D4.2 – Protocols for the Detection of Acute Viral Disease

WP number and title	**W4 Design, Development, and Delivery**
Deliverable number	D4.2
Lead Beneficiary	UIBK
Deliverable type	Report
Planned deliverable date	31 March 2025
Date of Issue	31 March 2025
Dissemination level	PUBLIC
Authors	UIBK (Anesu Chawaguta, Chris A. Mayhew and Lorenzo Peralia) HUA (Christos Sardianos) Solgenium (Matthias Scloegl) ROBO (Tim Gibson) Telesto (Chrysostomos Symvoulidis)

1. Introduction

This deliverable provides details on how a breath volatilome measured either with Gas Chromatography – Ion Mobility Spectrometry (GC-IMS) combined with AI/ML or with a gas sensor combined with multivariate statistical analysis can be used to provide a method to provide rapid decisions (minutes with GC-IMS or expected to be tens of minutes with the sensor device) in terms of deciding who is infected or not with a virus. This is the first step in containing a viral outbreak. A rapid response is needed to mitigate the impact of the entry of a novel virus into the human population. Acute inflammation volatile biomarkers present in exhaled breath can be used to determine whether a contagion is present or not.

By providing rapid triage, infected people can be isolated. This profiling in based on clinical studies undertaken by the UIBK and SMCH teams using patients breath profiles who are known to have viral infections compared to healthy control profiles. Once diagnosed, more detailed clinical analysis will be required to specify what the viral infection is (Task 4.5).

The concept is to provide a first-line of defense by providing near-to-real-time situational awareness to protect people, healthcare systems and economies.

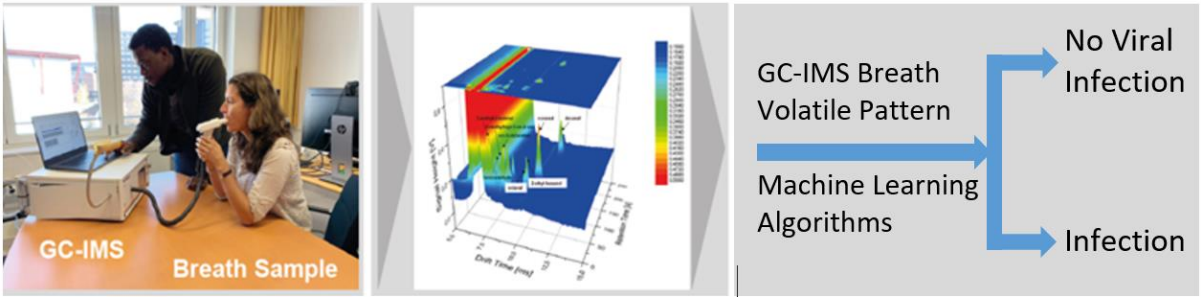
2. GC-IMS Breath Profiles and ML/AI Analysis (UIBK, HUA, Solgenium and Telesto)

Using GC-IMS **for breath pattern profiling** with artificial intelligence (AI) and machine learning (ML) algorithms, we developing a technological platform for **non-invasive breath infection-screening**.

2.1 Breath Collection Method and GC-IMS Breath Profiling

2.1.1 Conceptual Workflow

The conceptual workflow is illustrated below:



Exhaled breath can be sampled directly into the instrument. For the discovery programme, end-tidal orally exhaled breath sampling was undertaken. (However, see Appendix A for UIBK and G.A.S.'s recommendation for improving breath sampling for use in diagnosing viral infections.)

D4.2 has involved the following:

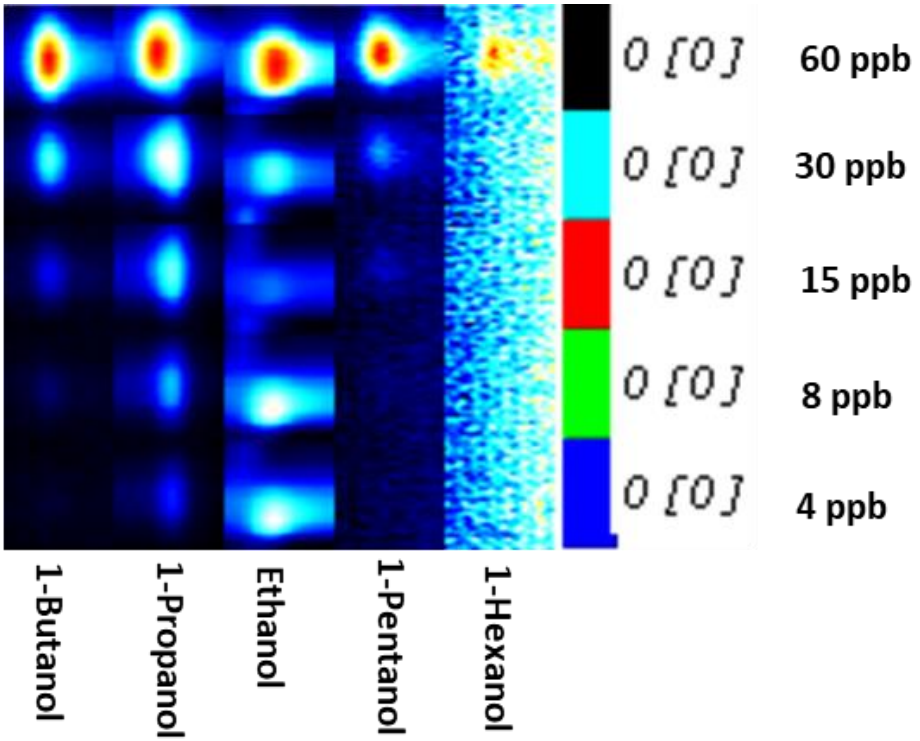
(i) Development and implementation of sampling procedures and rapid testing (e.g., temperature and breath volatile profiles) for the reliable, reproducible and rapid clinical analyses. This is illustrated above GC-IMS measurements using G.A.S. BreathSpec systems (see Appendix B).

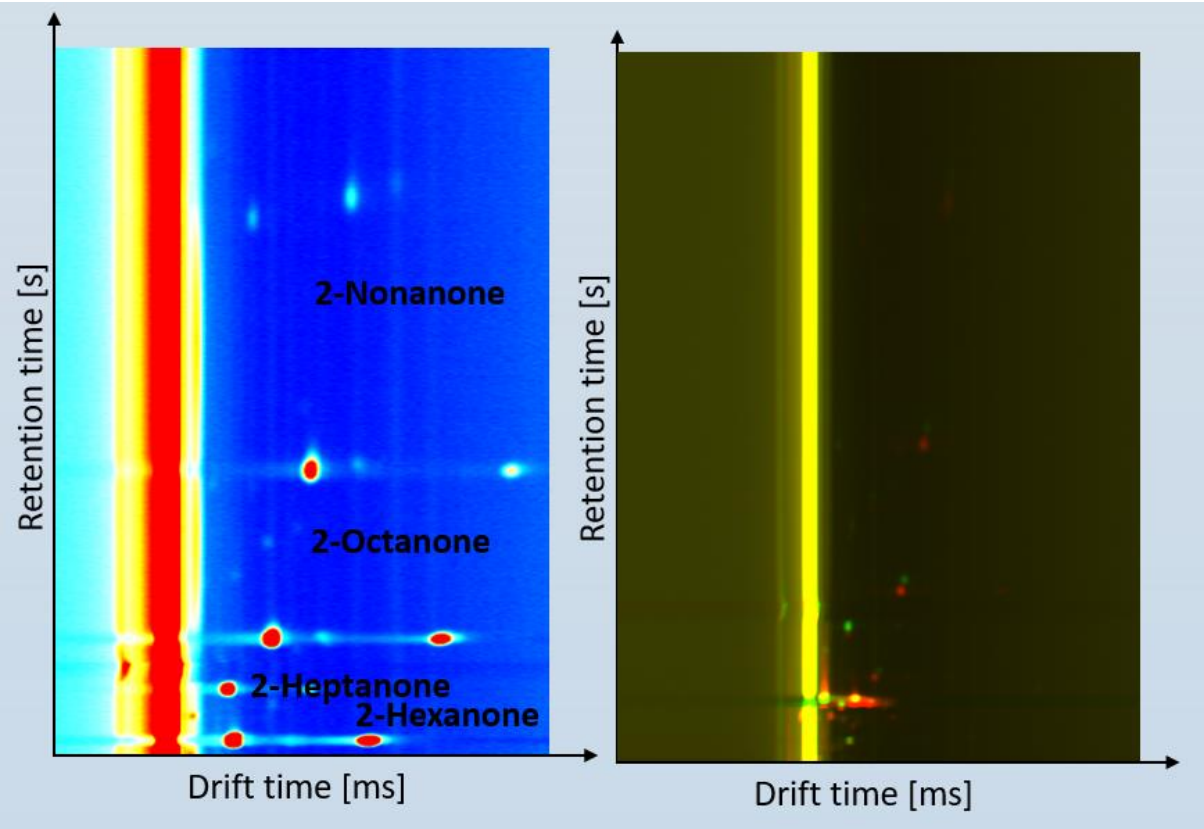
(ii) Development of AI/ML that combined biomarker "*breathprints*" can rapidly differentiate uninfected from infected people and then to discriminate infected people into those having either viral, bacterial, fungal or coinfections.

To assure quality data, quality checks are performed to ensure that the data are interoperable, structured correctly and of high quality. The GC-IMS breath data is collected and the anonymised analytical data are uploaded. For patient flow management, the RRML personnel can register incoming patients.

2.1.2 Calibration, Quality Control and Quality Assurance

Examples of calibrations for alcohols (left) and ketones (centre) are provided below. Topographic view of an aldehyde test undertaken everyday as part of our quality control procedures showing an overlay of the measurements across all sites (right).





2.1.3 Limits of Detection (LOD) and Quantification (LOQ)

LOD and LOQ values in parts per billion by volume (ppbv) were determined through volatile calibration (n = 25). The heavier alcohols, i.e., 1-heptanol and 1-nonanol, can be detected with GC-IMS, but require a longer runtime.

Substance class		GC-IMS	
		LOD [ppb_v]	LOQ [ppb_v]
Ketone	2-Butanone	10	30
	2-Pentanone	7	21
	2-Hexanone	2	6
	2-Heptanone	2	6
	2-Octanone	10	30
	2-Nonanone	16	48
Alcohol	1-Pentanol	9	27
	1-Hexanol	15	45
	1-Heptanol	-	-
	1-Nonanol	-	-

2.2 Data Handling, ML/AI Analysis, and Diagnostic Feedback Workflow

The following text provides a high-level overview of the GC-IMS data collection, analysis, and LIMS integration workflows. For a comprehensive and detailed technical description, please refer to **D4.1 - "Data Map and Integration"** [D4.1].

The process of collecting GC-IMS Spectra and analysis is enabled by implementation of a centralized library information management service (C-LIMS) allowing for data upload and quality control measures. This streamlines all procedures in the ONELAB RRML. The process to combine multi-centre data to support a standardized dataset ready for AI analysis is also handled in this platform.

The C-LIMS is a containerized Python application hosted on a Google Cloud Platform (GCP) virtual machine (EU-based servers), with user authentication via nginx and code maintained in a Solgenium owned GitLab repository. It supports deployment in both cloud and local environments. The system features a Streamlit-based frontend for controlled access by lab personnel and quality managers. Uploaded data are stored in raw form, then converted to standardized formats (SQLite, HDF5). It also includes trained models for viral infection detection using GC-IMS VOC data. The system also features an API service to communicate with local LIMS services (L-LIMS).

For classification models to detect infections from breath VOCs, supervised machine learning models were developed and integrated into the LIMS to enable automated infection classification based on GC-IMS data. Machine learning helps process and analyze complex, high-dimensional biomedical data. To support accurate classification (e.g., infected vs. non-infected), a variety of supervised ML models were tested after applied pre-processing steps. These AI models and preprocessing pipelines enable the LIMS to automatically analyze incoming GC-IMS data in real time, supporting decision-making in both cloud-based and local laboratory settings.

The L-LIMS is a local version deployed within the ICT system of a RRML setup, capable of operating offline and syncing with C-LIMS when internet is available. An API enables communication with the Patient Registration Mobile App, GC-IMS measurement devices and synchronization with central services. The system has reached Technology Readiness Level (TRL) 6, having proven reliable in a real-world pilot (FTX), with successful integration, consistent performance, and positive user feedback.

2.3 Preliminary Results – indicative sensitivity and specificity for distinguishing infected versus non-infected

Using clinical data obtained by UIBK only, the table below shows classification results for viral versus no infection. The metrics provided give the mean, standard deviation (std), min and max from 30 train/test shuffle splits. The sample used included 247 breath samples, of which 116 were associated with people who had confirmed viral infections.

Dataset	Mean	Std	Min	Max
Random Forest Classifier				
accuracy	0.87	0.04	0.76	0.92
sensitivity	0.85	0.06	0.68	0.94
specificity	0.88	0.07	0.64	0.94

3. Sensor measurements and statistical analysis and ML/AI Analysis (ROBO)

3.1 Profile of the technology

The Roboscientific 720 VOC Analyser is a portable, handheld instrument that detects volatile organic compounds (VOCs) in the gas phase and records the sensor responses to the VOCs. It is battery powered and may be used in laboratory or field settings as required.

The instrument contains 24 proprietary sensors, configured into a sensor array, that are sensitive to many different VOCs which gives a high level of flexibility for different sample types. A sensor array designed to detect viruses, originally aimed towards the accurate detection of SARS-CoV-2 (COVID-19 disease) is fitted into the instruments dedicated to the OneLab project. The model 720 is supplied in a rugged NANUK case which keeps the instrument safe and secure when not in use.

Accessing the data from the sensors requires interfacing with a laptop or desktop PC. The model 720 is fitted with a micro-USB port enabling data transfer and charging of the battery. When connected to a PC, the model 720 is displayed as an external hard drive on Windows Explorer and data can be transferred in the usual way. VOC Sampling: The analytical sequence of sampling is as follows:

Baseline – clean air from the side intake,

Absorption – sample from the front sample port,

Pause – all air flow is paused for a timed period,

Desorption – clean air is passed over the sensors to start desorbing the VOCs from the sensor materials,

Flush – clean air at maximum flow is passed over the sensors to return to baseline values.

Multiple repeat samples are usually taken from a single sample to allow sensor response averaging to take place, therefore a 'Wait' function is included, with a timed wait between repeat samples. The term given to this overall sequence of events is 'The Profile'. A Profile of 2-5-1-5-15-3 is defined as follows:

Baseline	2 seconds
Absorption	5 seconds
Pause	1 second
Desorption	5 seconds
Flush	15 seconds
Wait	3 seconds

It is usual to have the Desorption and Flush timings 4 to 5 times the length of the Absorption timing to ensure the sensor responses return to baseline values and any VOCs measured are flushed from the sensor surfaces. Expanded details of the events on performing a sample are described below:

Baseline: When a sample reading is initiated, the instrument records the baseline values of the sensor outputs for a selected number of seconds before the sample headspace is drawn through the chamber.

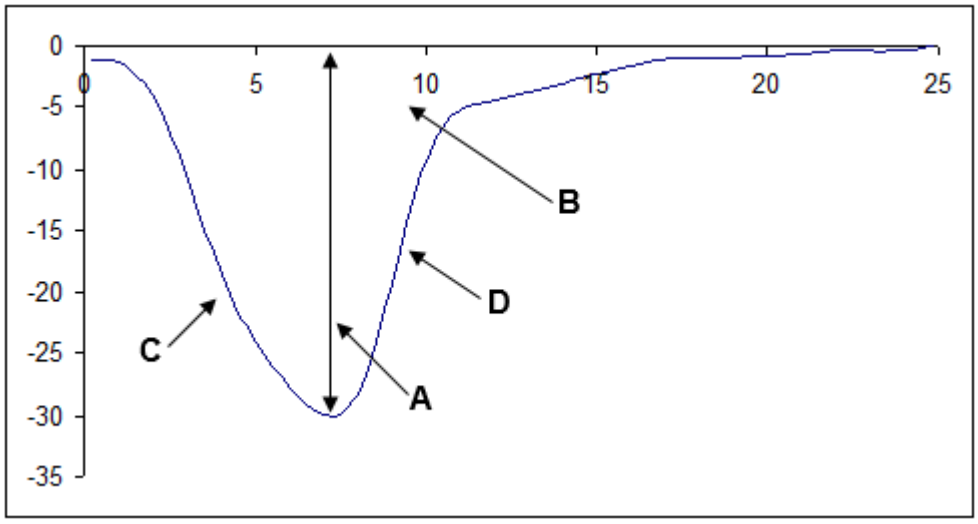
Absorption: During exposure to a volatile sample the composite electronic characteristics of the sensor materials of each sensor type changes to give defined sensor outputs that are recorded and stored. These changes are brought about by the interaction of volatile organic compounds (VOCs) with the active surface of each sensor. The actual responses observed for each sensor, which are usually quite different in character, are due to the complex interaction of the different VOCs in the sample mixture with different sensor materials used in the sensor array fitted.

Pause: A pause, during which the flow of sample over the sensors ceases, is a variable timed function and may give information in certain applications as a dynamic equilibrium occurs between the VOCs in the sample and the sensor materials.

Desorption: After the timed 'pause', clean external air is again passed over the sensors and the sensor outputs starts to return to the baseline values as the VOCs leave the sensor surface.

Flush: A final maximum air flow at the end of the desorption time is added to fully remove VOCs from the sensor surfaces, completing the sampling sequence and making sure the sensors are at the baseline values ready for the next sample.

Characteristics of the Sample Data: The Profile chosen for the samples taken will form a transient curve of response for each sensor. The figure below indicates this for a single sensor.



Key:

A: Divergence or Peak Height (maximum step response - positive or negative)

B: Area under the curve

C: Absorption (positive rate of change)

D: Desorption (negative rate of change)

The divergence, slope profiles and area are characteristic of both the sensor and the particular headspace volatiles being sampled. The differential responses in an array can, therefore, be used to digitally portray the VOCs, giving an Odour Fingerprint of that particular VOC mixture.

The output of the sensor array is not an absolute value of specific VOCs. It is a comparative output, giving the capability of discrimination of infected vs non-infected and as such the system is usually ‘trained’ on known samples so as to give a rapid positive / negative analysis in the field.

3.2 Breath sampling and collection

Breath analysis is an extremely useful non-invasive route for diagnosis and over the last few years has been the subject of intense research and development of new techniques, with significant funding for a plethora of scientific projects. The majority of the work has focused on traditional analytical techniques such as GC-MS, GC-IMS and PTR-MS, however sensor research is not too far behind.

Breath as a diagnostic sample is challenging. It is 100% saturated in water vapour on voiding from the lungs and contains hundreds or VOCs, most of which are present in the low parts per billion (ppb) range (ug per litre). The principle of the 720 VOC analyser is to measure the changes in the VOCs when persons are infected with a virus and report that change. Such is the challenge to be addressed.

This combination of VOCs and saturated water content has been addressed by using a simple breath collection device derived from a particular adsorbent fibre. The fibre is microporous and massively adsorbent for a huge range of VOCs, which can be released from the fibre by heat. It is physically superior to granular adsorbents as it lends itself to make simple planar VOC adsorbing filters with virtually no back pressure issues.

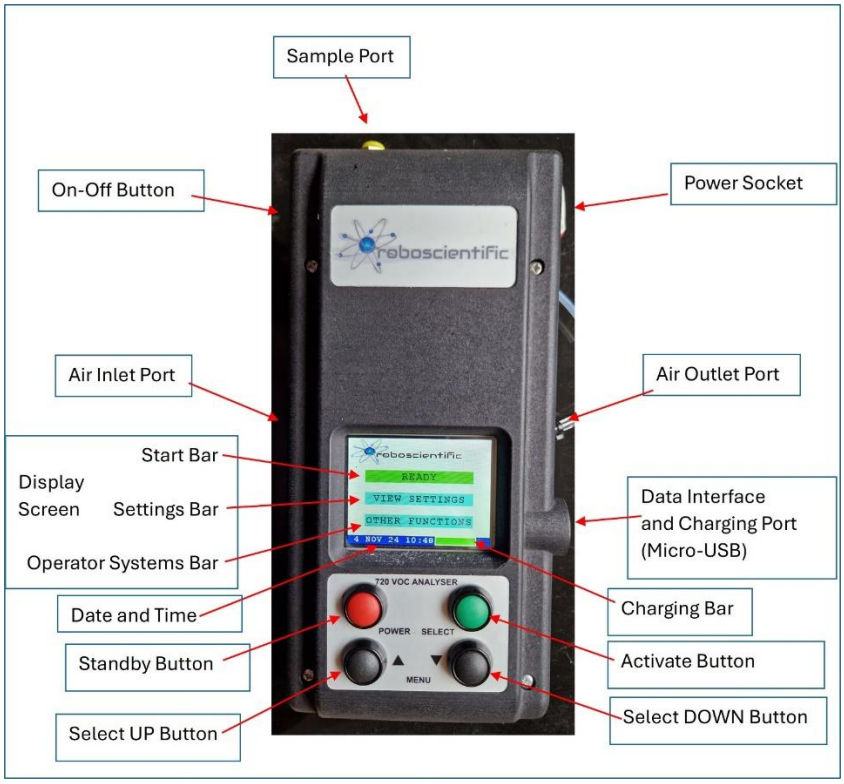
Importantly, it pre-concentrates low level VOCs and tends to remove background interferences, giving a route to produce measurable levels of key VOCs which are disease indicators, particularly those that are associated with viral infections. VOCs and a small amount of water vapour is adsorbed by the adsorbent material, whilst the rest of the water vapour, plus the breath gases, passes through the material to waste.

The adsorbed VOCs on the material pad is then removed from the mouthpiece and placed into a metallized bag with a grip seal that is then closed tightly. The sample is then transferred to the measuring area and a benchtop heat sealer is used to fully seal

the bag. At this stage the sample may be stored if required (fridge temperature will preserve the adsorbed VOCs for months) to form a batch of samples for analysis. Alternatively direct analysis may be performed, by inflating the bag using a small air pump (100 to 140ml air), heating the inflated bag with the sample at 65oC for a minimum of 20 minutes to release the adsorbed VOCs and analyzing with the pre-stabilized Model 720 VOC Analyser.

3.3 Operating the 720 VOC Analyser

The operational areas of the model 720 are shown below.



The operation of the 720 is semi-automatic.

1. Turn on the 720 using the Red button and run 3 sets of water vapour samples to stabilize the sensors. OneLab check list.
2. In parallel with sensor stabilization, heat the collected samples in the bags for a minimum of 20 minutes to desorb the VOCs with an oven set at 65oC. Remove the bagged first sample collected from the oven and cool for 2 minutes.
3. Connect the bag to the 720 using an 18g needle and PTFE tube.
4. Press the green button to start the analysis sequence, whilst simultaneously removing the next bagged sample from the oven. This next sample cools down

ready for analysis during the time it takes for the analysis sequence to run its course.

5. Remove the previous sample from the model 720 and attach the next bagged sample and press the green button.
6. Repeat points 7 to 9 to analyse all the stored bagged samples.
7. After all samples are analysed, wash the sensor array with butan-2-ol vapour. OneLab Check list - system shutdown.
8. All sample data are stored automatically on the SD card contained in the 720.
9. The data can be downloaded onto a standard laptop using a micro-USB to USB A connector (supplied with the instrument).
10. After the data are downloaded the 720 can be shut down by holding the Red button for over 3 seconds.

NOTE: The process described above is the standard proven methodology to generate VOC headspace samples from breath collection devices. Streamlining this workflow into a field methodology is in progress.

A new Thermal Desorption Unit (TDU) is being built to enable the rapid generation of VOC headspaces under field conditions as well as in the lab. This will negate the need for bagging samples, inflating the bag and incubation, thereby reducing the time of analysis and making it possible to work uninterrupted in the field. The TDU will become available in 2025. An updated VOC Analysis protocol will be released at a later date to cover the use of the TDU in the field.

3.3 Statistical analysis of array measurement data

To date only multivariate statistical analysis has been used to discriminate the data obtained from the 720 analyser. Principal Component Analysis (PCA) and Discriminant Analysis (LDA) have been used through the Microsoft Excel Add-On XLStat. The data from the 720 are saved in individual folders for each sample analysed. These are labelled '*Test_Results*' with an added time stamp. Within each folder the data files correspond to the number of repeat iterations of the same sample. The usual and most common number of repeats is 4, analysis is performed on the last 3 to prevent any carry over from the previous sample.

Data from a batch of samples are extracted using a proprietary software package to give a single folder with all the samples included. A second software package designed for the Benchtop VOC Analyser (Model 307B) allows data extraction of 4 parameters. Peak Height; Area; Absorption Rate and Desorption Rate as per the diagram on page 3.

Importing this data into Microsoft Excel is simple as the exported file is a text tab delimited file format. Once imported the data are labelled correctly to give Control samples and the different Experimental samples. PCA and LDA may be then carried out using XLStat. Alternatively, the data may be imported into any statistical analysis software and processed as required.

AI / ML treatment of the data has so far not been performed. However, Roboscientific has ongoing development of more advanced data processing software that will enable rapid Pass / Fail outputs in the future.

3.4 Diagnostic feedback/sharing information to decision makers (pass/fail) - total time information for whole process

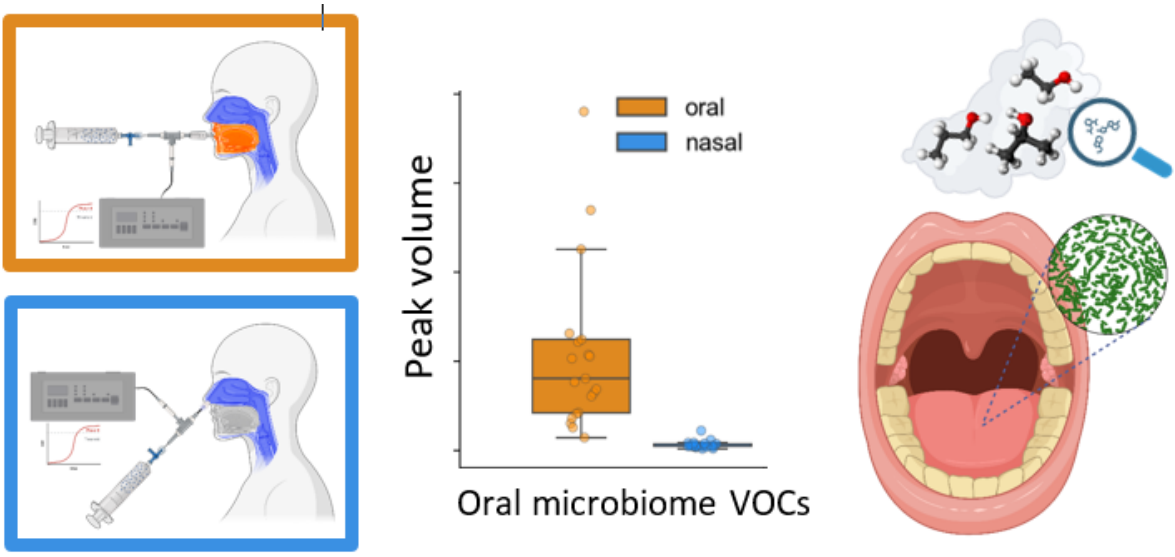
Using the standard proven methodology already described, the time taken to get a definitive answer if a breath sample is a Pass or a Fail will be close to an hour. Breath sampling takes approximately 5 to 7 minute, less if done in parallel on multiple patients. Bagging the samples, labelling and inflating the bag takes another 2 to 3 minutes. Incubation of the samples is a minimum 20 minutes. VOC Analysis of samples is 2 minutes for each sample. The extraction of data takes ~ 5 minutes. The labelling of each sample takes ~ 10 minutes PCA or LDA of samples and output of results takes another 10 minutes. Total time involved is of the order of 1 hour.

With the TDU and new software the time to decision making should be less. However at this stage of the development, the field deployable TDU / Pass / Fail system is still under construction, so to date the time to getting information to decision makers from a field system remains unclear. Still it is likely to be much shorter than the standard proven methodology described and in the above time analysis.

4. Concluding Remarks

Using merged GC-IMS data from two clinical sites, one within Europe (UIBK) and the other in Japan (SMCH), ML/AI algorithms have been developed and trained to stratify infections. Rapid breath tests (endogenous volatiles with GC-IMS with AI/ML algorithms and/or sensors with multivariate statistics) are now being developed to non-invasively determine uninfected from viral infected people. These will be tested in a blind study in the final months. Nevertheless, we have demonstrated that breath analysis has the potential to differentiate between people who have a viral infection from those who do not. Clinical data are still being collected (SMCH) to train further the AI/ML systems to obtain a higher diagnostic specificity.

Appendix A: Comments on Breath Sampling



All breath samples for the ONELAB project were collected using exhalation through the mouth. However, it is known that the mouth contains bacteria that produce volatiles [1-5]. These can cause confounders in the breath analysis, which could lead to incorrect diagnosis for viral infection. Owing to these concerns, during the last 12 months and during the clinical measurements, two partners within the consortium, UIBK and G.A.S., undertook a separate and detailed study of oral versus nasal exhaled breath sampling involving twenty-one volunteers. The evidence we have found is clear. In order to reduce the influence of the oral microbiome end-tidal exhaled nasal breath samples provide the best breath samples (independent of the inhalation process), otherwise with oral exhalation oral microbial volatiles could be falsely identified as biomarkers. This is particularly important given the increasing use of machine learning algorithms and artificial intelligence to identify variations in volatilome for diseases, including viral infections. We therefore recommend that in future exhaled nasal breath samples are collected for the detection of viral infection. However, this requires the development and commercialization of simple, user-friendly and comfortable end-tidal exhaled nasal sample collection devices so that nasal sampling could become widely adopted.

1. Khalid, T.Y.; Saad, S.; Greenman, J.; De Lacy Costello, B.; Probert, C.S.J.; Ratcliffe, N.M. Volatiles from oral anaerobes confounding breath biomarker discovery. J. Breath Res. 2013, 7, 017114. 10.1088/1752-7155/7/1/017114.

2. Sharma, A.; Kumar, R.; Varadwaj, P. Smelling the Disease: Diagnostic Potential of Breath Analysis. Mol. Diagn. Ther. 2023, 27, 321–347. 10.1007/s40291-023-00640-7.
3. Wang, T.; Pysanenko, A.; Dryahina, K.; Španěl, P.; Smith, D. Analysis of breath, exhaled via the mouth and nose, and the air in the oral cavity. J. Breath Res. 2008, 2, 037013. 10.1088/1752-7155/2/3/037013.
4. Sukul, P.; Oertel, P.; Kamysek, S.; Trefz, P. Oral or nasal breathing? Real-time effects of switching sampling route onto exhaled VOC concentrations. J. Breath Res. 2017, 11, 027101. 10.1088/1752-7163/aa6368
5. Smith, D.; Wang, T.; Pysanenko, A.; Španěl, P. A selected ion flow tube mass spectrometry study of ammonia in mouth- and nose-exhaled breath and in the oral cavity. Rapid Commun. Mass Spectrom. 2008, 22, 783–789. 10.1002/rcm.3434.

Appendix B: Gas Chromatograph Ion Mobility Spectrometer

All exhaled breath and room air samples were analysed using a BreathSpec Gas Chromatograph-Ion Mobility Spectrometer (GC-IMS) (G.A.S., Dortmund, Germany [https://www.gas-dortmund.de/Technology/Software-/1_461.html 1]). This purpose-built instrument for breath analysis provides the high chemical selectivity of gas chromatography through pre-separation of volatiles in complex sample matrices combined with the high sensitivity of an IMS, with detection limits typically in the low ppb_v level with no pre-concentration. The instrument combines an isothermal and flow programmable GC with an IMS. The combination of GC and ion mobility provides an analytical separation in two-time dimensions. The first is according to the GC retention times (in seconds) of the volatiles. The second is associated with the drift times (in milliseconds) of the volatiles' product ions in the drift region of the IMS.

The GC was equipped with a 30 m polar chromatographic column (MXT-WAX, L: 30.00 m, ID: 0.53 mm, FT: 0.50 µm), which was maintained at 50 °C. The IMS has a 5 cm drift tube operating at a voltage 2.7 kV (positive ion mode) and also maintained at a temperature of 50 °C. 370 MBq of radioactive tritium (3H) is the ionisation source of the IMS through the emission of electrons with a mean energy of 5.7 keV. High-purity nitrogen (5.0, 99.999%) served as both the carrier flow for the GC and the gas flow for the drift tube of the IMS.

A breath or room air sample was introduced via an internal pump, passing a 6-port-valve featuring a 1 mL sample loop. For each measurement, and before pressing run on the instrument, 10 mL of the breath or room sample were admitted through a heated (50 °C) sampling line, this being more than sufficient for a high quality BreathSpec measurement and reduces full analysis time. The rest of the sample in the syringe was discarded, because repeat analysis would be too time intensive and prone to sample degradation. After the analysis, the next sample was injected in the BreathSpec. We ensured that no carryover from a previous measurement would affect the following measurement. Following all measurements, syringes were cleaned in hot water containing a detergent, then rinsed with distilled water, and finally placed in an over at 100 oC for six hours prior to their use in any further sampling.

The GC-IMS spectra were parsed, processed and analysed using VOCal software (version 0.2.9, G.A.S., Dortmund, Germany [https://www.gas-dortmund.de/Technology/Software-/1_461.html]), which provides tools for a seamless

signal processing/feature extraction/analysis workflow. A qualitative analysis of the volatile compounds was conducted by comparing retention indices and drift times obtained using the GC-IMS results with external libraries (NIST2020 RI) and custom-built databases. Retention indices of volatile compounds were determined using measurements of the retention of gas standards. IMS drift times are expressed as relative to the position of the reactant ion peak (RIP) as internal standard to compensate for ambient pressure variations. Briefly, peak areas or region of interests (ROIs) were identified on all the spectra simultaneously using a 0.95 percentile criterion. Data were denoised, baseline correction and global cut-off values were adopted.