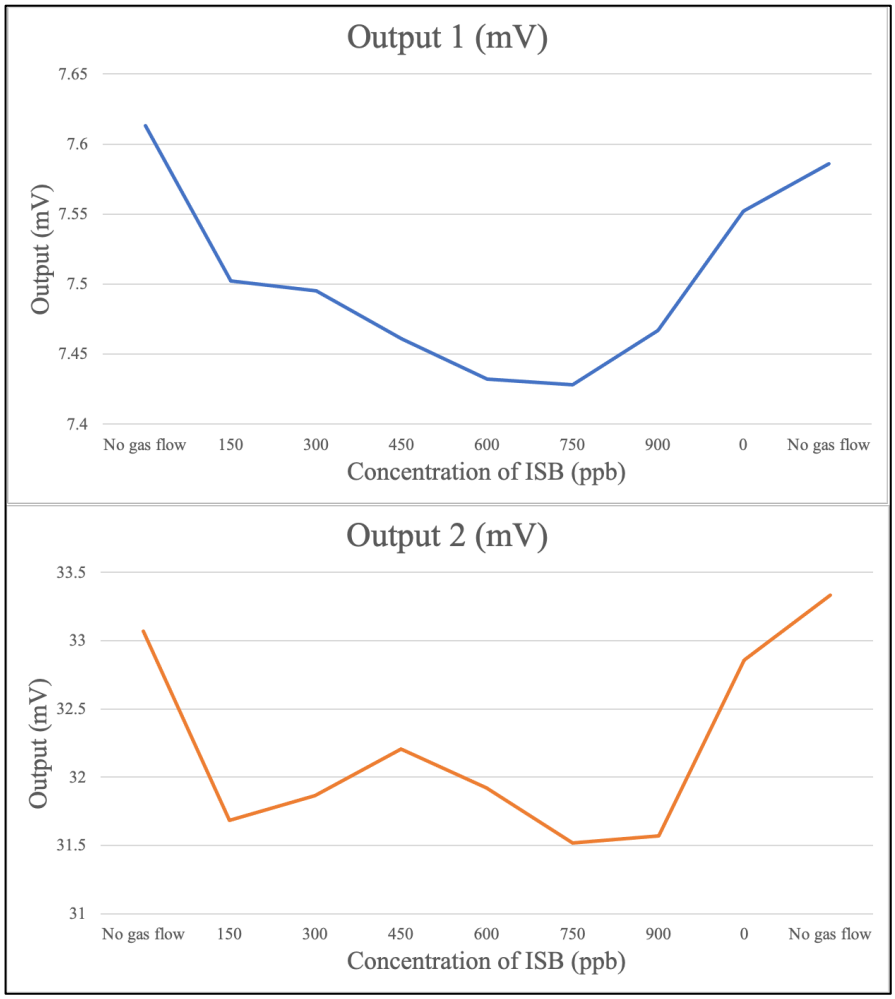


Figure 8.2-6: shows the response generated by the first generation PID+ prototype for different bias voltages.

Test five was conducted to understand the effect of bias voltage with respect to the ground voltage. It was observed that when the bias of the electrode was increased from 0 Volts to a higher voltage, it leads to an increase in the response of the detector. Table 8.2-5 contains the data recorded during this test.

Table 8.2-5: Effect of pulsated bias voltage on the PID+.

Bias voltage (V)	Output 1 (mV)	Output 2 (mV)
AIR	7.613	32.978
0	7.601	31.132
150	7.588	31.376
0	7.595	31.324
300	7.579	31.751
0	7.581	31.335

450	7.569	31.63
0	7.575	31.153
600	7.564	31.578
0	7.569	31.172
750	7.557	31.714
0	7.571	31.13
900	7.585	31.544
0	7.587	30.966
AIR	7.601	33.289

The output generated showed that the voltage decreased with the increase in the bias voltage. The highest signal response was observed at the bias voltage of 750 Volts. This signifies that the highest amount of ionised chemicals was attracted and absorbed by the electrode's plate at this voltage. Figure 8.2-7 shows the plots of the output responses of the PID+ during the test.

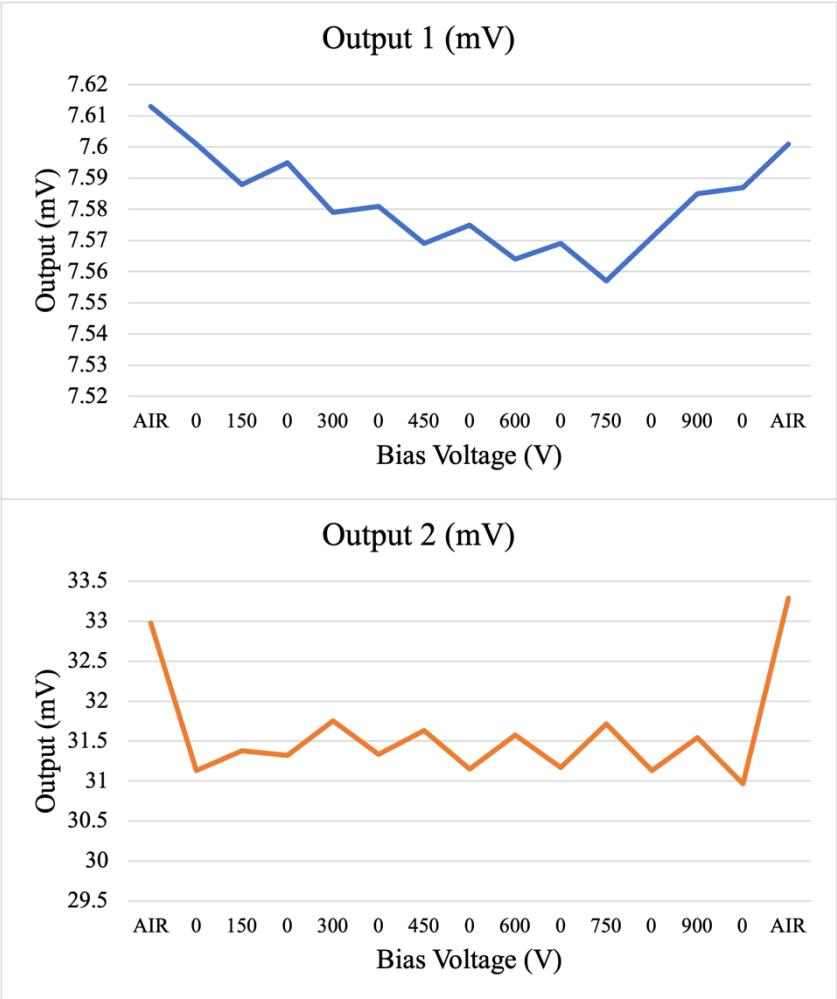


Figure 8.2-7: Shows the response generated by the first generation PID+ prototype for different bias voltages applied as a pulse.

The sixth experiment was conducted on the first generation PID+ prototype to understand the effect of different concentrations of ISB on the device. All the tests performed so far indicated that the highest effect of ISB was seen at a concentration of 5 ppm. In order to verify this, test 6 was conducted. The results obtained from test 6 are presented in Table 8.2-6 and Table 8.2-7.

Table 8.2-6: Output 1 of the PID+ for a changing flowrate at a constant concentration of ISB.

Flow rates (LPM)	Output 1 (mV) at 0.5 ppm	Output 1 (mV) at 1 ppm	Output 1 (mV) at 5 ppm
Air	4.523	4.483	4.478
0.2	4.502	4.478	4.461
0.5	4.486	4.463	4.456
1	4.478	4.457	4.443
1.2	4.48	4.469	4.453
1.5	4.492	4.482	4.462
2	4.499	4.49	4.467
Air	4.512	4.489	4.488

Table 8.2-7: Output 1 of the PID+ for a changing flowrate at a constant concentration of ISB.

Flow rates (LPM)	Output 2 (mV) at 0.5 ppm	Output 2 (mV) at 1 ppm	Output 2 (mV) at 5 ppm
Air	37.862	36.432	35.587
0.2	37.406	36.054	34.671
0.5	36.982	35.394	33.466
1	35.507	33.846	31.893
1.2	35.957	35.681	31.893
1.5	36.059	35.124	32.169
2	36.464	34.909	32.626
Air	37.967	34.849	33.674

The results shown in Table 8.2-6 and Table 8.2-7 indicate that the PID+ first generation prototype produced the most output to ISB at a concentration of 5 ppm and produced a moderate response at 1 ppm. The output generated during this test is shown Figure 8.2-8 and Figure 8.2-9. Figure 8.2-10 illustrates the output generated by electrode one of the detectors over the time period for different concentrations of ISB at different flow rates.

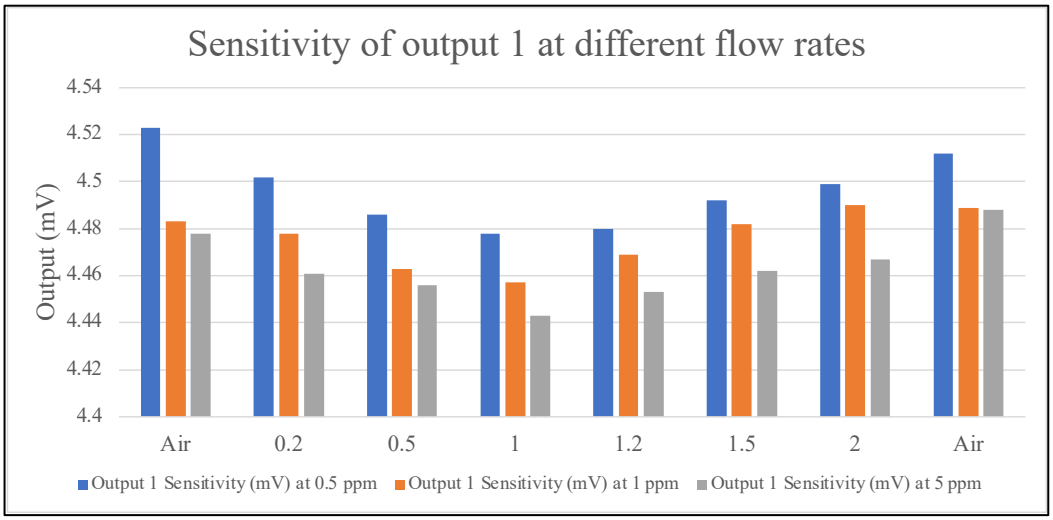


Figure 8.2-8: Shows the response generated by the first generation PID+ prototype for a different bias voltage applied as a pulse.

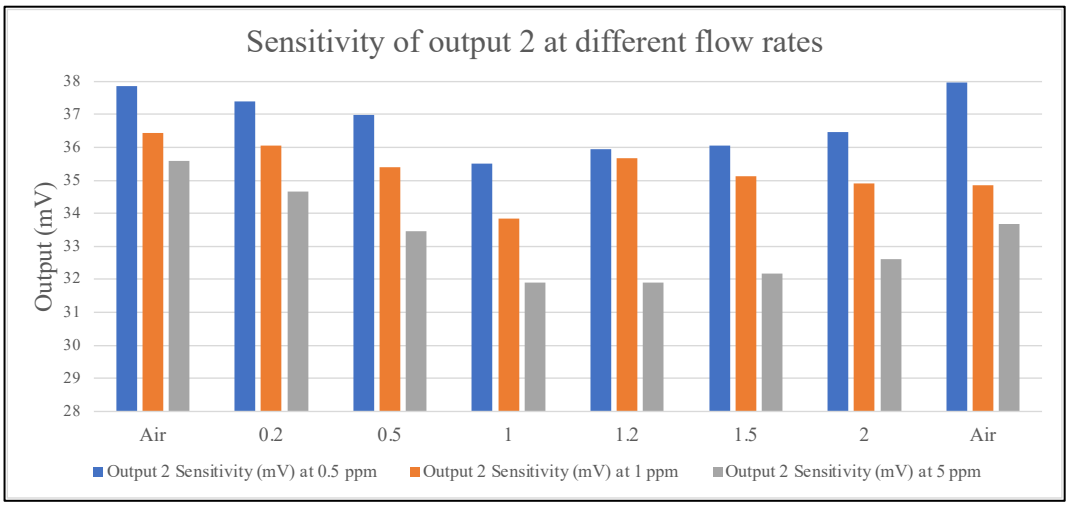


Figure 8.2-9: Shows the response generated by the first generation PID+ prototype for a different bias voltage applied as a pulse.

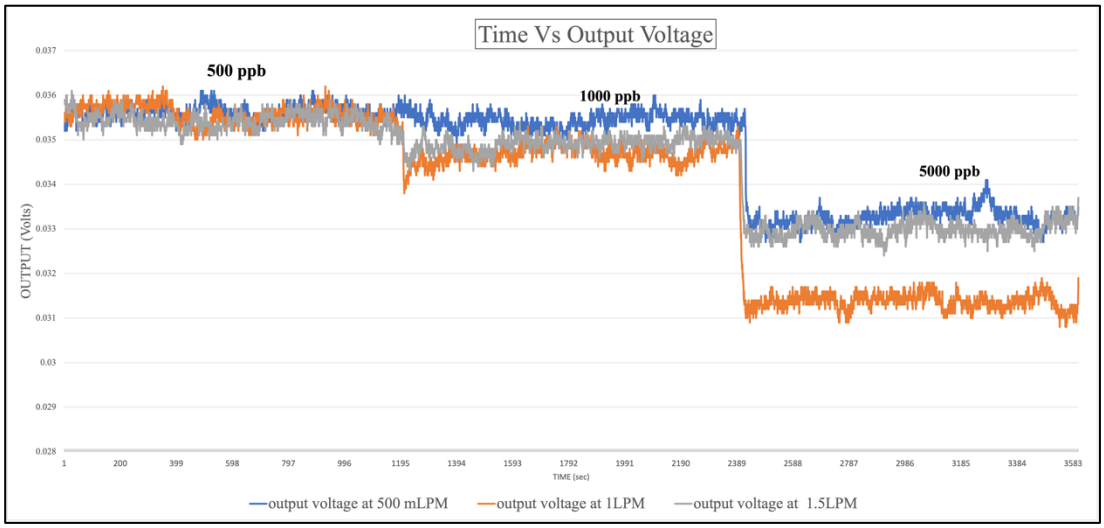


Figure 8.2-10: Shows the three concentrations at which the response from the device is maximum, i.e., 500ppb, 1 ppm and 5 ppm at three different flowrates of 500 mLPM, 1 LPM and 1.5 LPM with respect to the time in seconds.

8.3. Experimental Setup for the Second Version of the PID+:

The experimental procedure used for testing the performance of the second version of PID+ was similar to the first version. The second version of the PID+ prototype was powered up with a USB and the data was recorded using a serial monitor.

The experiments performed on the second generation PID+ are as follows:

1. The test performed on the second generation PID+ prototype included changing the gain of the amplifier used to amplify the signal detected by the detector plates while keeping the concentration and the flow rate of ISB constant. The feedback resistor used in the amplification stage was a 1 MΩ potentiometer. The resistance of this resistor was changed and hence led to a change in the gain of the amplifier. In this test, the PID+ was exposed for 20 mins to air, followed by 20 mins to ISB with 1 ppm concentration and lastly, another 20 mins of air at different gain values of 50, 100 and 150.

2. After adjusting the gain, further tests were conducted on the second generation PID+ in which the gain was kept constant, and the bias of the electrode plates was changed by using

a microcontroller. One of the Analog pins of the microcontroller was connected to the input of the op-amp and the output generated by the op-amp was used as input to a high voltage isolated DC to DC converter. During this test, the flow rate of the gas was kept constant at 1LPM and the PID+ was exposed to air, followed by 1 ppm of ISB, then 5 ppm of ISB and lastly air, each for 20 mins at different bias voltages of 0 V, 500 V, 650 V, 800 V and 1000 V.

3.The third test was performed in order to understand the best flow rate required in order to generate the most output from the second generation PID+. The procedure used during the third test included setting the dilutor to flow rates of 0.5 LPM, 1 LPM, 1.5 LPM and 2 LPM. During each flow rate, the PID+ was first exposed to air for 20 mins, then 20 mins exposure of ISB with 1 ppm concentration, followed by 20 mins exposure of ISB with 5 ppm concentration and finishing up with 20 mins of air.

4. The fourth test was the final test performed on the second generation of PID+ performed to optimise the ability of the PID+ in correspondence to the different concentration of ISB. During this test, the concentration of the ISB was varied from 150 ppb to 5 ppm and the response of the PID+ to the different concentrations was observed. The test was performed as follows. The PID+ was subjected to air for 20 mins, followed by 150 ppb of ISB, 300 ppb of ISB, 500 ppb of ISB, 1 ppm of ISB, 2 ppm of ISB, 5 ppm of ISB and air each for 20 mins.

8.4. Experimental results:

The best response from the second generation PID+ prototype was characterised using different criteria. The first step used to determine the performance of the second generation PID+ was to modulate the ability of the amplification of the output signal in the circuit in order to maximise the output signal. This was achieved by adjusting the gain of the amplifier stage and recording the output signal. The output obtained from the first test performed on the second generation PID+ is represented in Table 8.4 1 and Table 8.4 2.

Table 8.4-1: Output 1 from the second generation PID+ for different gain settings of the amplifier feedback resistor

PROCESS	OUTPUT 1 (V) @GAIN=50	OUTPUT 1 (V) @GAIN=100	OUTPUT 1 (V) @GAIN=150
AIR	0.56248	1.10184	1.64997
1 PPM ISB	0.56246	1.10274	1.64997
AIR	0.56241	1.10238	1.64997

Table 8.4-2: Output 2 from the second generation PID+ for different gain settings of the amplifier feedback resistor

PROCESS	OUTPUT 2 (V) @GAIN=50	OUTPUT 2 (V) @GAIN=100	OUTPUT 2 (V) @GAIN=150
AIR	0.50063	1.03772	1.50808
1 PPM ISB	0.50047	1.03882	1.50872
AIR	0.50071	1.03900	1.50891

The results presented in table above shows that the detector plates generated the highest amount of signal when the gain of the amplifiers was set to 100. The results obtained from the test are shown in Figure 8.4-1. It was seen that the outputs generated with a gain value of 50 and 150 were not very responsive to the change in the type of gas flow.

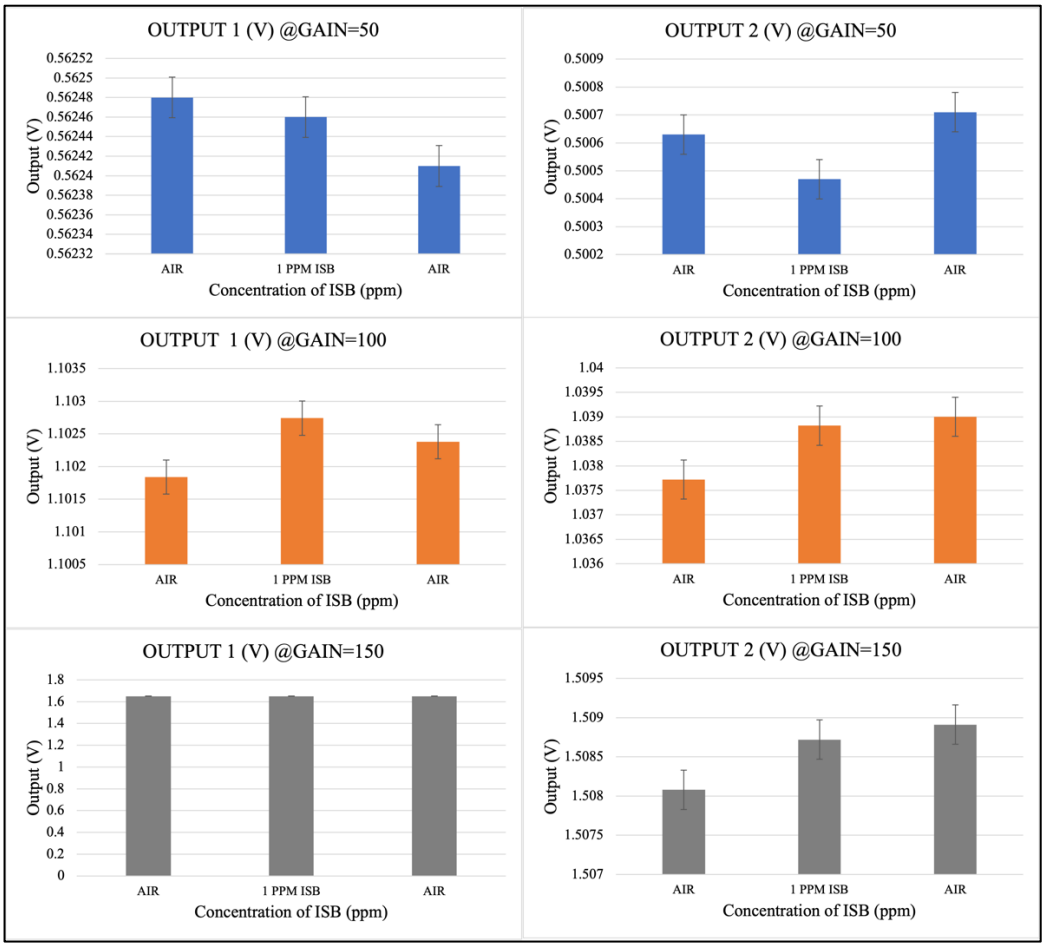


Figure 8.4-1: Outputs at the different output plates at different amplifier gains.

The further testing of the second prototype was done by keeping the gain constant at 100 to obtain the maximum output and varying the concentration of ISB. After adjusting the gain, the PID+ prototype was tested with different bias voltages. The different voltages were applied at the electrodes of the detector and the response was generated as shown in Table 8.4-3 and Table 8.4-4.

Table 8.4-3: Output 1 from the second generation PID+ for different bias voltage settings at the electrode plates.

Bias Voltage	Output 1 (V) with Air	Output 1 (V) at 1ppm ISB	Output 1 (V) at 5 ppm ISB	Output 1 (V) with Air
0 V	1.1444	1.1443	1.1440	1.1438
500 V	1.1686	1.1900	1.1915	1.1712
650 V	1.1677	1.1888	1.1897	1.1673
800 V	1.1656	1.1777	1.1810	1.1677
1000 V	1.1586	1.1577	1.1586	1.1584

Table 8.4-4: Output 2 from the second generation PID+ for different bias voltage settings at the electrode plates.

Bias Voltage	Output 2 (V) with Air	Output 2 (V) at 1ppm ISB	Output 2 (V) at 5 ppm ISB	Output 2 (V) with Air
0 V	0.97742	0.97745	0.97730	0.97697
500 V	0.99255	1.01188	1.01299	0.99020
650 V	0.99054	0.99891	1.00193	0.99041
800 V	0.99134	0.99787	0.99948	0.99099
1000 V	0.99111	0.99000	0.99067	0.99172

The results obtained from the second test showed that the detector plates generated the maximum output when the applied bias voltage was around 500 V to 650 V. The output response was recorded for two ISB concentrations, i.e., 1 ppm and 5 ppm and it was observed that both for 1 ppm and 5 ppm, the maximum output was generated when the bias voltage on the electrode plate was 500 to 650 V as shown in Figure 8.4-2.

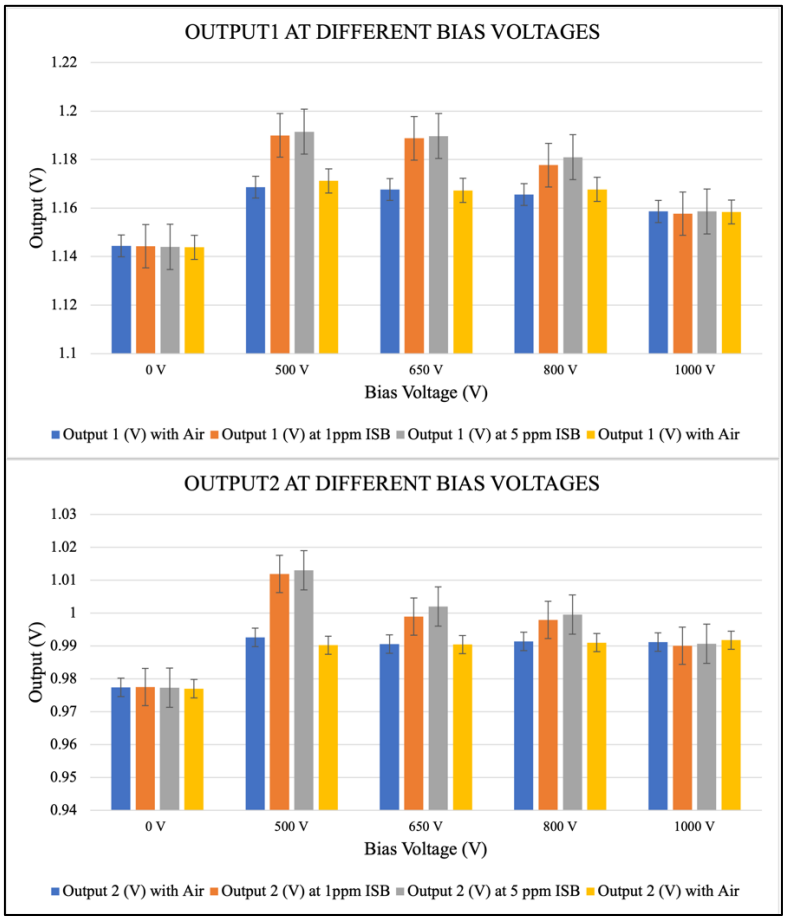


Figure 8.4-2: Responses at the different output plates at different bias voltages.

The efficiency of the second generation PID+ prototype was further tested by keeping the gain and the bias voltage at 100 and 550 V respectively. In order to optimise the performance of the second generation PID+ prototype, further tests were performed. The third test was performed with different flow rates of ISB generated by a calibration dilutor at two different ISB concentration. The different flow rates used for the test were 0.5 LPM, 1 LPM, 1.5 LPM and 2 LPM. The output obtained from the tests are tabulated in Table 8.4-5 and Table 8.4-6.

Table 8.4-5: Output 1 of the second generation PID+ for different flow rates of ISB.

Process	Output 1 (V) at FR=0.5LPM	Output 1 (V) at FR=1 LPM	Output 1 (V) at FR=1.5 LPM	Output 1 (V) at FR=2 LPM
Air	1.17968	1.17932	1.17892	1.17925
ISB-1ppm	1.17919	1.17954	1.17923	1.17902
ISB-5ppm	1.17940	1.17943	1.17921	1.17885
Air	1.17935	1.17914	1.17915	1.17884

Table 8.4-6: Output 2 of the second generation PID+ for different flow rates of ISB.

Process	Output 2 (V) at FR=0.5LPM	Output 2 (V) at FR=1 LPM	Output 2 (V) at FR=1.5LPM	Output 2 (V) at FR=2 LPM
Air	1.01133	1.01138	1.01131	1.01140
ISB-1ppm	1.01124	1.01148	1.01137	1.01141
ISB-5ppm	1.01146	1.01141	1.01135	1.01135
Air	1.01141	1.01136	1.01129	1.01129

The results obtained from the third test showed that the second generation PID+ was most responsive with the flow rate of 1 LPM and 1.5 LPM. Out of the two flow rates, 1 LPM generated the most output. Figure 8.4-3 shows the output generated during test 3.

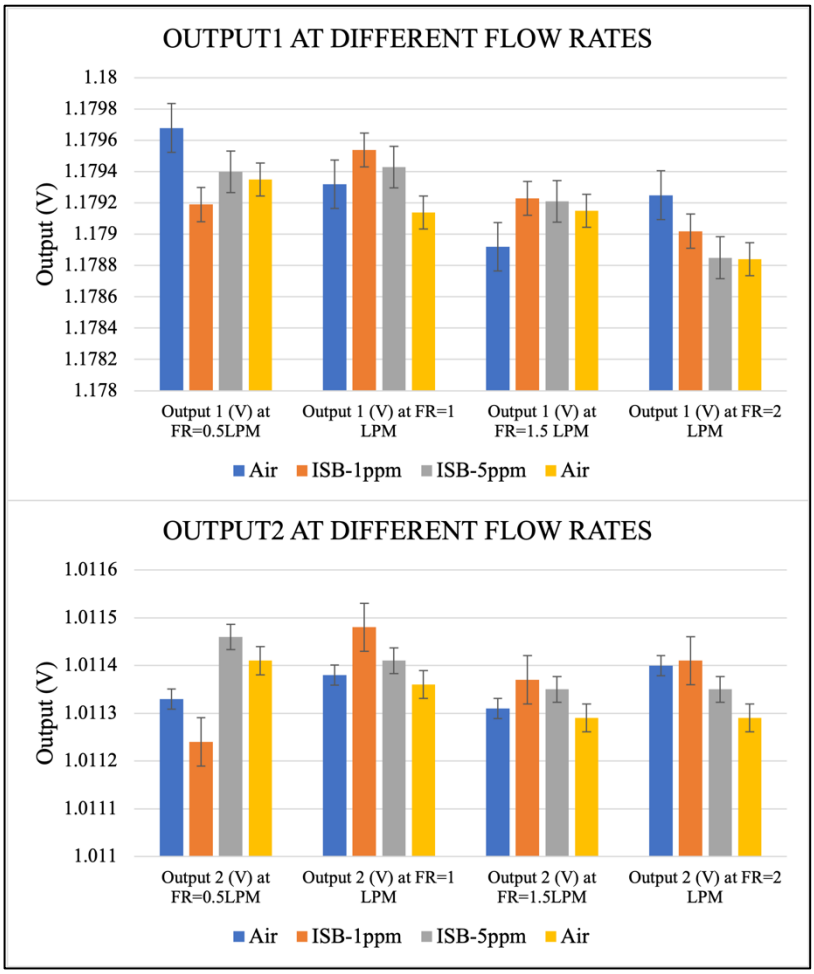


Figure 8.4-3: Sensitivities at the different output plates at different bias voltages.

The final testing was conducted with a gain of 100, bias of 650V and flow rate of 1 LPM for different concentrations of ISB. The lowest concentration obtained using a dilution calibrator was 150 ppb. So, the concentration was varied from 150 ppb to 5 ppm.

The outputs obtained from the tests are shown in Figure 8.4-4 and Figure 8.4-5. The output obtained from the PID+ version 2 shows that the second generation PID+ was sensitive to the presence of ISB at as low as 150 ppb. The response obtained at low concentration (150 ppb) and very high concentrations of 2 ppm to 5 ppm was negative which meant that voltage decreased below the baseline at these concentrations, whereas the response obtained at concentrations from 300 ppb to 1 ppm was positive which means that the voltage increased above the baseline at these concentrations.

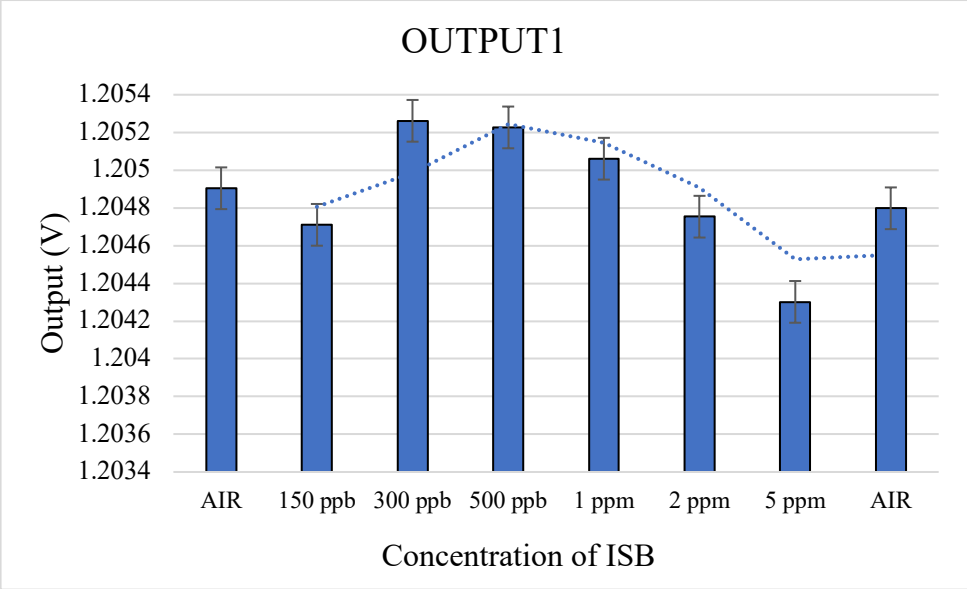


Figure 8.4-4: Response at output plate 1 for different concentrations of ISB.

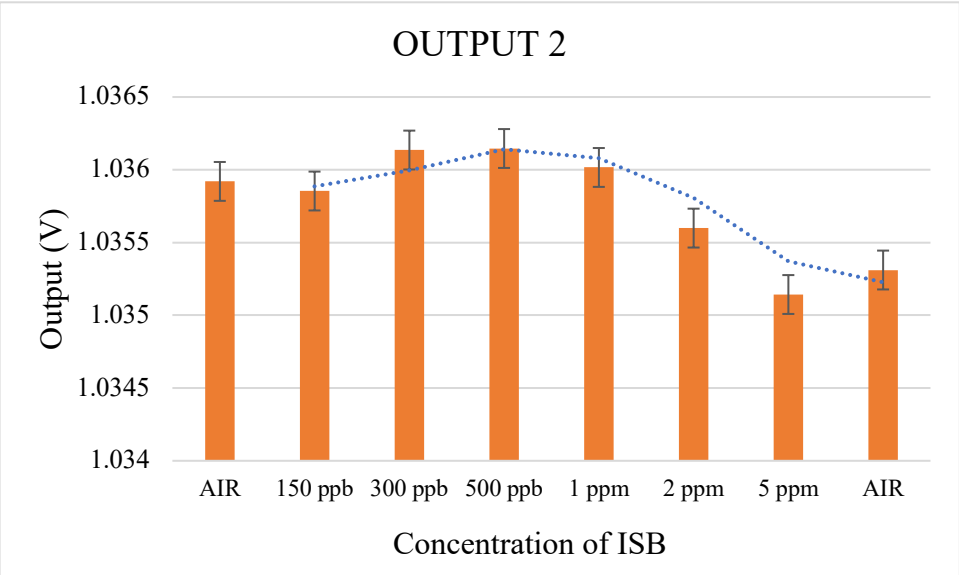


Figure 8.4-5: Response at output plate 2 for different concentrations of ISB.

8.5. Discussion:

Ion mobility is a highly efficient technique used for the detection of volatile organic compounds and the mobility of ions depends on its mass and charge [1]. In this research, the technique and the benefits of ion mobility were exhibited by the method of photoionization. This chapter presents the performance of the different prototypes of the PID+. The output

voltages reported in the study were directly proportional to the concentration of the ISB gas. The small changes in output voltages in response to varying concentrations demonstrate the ability of PID+ to detect and quantify gas concentrations. The impact of flow rate and bias voltage on the PID+ output is evident, indicating that these parameters need to be carefully controlled and optimized in practical applications. The data suggest that the performance of the first version PID+ is most reliable at a concentration of 5 ppm of ISB and a flow rate of 1 LPM, whereas the performance of the second version PID+ indicates most reliable output at a concentration of 1 ppm of ISB. This information is valuable for selecting the appropriate operational conditions for accurate measurements. However, replicating each test multiple times would have allowed for the calculation of statistical measures such as mean, standard deviation, and confidence intervals, providing a clearer understanding of the PID+'s performance.

During the testing of the first prototype of the PID+, different tests were conducted on the PID+ in order to understand and enhance the behaviour of the PID+. The first test was conducted in order to understand the response of the PID+ to different concentrations of ISB gas at a flow rate of 1 LPM as shown in Figure 8.2-1. The data indicate that as the concentration of ISB gas increases, the output voltages from both output1 and output2 electrodes generally decrease. This suggests that the PID+ is responsive to the presence of ISB gas and exhibits changes in its electrical output corresponding to different concentrations. The "Air" readings before and after the test runs show the baseline response of the PID+ when it is exposed to clean air with no ISB gas. These baseline values can be used as reference points for subsequent measurements to isolate the effect of ISB gas. The results suggests that the PID+ is capable of detecting ISB gas concentrations ranging from 50 ppb to 5000 ppb. However, the response becomes less significant at higher concentrations, as shown by the relatively small changes in output voltage between 1000 ppb and 5000 ppb. This shows that the PID+ was able to achieve a better response as low as parts per billion. A study conducted by Bilek *et. al* showed that the output of the PID was dependent on the concentrations of the VOCs [2].

The second tested conducted was similar to the first test except that now the flow rate was changed to 1.5 LPM. The PID+ demonstrated sensitivity to ISB concentrations, but the changes in output voltages between different concentrations were slightly smaller compared

to the 1 LPM flow rate, suggesting a reduction in sensitivity at this higher flow rate as shown in Figure 8.2-2. In order to test this, a third test was conducted in which the flow rate was changed to 0.5 LPM. Similar to the previous results, the output voltages from both output1 and output2 electrodes decreased as the concentrations of ISB gas increased. However, the PID+ response at this flow rate seemed to exhibit further reduced sensitivity compared to the higher flow rates. The reason for the correlation between the response of the PID+ and changes in flow rate can be explained by the relationship between the UV bulb's ionization capability and the number of molecules. A lower flow rate of ISB gas results in a lower output from the PID+ due to the decreased availability of ionizable molecules. As the flow rate increases, the ionization output from the UV bulb increases proportional to the number of molecules in the gas. An illustration of this process is shown in Figure 8.5-1.

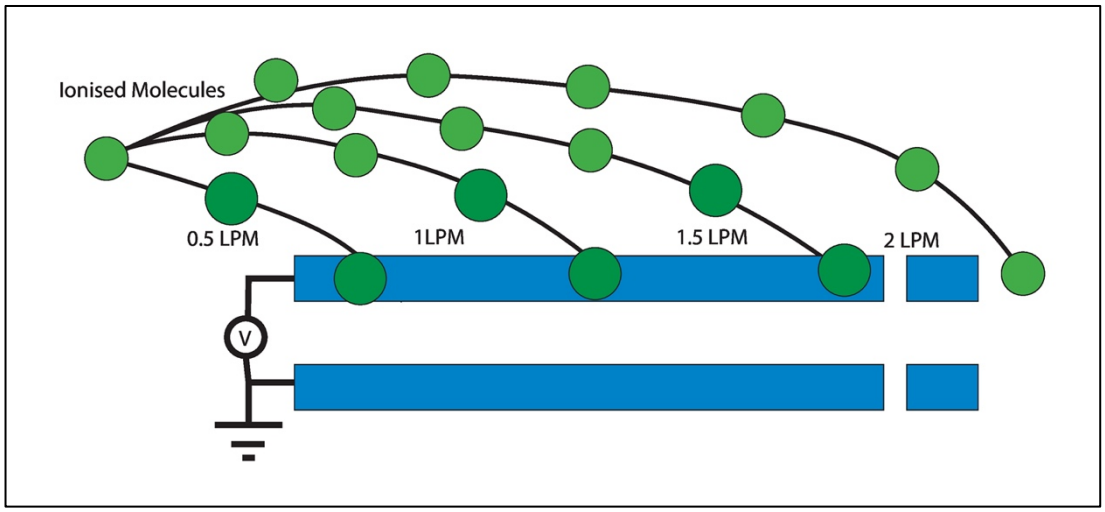


Figure 8.5-1: Illustration of the impact on ion path with change in the flowrate of the sample.

However, further increases in the flow rate caused the gas to pass more quickly under the UV light, resulting in a lower amount of time for absorption of UV light by the molecules contributing to the ionization current. Contrary to our finding, Freedman *et al.* [3] found in their studies that the response of a PID was not dependent on a change in the flow rate of the gas with constant concentration.

The fourth test was carried in order to understand the relationship between the applied voltage on the detector to the output. It was found that, as the magnitude of the bias voltage increased, the output generated by the first generation PID+ also increased. The output

generated was between 600 V to 750 V yielded the highest output response signifying that the optimal detector voltage for PID+ was 600V to 750 V. The reason for this is that as the voltage across the electrode plates increases, it increases the capability of the plates to collect charged particles and the flow of these charged particles collected by the electrodes results in the photoionization current. However, it was observed that as the voltage was increased above 750 V, then this led to a decrease in the output signal. This was due to the limit of the detector or could possibly be stated as the saturation point of the detector voltage. The experiment conducted by Mergemeier *et al.* [4] showed the dependence of photoionization currents on the applied voltage. They also showed that an increase in the electrode voltage resulted in an increase in the response of the electrodes.

Another test was conducted to provide insights into the behaviour of the PID+ when subjected to pulsating bias voltages. The results are shown in Figure 8.2-7. The table revealed that the PID+ responded to changes in bias voltage by exhibiting slight variations in output voltages. These variations indicated that the PID+ was sensitive to the applied voltage and that the output response could be modulated by adjusting the bias voltage. The fact that the output voltages returned to values close to the "AIR" readings during the intervals of zero bias voltage suggested that the PID+ had a relatively fast response time and could quickly recover its baseline state when the bias voltage was removed. Overall, the results demonstrated the PID+'s ability to detect and respond to changes in bias voltage, which was an essential characteristic for its operation as a gas sensor.

The tests conducted so far indicated that the first prototype of PID+ demonstrated the highest response at an electrode bias voltage of 600V to 750V and flow rate and concentration of 1LPM and 5ppm respectively, as shown in Table 8.2-4. In order to conclude the tests and verify these finding, a final test was conducted. The findings of this experiment demonstrated that the PID+ did not have the capacity to detect ISB gas for 500 ppb and 1 ppm concentrations at flow rates of 500 mLPM and 1.5 LPM, whereas, at a flow rate of 1 LPM, the PID+ exhibited a considerable response to the shift in the ionization current.

However, there were some challenges faced in the course of the testing of the first generation of the PID+ prototype. The first initial challenge was the lack of ground around the bulb in order to provide resonance. The bulb used during this research was an AC electric field

excited UV bulb. For the bulb to emit UV light, it needed ground in order to create the resonance. This was solved by placing a ring-shaped conductor connected to the ground plane of the PCB. Then the high voltage circuit was not able to produce as much voltage as the IC can provide out of the circuit which is needs to be fixed. It has been observed that the response was initially increasing with the increase in the bias and then decreasing, which implied that the electrode size was not optimal. The results obtained have been affected by the problems in high voltage circuits and failures of op-amp ICs in amplifier circuits from time to time. Lastly, since the output of the transimpedance amplifier was applied to the inverting terminal of the voltage amplifier, this led to a reduction in the signal below the baseline signal while detecting ISB.

These challenges were overcome in the second generation PID+. The goal for the second part was to build the next version for the PID+ and test it to increase the response and selectivity of the device for the detection of gases. The circuit for the second version of the PID+ was edited taking into consideration the outputs reported from the first version of the PID+. The changes implemented in the second version included changes to the bulb circuit. The reason for the change in this circuit was to provide a more stable signal required for the excitation of the UV bulb. The new bulb circuit included a signal generator, amplifiers, and a step-up transformer to provide an AC voltage with the ability to match the resonant frequency of the UV bulb. Moreover, in the second prototype, the microcontroller unit utilized in version 2 was fitted with the SAMD21 ARM® Cortex®-M0+ based microcontroller instead of using Arduino micro-controller boards in order to decrease the size of the PID+. Furthermore, in order to rectify the negative signal, the output from transimpedance signal was applied to the non-inverting terminal of the voltage amplifier.

The testing of the second generation PID+ began by selecting the optimal gain for the amplifier for the output signal as shown in Figure 8.4-1. It was important to select the right gain for the amplifier because this directly affected the sensitivity and accuracy of the detector. It was observed that the amplifier gain was too low, the signal was too weak to be accurately measured and was not able to detect ISB gas. Conversely, if the amplifier gain was too high, it resulted in saturating the output signal. In order to achieve the best results, the gain of the amplifier was set to 100.

Furthermore, the second test was carried out in order to identify the ideal bias voltage that would generate the greatest level of output for the electrode. During this test, the bias voltage was varied, and it was observed that the highest level of response was obtained at a much lower voltage than for the first prototype PID+. In the initial iteration, the optimal voltage was determined to be within the range of 600V to 750V, while in the subsequent iteration, the optimal voltage range was found to be between 500V to 600V as shown in Table 8.4 1 and Table 8.4 2. A third test was further conducted by keeping the bias voltage to 500V and gain of the amplifier to 100 in order to determine the flow rate at which the second generation PID+ showed the highest response. The findings derived from the third experiment, as depicted in Figure 8.4-3, demonstrated that the PID+ exhibited the generation of photoionization current, which varied in response to the flow rate of the ISB gas. The maximum output was observed at a flow rate of 1LPM, which was consistent with the outcomes obtained during the first generation PID+ testing.

Finally, a fourth experiment was carried out with the objective of determining the concentration of ISB gas at which the second-generation PID+ yielded the maximum photoionization current. The test results shown in Figure 8.4-4 demonstrate that the PID+ showed a response to ISB gas with a concentration as low as 150 ppb. The dynamic range for the second generation PID+ ranged from 150 ppb to 5 ppm. This proved that the response of the second generation PID+ was improved in comparison to the first version of the PID+.

During the testing of the two versions, the PID+ showed sensitivity to changes in gain settings, bias voltage settings, and flow rates, as evidenced by the varying output voltages. However, the effect of these factors was not dramatic, as indicated by the relatively small changes in output voltages and the presence of overlapping error bars. The error bars demonstrated the variability in the data and provided important information about the statistical significance of the observed differences. While the PID+ exhibited sensitivity to different experimental parameters, the extent of these effects might require further investigation and optimization for specific application requirements. While the PID+ showed some sensitivity to changes in input values, it could be considered relatively insensitive when compared to the wide range of input values tested. This information could be useful for further optimization and calibration of the PID+ to achieve the desired sensitivity and response characteristics for specific applications.

Lastly, some of the limitations faced during the testing of the second generation PID+ were that, first, the bootloader was not working properly with the device, so the code was updated. Second, there were some faults in the crystal oscillator and voltage regulators of the bulb circuit which were rectified.

The testing of the third version of the PID+ was not able to be conducted due to a fault in the PCB design and due to COVID pandemic and related restrictions, my ability to test the third version of the PID was significantly impacted. Due to this, the testing of the third iteration of the PID+ device was unfeasible. During preliminary testing, the bulb circuit functioned correctly and was able to generate a precise signal for the UV bulb excitation, as evidenced in Figure 8.5-2. However, complications with the 5V power line resulted in amplifier malfunction. Despite replacing the amplifiers and restoring the 5V power supply, a subsequent undervoltage issue caused microcontroller brownout. Additional testing is therefore required to evaluate the efficacy of the third version of the PID+ device.

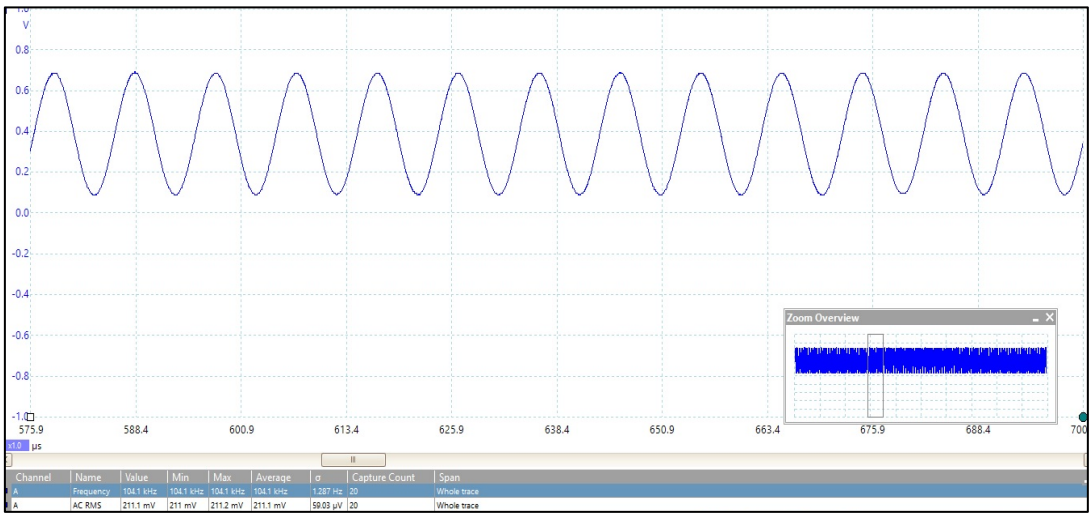


Figure 8.5-2: sinewave generated by the AD5932 of the bulb circuit.

Despite these limitations, our study provides valuable insights into the design and optimization of PID+ electronics. By improving the response and accuracy of the PID+ prototype, we can enhance their ability to detect low-level VOCs and improve the early detection of diseases and other health issues. Moving forward, further research is needed to explore the potential of novel electronic components and configurations for improving the performance of PID. Additionally, the use of advanced data processing techniques, such as

machine learning algorithms, could help to further improve the accuracy and specificity of VOCs detection.

8.6. Conclusion:

In this chapter, we investigated the electronics of a PID+ device and its impact on response and accuracy. Specifically, we tested the performance of two versions of a PID+ prototype at different flow rates, different concentrations of ISB gas, different bias voltages of detector plates, and different gains for the amplifier circuits.

Our results have shown that the response and accuracy of the PID are highly dependent on these variables. We found that increasing the flow rate of the sample gas improved the response of the detector but also that too high of a flow rate led to a decreased response due to reduced ion-molecule collision rates. In our study, we found that 1 LPM flow rate generated the highest photoionization response. Additionally, we found that the concentration of the target gas, in this case ISB gas, was a crucial factor that affects the response of the detector. It was discovered that through the optimization of the PID+ performance, the dynamic range of the response was enhanced. The initial generation of the PID+ had a dynamic range of 500 ppb to 5 ppm, while the second iteration's dynamic range was increased to 100 ppb to 5ppm.

Furthermore, we found that the bias voltage of the detector plates and the gain of the amplifier circuit had significant effects on the response and accuracy of the detector. Increasing the bias voltage led to a higher response, but excessive high voltage resulted in saturation. This, in turn, caused either a decline in output or no further increase. The gain of the amplifier circuit should also be carefully selected to balance the trade-off between response and noise.

Overall, our study provides valuable insights into the design and optimization of ion-mobility. By carefully selecting the optimal flow rate, concentration, bias voltage, and amplifier gain, we can improve the response and accuracy of PID for the early detection of diseases and other health issues. However, further research is needed to validate our findings across different types of VOCs. Additionally, future studies should explore the potential of

novel electronic components and configurations for improving the performance of PID and enhancing their ability to detect low-level VOCs.

8.7. References:

[1] G. Paglia, A. J. Smith, and G. Astarita, "Ion mobility mass spectrometry in the omics era: Challenges and opportunities for metabolomics and lipidomics," *Mass spectrometry reviews,* vol. 41, no. 5, pp. 722-765, 2022.

[2] J. Bílek, P. Maršolek, O. Bílek, and P. Buček, "Field Test of Mini Photoionization Detector-Based Sensors—Monitoring of Volatile Organic Pollutants in Ambient Air," *Environments*, vol. 9, no. 4, doi: 10.3390/environments9040049.

[3] A. N. Freedman, "The photoionization detector: Theory, performance and application as a low-level monitor of oil vapour," *Journal of Chromatography A,* vol. 190, no. 2, pp. 263-273, 1980/04/04/ 1980, doi: https://doi.org/10.1016/S0021-9673(00)88229-1.

[4] S. Mergemeier, I. Ebner, and F. Scholz, "Basic experimental studies on the operation of photoionization detectors," *Fresenius' Journal of Analytical Chemistry,* vol. 361, no. 1, pp. 29-33, 1998/04/01 1998, doi: 10.1007/s002160050829.

Chapter 9. Conclusions and Further work

9.1. Conclusions:

In this research, we examined the potential of VOCs as biomarkers for the diagnosis and monitoring of diseases using analytical instruments and developed an in-house point of care device based on ion- mobility (PID+). We investigated the potential of analytical techniques, such as, gas chromatography-ion mobility spectrometry (GC-IMS) and gas chromatography-time of flight mass spectrometry (GC-TOF-MS), and electronic nose (eNose) technology for the detection and identification of volatile organic compounds (VOCs) in diseases. We also examined the performance and capabilities of the photoionization detector (PID+) as an ion mobility technique for the measurement of volatile organic compounds (VOCs) and investigated the feasibility of using a PID+ as a diagnostic tool for disease detection. Our results suggest that both in-house devices and analytical instruments had advantages and disadvantages.

The first aim of our research was to evaluate the capabilities of the analytical devices, such as GC-IMS, GC-TOF-MS, and eNoses, in distinguishing between different disease groups as well as healthy controls. In this study, we compared urine samples for different cancers patients, such as colorectal cancer, prostate cancer, bladder cancer and hepatocellular cancer and patients with urinary tract infection disease with healthy individuals. The devices demonstrated their effectiveness in differentiating between the various disease groups and healthy individuals based on the VOCs patterns detected in the urine samples. The results indicate that these advanced analytical techniques hold promise as potential diagnostic tools for early disease detection and differentiation. We found that GC-IMS was able to distinguish between BCa samples and healthy control samples with an AUC of 95%, CRC samples and healthy control samples with AUC of 83% and PCa and healthy control samples with an AUC of 89%. Similarly, eNose PEN3 was able to distinguish between BCa samples and healthy control samples with an AUC of 92%, CRC samples and healthy control samples with an AUC of 75% and PCa and healthy control samples with an AUC of 89%. Moreover,

we found that all the analytical instruments were able to distinguish between cancer and UTI patients' urine samples from healthy controls with high sensitivity proving that these techniques can be used as an early diagnosis of cancer or disease. However, during the testing of urine samples of UTI disease, these analytical instruments were not as successful in distinguishing UTI positive samples from negative samples. Previous research has shown much higher sensitivity and specificity [2-5]. The reason behind this was that the urinary samples used were reported to be mishandled during storage which may have resulted in the dilapidation of the metabolic characteristics of these samples [6]. These findings suggest that VOCs analysis has the potential to be a valuable diagnostic tool for a wide range of diseases. Both GC-IMS and GC-TOF-MS can be effective techniques for the analysis of complex VOCs mixtures. While GC-IMS offered high sensitivity and rapid analysis times, GC-TOF-MS provided higher resolution, allowing for more confident compound identification. In our study, we found that the GC-IMS, GC-TOF-MS, AlphaMOS FOX 4000 and PEN3 eNose were able to separate different cancer groups and disease groups from each other as well as non-cancerous group.

The second aim of our research was to generate and produce VOCs fingerprints specific to different diseases investigated in the study. By analysing urine samples from patients with various diseases and healthy controls using GC-TOF-MS, we successfully identified distinct VOCs patterns associated with each disease. We identified and characterised VOCs profiles of different cancers and diseases through various analytical techniques and electronic nose (eNose) technology. Throughout the course of this research, we conducted an analysis of the chemical composition of bladder cancer and detected 13 VOCs out of which we found 3 noteworthy VOCs, Biphenyl, Heptanal, and 2, 6, 10, 14-tetramethyl- Pentadecane that did not overlap with other studies. Biphenyl has been identified as the most significant biomarker in our study. Biphenyl has been linked to various diseases, including carcinoma. It has been proven that Biphenyl is a promoter of BCa in rats [1]. We also characterized the chemical signature of colorectal cancer and detected 17 VOCs during experimentation. All 17 VOCs were reported before as colorectal cancer biomarkers. A total of 46 VOCs were found to be relevant for identifying these cancer groups, with several VOCs distinct to each cancer. These VOCs fingerprints served as unique markers for specific diseases, enabling accurate differentiation between them. The identification of disease-specific VOCs profiles is crucial as it provides valuable insights into the underlying biochemical processes and

metabolic changes associated with different health conditions. Moreover, these VOCs fingerprints have the potential to enhance disease diagnosis and monitoring, offering a non-invasive and early detection approach for a wide range of medical conditions. The successful generation of disease-specific VOCs fingerprints has opened up new avenues for future research and clinical applications, paving the way for the development of innovative diagnostic tools and personalized healthcare strategies. Our results demonstrate that the analytical techniques used in this study have the potential to detect and identify VOCs associated with various diseases, including cancer and other diseases, and that they can be valuable diagnostic tools for the early detection and monitoring of these diseases.

In addition to evaluating existing analytical devices such as GC-IMS, GC-TOF-MS, and eNoses, the third and final aim of our research involved the development of an in-house device based on ion-mobility techniques. This endeavour aimed to address the limitations of the existing analytical instruments and create a more affordable, portable, and efficient alternative. The new device leverages ion-mobility technology to analyse VOCs in a highly sensitive and selective manner, offering improved detection capabilities compared to conventional methods. By designing this device in-house, we had the flexibility to customize and optimize its performance according to our specific research needs. We successfully achieved accurate and reliable VOCs (ISB) measurements with the PID+ by carefully designing and optimizing its electronics, which encompassed the power supply, amplifier, and signal processing circuits. Our research revealed that the response of the PID+ could be influenced by various factors, such as environmental conditions and the composition of the sample under analysis. Recognizing the pivotal role of the PID+'s electronics in its performance, we emphasized the importance of meticulous design and optimization to ensure precise measurements. Moreover, we identified and implemented several techniques that led to significant enhancements in the accuracy and sensitivity of the PID's measurements. This achievement marks a significant advancement in VOCs analysis, paving the way for improved detection capabilities and broader applications of the PID+ in various scientific and industrial settings. The developments achieved in this research have paved the way for potential advancements in VOCs analysis and its broader applications. The in-house device we worked towards developing effectively overcame some of the limitations faced with commercial instruments, such as high costs and limited portability, thus enhancing accessibility and versatility for research and potential clinical use. This novel

in-house device represents a significant step forward in the field of VOCs analysis, offering innovative solutions for diverse research and diagnostic endeavours. The ongoing improvement and refinement of this device hold the promise of future breakthroughs in VOCs analysis, contributing to the progress of medical diagnostics and disease monitoring. By fostering a deeper understanding of volatile organic compounds and their correlations with various health conditions, the advancements achieved in this research open new possibilities for enhanced disease detection, monitoring, and personalized medical interventions.

However, we identified various challenges and limitations in this study. We identified that VOCs tend to be influenced by factors like sampling and analysis protocols, the potential for interference from confounding factors, and the need for larger and more diverse patient cohorts to validate the diagnostic utility of VOCs. We also found that the analytical techniques used for VOCs detection need careful sample preparation and analysis to avoid interference from matrix components, the potential for sample degradation or alteration during sample handling and collecting these was expensive and needed specialized equipment and trained personnel. These limitations suggested that though these analytical techniques were highly sensitive, these techniques could not be used as a long-term method for VOCs detection.

On the other hand, we also identified several limitations and challenges associated with the use of a PID+. These included the potential for interference from other compounds, variations in detector response, and the need for regular calibration and maintenance. These limitations suggest that further research is needed to optimize the use of a PID+.

Despite these limitations, our study has important implications for the future use of VOCs as biomarkers for diseases. By identifying the strengths and weaknesses of the techniques used for the detection of VOCs, we have contributed to the growing body of research on precision medicine and personalized care. Our findings suggest that VOCs analysis has the potential to improve disease diagnosis, monitoring, and treatment. Our findings also suggest that the PID can be a valuable tool in the identification and measurement of VOCs, and that it can help to improve our understanding of the impact of VOCs on public health.

Moving forward, we recommend that future studies investigate the use of ion mobility in larger patient cohorts and across multiple cancer types. This will allow for a more comprehensive evaluation of the technique's diagnostic utility and will help to identify any limitations or challenges associated with its implementation in a clinical setting. Further research is also needed to optimize the performance and reliability of the PID, and to develop strategies for addressing the limitations and challenges associated with its use. We also recommend that future research explores new approaches for improving the performance and reliability of the PID's electronics, and that they investigate the use of a PID in combination with other analytical techniques for more comprehensive VOCs analysis. Further research is also needed to optimize the design and performance of the PID's electronics for specific applications, and to develop strategies for addressing the challenges associated with their use.

Overall, our study provides important insights into the potential of VOCs as biomarkers for diseases and the use of GC-IMS and GC-TOF-MS for the detection and identification of VOCs. Furthermore, our study provides important insights into the electronics and signal processing of the PID+ for the measurement of VOCs. While further research is needed to fully evaluate their diagnostic utility and address any limitations, we believe that our findings have important implications for the future of precision medicine and personalized care.

9.2. Further work:

While the results of this thesis demonstrated the potential of photoionization detectors (PID) for the detection of volatile organic compounds (VOCs), there is still much to be done to improve their performance and expand their applications. In this section, we will discuss several areas for future work that could build upon the research presented in this thesis.

1. Development of a more sensitive PID+: One of the major limitations of PID is their limited sensitivity for certain VOCs. PID are limited by the proton affinity of the VOCs and the ionisation capability of the UV source. Therefore, one way of achieving a better range of VOCs detection is to use a 11.3 eV bulb over a 10.6 eV. However, UV bulbs with a 11.3 eV potential have a relatively short lifespan

compared to other types of bulbs, meaning that they will need to be replaced frequently, adding to the cost of the equipment. To overcome this, future work could focus on the development of a more sensitive PID+. One of the suggestions is to vary the form factor of the detector plates in order to develop the high field asymmetric electric field that separates gas-phase ions based on their characteristic differences in mobility in high and low electric fields. Another suggestion is to increase the electric field strength or optimise the geometry of the ionisation chamber in order to increase the efficiency of the ionization chamber. In our study, we found that the spacing between the electrode plates was smaller than required which led to electric arc. This resulted in the use of low voltages which affected the ionization efficiency. Therefore, both the electric field and correct detector plates format can result in more efficient ionization of the VOCs, which in turn can improve the sensitivity of the detector.

2. Improve the collection and amplification of the signal: The signal generated by the ionized VOCs can be amplified using various methods, such as increasing the gain of the detector circuit or using a more sensitive operational amplifier.
3. Reduce background interference: finally, to improve sensitivity it is important to minimize interference from background noise and other sources of interference. This can be achieved by carefully selecting the operating parameters of the detector, such as the flow rate and temperature, use fences around the electrodes to provide proper ground and remove humidity resistance, and by using appropriate filters or preconcentration techniques to remove interfering compounds. These technologies could improve the sensitivity and selectivity of the PID+, making them more effective for detecting low-level VOCs.
4. Improvements in the electronics of the PID+: the PID+ electronics produced in this study were tested and experimented by carefully balancing factors such as the UV bulb excitation method, the UV lamp energy, signal collection and amplification etc., and it was possible to achieve a more sensitive and reliable detection of VOCs. Though there were some parts that could be improved such as adding reverse polarity, adding a screen to the enclosure of the PID+, adding power switches and various voltage monitoring, adding over voltage, under voltage protection to the circuit. A new PID+ schematic was created, shown in Figure 9.2 1, Figure 9.2 2,

Figure 9.2 3 and Figure 9.2 4 in order to incorporate all the improvements suggested in this study and would have been implemented if not for COVID-19.

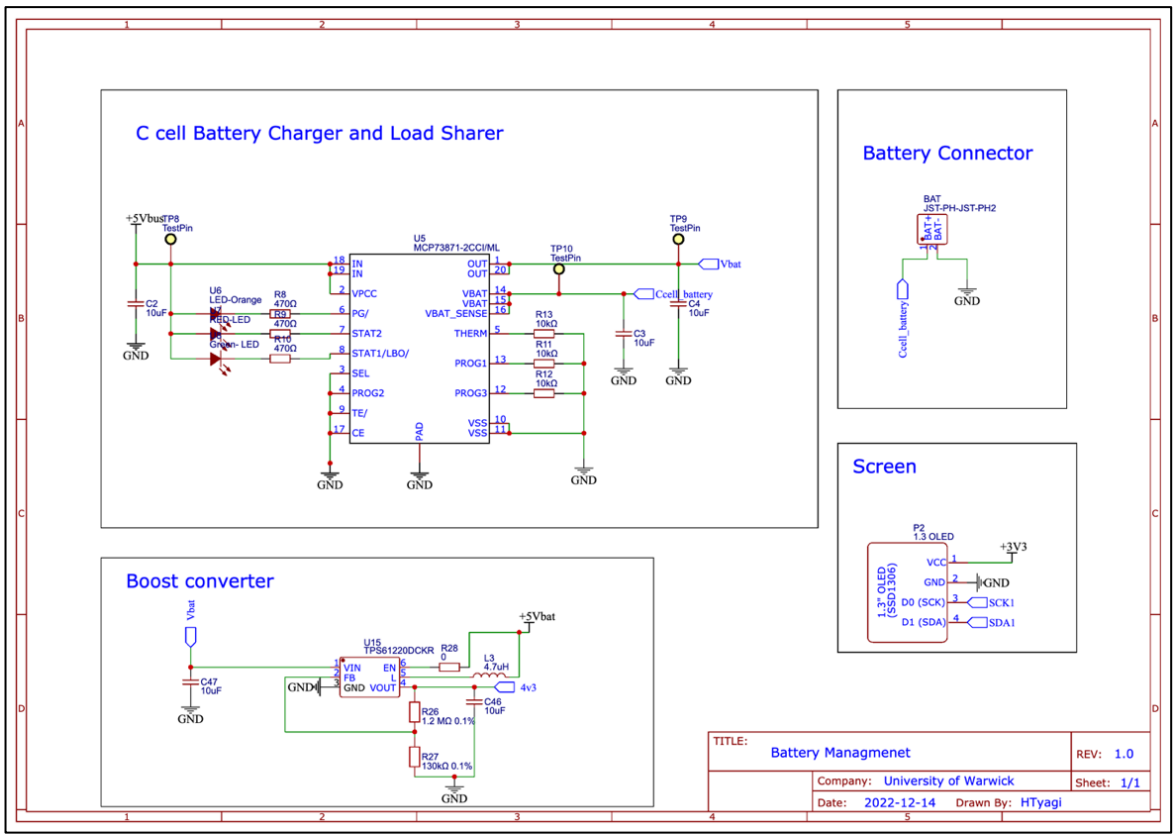


Figure 9.2-1: Shows the schematics for the rechargeable battery and charger circuit. The IC is selected here such that it can be used as load sharer which means when the battery dies, a 5 V bus from the micro-USB-B could be used both for charging the battery as well as to power the PID+.

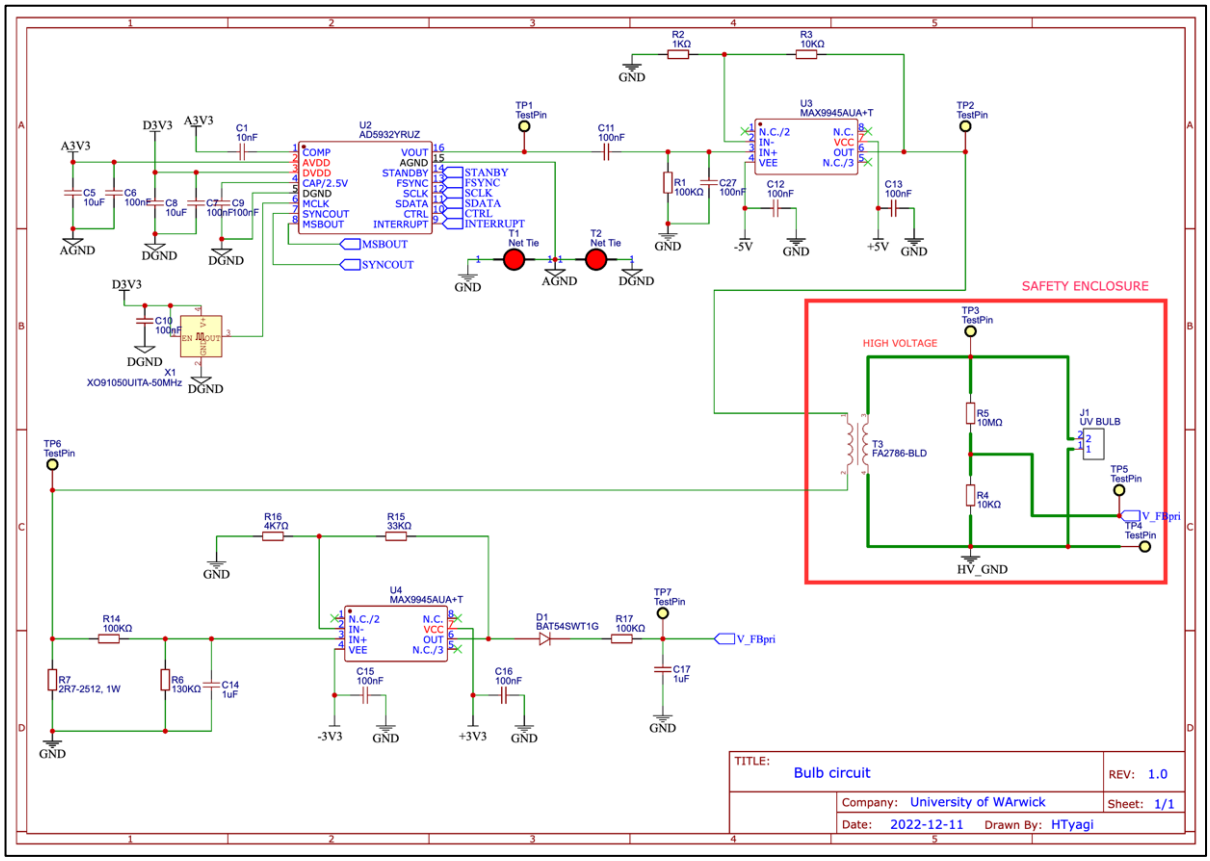


Figure 9.2-2: Shows the schematics for the bulb circuit with more stable signal generation and noise reduction.

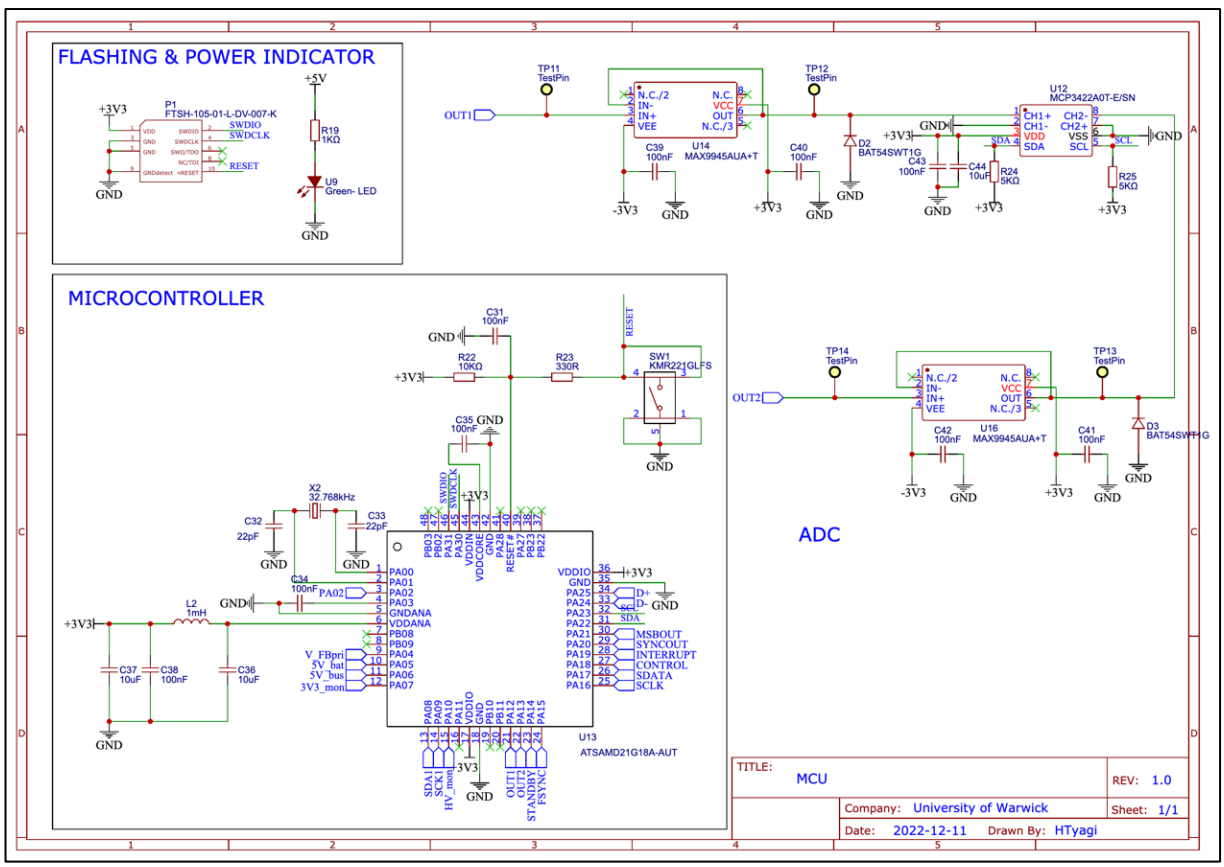


Figure 9.2-3: Shows the schematics for the microcontroller, ADC, and flashing circuit.

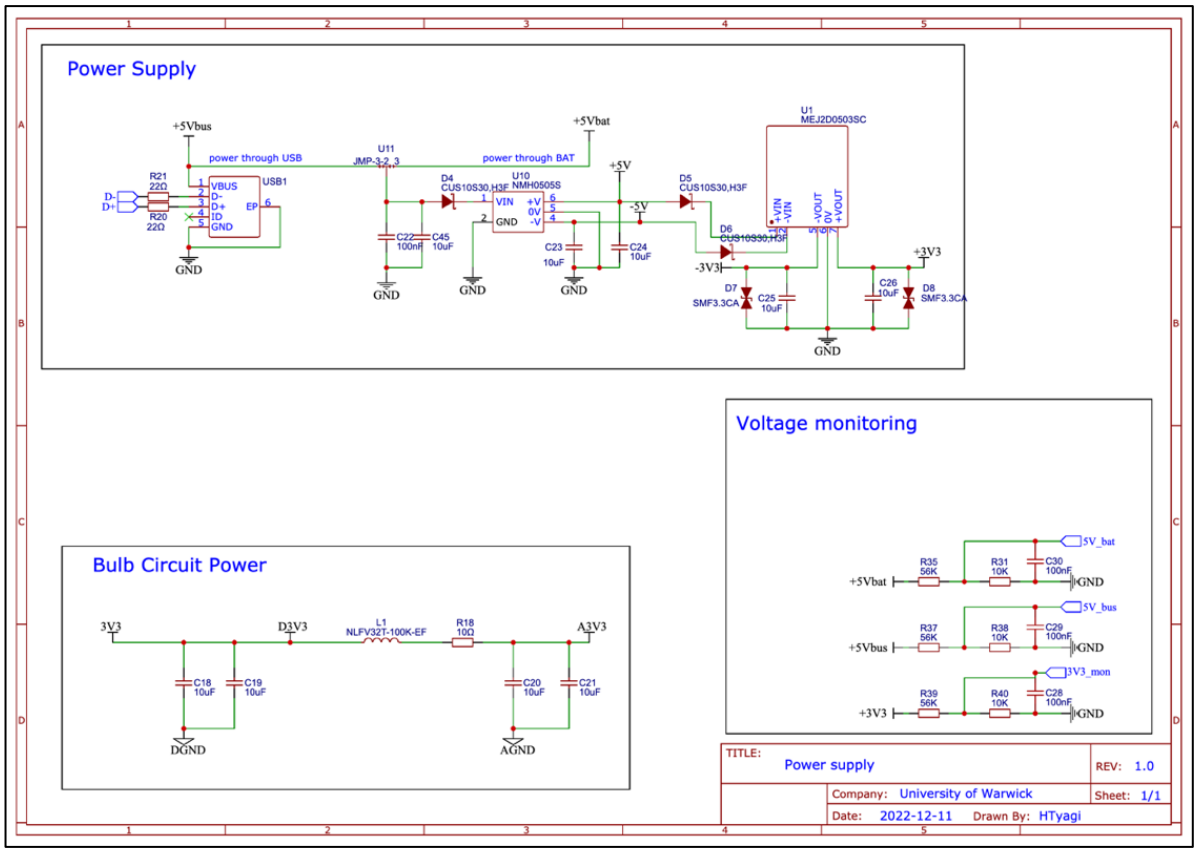


Figure 9.2-4: Shows the schematics for the various power supplies used in the circuit. The power supply was made more resilient to noise by adding extra reverse voltage protection and decoupling capacitors to reduce EMI. Various voltage monitoring circuits were also added in the schematics.

5. Miniaturization and portability of PID: Another area for future work is the development of miniaturized and portable PID. Currently, the PID+ was typically bulky and requires external power sources or frequent charging, which limited its usage. The development of miniaturized and portable PID could enable their use in a wider range of settings, including in point-of-care diagnostics and in remote monitoring applications.

A PID is a versatile device capable of detecting a wide range of VOCs in various applications such as disease detection, environmental monitoring, industrial safety, and indoor air quality. Its sensitivity and selectivity make it an ideal choice for detecting low-level concentrations of VOCs, making it an essential tool for researchers, first responders, and industrial professionals. PID can detect VOCs in concentrations as low as parts per billion (ppb) and can provide real-time monitoring, making it a valuable asset in identifying and mitigating VOCs exposure risks.

9.3. References:

[1] M. Cauchi *et al.*, "Evaluation of gas chromatography mass spectrometry and pattern recognition for the identification of bladder cancer from urine headspace," in *Analytical Methods* vol. 8, ed: Royal Society of Chemistry, 2016, pp. 4037-4046.

[2] A. Roine *et al.*, "Rapid and Accurate Detection of Urinary Pathogens by Mobile IMS-Based Electronic Nose: A Proof-of-Principle Study," *PLOS ONE,* vol. 9, no. 12, p. e114279, 2014, doi: 10.1371/journal.pone.0114279.

[3] A. K. Pavlou *et al.*, "Use of an electronic nose system for diagnoses of urinary tract infections," *Biosensors and Bioelectronics,* vol. 17, no. 10, pp. 893-899, 2002/10/01/ 2002, doi: https://doi.org/10.1016/S0956-5663(02)00078-7.

[4] V. S. Kodogiannis, J. N. Lygouras, A. Tarczynski, and H. S. Chowdrey, "Artificial Odor Discrimination System Using Electronic Nose and Neural Networks for the Identification of Urinary Tract Infection," *IEEE Transactions on Information Technology in Biomedicine,* vol. 12, no. 6, pp. 707-713, 2008, doi: 10.1109/TITB.2008.917928.

[5] K. Hewett *et al.*, "Towards the Identification of Antibiotic-Resistant Bacteria Causing Urinary Tract Infections Using Volatile Organic Compounds Analysis—A Pilot Study," *Antibiotics*, vol. 9, no. 11, doi: 10.3390/antibiotics9110797.

[6] M. Rotter *et al.*, "Stability of targeted metabolite profiles of urine samples under different storage conditions," *Metabolomics,* vol. 13, no. 1, p. 4, 2016/11/28 2016, doi: 10.1007/s11306-016-1137-z.

Appendix A

Appendix A contains the firmware used for the generation of AC signal for the excitation of UV bulb and for data collection.

```
int SDATA = 13;
int SCLK = 11;
int FSYNC = 5;
int STANDBY = 2;
int INTERRUPT = 12;
int CTRL = 10;
int SYNCOUT = 6;
/*      end of Pin setup for AD5932    */
 float out1 = 9;
 float out2 = 19;
 float vprim = A3;
 int vbias = A0;



/*      setting clock and byte configuration for AD5932
*/
void myShiftOut(uint8_t dataPin, uint8_t clockPin,
uint8_t bitOrder, byte val)
{
 int i;
 for (i = 0; i < 8; i++)  {
  if (bitOrder == LSBFIRST)
   digitalWrite(dataPin, !!(val & (1 << i)));
  else
   digitalWrite(dataPin, !!(val & (1 << (7 - i))));
  digitalWrite(clockPin, LOW);
  digitalWrite(clockPin, HIGH);
 }
}
/*      end of clock and byte configuartion for AD5932
*/



void setup() {


 SerialUSB.begin(19200);
 analogWriteResolution(10);
 pinMode(vbias,OUTPUT); //BIAS VOLTAGE
 pinMode(vprim, INPUT);   //feedback from primary
transformer CH3


 pinMode(out1 , INPUT);
 pinMode(out2 , INPUT);
 pinMode(A3 , INPUT);


/*  Start of Code for AD5932     */
 pinMode(CTRL , OUTPUT);
 pinMode(SCLK , OUTPUT);
 pinMode(SDATA , OUTPUT);
 pinMode(FSYNC , OUTPUT);
 pinMode(INTERRUPT , OUTPUT);
 pinMode(STANDBY , OUTPUT);


 digitalWrite(INTERRUPT , LOW);
 digitalWrite(STANDBY , LOW);
 digitalWrite(CTRL , LOW);
 digitalWrite(FSYNC , HIGH);
 digitalWrite(SCLK , HIGH);
 delay(10);


 //Control Register 16-bit
 digitalWrite(FSYNC , LOW);
 delay(1);
 myShiftOut(SDATA , SCLK , MSBFIRST , 0x0F);
 myShiftOut(SDATA , SCLK , MSBFIRST , 0xD3);
 delay(1);
 digitalWrite(FSYNC , HIGH);
 delay(10);



 //Fstart LSB 16-bit
 digitalWrite(FSYNC , LOW);
 delay(1);
 myShiftOut(SDATA , SCLK , MSBFIRST , 0xC8);
 myShiftOut(SDATA , SCLK , MSBFIRST , 0X0B);
 delay(1);
 digitalWrite(FSYNC , HIGH);
 delay(10);



 //Fstart MSB 16-bit
 digitalWrite(FSYNC , LOW);
 delay(1);
 myShiftOut(SDATA , SCLK , MSBFIRST , 0xD0);
 myShiftOut(SDATA , SCLK , MSBFIRST , 0x09);
```

```
delay(1);
digitalWrite(FSYNC , HIGH);
delay(10);

//Fsatrt delta f 24-bit <- input 32-bit (frame4 + frame5)
digitalWrite(FSYNC , LOW);
delay(1);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x20);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x00);
delay(1);
digitalWrite(FSYNC , HIGH);
delay(10);

digitalWrite(FSYNC , LOW);
delay(1);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x38);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x00);
delay(1);
digitalWrite(FSYNC , HIGH);
delay(10);

//Number of Increments 16-bit
digitalWrite(FSYNC , LOW);
delay(1);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x10);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x02);
delay(1);
digitalWrite(FSYNC , HIGH);
delay(10);



//Increment Interval
digitalWrite(FSYNC , LOW);
delay(1);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x63);
myShiftOut(SDATA , SCLK , MSBFIRST , 0xE8);
delay(1);
digitalWrite(FSYNC , HIGH);
delay(10);

digitalWrite(CTRL , HIGH);
delay(1);
digitalWrite(CTRL , LOW);
delay(25);
/*  END OF CODE FOR AD5932     */

 /* code for optimum frequency    */

 float feedback_voltage[40];
 int j=0;
 feedback_voltage[j]=vprim;
 if(j<40)
 {if(feedback_voltage[j]!= vprim)
 {j++;
 feedback_voltage[j]=vprim;
 }

  int resonance_freq=0;
  int lowest_voltage= feedback_voltage[0];
 for(j=1;j<40;j++)
 {if(lowest_voltage > feedback_voltage[j])
 {lowest_voltage=feedback_votage[j];
  resonance_freq=90+j;
 }
 }

 //Control Register 16-bit

 digitalWrite(FSYNC , LOW);
 delay(1);
 myShiftOut(SDATA , SCLK , MSBFIRST , 0x0F);
 myShiftOut(SDATA , SCLK , MSBFIRST , 0xD3);
 delay(1);
 digitalWrite(FSYNC , HIGH);
 delay(10);



 //Fstart LSB 16-bit
 digitalWrite(FSYNC , LOW);
 delay(1);
 myShiftOut(SDATA , SCLK , MSBFIRST , 0xC8);
 myShiftOut(SDATA , SCLK , MSBFIRST , 0X0B);
 delay(1);
 digitalWrite(FSYNC , HIGH);
```

```
delay(10);



//Fstart MSB 16-bit
digitalWrite(FSYNC , LOW);
delay(1);
myShiftOut(SDATA , SCLK , MSBFIRST , 0xD0);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x09);
delay(1);
digitalWrite(FSYNC , HIGH);
delay(10);

//Fsatrt delta f 24-bit <- input 32-bit (frame4 + frame5)
digitalWrite(FSYNC , LOW);
delay(1);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x20);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x00);
delay(1);
digitalWrite(FSYNC , HIGH);
delay(10);

digitalWrite(FSYNC , LOW);
delay(1);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x38);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x00);
delay(1);
digitalWrite(FSYNC , HIGH);
delay(10);

//Number of Increments 16-bit
digitalWrite(FSYNC , LOW);
delay(1);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x10);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x02);
delay(1);
digitalWrite(FSYNC , HIGH);
delay(10);



//Increment Interval
digitalWrite(FSYNC , LOW);
delay(1);
myShiftOut(SDATA , SCLK , MSBFIRST , 0x63);
myShiftOut(SDATA , SCLK , MSBFIRST , 0xE8);
delay(1);
digitalWrite(FSYNC , HIGH);
delay(10);

digitalWrite(CTRL , HIGH);
delay(1);
digitalWrite(CTRL , LOW);
delay(25);

}


void loop() {

 SerialUSB.println("PID+ is Ready");

/*    DAC OUTPUT SETUP      */
 int val=0.5;
 analogWrite(vbias,0); //initalise DAC to zero
 int i=217;  //control pin minimum value 0.7V
 while(i<1024)
 {
  analogWrite(vbias,i);
  i++;
  delay(1000);
  if(i==1023)
  {  i==217;
 }
 }


 SerialUSB.println("PID+ is Ready");
 float x= analogRead(out1);
 float y= analogRead (out2);
 float z = analogRead (A3);
 float Vout1 =(x*5)/1023;
 float Vout2=(y*5)/1023;
 float fbout=(z*5)/1023;
 SerialUSB.println( "OUTPUT 1 is:");
 SerialUSB.println (Vout1);
```

```
  SerialUSB.println( "OUTPUT 2 is:");
  SerialUSB.println (Vout2);
  SerialUSB.println( "OUTPUT from transformer
feedback is:");
  SerialUSB.println (fbout);
  delay(1000);
 /*


 /* if(counter > 2000){
    Serial.println("counter reset");
    digitalWrite(CTRL , HIGH);
    delay(1);
    digitalWrite(CTRL , LOW);
    delay(25);
    counter = 0;
  }*/


  counter++;



//  int sensorValue = analogRead(A0);
//  float voltage = sensorValue * (5.0 / 1023.0);
//  Serial.println(voltage);*/
```

Appendix B

Appendix B contains the link to the GitHub library of the R program used in this study to analysis the urine samples.

https://github.com/JimSkinner/toftools