D4.1 – Data Map and Integration

WP number and title	WP4 – Design, Development and delivery
Deliverable number	D4.1
Version Number	1.0
Document Reference	D4.1
Lead Beneficiary	TEL
Deliverable type	DEM
Planned deliverable date	30/09/2024
Date of Issue	
Dissemination level	PU
Authors	TEL, SOL, HUA
Contributor(s)	Chrysostomos Symvoulidis (TEL), Matthias Schloegl (SOL), Christos Sardianos (HUA), Athanasios Kalogeropoulos (TEL)
Keywords	Laboratory Information Management System, Rapid Response Mobile Laboratory, Communication, Machine Learning

This project has received funding from the European Union's HORIZON EUROPE Research and Innovation programme under grant agreement No. 101073924

Participant N°	Participant organisation name	Short Name	Type	Country
1	Amsterdam University Medical Centre	AMC	Clinical	Netherlands
2	IANUS Consulting Ltd	IANUS	SME	Cyprus
3	Telesto	TEL	SME	Greece
4	Saitama Medical Centre Hospital	SMCH	Clinical	Japan
5	Harokopio University	HUA	University	Greece
6	Innsbruck University	UIBK	University	Austria
7	Jurrisk	JUR	SME	Belgium
8	IMSPEX	IMSPEX	SME	United Kingdom
9	University of Cyprus	UCY	University	Cyprus
10	University of Pécs	UP	University	Hungary
11	RoboScientific	ROBO	SME	United Kingdom
12	Solgenium OG	SOL	SME	Austria
13	Ghent University	UGent	University	Belgium
14	Panou	PANOU	SME	Greece
15	Institute of International Sociology	ISIG	Research	Italy
16	Johanniter Österreich Ausbildung und Forschung gemeinnützige GmbH	JOAFG	First Responders	Austria
17	Bavarian Red Cross	BRK	First Responders	Germany
18	Bioxhale Ltd	BIOX	SME	United Kingdom
19	Municipality of Hasselt	HASS	Local Authority	Belgium
20	Hungarian Red Cross	HRC	First Responders	Hungary
21	Region of Western Greece	RWG	Local Authority	Greece
22	Gesellschaft für analytische Sensorsysteme mbH	GAS	SME	Germany

Version History

Version No.	Date	Editor	Description of action
V0.1	05/06/2024	Chrysostomos Symvoulidis (TEL)	Provided draft Table of Contents
V0.2	17/10/2024	Chrysostomos Symvoulidis (TEL), Matthias Schloegl (SOL), Christos Sardianos (HUA)	Provided input on Sections 1, 2, and 3
V0.3	30/10/2024	Chrysostomos Symvoulidis (TEL), Athanasios Kalogeropoulos (TEL)	Input on Patient Registration Mobile Application section and Synchronization service, Updates on Mobile Application section
V0.4	08/01/2025	Matthias Schloegl (SOL), Christos Sardianos (HUA)	Updates on use cases section, Include Frontend description, Include API description, Updates on AI Models and Data Analysis services section
V0.5	12/02/2025	Matthias Schloegl (SOL)	Included information about data preprocessing and first draft on results of FTX "Hard Winter”
V0.6	12/02/2025	Matthias Schloegl (SOL)	Completed information on FTX “Hard Winter”
V0.7	13/02/2025	Chrysostomos Symvoulidis (TEL)	Finalisation of document
V0.8	14/02/2025	Chrysostomos Symvoulidis (TEL)	Submitted for review
V0.9	26/02/2025	Gino Claes (External Reviewer), Julie Turner (External Reviewer), Kathleen Vanheuverswyn (JUR)	Review received
V1.0	27/02/2025	Chrysostomos Symvoulidis (TEL), Matthias Schloegl (SOL), Christos Sardianos (HUA)	Addressed review comments and finalised deliverable

Table of Contents

Table of Figures

Table of Tables

Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
API	Application Programming Interface
C-LIMS	Centralized Laboratory Information Management System
CNN	Convolutional Neural Network
CV	Cross Validation
DL	Deep Learning
EU	European Union
FTX	Field Technical Exercise
GC	Gas Chromatography
GC-IMS	Gas Chromatography - Ion Mobility Spectrometry
GCP	Google Cloud Platform
GIS	Geographic Information System
HDF5	Hierarchical Data Format version 5
ICT	Information and Communication Technology
IMS	Ion Mobility Spectrometry
LC-MS	Liquid chromatography–mass spectrometry
L-LIMS	Local Laboratory Information Management System
LIMS	Laboratory Information Management System
ML	Machine Learning
PCA	Principal Component Analysis
PLS-DA	Partial Least Squares-Discriminant Analysis
PRMA	Patient Registration Mobile Application
RRML	Rapid Response Mobile Laboratory
SVC	Support Vector Classifier
SVM	Support Vector Machines
TRL	Technology Readiness Level
VOC	Volatile Organic Compounds
W-LAN	Wireless Local Area Network

Executive Summary

This document presents a comprehensive analysis of the ONELAB LIMS solution, providing detailed use cases, user requirements, system architecture, and outcomes from a practical exercise. The structure of the document covers key elements critical to understanding and implementing the ONELAB LIMS solution.

The report starts with a thorough **Use Cases and Requirements Analysis** (Section 3), identifying the specific needs of various users within the lab environment. The identified use cases and user requirements form the basis for designing and implementing the ONELAB LIMS solution.

In **Section 4**, the document delves into the core of the **ONELAB LIMS Solution**, starting with an explanation of what it is and how it addresses the requirements identified earlier. It outlines the overall system architecture and provides an in-depth analysis of its key components and software applications. These include (i) the **LIMS** itself, which serves as the central tool for managing all Rapid Response Laboratory (RRML) workflows, (ii) the **AI Models and Data Analysis services** that bring advanced computational power to data processing, (iii) the **Synchronization service** to ensure data and AI models consistency, and finally (iv) the **Patient Registration Mobile Application**, used for the registration of the patients of the RRML.

Finally, **Section 5** discusses the outcomes of the **FTX Exercise**, a practical scenario in which the ONELAB LIMS solution was deployed, while **Section 6** concludes the document.

1. Introduction

LIMSs play a crucial role during the operations of any modern laboratory, providing streamlined solutions for data management, workflow automation, and enhanced decision-making through advanced analytics. In the context of RRMLs, such systems are even more critical, as they must support high efficiency, accuracy, and mobility in environments where timely and precise data management can significantly impact outcomes.

This document presents an in-depth analysis of the **ONELAB LIMS solution**, a robust system designed to address the specific needs of RRMLs. The ONELAB LIMS integrates core functionalities, such as patient and sample management, AI-powered data analysis, and synchronization services, making it a powerful tool for improving laboratory efficiency. Additionally, the Patient Registration Mobile Application is designed to facilitate seamless patient registration and data capture, further enhancing the operational capabilities of the ONELAB RRML.

Through a comprehensive examination of use cases, user requirements, system architecture, and the ONELAB Field Trial Exercise (FTX), this document provides critical insights into how the ONELAB LIMS can be effectively deployed. This document also highlights the system's ability to streamline processes, ensure data consistency, and leverage AI models to support rapid and accurate decision-making in mobile lab environments.

2. Purpose of document

The purpose of this document is to provide a detailed analysis and overview of the ONELAB LIMS solution. It aims to explain how the solution addresses the specific needs of laboratory environments by outlining relevant use cases, user requirements, and the system’s overall architecture. Additionally, this document explores the core components of ONELAB LIMS, including the LIMS per se, the AI models and data analysis services, the synchronization service, and the Patient Registration Mobile Application, to showcase its full capabilities.

3. Use Cases and Requirements Analysis

The developed tools and services described in this deliverable are required to enable the following use cases:

(1) Structured and controlled data collection for a multi-centre clinical study collecting breath samples with various measurement devices and corresponding meta data in accordance with technical requirements defined in D5.1 [D5.1].

(2) Training of classification models for distinction of viral infections vs. no infection based on the collected clinical data including biomarker detection to support the requirements in FTX “Pre-Emptive” (D5.3 [D5.3]).

(3) Manage pandemic monitoring and intervention by integrating the LIMS in the ICT system of individual RRMLs as defined in D4.4 [D4.4] in accordance with technical requirements defined in D5.1. In more detail this involves:

(a) managing the patient and sample collection workflow,

(b) integrating trained classification models for viral detection from breath VOCs,

(c) visualization of GIS information

(4) Enable data and classification model synchronization between an arbitrary number of deployed RRMLs and a centralized LIMS.

3.1. Use Cases and Requirements Gathering Methodology

The following methodology was employed for gathering user requirements to support the data collection during the study phase as well as the field technical exercises. The process of gathering user requirements and defining use cases followed a structured and iterative approach. Initially, end users were consulted to provide their requirements, which were then reviewed and refined by technical partners. The enhanced list was shared back with the users for further feedback to ensure alignment with their needs. Throughout this process, continuous interactions were maintained through calls, interactive sessions and live demonstrations to collect additional insights. To facilitate hands-on evaluation, a test service was provided, allowing users to test the applications and services and submit their feedback.

The requirements were also aligned with specifications provided through D6.1 [D6.1]. Additionally, technical partners were engaged through scheduled biweekly meetings to monitor the progress of development and address any emerging challenges. For the use cases, inputs were gathered from WP5 partners and pilot project descriptions. These inputs were analyzed and refined to ensure that the use cases accurately reflected real-world scenarios while addressing both technical and operational requirements. In 06/2024 a end-user workshop was held at JOAFG to collect final feedback and give the possibility for amendments before the FTX in 11/2024. By following this methodology, the process ensured a comprehensive approach to requirements gathering and use case development, integrating user needs with technical feasibility while maintaining an ongoing feedback loop.

3.2. Use Cases

Table 1: Study Phase Data Collection.

ID	UC1_1
Unique Name / Title	Study Phase Data Collection
Priority	High
Brief Description	Web service for data upload
Rationale	In order to train first classification model candidates' measurement and meta data needs to be acquired from multiple measurement sites in a controlled way allowing for QA/QC interventions and user management.
Validation method	Validation by successful data to AI transfer.

Table 2: Pandemic monitoring and intervention

ID	UC1_2
Unique Name / Title	Pandemic monitoring and intervention
Priority	High
Brief Description	Services for patient and sample collection flow at individual RRML
Rationale	Patient information needs to be documented and aligned electronically with collected samples to facilitate automatic test result entries.
Validation method	Validation in D5.2 – FTX "Hard Winter" [D5.2]

Table 3: Synchronization

ID	UC1_3
Unique Name / Title	Data and model synchronization / reporting
Priority	High
Brief Description	Services to synchronize between individual RRMLs and centralized service
Rationale	Individual RRML must be able to operate in offline mode. However, data synchronization whenever possible needs to enable reporting for logistics coordinators, public health authorities and patients
Validation method	Validation in D5.2 – FTX "Hard Winter" [D5.2]

3.3. User Requirements

Based on the Use Cases described in section 3.1, the User Requirements that had to be implemented were defined. These requirements are related to the applications and the components related to the LIMS and described in detail in the following tables.

The requirements are either Functional or Non-Functional. Functional requirements can be any requirement related to the behavior of the software application in correspondence of specific events (including user interactions). On the other hand, Non-Functional requirements are constraints that must be fulfilled by the implementation of several functional requirements or by the entire application.

Each user requirement has a unique ID, which can state the number of user requirement and the related software application. For instance, R1_1 is the first requirement related to software component 1 which is the LIMS, while R2_1 is the first requirement related to software component 2 which is the Analysis service. All software components will be described in detail in the following Chapter. The list of the software components is the one described below:

1. LIMS (C-LIMS and L-LIMS),
2. Analysis Service,
3. Synchronization Service,
4. Patient Registration Mobile Application

The roles / actors interacting with the system are the following:

1. RRML Personnel: The personnel working on the RRML,
2. Patient: The patient that arrived at the RRML in order to be examined,
3. In case the requirement is not triggered by one of the abovementioned roles the role is described as N / A (i.e., Not Applicable).

It should be noted that the LIMS can be split into two main components: the C-LIMS (Central LIMS) and the L-LIMS (Local LIMS). In most cases, the user requirements described below are needed for both applications, yet in some cases (i.e., the training of the AI models can only happen in the centralized infrastructure where the C-LIMS is deployed).

ID	R1_1
Unique Name / Title	The user logs in the LIMS using their credentials
Type	Functional
Priority	High
Mandatory	Mandatory
Actor	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user logs into the LIMS in order to be able to access the provided features.
Rationale	Security
Validation method	User testing

ID	R1_2
Unique Name / Title	The user uploads GC-MS data
Type	Functional
Priority	High
Mandatory	Mandatory
Actor	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user uploads GC-MS samples which are stored in the LIMS.
Rationale	Usability
Validation method	User testing

ID	R1_3
Unique Name / Title	The user uploads GC-IMS data
Type	Functional
Priority	High
Mandatory	Mandatory
Actor	RRML Personnel
Related Use Case	All

Related Component	C-LIMS, L-LIMS
Brief Description	The user uploads GC-IMS samples which are stored in the LIMS.
Rationale	Usability
Validation method	User testing

ID	R1_4
Unique Name / Title	The user uploads LC-MS data
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user uploads LC-MS samples which are stored in the LIMS.
Rationale	Usability
Validation method	User testing

ID	R1_5
Unique Name / Title	The user uploads e-Nose data
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user uploads e-Nose samples which are stored in the LIMS.
Rationale	Usability
Validation method	User testing

ID	R1_6
Unique Name / Title	The user labels a measurement
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	Upon upload of the GC-IMS measurement and CRF, the user can label the corresponding UUID with its infection type (no infection, viral, bacterial, mixed)
Rationale	Usability
Validation method	User testing

ID	R1_7
Unique Name / Title	The user downloads a measurement
Type	Functional
Priority	High

Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	Download all measurements per type, per infection type or download measurements and correlated CRF / Blood sample analysis files.
Rationale	Usability
Validation method	User testing

ID	R1_8
Unique Name / Title	The user views statistical information of the stored samples
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user has the ability to view the number of collected samples, the number of collected samples per infection type, the number of collected samples per data type, and labels of each collected sample
Rationale	Usability
Validation method	User testing

ID	R1_9
Unique Name / Title	The user views visualized information of a GC-MS sample
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user has the ability to select and view in a graphical way, information about the GC-MS sample.
Rationale	Usability
Validation method	User testing

ID	R1_10
Unique Name / Title	The user views visualized information of a GC-IMS sample
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user has the ability to select and view in a graphical way information about the GC-IMS sample.
Rationale	Usability
Validation method	User testing

ID	R1_11
Unique Name / Title	The user views visualized information of a LC-MS sample
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user has the ability to select and view in a graphical way, information about the LC-MS sample.
Rationale	Usability
Validation method	User testing

ID	R1_12
Unique Name / Title	The user views visualized information of a e-Nose sample
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user has the ability to select and view in a graphical way, information about the e-Nose sample.
Rationale	Usability
Validation method	User testing

ID	R1_13
Unique Name / Title	The user views the outcomes / predictions of the Classification model
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The classification model is deployed and can be utilized by the user within the LIMS by selecting a measurement for inference.
Rationale	AI Analysis
Validation method	User testing

ID	R1_14
Unique Name / Title	The user views the outcomes / predictions of the Biomarkers Identification model
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All

Related Component	C-LIMS, L-LIMS
Brief Description	The Biomarkers Identification model is deployed and can be utilized by the user within the LIMS by selecting a measurement for inference.
Rationale	AI Analysis
Validation method	User testing

ID	R1_15
Unique Name / Title	Logging of user actions
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	All user actions must be logged by a logging tracker
Rationale	Security
Validation method	User testing

ID	R1_16
Unique Name / Title	Logging of data actions
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	All actions performed on the data (Create – Read – Update – Delete) must be logged by a logging tracker
Rationale	Security
Validation method	User testing

ID	R2_1
Unique Name / Title	Training of Infection type classification model with the collected GC-IMS samples
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	Analysis service, C-LIMS
Brief Description	The classification model is trained using the available GC-IMS samples. The process includes Data preprocessing and cleaning, Training of ML / DL model, Evaluation of trained model, Re-training using new collected data
Rationale	Measurements Analysis
Validation method	User testing

ID	R2_2

Unique Name / Title	Training of Biomarkers Identification model
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	Analysis service, C-LIMS
Brief Description	The biomarkers classification model is trained for the evaluation of clinical relevance of the identified biomarkers, and validated in diverse patient populations
Rationale	Measurements Analysis
Validation method	User testing

ID	R3_1
Unique Name / Title	Synchronization data stored in the L-LIMS with the C-LIMS
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	L-LIMS, C-LIMS, Synchronization Service
Brief Description	The data (i.e., measurements, samples, etc.) that is initially stored in the L-LIMS are backed up in the centralized infrastructure. The synchronization procedure is performed iteratively every 1 hour.
Rationale	AI Analysis and Storage
Validation method	Data integrity and logging monitoring

ID	R3_2
Unique Name / Title	AI Models synchronization between the C-LIMS and the L-LIMS
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	L-LIMS, C-LIMS, Synchronization Service
Brief Description	The trained models trained with the data found in C-LIMS are transferred to the L-LIMS for local inference. The synchronization procedure is performed iteratively every 1 hour.
Rationale	AI Analysis and Storage
Validation method	Models' integrity and logging monitoring

ID	R4_1
Unique Name / Title	Patient registration including preregistration
Type	Functional
Priority	High

Mandatory	Mandatory
Role	RRML Personnel, Patient
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The patient is registered by the RRML personnel using the mobile application. Information from preregistered patients can be retrieved.
Rationale	Registration of patient in order to proceed to the RRML for measurement sampling.
Validation method	User testing

ID	R4_2
Unique Name / Title	Automated assessment based on patient's responses to Triage Questionnaire
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The RRML Personnel fills in the Triage questionnaire and the patient's priority is calculated according to their answers.
Rationale	Registration of patient in order to proceed to the RRML for measurement sampling.
Validation method	User testing

ID	R4_3
Unique Name / Title	The user inserts the patient's contact person information
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The patient's emergency contact person information is collected.
Rationale	Patient tracking
Validation method	User testing

ID	R4_4
Unique Name / Title	The user inserts the patient's information
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The patient's personal information is collected.
Rationale	Patient tracking
Validation method	User testing

ID	R4_5
Unique Name / Title	The user links a measurement sample with patient
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The sample is linked to the patients UUID.
Rationale	Measurement tracking
Validation method	User testing

ID	R4_6
Unique Name / Title	The user views classification results for a given sample
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	By scanning the patient's QR code, the RRML personnel can view the classification results on whether they are infected or not
Rationale	Measurement tracking
Validation method	User testing

ID	R4_7
Unique Name / Title	The user views a specific patient's contact information
Type	Functional
Priority	High
Mandatory	Recommended
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	By scanning the patient's QR code, the RRML personnel can view the patient's contact information.
Rationale	Measurement tracking
Validation method	User testing

4. The ONELAB LIMS Solution

4.1. ONELAB LIMS Solution overview

The ONELAB LIMS platform is used to streamline all procedures in the ONELAB RRML (a high-level view of the ONELAB LIMS solution can be seen in Figure 1), from data collection, data quality assurance, patient registration, to AI training using the collected data, and the utilization of these models for inference. The core procedures are related to the following concepts:

- **Data Collection:** The ONELAB LIMS platform allows its users to upload anonymized analytical data (i.e., patient measurements, clinical data, etc.) seamlessly. This data can be then utilized for further analysis within the platform, while it can also be used for training AI models.
- **Data Quality Assurance:** Upon the upload of any data type to the ONELAB LIMS, specific quality checks are performed in order to to ensure that the data is interoperable, structured correctly and of high quality.
- **Provision of data for AI training:** The ONELAB LIMS platform, as already described above, allows the utilization of the data collected for AI training. This is performed via dedicated functionalities that allow its users to access this data for training purposes.
- **Utilization of AI models for inference:** Through the ONELAB LIMS, its users will be also able to utilize the AI models that will be trained with the collected data for inference.
- **Patient Registration:** Via a dedicated component of the ONELAB LIMS (i.e., the Patient Registration Mobile Application) the RRML personnel can register incoming patients and link them with the obtained measurements.

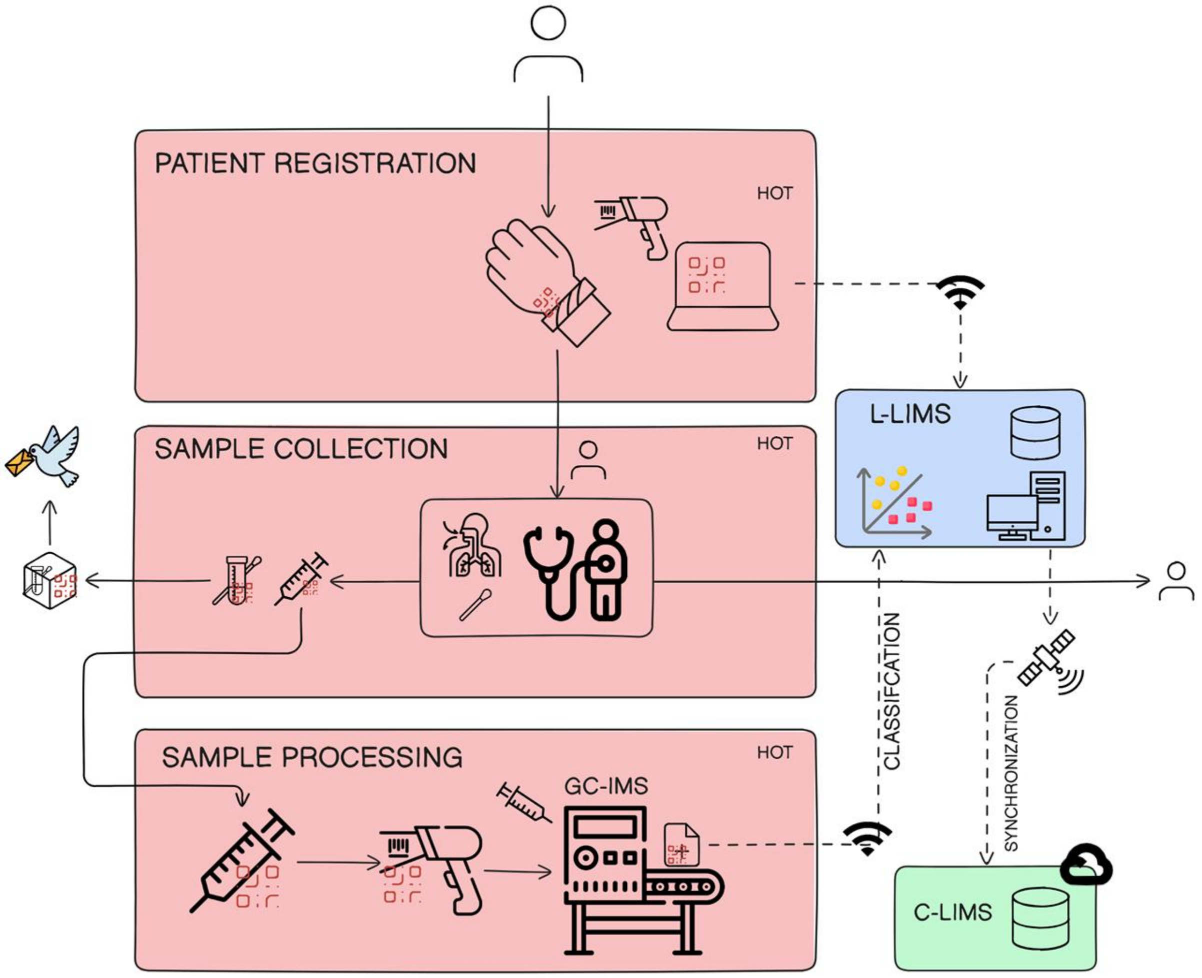

Figure 1: ONELAB LIMS Solution overview

4.2. Overall architecture

The above-mentioned functionalities are available to the users of the ONELAB LIMS platform. This is actually a suite of services that can be split into two main components: (i) the centralized component (C-LIMS) which can be used mainly for data collection from various institutions and organizations, as well as for training of AI models using the collected data and the decentralized component or local deployment (L-LIMS) which will be deployed within a RRML, (ii) the Patient Registration Mobile Application used for the patients' registration and additional functionalities such as linking the patients to their samples, or answering the Triage Questionnaire, (iii) the Synchronization service used for the synchronization between any deployed L-LIMS and the C-LIMS, and finally (iv) the AI Models and Data Analysis services, which handles all functionalities related to the training of the AI models and provision of those models for inference. In the following sections, these components will be described in detail.

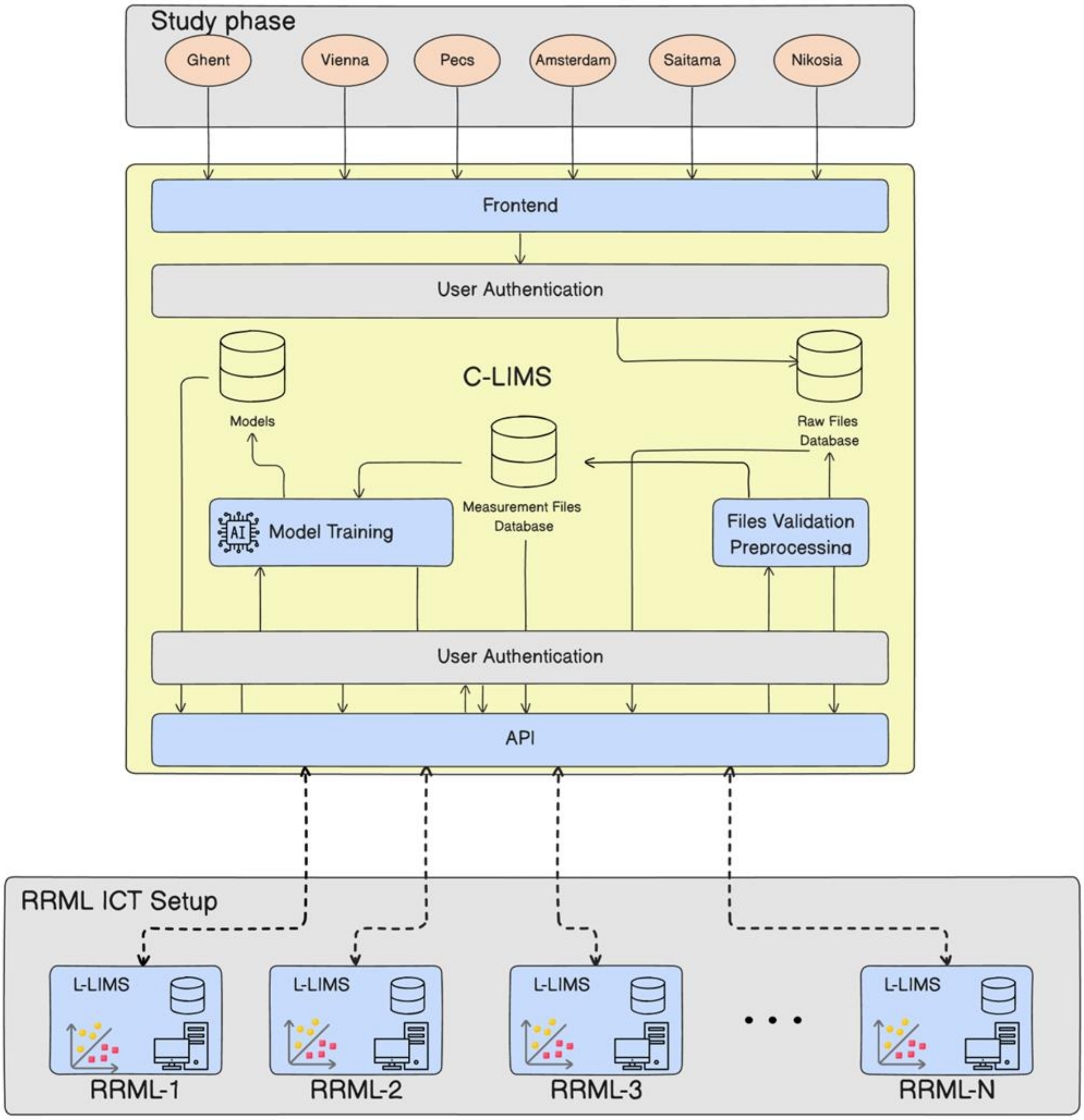

Figure 2: ONELAB LIMS High-level architecture

As also seen in Figure 2, there are study sites (e.g., hospitals) involved in collecting and uploading data to the platform. Each study site interacts with the C-LIMS via its **Frontend** layer, which serves as the user interface for uploading, monitoring, and accessing the system. Once the data is validated and stored, it is used for Model Training, where ML / DL models are trained on the available data.

The models are then stored and can be made available for inference in the L-LIMS deployments of any connected RRML (as shown at the bottom of Figure 2). Within each RRML, an L-LIMS is deployed which is used to manage local datasets (i.e., data collected by incoming patients) while it can also handle inference tasks using the trained ML / DL models.

All procedures are performed by authenticated users.

4.3. Components analysis

4.3.1. Laboratory Information Management System (LIMS)

4.3.1.1. Overview

The C-LIMS is a containerized Python-based application deployed on a virtual machine in Google Cloud Platform (GCP) with servers located in the EU. User authentication is managed through Nginx, and the codebase is hosted in a on-premises-owned GitLab [GitLab] repository. The version of the application can be deployed on both cloud-based and local infrastructures. The L-LIMS on the other hand is a local deployment of the LIMS. In more detail, it is deployed within the RRML and can operate independently from the C-LIMS, so that the RRML can operate without any issues even if deployed in areas where internet connection is absent. Yet the two systems are in sync when internet connection is available. More on this will be presented in Section 4.3.3 – 'Synchronization service'.

The system provides controlled access for laboratory personnel and quality control managers via a frontend built with Streamlit [Streamlit]. Data uploaded to the system is stored in raw format and, after data checks, converted to standardized database formats (SQLite [SQLite] and HDF5).

The LIMS also hosts trained classification models for viral infection detection using GC-IMS VOC data from the study phase. Additionally, it features an API with endpoints for:

- Communication with the Patient Registration Mobile Application (PRMA) in the RRML setting.
- Synchronization between the central service and individual RRML patient preregistration.
- GIS mapping for enhanced data visualization and integration.

The LIMS is considered Technology Readiness Level (TRL) 6, since it is a prototype that has been demonstrated in relevant environment (in this case the relevant environment is the FTX), in which the LIMS' ability to perform reliably in real-world conditions was reflected. The LIMS was successfully deployed, integrated and used into the pilot-scale environment where its functionality was proved, interoperability with existing workflows was successfully performed, and its capacity to handle real data under expected conditions was validated. Throughout the pilot deployment during the FTX, the LIMS demonstrated consistent performance, efficiently managing data acquisition, processing, and storage while ensuring data integrity and traceability. User feedback also confirmed its usability, while stress testing under operational conditions revealed no critical failures.

4.3.1.2. Related User Requirements

The user requirements that are related to the LIMS are listed below:

ID	R1_1
Unique Name / Title	The user logs in the LIMS using their credentials
Type	Functional
Priority	High
Mandatory	Mandatory

Actor	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user logs into the LIMS in order to be able to access the provided features.
Rationale	Security
Validation method	User testing

ID	R1_2
Unique Name / Title	The user uploads GC-MS data
Type	Functional
Priority	High
Mandatory	Mandatory
Actor	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user uploads GC-MS samples which are stored in the LIMS.
Rationale	Usability
Validation method	User testing

ID	R1_3
Unique Name / Title	The user uploads GC-IMS data
Type	Functional
Priority	High
Mandatory	Mandatory
Actor	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user uploads GC-IMS samples which are stored in the LIMS.
Rationale	Usability
Validation method	User testing

ID	R1_4
Unique Name / Title	The user uploads LC-MS data
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user uploads LC-MS samples which are stored in the LIMS.
Rationale	Usability
Validation method	User testing

ID	R1_5
Unique Name / Title	The user uploads e-Nose data
Type	Functional

Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user uploads e-Nose samples which are stored in the LIMS.
Rationale	Usability
Validation method	User testing

ID	R1_6
Unique Name / Title	The user labels a measurement
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	Upon uploading the GC-IMS measurement and CRF, the user can label the corresponding UUID with its infection type (no infection, viral, bacterial, mixed)
Rationale	Usability
Validation method	User testing

ID	R1_7
Unique Name / Title	The user downloads a measurement
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	Download all measurements per type, per infection type or download measurements and correlated CRF / Blood sample analysis files.
Rationale	Usability
Validation method	User testing

ID	R1_8
Unique Name / Title	The user views statistical information of the stored samples
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user has the ability to view the number of collected samples, the number of collected samples per infection type, the number of collected samples per data type, and labels of each collected sample

Rationale	Usability
Validation method	User testing

ID	R1_9
Unique Name / Title	The user views visualized information of a GC-MS sample
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user has the ability to select and view in a graphical way, information about the GC-MS sample.
Rationale	Usability
Validation method	User testing

ID	R1_10
Unique Name / Title	The user views visualized information of a GC-IMS sample
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user has the ability to select and view in a graphical way information about the GC-IMS sample.
Rationale	Usability
Validation method	User testing

ID	R1_11
Unique Name / Title	The user views visualized information of an LC-MS sample
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The user has the ability to select and view in a graphical way, information about the LC-MS sample.
Rationale	Usability
Validation method	User testing

ID	R1_12
Unique Name / Title	The user views visualized information of a e-Nose sample
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All

Related Component	C-LIMS, L-LIMS
Brief Description	The user has the ability to select and view in a graphical way, information about the e-Nose sample.
Rationale	Usability
Validation method	User testing

ID	R1_13
Unique Name / Title	The user views the outcomes / predictions of the Classification model
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The classification model is deployed and can be utilized by the user within the LIMS by selecting a measurement for inference.
Rationale	AI Analysis
Validation method	User testing

ID	R1_14
Unique Name / Title	The user views the outcomes / predictions of the Biomarkers Identification model
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The Biomarkers Identification model is deployed and can be utilized by the user within the LIMS by selecting a measurement for inference.
Rationale	AI Analysis
Validation method	User testing

ID	R1_15
Unique Name / Title	Logging of user actions
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	All user actions must be logged by a logging tracker
Rationale	Security
Validation method	User testing

ID	R1_16
Unique Name / Title	Logging of data actions

Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	All actions performed on the data (Create – Read – Update – Delete) must be logged by a logging tracker
Rationale	Security
Validation method	User testing

ID	R2_1
Unique Name / Title	Training of Infection type classification model with the collected GC-IMS samples
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	Analysis service, C-LIMS
Brief Description	The classification model is trained using the available GC-IMS samples. The process includes Data preprocessing and cleaning, Training of ML / DL model, Evaluation of trained model, Re-training using new collected data
Rationale	Measurements Analysis
Validation method	User testing

ID	R2_2
Unique Name / Title	Training of Biomarkers Identification model
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	Analysis service, C-LIMS
Brief Description	The biomarkers classification model is trained for the evaluation of clinical relevance of the identified biomarkers, and validated in diverse patient populations
Rationale	Measurements Analysis
Validation method	User testing

ID	R3_1
Unique Name / Title	Synchronization data stored in the L-LIMS with the C-LIMS
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	L-LIMS, C-LIMS, Synchronization Service
Brief Description	The data (i.e., measurements, samples, etc.) that is

	initially stored in the L-LIMS are backed up in the centralized infrastructure. The synchronization procedure is performed iteratively every 1 hour.
Rationale	AI Analysis and Storage
Validation method	Data integrity and logging monitoring

ID	R3_2
Unique Name / Title	AI Models synchronization between the C-LIMS and the L-LIMS
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	L-LIMS, C-LIMS, Synchronization Service
Brief Description	The trained models trained with the data found in C-LIMS are transferred to the L-LIMS for local inference. The synchronization procedure is performed iteratively every 1 hour.
Rationale	AI Analysis and Storage
Validation method	Models' integrity and logging monitoring

ID	R4_1
Unique Name / Title	Patient registration including preregistration
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel, Patient
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The patient is registered by the RRML personnel using the mobile application. Information from preregistered patients can be retrieved.
Rationale	Registration of patient in order to proceed to the RRML for measurement sampling.
Validation method	User testing

ID	R4_2
Unique Name / Title	Automated assessment based on patient's responses to Triage Questionnaire
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The RRML Personnel fills in the Triage questionnaire and the patient's priority is calculated according to their answers.
Rationale	Registration of patient in order to proceed to the RRML for measurement sampling.

Validation method	User testing

ID	R4_3
Unique Name / Title	The user inserts the patient's contact person information
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The patient's emergency contact person information is collected.
Rationale	Patient tracking
Validation method	User testing

ID	R4_4
Unique Name / Title	The user inserts the patient's information
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The patient's personal information is collected.
Rationale	Patient tracking
Validation method	User testing

ID	R4_5
Unique Name / Title	The user links a measurement sample with patient
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The sample is linked to the patien's UUID.
Rationale	Measurement tracking
Validation method	User testing

ID	R4_6
Unique Name / Title	The user views classification results for a given sample
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	By scanning the patient's QR code, the RRML personnel can view the classification results on whether they are infected or not

Rationale	Measurement tracking
Validation method	User testing

ID	R4_7
Unique Name / Title	The user views a specific patient's contact information
Type	Functional
Priority	High
Mandatory	Recommended
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	By scanning the patient's QR code, the RRML personnel can view the patient's contact information.
Rationale	Measurement tracking
Validation method	User testing

4.3.1.3. Interfaces

The LIMS can be split into two main services; (i) the frontend service which allows users to interact with the system via a user interface and perform actions such as uploading data to the system, and (ii) a backend API which allows the communication of the LIMS with other systems and services such as the synchronization service, and the communication between a L-LIMS and the C-LIMS, etc.

Study-Phase Frontend-Service

Starting with the frontend service, it is a featured cloud hosted web application reachable via the internet[1] provides several sub-pages to support the aforementioned requirements and are detailed below:

[1] ONELAB LIMS Platform, Available at: https://infection-detection-platform.solgenium.com/

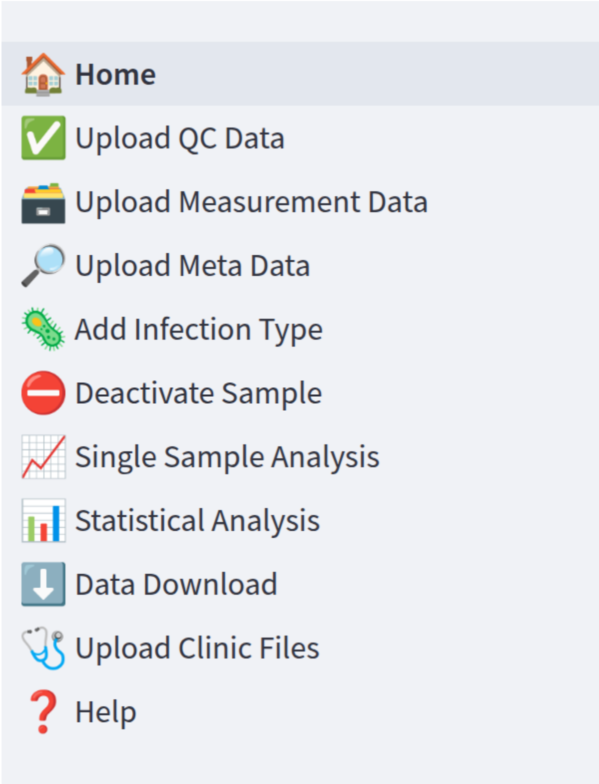

Figure 3: LIMS Sidebar: Provides a comprehensive list of the available functionalities and services.

1. Home Page

The Home Page of the LIMS provides the following functionalities:

- generation of 6-digit unique IDs
- download of CRF files (pdf/excel)
- download of SOPs for GC-IMS operation and data upload

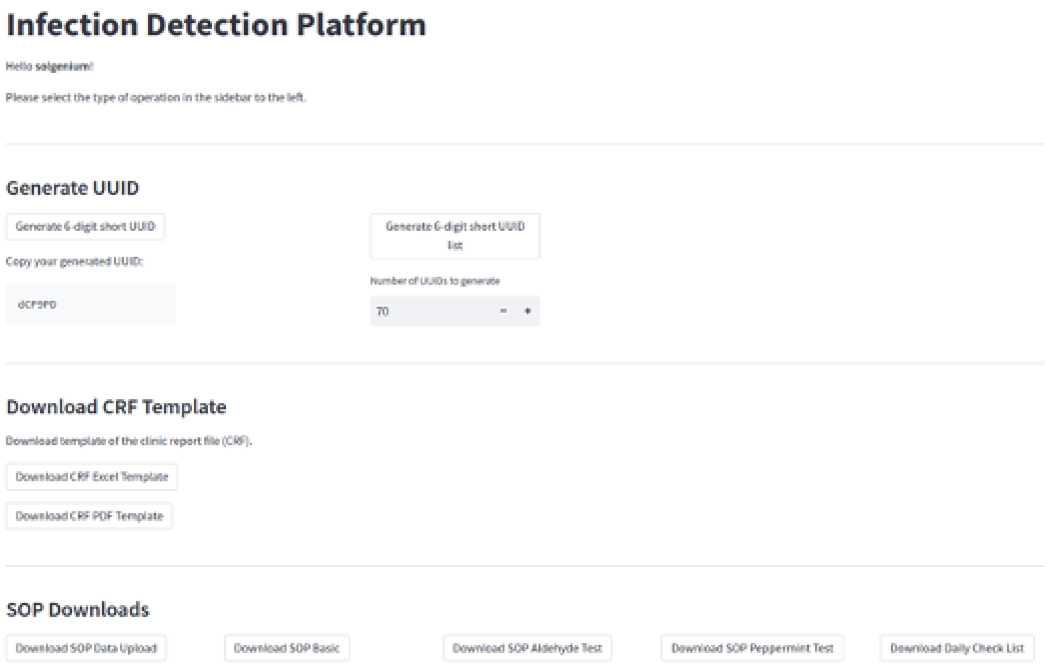

Figure 4: LIMS Home Page.

2. QC Data uploaded

Provides upload functionality of Aldehyde, Ketonmix and Peppermint QC data, as shown in the following figure:

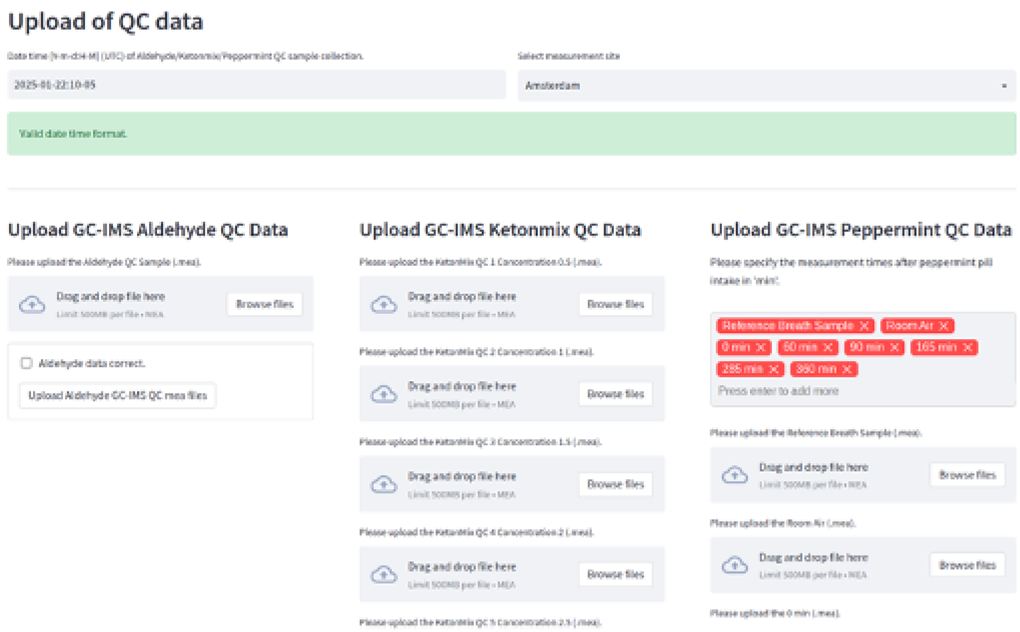

Figure 5: Quality-Checked Data upload page.

3. Upload of Measurement Data

Provides upload functionalities for the following data types:

- GC-IMS, CRF,
- GC-MS
- LC-MS
- E-Nose data

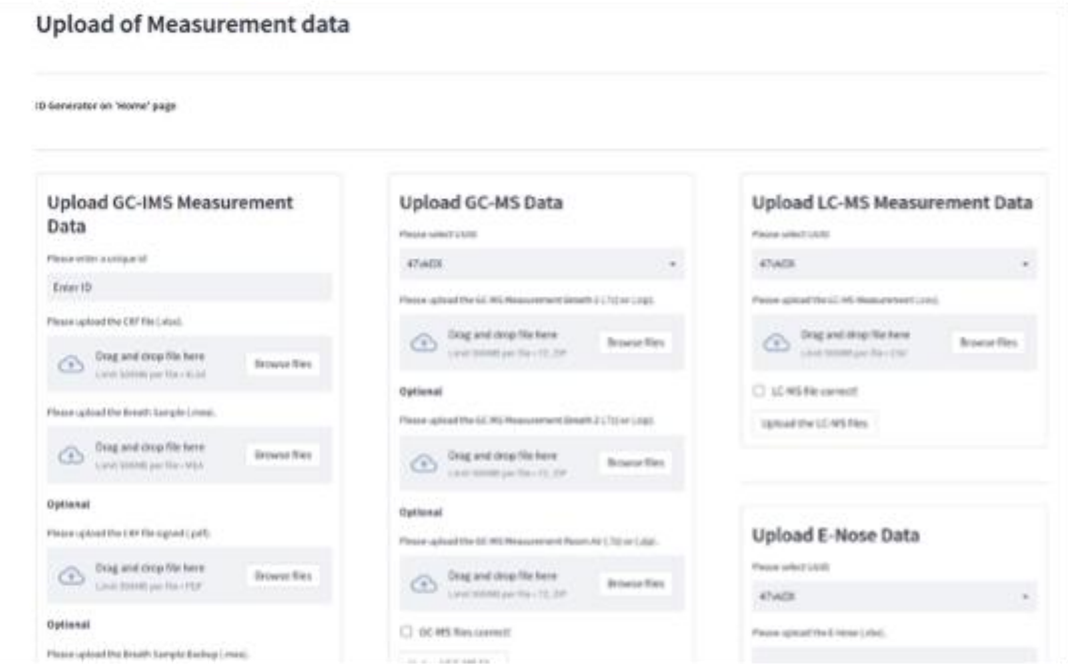

Figure 6: Upload of Measurement Data page.

2. Infection-Detection-API

Following a comprehensive list of the endpoints available in the LIMS API are depicted. These endpoints, as already discussed, allow the communication of any LIMS (either the C-LIMS or a L-LIMS) with other key components of the overall ONELAB LIMS solution.

Name	Create RRML Registration
URL	/register_RRML/
HTTP Method	POST
Description	Create database entry for new RRML
Header	None

Arguments	{
	"rrml": "string",
	"address": {
	"street": "Dr. Karl Renner-Ring",
	"number": "3",
	"city": "Vienna",
	"state_province": "Vienna",
	"postal_code": "1017",
	"country": "AT"
	}
	}

Request Body	application/json
Return Value	200, 422
HTTP Return Codes	200: Successful Response 422: Validation Error
Exceptions	422
Precondition	N/A

Name	Get Registered RRMLs
URL	/RRML_ids/
HTTP Method	GET
Description	Get registered RRMLs
Header	None
Arguments	None
Request Body	None
Return Value	["string"]
HTTP Return Codes	200: Successful Response
Exceptions	N/A
Precondition	N/A

Name	Read Registration RRML
URL	/RRML_per_uuid/{uuid}
HTTP Method	GET
Description	Get RRML registered for the UUID
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Delete RRML
URL	/delete_RRML/{uuid}
HTTP Method	DELETE
Description	Delete RRM registered for the UUID
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Get Valid Questions
URL	/valid_triage_questions/
HTTP Method	GET

Description	Retrieve valid triage questions
Header	None
Arguments	None
Request Body	None
Return Value	200: Successful Response
HTTP Return Codes	200
Exceptions	N/A
Precondition	N/A

Name	Get Triage Ids
URL	/get_triage_ids
HTTP Method	GET
Description	Retrieve valid triage question ids
Header	application/json
Arguments	None
Request Body	None
Return Value	200: Successful Response
HTTP Return Codes	200
Exceptions	N/A
Precondition	N/A

Name	Get Triage Score
URL	/get_triage_score/{uuid}
HTTP Method	GET
Description	Retrieve triage score for patient with UUID uuid
Header	application/json
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Create Triage Question
URL	/triage_questions/
HTTP Method	POST
Description	Create a new set of triage questions for patient with UUID uuid in RRML rrml
Header	None

Arguments	{
	  "uuid": "string",
	  "rrml": "string",
	  "questions": [
	    "string"
	  ]
	}

Request Body	application/json
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422

Exceptions	422
Precondition	N/A

Name	Create Preregistration
URL	/pre_register_user/
HTTP Method	POST
Description	Create a new preregistration
Header	None

Arguments	{
	"rrml": "string",
	"patient": {
	"name": {
	"first_name": "Max",
	"last_name": "Müller"
	},
	"address": {
	"street": "Dr. Karl Renner-Ring",
	"number": "3",
	"city": "Vienna",
	"state_province": "Vienna",
	"postal_code": "1017",
	"country": "AT"
	},
	"phonenumber": "+43123456789",
	"email": "user@example.com",
	"uuid": "string",
	"sex": "male",
	"birthdate": {
	"year": 0,
	"month": 0,
	"day": 0
	},
	"document_id": "string"
	}
	}

Request Body	application/json
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Get Registration Ids
URL	/pre_registration_ids/
HTTP Method	GET
Description	Get all registered UUIDs
Header	None
Arguments	None
Request Body	None
Return Value	200: Successful Response
HTTP Return Codes	200
Exceptions	N/A

Precondition	N/A

Name	Read Preregistration Patient
URL	/pre_registrations_per_uuid/{uuid}
HTTP Method	GET
Description	Get patient registered for the UUID
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Create Registration
URL	/register_user/
HTTP Method	POST
Description	Register new patient with UUID uuid in RRML rrml
Header	None

Arguments	{
	"rrml": "string",
	"patient": {
	"name": {
	"first_name": "Max",
	"last_name": "Müller"
	},
	"address": {
	"street": "Dr. Karl Renner-Ring",
	"number": "3",
	"city": "Vienna",
	"state_province": "Vienna",
	"postal_code": "1017",
	"country": "AT"
	},
	"phonenumber": "+43123456789",
	"email": "user@example.com",
	"uuid": "string",
	"sex": "male",
	"birthdate": {
	"year": 0,
	"month": 0,
	"day": 0
	},
	"document_id": "string"
	}
	}

Request Body	application/json
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422

Precondition	N/A

Name	Get Registration Ids
URL	/registration_ids/
HTTP Method	GET
Description	Get all registered patient UUIDS
Header	None
Arguments	None
Request Body	None
Return Value	200: Successful Response
HTTP Return Codes	200
Exceptions	N/A
Precondition	N/A

Name	Read Registration Dates
URL	/registration_dates/
HTTP Method	GET
Description	Get all registered UUIDs for certain day in YYYY-MM-DD format
Header	None
Arguments	None
Request Body	None
Return Value	200: Successful Response
HTTP Return Codes	200
Exceptions	N/A
Precondition	N/A

Name	Read Registration RRML
URL	/registrations_per_RRML/{RRML}
HTTP Method	GET
Description	Get all patients registered for the RRML
Header	None
Arguments	RRML
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Read Registration Patient
URL	/registrations_per_uuid/{uuid}
HTTP Method	GET
Description	Get patient registered for the UUID
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422

Exceptions	422
Precondition	N/A

Name	Read Registration Address
URL	/registration_address/{uuid}
HTTP Method	GET
Description	Get patient registered for the UUID
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Read Registration Date
URL	/registration_date/{uuid}
HTTP Method	GET
Description	Get registration timestamp for the UUID
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Read Registration Timestamp
URL	/registrations_per_date/{input_date}
HTTP Method	GET
Description	Get patient registered for certain day in YYYY-MM-DD format
Header	None
Arguments	input_date
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Read Registration Timestamp RRML
URL	/registrations_per_RRML_date/{RRML}/{input_date}
HTTP Method	GET
Description	Get patient registered for certain day in YYYY-MM-DD format
Header	None
Arguments	RRML

	input_date
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Delete Registration
URL	/delete_registration/{uuid}
HTTP Method	DELETE
Description	Delete a registered patient with UUID uuid
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Create Contact
URL	/add_contact_information/
HTTP Method	POST
Description	Create contact person information for patient with UUID uuid
Header	None

Arguments	{
	  "rrml": "string",
	  "contact": {
	    "name": {
	      "first_name": "Max",
	      "last_name": "Müller"
	    },
	    "address": {
	      "street": "Dr. Karl Renner-Ring",
	      "number": "3",
	      "city": "Vienna",
	      "state_province": "Vienna",
	      "postal_code": "1017",
	      "country": "AT"
	    },
	    "phonenumber": "+43123456789",
	    "email": "user@example.com",
	    "contact_date": {
	      "year": 0,
	      "month": 0,
	      "day": 0
	    },
	    "contact_uuid": "string"
	  }
	}

Request Body	application/json
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Read Contact IDs
URL	/get_contact_ids/
HTTP Method	GET
Description	Retrieve all registered contact persons uuids for patient with UUID uuid
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response
HTTP Return Codes	200
Exceptions	N/A
Precondition	N/A

Name	Read Contact
URL	/get_contact_information/{uuid}
HTTP Method	GET
Description	Retrieve contact details of contact person with UUID uuid
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Delete Contact
URL	/delete_contact/{uuid}
HTTP Method	DELETE
Description	Delete contact detail information for contact persion with UUID uuid
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Register Patient Sample
URL	/register_sample/
HTTP Method	POST

Description	Register sample of sample type and patient in RRML
Header	None
Arguments	{ "rrml": "string", "sample_id": "string", "patient_id": "string", "sample_type": "string" }
Request Body	application/json
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Read Sample Per Patient
URL	/read_sample_per_patient/{patient_id}
HTTP Method	GET
Description	Retrieve all registered samples for patient with UUID patient_id
Header	None
Arguments	patient_id
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Read Sample ID Per Type
URL	/read_sample_id_per_type/{sample_type}
HTTP Method	GET
Description	Retrieve all registered sample Ids of specific sample type
Header	None
Arguments	sample_type
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Get Patient ID Per Sample
URL	/get_patient_id_per_sample/{sample_id}
HTTP Method	GET
Description	Retrieve all registered patient Ids for for sample with sample_id
Header	None
Arguments	sample_id
Request Body	None
Return Value	200: Successful Response

	422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Read Sample Per Type
URL	/read_patient_per_type/{sample_type}
HTTP Method	GET
Description	Retrieve list of patients Ids for specific sample_type
Header	None
Arguments	sample_type
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Read Sample Date
URL	/gcims_sample_date/{sample_type}/{sample_id}
HTTP Method	GET
Description	Get sample timestamp for the UUID and sample-type
Header	None
Arguments	sample_type sample_id
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Delete Sample
URL	/delete_sample_registration/{sample_id}
HTTP Method	DELETE
Description	Delete registered sample entry with sample_id
Header	None
Arguments	sample_id
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Get GC-IMS Ids
URL	/get_gcims_ids
HTTP Method	GET
Description	Retrieve all registered ID for GC-IMS samples
Header	None

Arguments	None
Request Body	None
Return Value	200: Successful Response
HTTP Return Codes	200
Exceptions	N/A
Precondition	N/A

Name	Store GC-IMS
URL	/store_gcims/{uuid}
HTTP Method	POST
Description	Store GC-IMS file provided as binary string with UUID uuid
Header	None
Arguments	uuid, gcims_file [string<binary>]
Request Body	multipart/form-data
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Store And Predict
URL	/store_and_predict/{uuid}
HTTP Method	POST
Description	Store GC-IMS file provided as binary string with UUID uuid and invoke classification in backend
Header	None
Arguments	uuid, gcims_file [string<binary>],
Request Body	multipart/form-data
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Delete GC-IMS
URL	/delete_gcims/{uuid}
HTTP Method	DELETE
Description	Delete GC-IMS file with UUID uuid
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Predict
URL	/predict/{uuid}
HTTP Method	GET

Description	Retrieve prediction result for patient with UUID uuid
Header	None
Arguments	uuid
Request Body	None
Return Value	200: Successful Response 422: Validation Error
HTTP Return Codes	200, 422
Exceptions	422
Precondition	N/A

Name	Alive
URL	/alive
HTTP Method	GET
Description	Retrieve alive status of API
Header	None
Arguments	None
Request Body	None
Return Value	200: Successful Response
HTTP Return Codes	200
Exceptions	N/A
Precondition	N/A

4.3.2. AI Models and Data Analysis services

4.3.2.1. Overview

Machine learning (ML) is increasingly being used in **healthcare and biomedical research**, particularly in areas requiring large-scale data processing such as biomarker detection, disease classification, and predictive analytics. Traditional methods often struggle with the complexity and volume of modern biomedical datasets, whereas ML algorithms can identify patterns and insights that improve diagnosis and treatment.

Machine learning algorithms can be broadly categorized into **supervised learning** and **unsupervised learning**:

- **Supervised learning** involves labelled data, where the model is trained to map input data to known output labels. Examples include classification and regression models.
- **Unsupervised learning** is used when labels are not available, relying on clustering and dimensionality reduction techniques to find hidden patterns in data.

In ONELAB, we employed **supervised learning models** for **infection classification** based on GC-IMS data, enabling automated decision-making.

Gas Chromatography - Ion Mobility Spectrometry (GC-IMS) is an analytical technique used in chemistry and chemical analysis to separate and identify compounds in a mixture [2, 3, 5]. It combines two powerful techniques: Gas Chromatography (GC) and Ion Mobility Spectrometry (IMS). Each of these methods are described below:

Gas Chromatography (GC) Data: GC is used to separate a mixture of compounds into individual components based on their chemical properties, such as volatility and affinity for a stationary phase. The data generated by GC includes:

- **Retention Time:** The time it takes for each compound to travel through the GC column, which is unique to each compound and can be used for identification.
- **Peak Intensity:** The intensity or abundance of each compound's signal as it elutes from the GC column, which is related to its concentration in the sample.

Ion Mobility Spectrometry (IMS) Data: IMS is used to further separate and identify compounds based on their size, shape, and charge. The data from IMS includes:

- **Drift Time:** The time it takes for ions generated from the separated compounds to travel through a drift tube under the influence of an electric field. Different ions experience different drift times based on their size and charge, allowing for additional separation.
- **Ion Mobility Spectra:** The resulting spectra, often represented as plots or histograms, showing the distribution of ion drift times. Peaks in the IMS spectra correspond to different ions and can aid in compound identification.

GC-IMS: The data from both techniques is typically combined to create two-dimensional plots or chromatograms that provide information about the compounds in the sample. These plots can show the retention time from GC on one axis and the drift time from IMS on the other axis, allowing for the visualization and identification of compounds based on both their chemical properties and size/charge characteristics.

Classification is a fundamental task in machine learning, where models are trained to categorize data points into predefined labels. In the case of ONELAB, models were developed to classify GC-IMS samples as infected or non-infected.

This section describes the AI models and data analysis services developed in the ONELAB project for infection detection using Gas Chromatography-Ion Mobility Spectrometry (GC-IMS) data. Machine learning models were integrated into the Laboratory Information Management System (LIMS) to facilitate automated data analysis and classification. For this reason, we tested a variety of machine learning techniques to assess their performance on GC-IMS data:

- Logistic Regression: A linear classification model that predicts probabilities for binary outcomes.
- Decision Trees: Rule-based classifiers that split data based on feature thresholds.
- Random Forest: An ensemble of decision trees that enhances prediction accuracy.
- Support Vector Machines (SVM): A classification model that finds the optimal decision boundary in high-dimensional space.
- Partial Least Squares-Discriminant Analysis (PLS-DA): A multivariate technique designed to handle high-dimensional biological datasets.

Preprocessing of raw GC-IMS data was conducted prior to the data upload on the LIMS platform and before data was made available for model training. The preprocessing pipeline included tasks like:

- **Artifact removal and noise reduction** to improve data quality.
- **Drift time alignment** to standardize retention times across samples and measurement sites
- **Feature extraction** methods to convert high-dimensional GC-IMS data into structured tabular formats.
- **Normalization and standardization** to prepare the data for machine learning models.

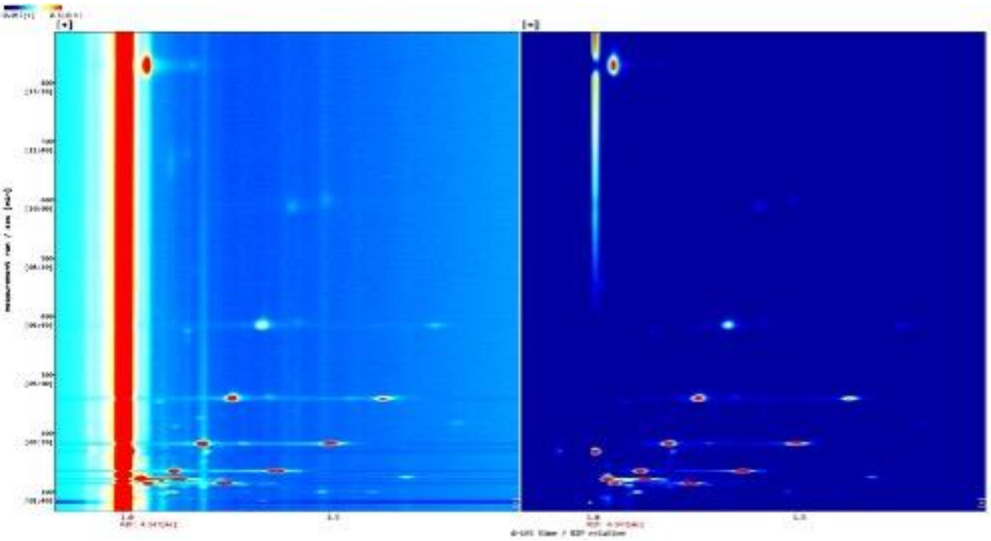

Figure 7: Example of application of individual denoising on GC-IMS data.

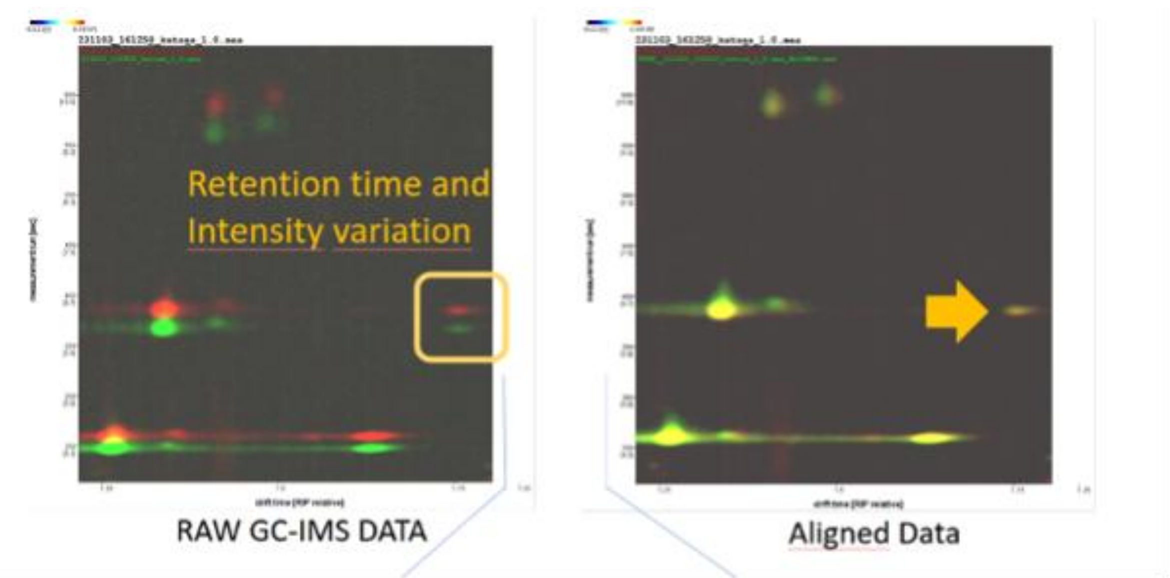

Figure 8: Example of retention time and intensity variation alignment between GC-IMS data.

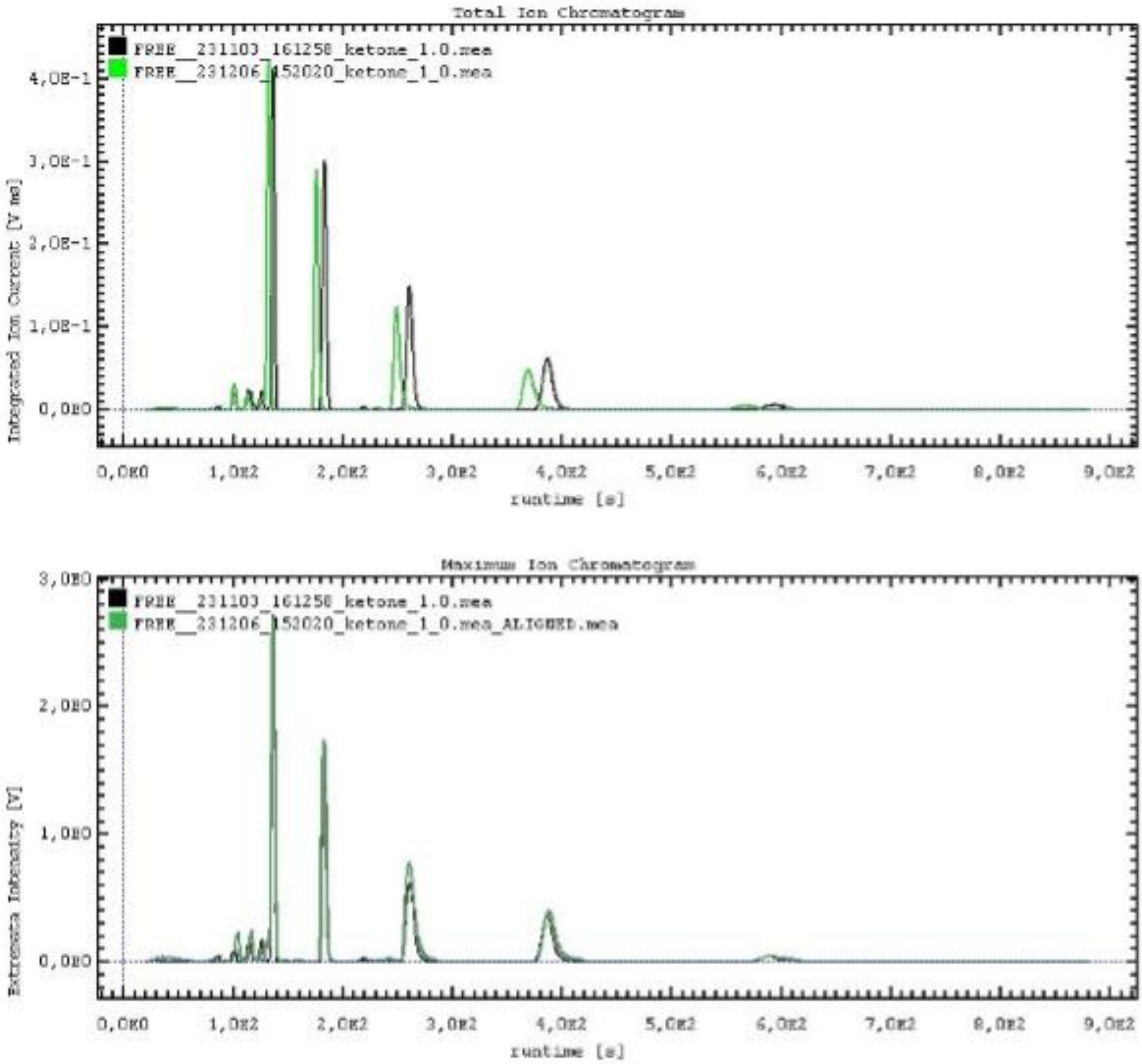

Figure 9: Example of reference alignment of GC-IMS measurements from different measurement sites based on KetonMix Calibration data to reference device based in Amsterdam.

Furthermore, the preprocessing formally includes manual checks as feature of the LIMS platform giving access to QA/QC managers to download, investigate and potentially deactivate samples based on defined criteria as described in D6.1 [D6.1].

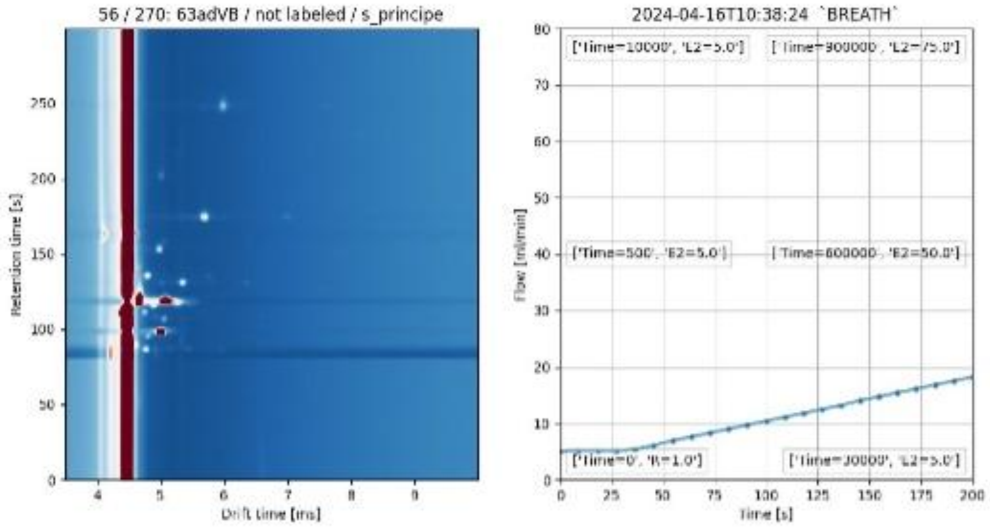

Figure 10: Example of routinely provided quality control data: GC-IMS spectra together with sampling flow from nitrogen generator.

Following preprocessing, to optimize performance, **Principal Component Analysis (PCA)** and feature importance ranking methods were used to reduce dimensionality while preserving relevant information, and **machine learning model development** tasks followed. This involved:

- **Dimensionality reduction techniques** such as **Principal Component Analysis (PCA)** to retain key features while reducing computational complexity.
- **Selection of key biomarkers** using feature importance ranking methods.
- **Testing multiple classification algorithms**, including:
    - Decision Trees
    - Logistic Regression
    - Partial Least Squares-Discriminant Analysis (PLS-DA)
    - Support Vector Machines (SVM)
    - Random Forest
- **Evaluation of classification accuracy**, with models achieving performance between 53% and 84% of accuracy.
- **Integration into the LIMS platform** for real-time inference and decision support.

Classification models

Many classification tasks benefit from a wide range of machine learning algorithms, and these algorithms have been applied in diverse fields. In the scope of the ONELAB project, we employ a selection of these algorithms, namely Decision Trees, Logistic Regression, Support Vector Classifier (SVC), Random Forest, and Partial Least Squares Discriminant Analysis (PLS-DA). In contexts related to Gas Chromatography – Ion Mobility Spectrometry (GC-IMS) data, these algorithms have garnered significant attention and recognition within the scientific community. In the subsequent paragraphs, we provide concise high-level descriptions of each of these algorithms.

Logistic regression

Logistic regression is a statistical modelling technique used in data analysis and machine learning to predict binary outcomes, such as yes/no or true/false decisions. It is particularly suited for situations where you want to understand the relationship between one or more independent variables (features) and a categorical dependent variable (target), which has two possible outcomes. Unlike linear regression, logistic regression employs a logistic function to transform the output into a probability score, which represents the likelihood of the event occurring. By estimating coefficients for each independent variable, logistic regression quantifies the impact of these features on the probability of the desired outcome, making it a valuable tool for classification tasks and decision-making in various fields, from finance to healthcare and beyond.

Decision trees

A decision tree is a versatile machine learning algorithm used for both classification and regression tasks. It constructs a tree-like structure where decisions are made at each internal node based on specific features, leading to different branches or outcomes. At the leaf nodes, final predictions or values are assigned. Decision trees are adept at capturing complex relationships between input features and the target variable, making them highly interpretable models. They are particularly valuable for feature selection, as they reveal which features are most influential in making predictions. This simplicity and effectiveness make decision trees a widely adopted tool in various industries, aiding in data-driven decision-making and enhancing our understanding of the underlying patterns in data.

SVC

The Support Vector Classifier (SVC) is a powerful and widely used machine learning algorithm for binary and multiclass classification tasks. It operates by finding an optimal hyperplane that best separates data points into distinct classes while maximizing the margin between them. SVC is particularly effective when dealing with complex and high-dimensional data, as it focuses on the most influential data points, known as support vectors. By using a mathematical concept called the kernel trick, SVC can also handle nonlinear relationships in the data. This versatility, coupled with its ability to handle outliers effectively, has made SVC a valuable tool in fields like image recognition, text classification, and bioinformatics.

Random forest

Random Forests is an ensemble learning method that combines multiple decision trees to improve predictive accuracy and robustness. Each tree in the forest is constructed using a random subset of the data and a random subset of the features, reducing the risk of overfitting and enhancing generalization. Random Forests excel at handling complex relationships, noisy data, and feature importance assessment. They are a popular choice for a wide range of applications, from finance and marketing to ecology and healthcare, due to their ability to deliver reliable and interpretable predictions.

PLS-DA

Partial Least Squares Discriminant Analysis (PLS-DA) is a multivariate statistical technique commonly employed in classification tasks, particularly in the analysis of high-dimensional data such as spectroscopy or genomics. PLS-DA seeks to find linear combinations of the original features (variables) that maximize the separation between different classes while considering their intercorrelations. This approach effectively reduces dimensionality and extracts informative features for classification. PLS-DA is valuable in fields like chemometrics, biology, and pharmaceuticals, where discriminating between different groups or classes is a critical objective, as it can provide insights into the most significant factors driving classification.

Experimental Setup

The data set used for training/testing purposes included 54 instances of GC-IMS (gas chromatography ion mobility spectrometry) data given in .hdf5 format without pre-processing as the previous sample. It was used in a similar way with the previous data. The difference is that in this case the Retention Time dimension of each instance is 6939. Each data instance is represented by a matrix (dim.: X, Y) in which the X dimension is the Drift Time and the Y dimension is the Retention Time while the actual values of the matrix represent the intensity of the concentration of each compound in each Drift-Retention Time point. The usual process that is followed in modelling such data is to unfold each sample into vector in which each dimension corresponds to a pair of Drift and Retention Time. Therefore, the final format of the data that will be utilized for modelling has a tabular format. Each row corresponds to one sample and each column to one pixel of the spectrum (Drift-Retention Time pair) and table values represent the corresponding intensities.

To ensure optimal model performance and generalization, the dataset was **split into training and testing sets**, with **80% of the data allocated for training and 20% for testing**. This division allowed for sufficient training while maintaining an independent test set for unbiased model evaluation.

To further enhance consistency across experiments, **standardization techniques were applied** to normalize the dataset, ensuring that each feature was scaled appropriately. This step was crucial in preventing features with larger ranges from dominating the learning process.

Feature selection was also a critical part of the workflow, with **various techniques evaluated** to improve model efficiency and performance. Principal Component Analysis (PCA), LASSO regression, and feature importance methods were applied to retain the most relevant attributes while reducing computational complexity.

To validate model robustness and prevent overfitting, we implemented **5-fold cross-validation (CV)**, a widely used method in machine learning. This technique involved partitioning the training data into five subsets, training the model on four subsets while using the remaining one for validation, and repeating the process five times. The averaged results provided a reliable estimate of model performance.

The machine learning models were evaluated on the **GC-IMS datasets** to classify samples into **infected vs. non-infected** categories. The table below presents an overview of the classification accuracy of different models.

Table 4: Implemented Models evaluation.

Model	Feature Selection	Accuracy	Sensitivity	Specificity
Logistic Regression	No	80%	82%	76%
Logistic Regression	Yes	60%	63%	57%
Decision Trees	No	60%	58%	62%
Decision Trees	Yes	40%	42%	37%
SVM	Yes	60%	55%	61%
Random Forest	No	60%	62%	59%
Random Forest	Yes	80%	84%	77%
PLS-DA	No	80%	81%	79%
PLS-DA	Yes	60%	59%	58%

These results indicate that **Random Forest and PLS-DA** were among the best-performing models, demonstrating their potential for infection classification in the ONELAB context. To further improve infection classification, a Convolutional Neural Network (CNN) model is being developed to analyze spectral images obtained from GC-IMS. Unlike traditional feature extraction and classification methods, CNNs are well-suited for detecting spatial patterns in biomedical imaging data, making them an ideal approach for biomarker identification.

The proposed CNN model follows a structured deep learning pipeline:

1. Input Layer: The model receives spectral images from the GC-IMS device, where pixel intensity variations represent molecular signatures.
2. Convolutional Layers: Multiple convolutional filters are applied to extract hierarchical features, capturing distinct spectral patterns relevant to infection classification.
3. Pooling Layers: Dimensionality reduction through max-pooling operations ensures the model retains essential features while discarding irrelevant noise.
4. Fully Connected Layers: High-level features extracted from convolutional layers are fed into fully connected layers to generate classification decisions (infected vs. non-infected).
5. Activation Maps (Grad-CAM): A heatmap overlay is generated using Gradient-weighted Class Activation Mapping (Grad-CAM), highlighting the most influential spectral regions contributing to the classification. This provides interpretability for clinical experts, allowing them to validate whether the model identifies meaningful biomarkers.

The use of CNNs for spectral image analysis is expected to significantly improve infection classification by leveraging spatial feature extraction capabilities that traditional models lack. This approach enables a more accurate identification of spectral patterns, leading to enhanced classification performance. Additionally, the use of activation maps facilitates biomarker visualization, allowing clinical experts to interpret and validate model predictions. By providing a deeper understanding of spectral variations, CNN-based analysis can enhance the reliability of automated infection detection. Moreover, the scalability of this architecture allows for potential future adaptations, such as differentiating between various infection types (bacterial vs. viral) or incorporating multi-modal data sources. These advancements make CNNs a promising tool for real-time spectral image interpretation within the ONELAB framework. Future work will focus on optimizing hyperparameters, enhancing model robustness, and integrating CNN-based predictions into the LIMS platform for real-time spectral image interpretation.

4.3.2.2. Related User Requirements

The development of the AI models and data analysis services aligns with the project requirements outlined in **Section 3** of this deliverable. Specifically, it addresses the following requirements:

ID	R1_13
Unique Name / Title	The user views the outcomes / predictions of the Classification model
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The classification model is deployed and can be utilized by the user within the LIMS by selecting a measurement for inference.
Rationale	AI Analysis
Validation method	User testing

ID	R1_14
Unique Name / Title	The user views the outcomes / predictions of the Biomarkers Identification model
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	C-LIMS, L-LIMS
Brief Description	The Biomarkers Identification model is deployed and can be utilized by the user within the LIMS by selecting a measurement for inference.
Rationale	AI Analysis
Validation method	User testing

ID	R2_1
Unique Name / Title	Training of Infection type classification model with the collected GC-IMS samples
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	Analysis service, C-LIMS
Brief Description	The classification model is trained using the available GC-IMS samples. The process includes Data preprocessing and cleaning, Training of ML / DL model, Evaluation of trained model, Re-training using new collected data
Rationale	Measurements Analysis
Validation method	User testing

ID	R2_2
Unique Name / Title	Training of Biomarkers Identification model
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	Analysis service, C-LIMS
Brief Description	The biomarkers classification model is trained for the evaluation of clinical relevance of the identified biomarkers, and validated in diverse patient populations
Rationale	Measurements Analysis
Validation method	User testing

ID	R4_6
Unique Name / Title	The user views classification results for a given sample
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	By scanning the patient's QR code, the RRML personnel can view the classification results on whether they are infected or not
Rationale	Measurement tracking
Validation method	User testing

4.3.2.3. Interfaces

The AI models are deployed within the **LIMS platform**, providing seamless interaction with clinical users and enabling efficient processing, analysis, and interpretation of infection-related data. The system offers an integrated environment where data can be ingested, analyzed, and reported in a structured manner, ensuring that medical professionals and researchers can leverage AI-driven insights effectively.

Key Interface Components:

- **Data Upload Portal**: This portal serves as the entry point for processed GC-IMS data, ensuring that structured and preprocessed information is readily available for AI analysis. It supports batch uploads, enabling large-scale dataset integration, and includes validation checks to ensure data consistency before proceeding to model training or inference.
- **Classification Output Module**: A crucial component of the system, this module provides real-time visual feedback on infection detection results. Once the AI models process incoming GC-IMS data, results are displayed in an intuitive format, indicating the probability of infection and other relevant classification metrics. This module supports automated reporting, enabling clinicians to make informed and timely as well as geographically placed decisions based on model predictions.
- **API Integration for Data Exchange**: The platform is designed with **RESTful API support**, allowing for seamless data exchange between the LIMS and external systems. This ensures interoperability with other laboratory management platforms, facilitating

remote access to classification results, and real-time synchronization of infection status across multiple healthcare facilities.

By integrating these interface components, the LIMS platform enables efficient AI-driven infection detection, ensuring scalability and usability in clinical settings while maintaining compliance with data security and privacy regulations.

4.3.3. Synchronization service

4.3.3.1. Overview

The Synchronization service allows the synchronization between any L-LIMS with the C-LIMS. In more detail, it is a two-component service; (i) the model synchronization component, and (ii) the data synchronization component. Both components are deployed in each L-LIMS.

The first component (i.e., the model synchronization) checks if there exist any AI models available in the C-LIMS that do not exist in the L-LIMS deployment and requests to download them. In addition, it checks whether there are updates on existing models and updates them as well.

On the other hand, the data synchronization component checks if there is any measurement data stored in the L-LIMS which is not yet uploaded on the C-LIMS and uploads them. This happens by checking the available UUIDs existing in the C-LIMS and the L-LIMS and uploads the data of those UUIDs which only exist in the L-LIMS, to the C-LIMS.

The synchronization is a Python service which runs in the background and performs those synchronization operations in an iterative manner (i.e., every 6 hours).

4.3.3.2. Related requirements

The requirements related to the Synchronization service are the following:

ID	R3_1
Unique Name / Title	Synchronization data stored in the L-LIMS with the C-LIMS
Type	Non-Functional
Priority	High
Mandatory	Mandatory
Role	N / A
Related Use Case	All
Related Component	L-LIMS, C-LIMS, Synchronization Service
Brief Description	The data (i.e., measurements, samples, etc.) that is initially stored in the L-LIMS are backed up in the centralized infrastructure. The synchronization procedure is performed iteratively every 1 hour.
Rationale	AI Analysis and Storage
Validation method	Data integrity and logging monitoring

ID	R3_2
Unique Name / Title	AI Models synchronization between the C-LIMS and the L-LIMS
Type	Non-Functional
Priority	High
Mandatory	Mandatory

Role	N / A
Related Use Case	All
Related Component	L-LIMS, C-LIMS, Synchronization Service
Brief Description	The trained models, trained with the data found in C-LIMS are transferred to the L-LIMS for local infernece. The synchronization procedure is performed iteratively every 1 hour.
Rationale	AI Analysis and Storage
Validation method	Models' integrity and logging monitoring

4.3.3.3. Interfaces

This is a service that runs in the background, so no interfaces are required in order to interact with the users.

4.3.4. Patient Registration Mobile Application (PRMA)

4.3.4.1. Overview

The Patient Registration Mobile Application is an Android-based application that is used by the RRML personnel in order to register the patients when they arrive at the RRML. The application can also be used in order to perform additional operations such as linking the collected samples to a patient, retrieving the classification results and the predictions made by the AI models for a given patient's sample measurement, while also perform administrative operations such as seeing all registered patients, deleting patients, etc.

4.3.4.2. Related requirements

The requirements related to the Patient Registration Mobile Application are the following:

ID	R4_1
Unique Name / Title	Patient registration including preregistration
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel, Patient
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The patient is registered by the RRML personnel using the mobile application. Information from preregistered patients can be retrieved.
Rationale	Registration of patient in order to proceed to the RRML for measurement sampling.
Validation method	User testing

ID	R4_2
Unique Name / Title	Automated assessment based on patient's responses to Triage Questionnaire
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel

Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The RRML Personnel fills in the Triage questionnaire and the patient's priority is calculated according to their answers.
Rationale	Registration of patient in order to proceed to the RRML for measurement sampling.
Validation method	User testing

ID	R4_3
Unique Name / Title	The user inserts the patient's contact person information
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The patient's emergency contact person information is collected.
Rationale	Patient tracking
Validation method	User testing

ID	R4_4
Unique Name / Title	The user inserts the patient's information
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The patient's personal information is collected.
Rationale	Patient tracking
Validation method	User testing

ID	R4_5
Unique Name / Title	The user links a measurement sample with patient
Type	Functional
Priority	High
Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	The sample is linked to the patients UUID.
Rationale	Measurement tracking
Validation method	User testing

ID	R4_6
Unique Name / Title	The user views classification results for a given sample
Type	Functional
Priority	High

Mandatory	Mandatory
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	By scanning the patient's QR code, the RRML personnel can view the classification results on whether they are infected or not
Rationale	Measurement tracking
Validation method	User testing

ID	R4_7
Unique Name / Title	The user views a specific patient's contact information
Type	Functional
Priority	High
Mandatory	Recommended
Role	RRML Personnel
Related Use Case	All
Related Component	L-LIMS, Patient Registration Mobile Application
Brief Description	By scanning the patient's QR code, the RRML personnel can view the patient's contact information.
Rationale	Measurement tracking
Validation method	User testing

4.3.4.3. Interfaces

The Patient Registration Mobile Application provides a set of user interfaces in order for the user to perform their respective actions. These interfaces are depicted in detail below:

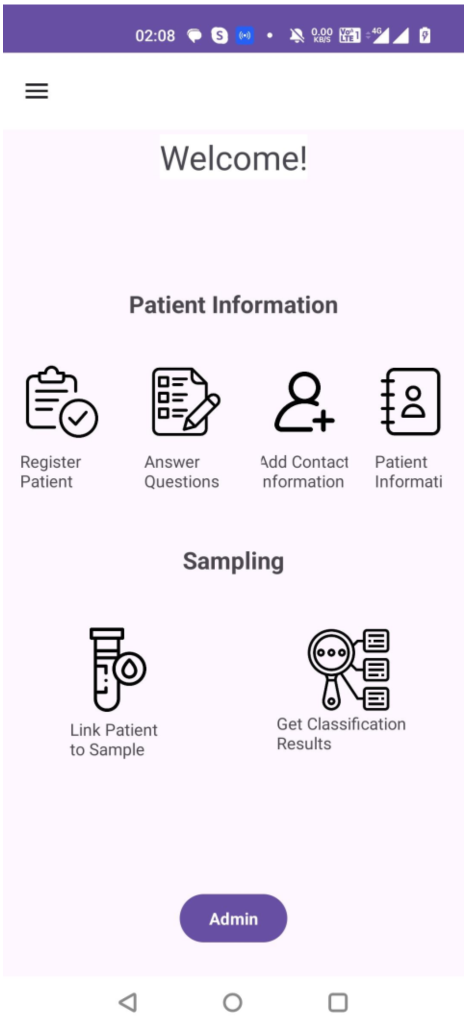

Figure 11: Home Page.

The Home Page of the Patient Registration Mobile Application. From here, the user can select any of the most common actions to perform. These include actions related to registering a patient, answering the triage questionnaire, adding the patient's contact information, and other actions related to sampling such as linking a patient ID to a sample measurement of retrieving the classification results for a patient's sample.

On the left side the user may select to open the sidebar by taping on the burger menu. This sidebar lists all available actions. The sidebar can be shown in the following figure.

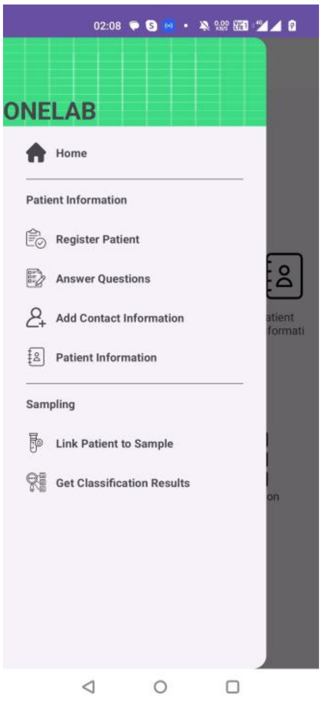

Figure 12: Sidebar.

When selecting the option to register a patient, the user is redirected to a page where they need to fill in some information about the citizen and scan the QR code of the patient in order to link them. The registration page along with the information that have to be inserted are shown in the following figure.

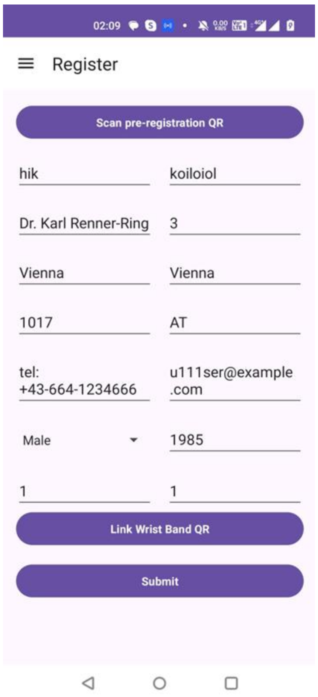

Figure 13: Patient registration page.

In addition to this information, the patient's QR code is scanned, and their record is linked with their UUID, when pressing the button 'Link Wrist band QR'.

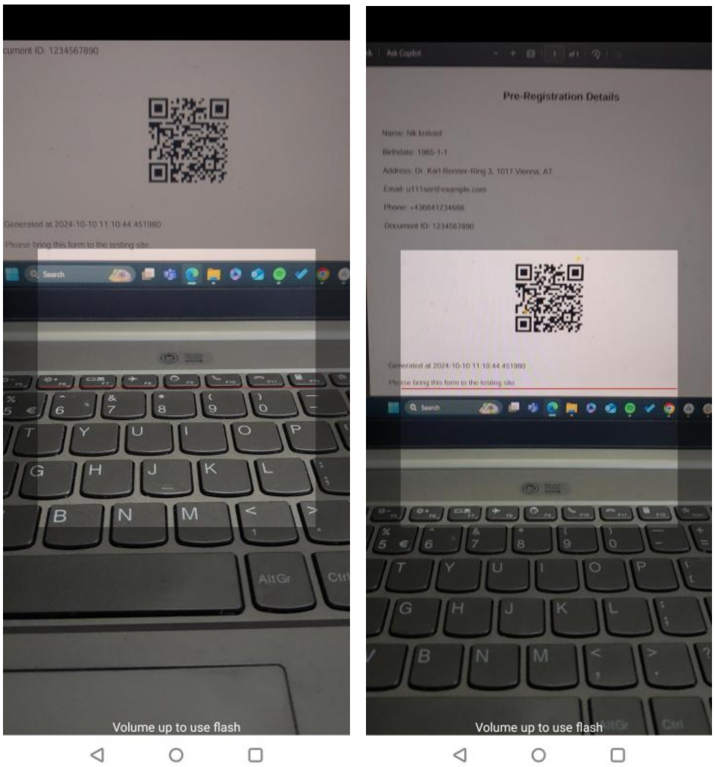

Figure 14: Link patient to UUID by scanning their wristband.

Upon the registration a notification is shown in order to let the RRML personnel know that the operation was successful.

The Triage page regards a list of questions that the patient needs to reply, as shown in the figure below.

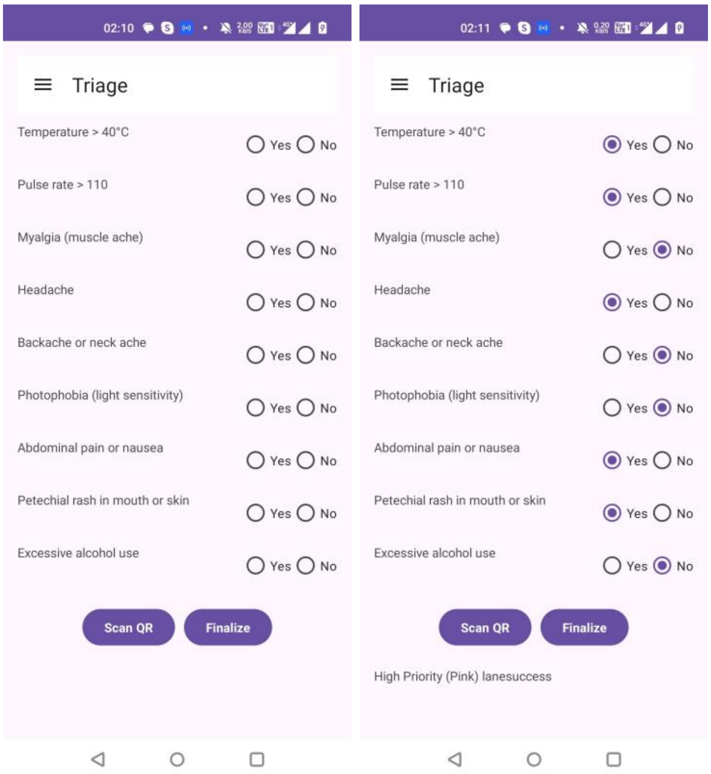

Figure 15: Triage questionnaire.

After the registration of the patient, it is important to add a contact person for each patient. This can be done, by selecting the 'Add Contact Information' option from the sidebar.

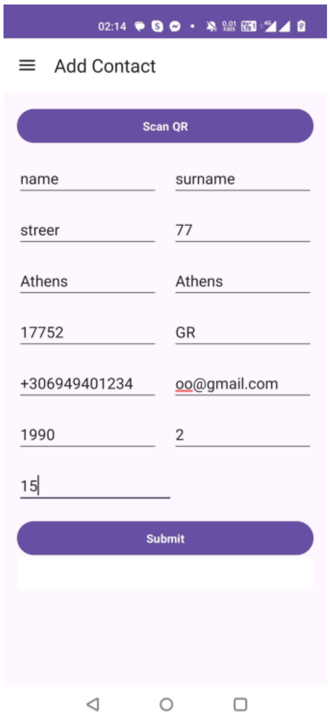

Figure 16: Contact information insertion page.

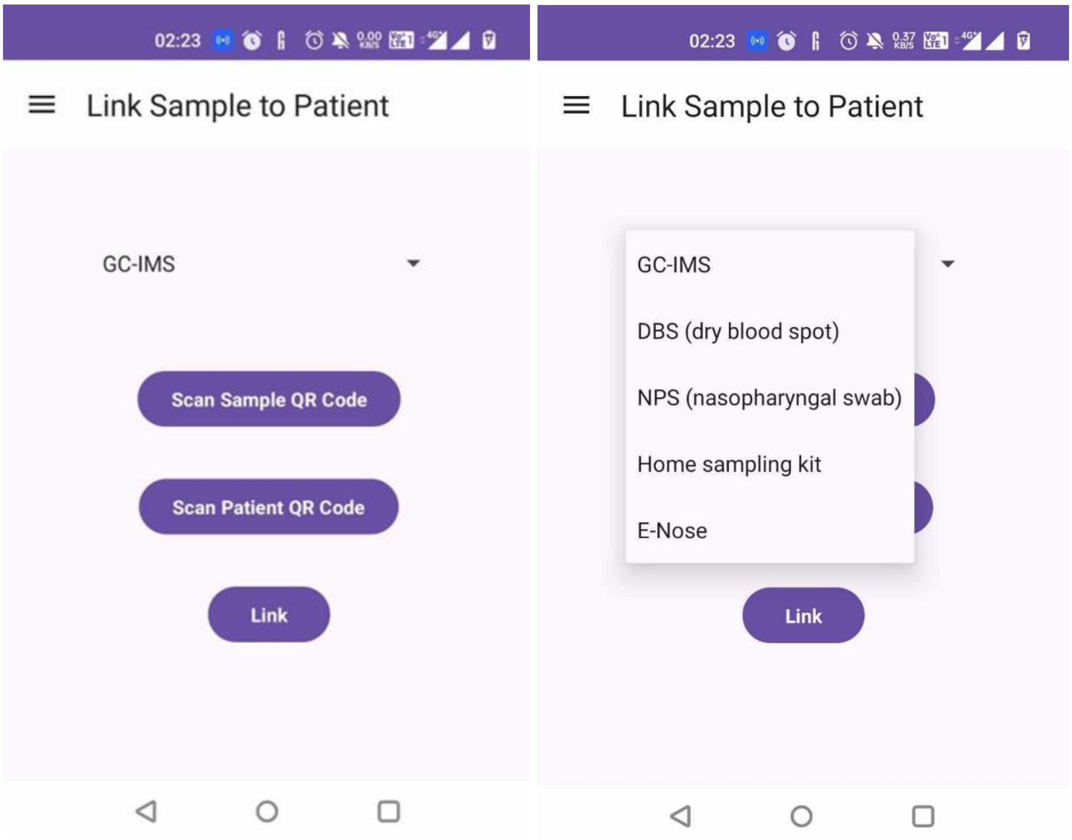

Figure 17: Link Sample to Patient

Using the ‘Link Sample to Patient’ page, the user can link any collected sample out of the available in the list (Figure 9b). Then they have to scan the sample’s QR code along with the patient’s QR code and the link between the two objects is performed.

5. FTX Exercise and Outcomes

According to the project proposal a field technical exercise (FTX) was designed, prepared and carried out at Johanniter Research and Innovation Center (JOAFG) in Vienna, Austria on 26th until 28th of October 2024 with the goal to evaluate the logistics of patient flow through a rapid response mobile laboratory (RRML). To this end the components detailed in this deliverable were combined to create the information and communication technology (ICT) system integrated in the hardware setup of the containers to form the technological basis for the logistics. For details about the overall concept, we refer to D5.2 - FTX “Hard Winter” [D5.2]. In the following only details about the relevant components for this deliverable are described.

In a first step a set of 30 patient actors were created and briefed by JOAFG each of them receiving a pre-registration sheet as depicted in Fig. 18. A defined set of patient actors were selected to act as infected patients and received a dose of eucalyptol liquid to gargle before entering the RRML area. We received a dummy training set for patients with and without eucalyptol before the FTX on which classification model was trained and deployed in the ICT using the setup described in Section 4.3.2. Since the classification is straightforward, i.e., signal at expected location in GC-IMS spectrum present or not, 100% classification accuracy was possible (see Figure 19).

The core components of the ICT connected within the private wireless local area network (W-LAN) are the following:

Table 5: Core components of the ICT connected within the private W-LAN.

Component	Locations
Patients' registration mobile application (PRMA)	Patient Registration and Triage Desk Sample container
L-LIMS	Linux Server in ICT Container

Patients were then passing the RRML along the following stations. The following table summarizes these stations together with the components used of the ICT.

Table 6: Stations and the related ICT components using in each one.

Location	Description	Used components
Patient registration and triage desk	Patient presents pre-registration sheet with QR code. After scanning the QR Code using the PRMA installed on Samsung Tablets patient details are verified registration is completed by sending the information to the L-LIMS backend. Triage questions	PRMA L-LIMS

	are answered using the PRMA and are also sent to the L-LIMS after completion. Patient moves on to sampling site.	
Sampling site	A breath sample is taken, and the sample is linked via QR code on sample tube to patients QR code using the PRMA. The sample is handed to personnel inside the measurement container.	PRMA L-LIMS
Measurement site	The sample QR code is scanned with a barcode scanner attached to GC-IMS device before injecting the breath sample. After sample processing a measurement artifact is generated and stored automatically on an attached NUC in the W-LAN along the sample-ID. A watch script automatically transfers the measurement file to the L-LIMS service	L-LIMS Synchronization service
ICT container	The sample received is automatically classified, and the classification result is stored in the L-LIMS	AI Model and data analysis service
ICT container	The local database is synchronized to the C-LIMS	Synchronization service

During the FTX the resulting flow of patients could be monitored continuously from the ICT container by technical personnel. Figure 20. depicts a dashboard developed for monitoring the continuous updates in the database. After completing the FTX it was then possible to evaluate the timing of patients between different locations of action as depicted in Figure 21. Finally, from the L-LIMS as well as from the C-LIMS (after synchronization) it was then possible to use GIS mapping to locate infected and non-infected patients throughout the city area of Vienna via provided dashboards as depicted in Figures 22 with individual locations and Figure 23 providing a heat-map with the main hot spots of simulated viral outbreak.

ONELAB Patient information

RRML ID: HardWinter

Patient Information

Name: Egon Estragon

Birthdate: 1943-4-16

Address: Bellgasse 2, Vienna, Vienna, 1210, AT

Fühlen Sie sich krank? Ja

Gliederschmerzen ?: Ja
Kopfschmerzen ?: Nein
Genick-/Rückenschmerzen ?: Ja
Bauchschmerzen ?: Nein
Alkohol in größeren Mengen ?: Nein

Temperatur: niedrig
Pulse: niedrig
Licht: Nein
Blutungen: Nein

Notes: Living with 4
Ref.: 1, **Code:** I-S, **Lane:** PINK, **Eucalyptol:** Y, **Pre-reg:** Y
LAP: 1,**PA Nr.:** 1, **Score:** 8

Figure 18: Preregistration sheets for patient actor in FTX.

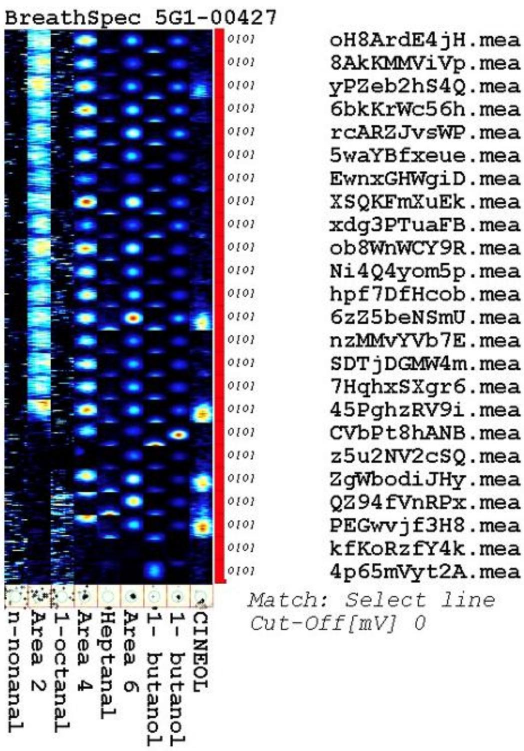

Figure 19: Depiction of signal peaks present for patients who received a dose of eucalyptol before entering RRML area (column CINEOL).

FTX Patients Pre-Registered GIS

	ID	Registration Date	GCIMS	GCIMS-Sample-ID	Sample Date	TimeDelta	NPS	HSK	Prediction
5	cc4e702d-c3ca-4042-a9f9-5fc24f48780b	2024-11-27 11:28:25+00:00	☐	NA	None	0	☐	☐	NA
42	326f8085-405e-4617-b1f7-854d602073a0	2024-11-27 11:25:10+00:00	☑	accRjyccLn	2024-11-27 11:27:48+00:00	0 days 00:00:02.632633150	☐	☑	no infection
2	a5434152-f466-45ab-88a6-59b606ed3b47	2024-11-27 11:22:16+00:00	☑	oH8ArdE4jH	2024-11-27 11:24:49+00:00	0 days 00:00:02.550988850	☐	☑	no infection
43	f2532b34-120b-4db9-8c57-88a20d44e3c9	2024-11-27 11:19:08+00:00	☑	cgGHYBZRHu	2024-11-27 11:21:28+00:00	0 days 00:00:02.333311600	☐	☑	no infection
16	51e7d7c1-e840-401b-a5b8-425026725e3a	2024-11-27 11:16:29+00:00	☑	04460626	2024-11-27 11:18:46+00:00	0 days 00:00:02.286877083	☐	☐	NA
12	2fdd146d-f9b7-47bf-84d1-a9e8f5c01dcb	2024-11-27 11:13:45+00:00	☑	r4JZNGbS3U	2024-11-27 11:16:05+00:00	0 days 00:00:02.344624300	☐	☑	no infection
37	a907bc08-8e86-4dd9-9c30-acbac01ec1b8	2024-11-27 11:10:17+00:00	☑	8AkKMMViVp	2024-11-27 11:12:57+00:00	0 days 00:00:02.667583016	☐	☑	no infection
45	70980a96-7878-4f25-9041-0c94d319009d	2024-11-27 11:07:41+00:00	☑	cgDLnP52ZM	2024-11-27 11:09:58+00:00	0 days 00:00:02.295295533	☐	☑	no infection
48	117734c9-8602-4852-8147-a2bf39b6ec90	2024-11-27 11:04:21+00:00	☑	yPZeb2hS4Q	2024-11-27 11:07:03+00:00	0 days 00:00:02.704012016	☐	☑	no infection
49	dc90737d-a67e-4dbe-8700-d44001e537f9	2024-11-27 11:01:24+00:00	☑	5aCsoZoP9h	2024-11-27 11:03:36+00:00	0 days 00:00:02.201044983	☐	☑	no infection
35	3057a96c-db80-428f-998d-03c90201281e	2024-11-27 10:57:38+00:00	☑	6bkKrWc56h	2024-11-27 11:00:50+00:00	0 days 00:00:03.208283616	☐	☑	no infection
8	b3fdee2f-ebc2-4c35-a347-6c55100f74c0	2024-11-27 10:53:42+00:00	☑	d43AxRHFaa	2024-11-27 10:57:27+00:00	0 days 00:00:03.752585850	☐	☑	no infection
30	c8cdba73-8168-4954-a702-72e493993c02	2024-11-27 10:50:23+00:00	☑	z6JwHofdki	2024-11-27 10:53:51+00:00	0 days 00:00:03.466298516	☐	☑	no infection
33	70d7911a-baf5-4064-bf38-ce5845635799	2024-11-27 10:48:05+00:00	☑	rcARZJvsWP	2024-11-27 10:54:40+00:00	0 days 00:00:06.586296200	☐	☑	no infection
9	e42fad03-3a07-4ad7-b323-84dc2dc0155f	2024-11-27 10:47:08+00:00	☑	5waYBfxeue	2024-11-27 10:50:16+00:00	0 days 00:00:03.130388733	☐	☑	viral infection

Figure 20: Dashboard for technical personnel within ICT container to monitor patient flow.

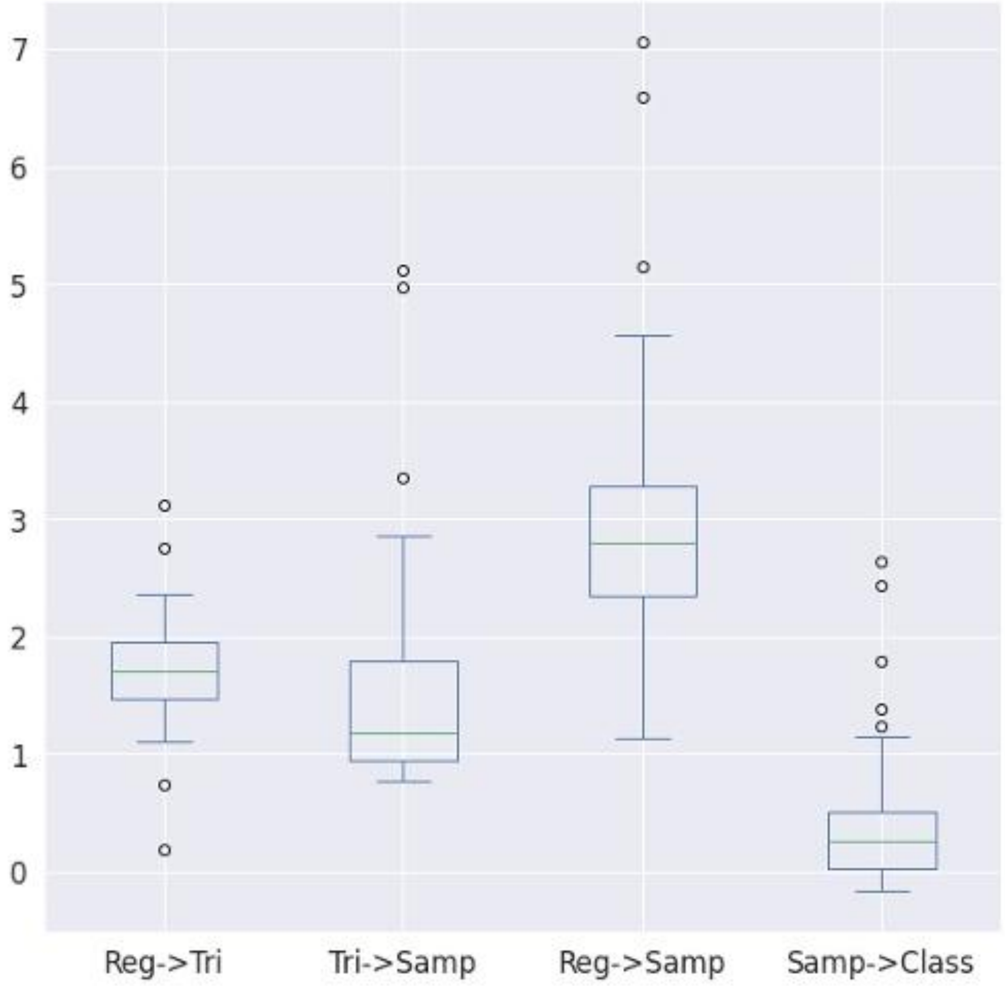

Figure 21: Statistical Evaluation of time durations for patients and samples during the FTX: (Reg->Tri) Time in minutes for patients from registration to complete triage questions (Tri->Samp) time in minutes for patients from triage desk to location of sample taking, (Reg->Samp.) total time of patient after registration to sample location (including sample taking). (Samp->Class) Time used to process sample in measurement container.

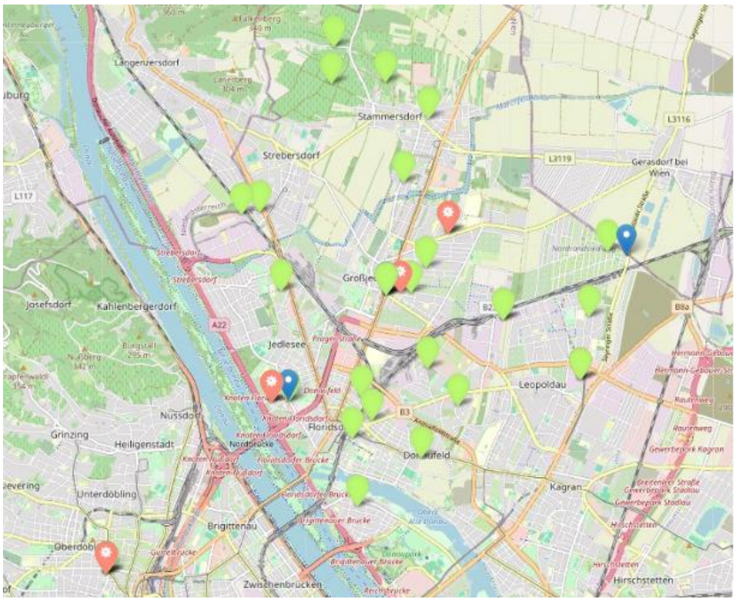

Figure 22: GIS mapping of fake patient actor locations after presentation in the RRML (green: not-infected, red: infected, blue: not classified).

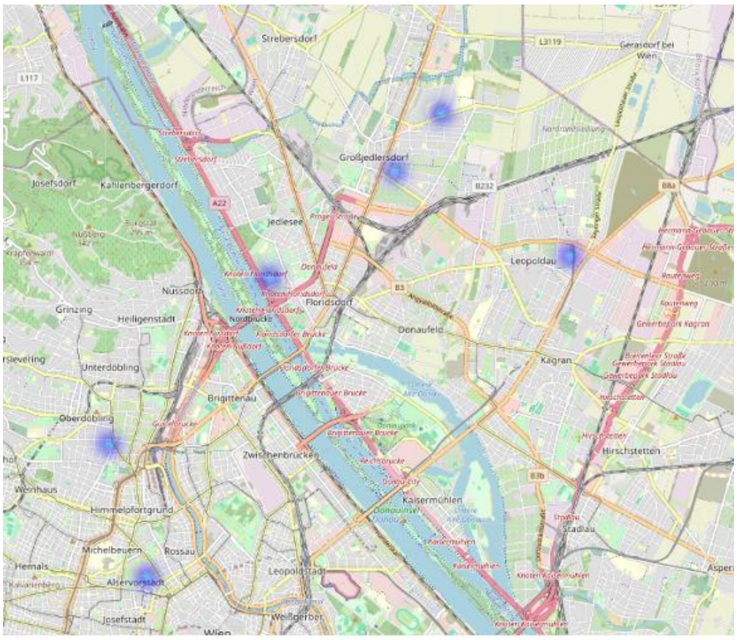

Figure 23: Heatmap based on GIS data for fake infected patients after

6. Conclusions

In conclusion, this document has provided a thorough examination of the ONELAB LIMS solution, detailing its architecture, core components, and alignment with the needs of the ONELAB use cases and field exercises.

In more detail a comprehensive analysis of the ONELAB LIMS solution was provided, focusing on its design, architecture, and practical implementation. It began with an in-depth review of use cases and user requirements, forming the foundation for the ONELAB LIMS solution.

The core components, including the LIMS, AI models, synchronization service, and Patient Registration Mobile Application, were thoroughly examined to demonstrate how the system meets lab management needs, particularly for the ONELAB RRML. Finally, the outcomes from the FTX exercise were discussed in which the LIMS was utilized and validated.

7. References

[D4.4] D4.4 - Modular Laboratory RRML Platform, ONELAB Consortium.

[D5.1] D5.1 – FTX Technical file, ONELAB Consortium.

[D5.2] D5.2 – FTX - “Hard Winter” (clinically based PoC), ONELAB Consortium.

[D5.3] D5.3 - FTX “Pre-emptive” (Pandemic simulation in collaboration with WPCTF), ONELAB Consortium.

[D6.1] D6.1 - Pandemic Testing Playbook, ONELAB Consortium.

[GitLab] GitLab, Available at: https://about.gitlab.com/

[SQLite] SQLite, Available at: https://www.sqlite.org/

[Streamlit] Streamlit, Available at: https://streamlit.io