D5.6 CDSS Evaluation Report

31/12/2022

Project title	Patients-centered SurvivorShip care plan after Cancer treatments based on Big Data and Artificial Intelligence technologies
Grant Agreement number	875406
Call and topic identifier	SC1-DTH-01-2019 - Big data and Artificial Intelligence for monitoring health status and quality of life after the cancer treatment
Funding schema	RIA
Coordinator	FUNDACION CENTRO TECNOLOXICO DE TELECOMUNICACIONS DE GALICIA (GRADIANT)
Website	www.projectpersist.com
Document keywords	CDSS, Inference Engine
Document Abstract	This document explains the specifications and features of clinical decision support in PERSIST and mClinician mobile application as its user interface and visualization of outputs.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

DOCUMENT	
Authors	Kadir Üğüdücü (EMO), Tunç Cerit (EMO)
Internal reviewers	SERGAS, UKCM
Work package	WP5 – Decision support system at the point of care
Task	T5.5 Integration and evaluation of CDSS
Nature	Report
Dissemination Level	Confidential

VERSION	DATE	CONTRIBUTOR	DESCRIPTION
0.1	01.12.2022	Tunç CERİT (EMO)	Initial draft
0.2	05.12.2022	Tunç CERİT (EMO), Kadir Üğüdücü (EMO)	CDSS Architecture
0.3	07.12.2022	Tunç CERİT (EMO), Kadir Üğüdücü (EMO)	Evaluation of CDSS Services
0.4	09.12.2022	Tunç CERİT (EMO), Kadir Üğüdücü (EMO)	Conclusion and review
0.5	11.12.2022	Tunç CERİT (EMO), Kadir Üğüdücü (EMO)	Figures and tables
0.6	12.12.2022	Tunç CERİT (EMO), Kadir Üğüdücü (EMO)	Review Version
1.0	22.12.2022	Tunç CERİT (EMO), Kadir Üğüdücü (EMO),Maja Molan (UKCM), Ignacio Serrano (GRAD), Alberto Sánchez (GRAD)	Final Version

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

DISCLAIMER

ACKNOWLEDGEMENT

This document is a deliverable of PERSIST project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 875406

INDEX

INDEX 4

List of figures 6

List of tables 7

Acronyms and Abbreviations 8

Executive Summary 9

Introduction 10

1. Work Package 5: Decision support system at the point of care 10

2. D5.6 CDSS Evaluation Report 11

CDSS Architecture 12

1. General Structure of CDSS 12

2. Process flow of CDSS 13

3. Components of CDSS in PERSIST 14

3.1. Knowledge base 14

3.2. Inference Engine 15

3.3. User Interface 15

3.4. Data Ingestion 16

3.5. Risk Assessment 17

3.6. Alert Mechanism 18

3.7. Trajectories 19

3.8. Recurrence Prediction 20

3.9. Disease Centric Discourse 21

3.10. Diary Analysis Results 22

Evaluation of the CDSS 23

1. Overview of the evaluation process 23

2. Evaluation of Risk Assessment Services 23

2.1. Breast Cancer Recurrence Risk Assessment Service 24

2.2. Colon Cancer Recurrence Risk Assessment Service 26

2.3. Cardiovascular Disease Risk Assessment Service 28

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

3. Evaluation of Alert Mechanism 30
3.1. Test Scenarios for alert trigger 30
3.2. User Testing 34
4. Evaluation of integrated services 35
4.1. Evaluation of Recurrence Prediction 35
5. Evaluation of integrated services 37
5.1. Evaluation of DCD integration 37
5.2. Evaluation of Diary Analysis integration 37
6. Evaluation of User Interface 38
Conclusions **39**
References **40**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

List of figures

Figure 1 Overview of CDSS Structure 12

Figure 2 Overview of CDSS flow 13

Figure 3 Cox Proportional Hazards for feature importance breast cancer patients 14

Figure 4 Components providing data in PERSIST 16

Figure 5 Data usage by CDSS in PERSIST 17

Figure 6 Alert mechanism request flow overview 18

Figure 7 Kaplan-Meier breast cancer patient trajectory grouped by cancer stage T from TNM 19

Figure 8 Overview of Recurrence Prediction service architecture 20

Figure 9 DCD Result displayed on mClinician application 21

Figure 10 Diary analysis result displayed in mClinician application 22

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

List of tables

Table 1 Test case results for breast cancer recurrence risk assessment service . 25

Table 2 Test case results for colon cancer recurrence risk assessment service .. 27

Table 3 Test case results for cardiovascular disease risk assessment service 29

Table 4 Test cases for alert mechanism 33

Table 5 Statistical results of user tests for the alert mechanism 34

Table 6 Evaluation outputs of colon cancer model of Recurrence Prediction 36

Table 7 Evaluation results of breast cancer model of Recurrence Prediction....... 36

Table 8 Test cases for the DCD integration 37

Table 9 Test case for the diary analysis integration 37

Table 10 Reported issues and feedbacks from user functionality tests between version releases 38

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Acronyms and Abbreviations

ACRONYM	TITLE
BLE	Bluetooth Low Energy
CDSS	Clinical Decision Support Systems
DCD	Disease Centric Discourse
FHIR	Fast Healthcare Interoperability Resources
GDPR	General Data Protection Regulation
HL7	Health Level Seven
HTTP	Hyper-Text Transfer Protocol
IoT	Internet of Things
LOINC	Logical Observation Identifiers Names and Codes
MQTT	Message Queuing Telemetry Transport
OHC	Open Health Connect
PHQ	Patient Health Questionnaire
PREM	Patient Reported Experiences
PROM	Patient Reported Outcomes
REST	Representational State Transfer
SNOMED	Systematized Nomenclature of Medicine
SSL	Secure Sockets Layer
TLS	Transport Layer Security
UI	User Interface

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Executive Summary

This deliverable report is written in the framework of WP5 – Decision support system at the point of care (Task 5.5 Integration and evaluation of CDSS) of PERSIST project under Grant Agreement No. 875406.

This deliverable highlights the end-to-end steps taken to develop an operating Clinical Decision Support System (CDSS) inference engine in PERSIST environment. Using data from clinical repositories and analyse data recorded in order to present outcomes to clinicians.

The goal of this deliverable is integration of CDSS with the system and other services and evolution of CDSS with its component.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Introduction

PERSIST aims at developing an open and interoperable ecosystem to improve the care of cancer survivors. The key results to be achieved by partners are:

- Related to **patients**: **increased self-efficacy and satisfaction** with care as well as reduced psychological stress for better management of the consequences of the cancer treatment and the disease.
- Related to **professionals**: **increased effectiveness in cancer treatment and follow-up** by providing prediction models from Big Data that will support decision-making and contribute to optimal treatment decisions.
- Related to **healthcare providers**: improved information and evidence to advance the efficacy of management, intervention, and prevention policies. The long-term result will be to **reduce the socio-economic burden** related to cancer survivors' care.

In this document, the first prototype of mHealth application will be explained. The task can be examined in three parts: Conceptual Design, General Concepts, and the User Manual.

1. Work Package 5: Decision support system at the point of care

The overall goal of WP5 is to provide and implement a toolkit of basic and cutting-edge analytical methods to build the data mining and knowledge discovery services for guidance and support of the decision making and personalization of the survivor care plan. This work package will deliver a knowledge-based decision support system. The data-driven tools to be implemented will analyse the information in the databases and give conclusions related to the usage of the self-management platform, reported adverse events and health issues, compliances, health status, quality-of-life, etc. These tools and services will be integrated into the big data platform and support the smart data analytics. The specific objectives are:

- To develop a clinical decision support system to support the clinicians for diagnosis, treatment, and follow-up of cancer patients
- To carry out data exploration, harmonization, visualization, and basic statistics in a way that uncovers new insights into the research questions being asked
- To evaluate the clinical decision support system in the relevant environment (validation of results of Machine Learning algorithms)

Role of each partner: EMO is the leader of WP5 and will be responsible for software requirements specification, software design, the implementation, integration, and evaluation of CDSS. SYMP contributes with its expertise in EHR normalization and processing to define and develop solutions to convert raw exported data records into a conclusive format. HESSO leads the patients' cohort and trajectory analysis that populates the knowledge base for the CDSS. All the involved partners are supported by the clinical partners of the consortium, which will contribute to the development of CDSS and the rule-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

based inference engine with their clinical expertise and perspectives in an iterative process. The CDSS is one of the main results of the project.

2. D5.6 CDSS Evaluation Report

The scope of this deliverable is to output evaluation process and results of the developed CDSS components and modules including inference engine, alert mechanism, recurrence prediction and user interface. First, describing the CDSS and its features in PERSIST, including high level explanations for the underlying technologies, knowledge base and user interface were given to summarize the work carried before evaluation.

The aim of this evaluation is to assess the technical accessibility, performance, and scalability of the CDSS implemented in PERSIST. The CDSS incorporates a variety of tools, including rule-based algorithms, notifications and alerts, and AI and machine learning. The evaluation will focus on the system's ability to accurately process and analyze clinical data and generate appropriate recommendations in a timely manner. It will also assess the system's scalability, user-friendliness, and potential for future expansion.

An overview of the CDSS evaluation process was described including the methods to assess the CDSS’s performance. Also as a conclusion notable strengths, weaknesses and challenges possible overall improvements for the CDSS, based on the findings of the evaluation were mentioned.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

CDSS Architecture

1. General Structure of CDSS

The CDSS in PERSIST is a medical software constructed with collection of services. that uses artificial intelligence technologies, algorithms, and advanced software services to help healthcare professionals make more efficient treatment plans for their cancer-survivor patients. CDSS generally has the following components:

- Knowledge base consisting of evidence-based guidelines and medical knowledge both already existing and provided by clinical partners, which the CDSS uses to provide predictions, recommendations, and insight to healthcare professionals. NCCN guidelines, scientific literature and high-risk markers from cohort and trajectory analysis creates the foundation for the knowledge base in PERSIST.

- A user interface, which allows healthcare professionals to input detailed patient data and receive outputs from the CDSS.

- An inference engine, which is the core component of the CDSS. It uses algorithms to analyse the patient data and generate recommendations and assessments based on the medical knowledge and guidelines in the knowledge base.

- A knowledge representation system is needed to encode the medical knowledge in a format that can be easily processed between the inference engine and coexisting systems and services. In PERSIST data is represented with HL7 standard FHIR format and using coding systems such as SNOMED and LOINC.

Overall, the CDSS in PERSIST designed to assist healthcare professionals in making more informed and patient-specific decisions about their patients' care, by providing them with real-time access to relevant medical knowledge.

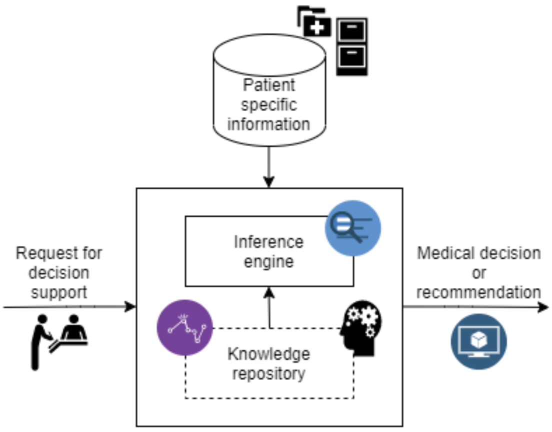

Figure 1 Overview of CDSS Structure

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

2. Process flow of CDSS

The process flow of the CDSS involves the following steps:

1. A healthcare professional inputs relevant patient data into the CDSS, such as the patient's symptoms, medical history, and laboratory test results, the multi-modal sensing network enables collecting patient data and provide additional information with use of AI technologies.

2. The CDSS uses its inference engine to analyse the patient data and generate recommendations based on the medical knowledge, AI and ML algorithms, and modelled guidelines in its database.

3. The CDSS presents the recommendations, assessment results or predictions to the healthcare professional in a user-friendly format, such as a list of possible diagnoses, underlying symptoms, negative and positive factors, missing information, possibility of having a condition or calling to further action.

4. The healthcare professional reviews the outputs and uses them to inform their diagnosis and/or treatment plan for the patient.

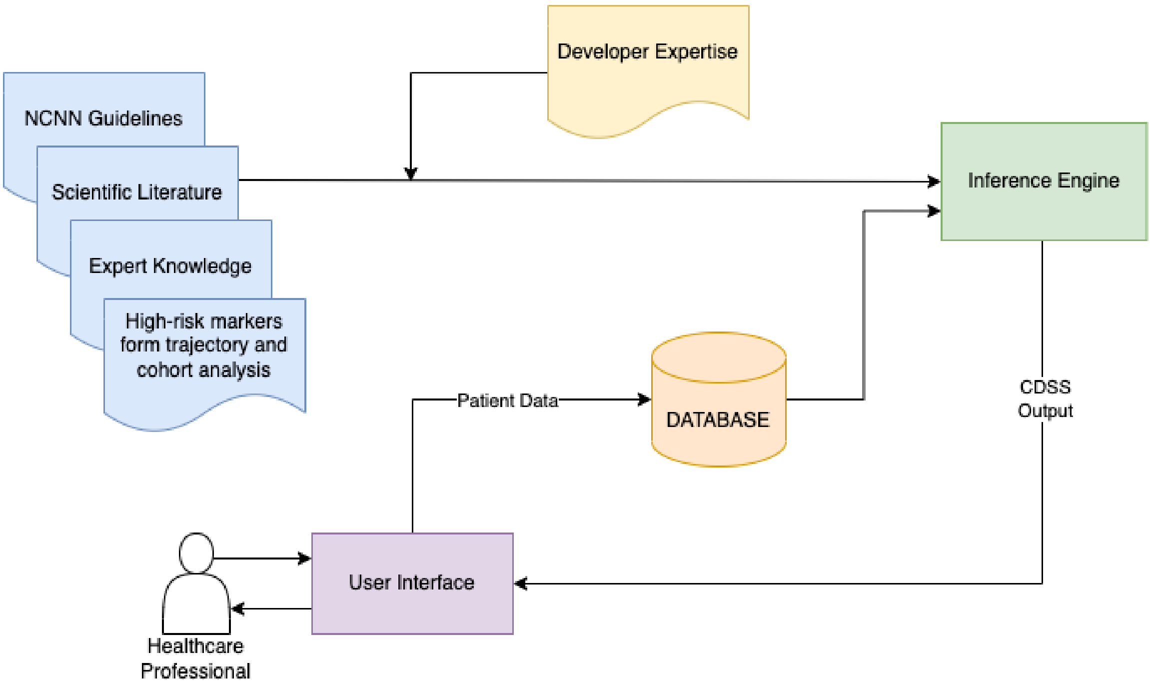

Figure 2 Overview of CDSS flow

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

3. Components of CDSS in PERSIST

3.1. Knowledge base

The knowledge base of a CDSS is a collection of medical knowledge and evidence-based guidelines that the CDSS uses to generate recommendations for healthcare professionals. This knowledge base typically includes information on diseases, treatments, diagnostic tests, and other relevant medical topics.

The knowledge base is typically structured in a way that allows the CDSS's inference engine to easily access and process the information it contains. This may involve using a specialized knowledge representation system, which encodes the knowledge in a format that can be easily processed by the CDSS's algorithms and machine learning models.

By constantly updating and expanding its knowledge base, a CDSS can help ensure that its recommendations are based on the most current and relevant medical evidence. This can be achieved in a few ways, first by manually updating the rules from feedbacks of medical experts or scientific literature. Second, by retraining the machine learning and artificial intelligence algorithms with.

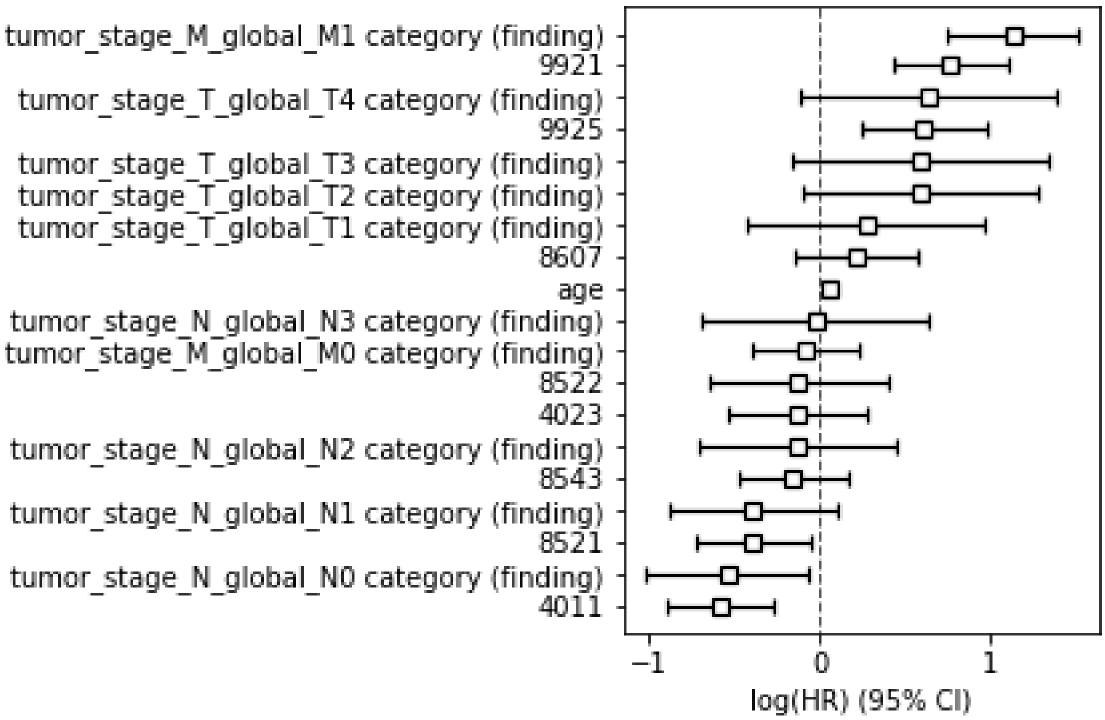

Figure 3 Cox Proportional Hazards for feature importance breast cancer patients

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

3.2. Inference Engine

The inference engine is the core component of a CDSS. It is responsible for analysing patient data and generating recommendations based on the medical knowledge and guidelines in the CDSS's knowledge base.

Typically, the inference engine uses a combination of algorithms and machine learning techniques to process the patient data and generate recommendations. For example, it may use rule-based algorithms to identify potential diagnoses based on the patient's symptoms, or it may use machine learning models to predict the likelihood of a certain condition based on the patient's medical history and test results.

The inference engine is designed to be fast and efficient, so that it can provide healthcare professionals with real-time access to relevant medical knowledge and guidelines. It is also designed to be flexible and adaptable, so that it can incorporate new knowledge and guidelines as they become available.

Overall, the inference engine is a key component of a CDSS, as it is responsible for providing healthcare professionals with the information, they need to make more accurate and evidence-based decisions about their patients' care.

3.3. User Interface

The user interface of CDSS in PERSIST is a mobile application designed to be easy to use and intuitive, so that healthcare professionals can quickly and easily access the outputs generated by the CDSS.

The user interface of CDSS in PERSIST is task-oriented and accessible, so that it can be used by healthcare professionals with different levels of technical expertise and experience. The interface of the mClinician mobile application is designed to support the work of healthcare professionals, by providing them with quick and easy access to the outputs generated by the CDSS, and the mClinician web application specialized to create patient data in cancer survivor context.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

3.4. Data Ingestion

In a healthcare context, data ingestion is the process of bringing patient data into a healthcare software system for further processing and analysis. This might involve capturing data from a variety of sources, such as electronic health records, medical devices, or other systems, and importing it into the software for storage and processing. Once the data is ingested, it can be used to support a variety of clinical and administrative processes, such as tracking patient health information, managing appointments and schedules, or generating reports and analytics. In PERSIST, we focused on developing tools and processes for efficient and effective data ingestion in a healthcare software system, with a focus on ensuring the accuracy and integrity of the data as it is imported into the system. This involved working closely with healthcare professionals to understand their needs and requirements, as well as developing algorithms and other tools to automate and optimize the data ingestion process.

Patient data is crucial for the effective operation of a CDSS. The CDSS uses this data to generate recommendations and advice for healthcare professionals, and the accuracy and relevance of these recommendations depend on the quality and completeness of the patient data. PERSIST contains different data ingestion sources, also algorithms and service outputs that becomes important factor on delivering more diversified recommendations.

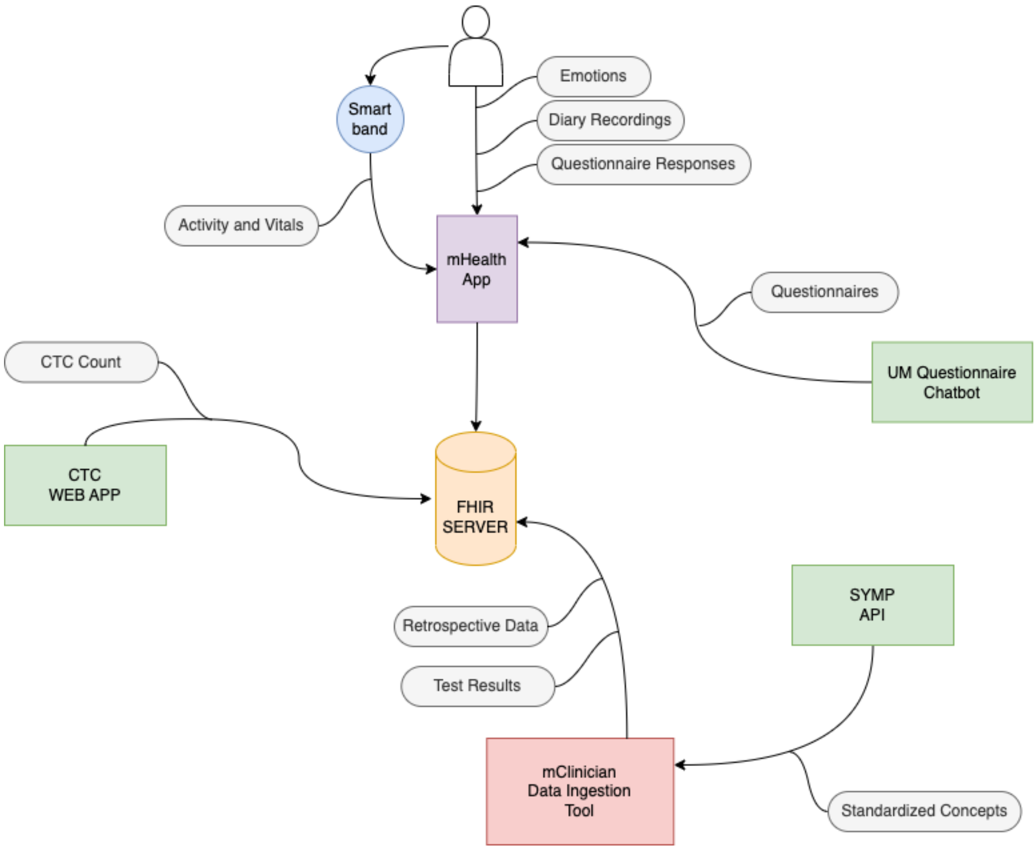

Figure 4 Components providing data in PERSIST

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

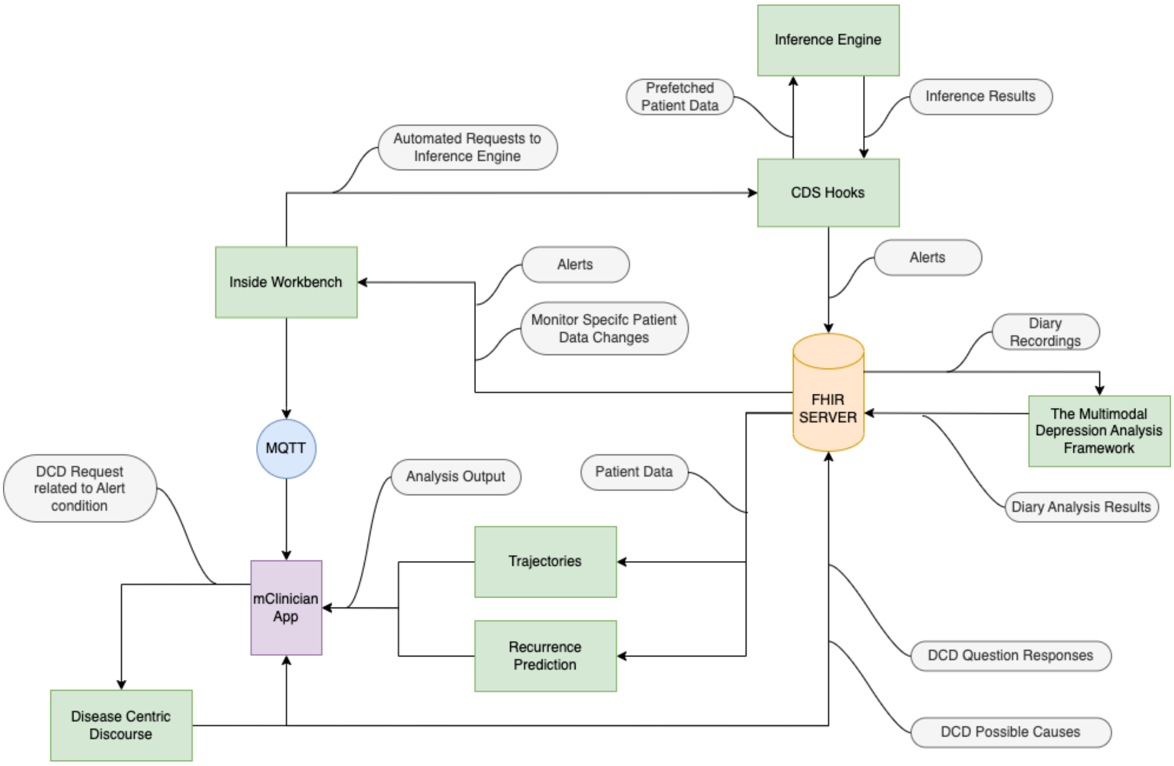

Figure 5 Data usage by CDSS in PERSIST

3.5. Risk Assessment

A rule-based clinical decision support service that assesses breast cancer recurrence, colon cancer recurrence, and cardiovascular disease risk for cancer survivors uses a set of predefined rules and algorithms for analysing patient data and generate output for healthcare professionals including negative and positive factors and risk level for cancer survivor patient.

The service is provided with the knowledge base for assessing these conditions in cancer survivors. The service also provided with patient data collected by multi-modal sensing network and retrospective data and test results recorded via mClinician data ingestion tool.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

3.6. Alert Mechanism

The OHC platform is designed to continuously monitor patient data. When new data is received, the platform determines whether to activate an alert mechanism. If the platform decides to run the alerts, it collects the necessary information and sends a request through the CDS Hooks API to the inference engine. The results are then returned to the OHC platform as JSON data in a CDS Card, which includes a Flag resource in FHIR format to save the alert, if one occurs. Alerts in PERSIST, using MQTT, enables efficient and reliable communication and notification in a variety of contexts, helping users to stay informed and respond quickly to important changes in patient data.

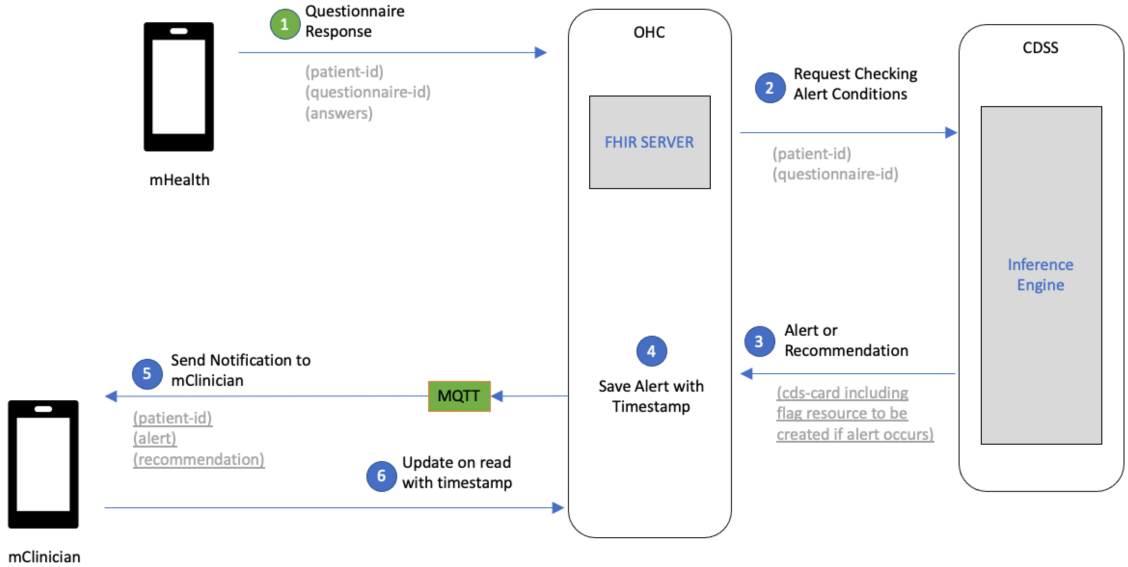

Figure 6 Alert mechanism request flow overview

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

3.7. Trajectories

Trajectory analysis involves studying the patterns of change or development over time for a particular group of individuals, such as patients with a certain condition. This analysis can help clinicians understand the factors that influence the progression of the condition and identify potential interventions that may improve outcomes for these individuals. In PERSIST trajectory analysis looks at factors such as the stage of the cancer at the time of diagnosis, TNM classification, the type of treatment the patients received, and the patients' overall health and lifestyle factors.

Trajectory and cohort analysis from T5.3 of WP5 outputs an API to access to the general and patient-specific results of the trajectories produced by machine learning models. The user interface of CDSS in PERSIST, mClinician, presents views to access these API endpoints and displays responses as charts to visualize trajectories.

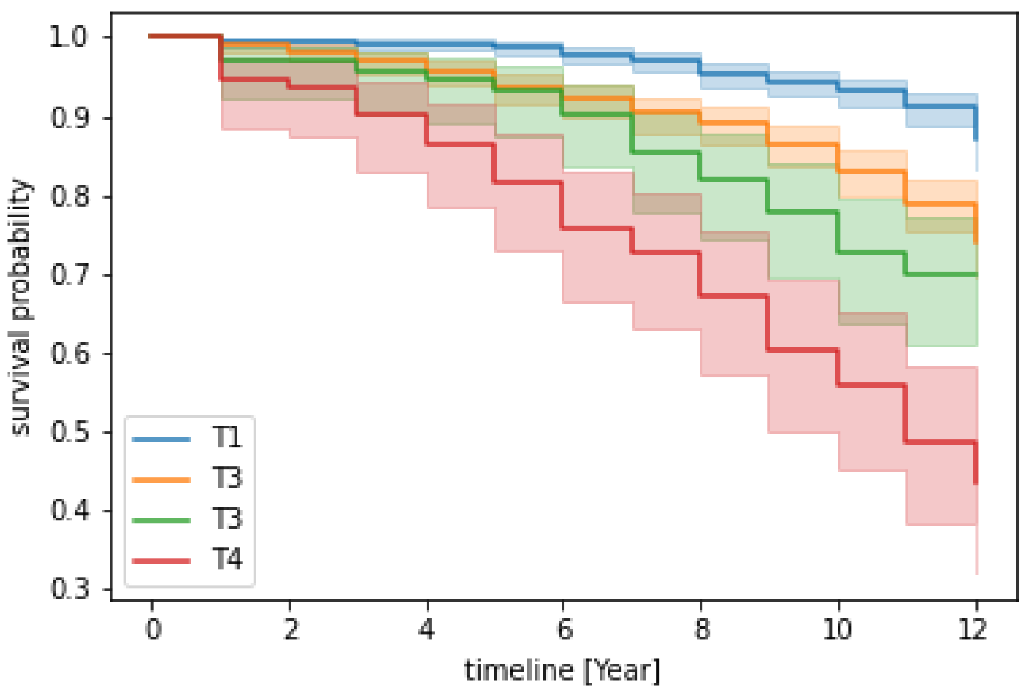

Figure 7 Kaplan-Meier breast cancer patient trajectory grouped by cancer stage T from TNM

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

3.8. Recurrence Prediction

An AI service that predicts the likelihood of relapse recurrence in the next five years for cancer survivors. It uses two separate prediction models for breast and colon cancer survivors.

The training data for the prediction model consists of 4,282 patients from CHU de Liége who were diagnosed with breast or colon cancer and were still alive five years after their diagnosis. The data includes general information about the patients, such as their gender, body mass index (BMI), smoking and drinking habits, and details about their tumor characteristics and hormone markers. The data also includes information about the patients' medical history, including their oldest and newest cancer diagnoses, any surgeries or other treatments they received, and any comorbidities they had.

For patients with breast cancer, the data includes information about comorbidities such as hypertension, diabetes, and obesity. For patients with colon cancer, the data includes information about comorbidities such as arrhythmia, valvular disease, hypertension, diabetes, renal failure, weight loss, and obesity.

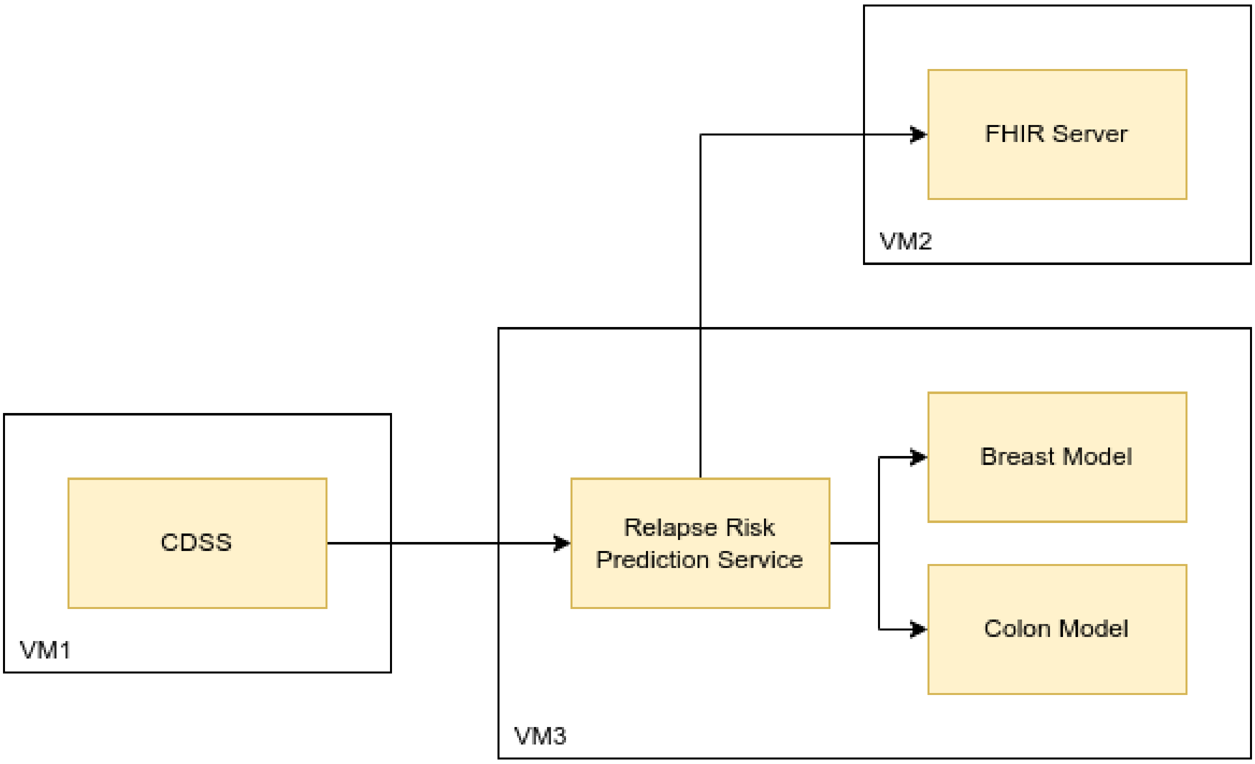

Figure 8 Overview of Recurrence Prediction service architecture

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

3.9. Disease Centric Discourse

Disease centric discourse is a dynamically triggerable, targeted patient reported outcome which send a series of questions according to a topic, and the given responses to the questions during the process, to get more detailed information. The concept and the specification of DCD is mentioned in Deliverable D4.2 Embodied conversational assistant and the implementation and communication details are described in Deliverable D.4.4.

From the CDSS perspective it provides further support to the clinicians upon an alert situation for a patient. Clinicians can choose to initiate DCD process by click of a button inside the current alert. After patient finishes the DCD process on his/her application. The responses are analysed and displayed as possible causes and patient responses to questions during DCD session.

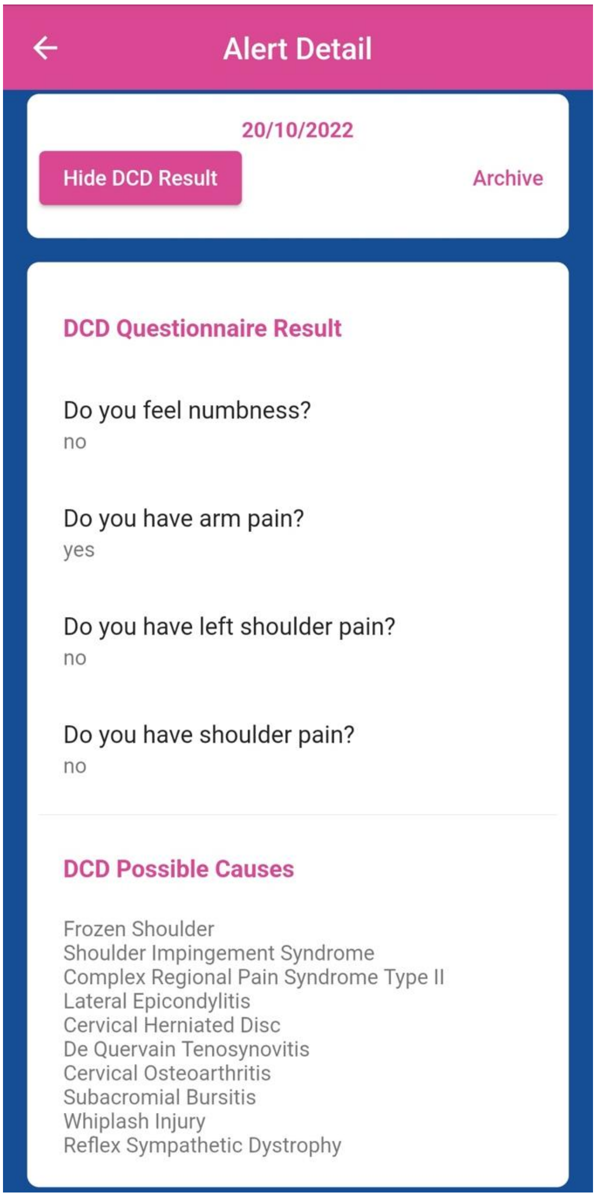

Figure 9 DCD Result displayed on mClinician application

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

3.10. Diary Analysis Results

PERSIST Depression Analysis Framework includes the multi-modal analysis of the patient video recordings. It mainly consists of those main steps: getting the patients video recordings, preparation for feature extraction, applying feature extraction algorithms and getting the prediction from machine learning algorithm.

After getting the patient videos as input to the framework, feature extraction algorithms for three different modalities (text, audio and facial) are applied. For audio feature extraction, audio from the video file is extracted as a first step. Then, this file is given as an input to the audio feature extraction algorithm. For audio feature extraction, transcription of the audio file is obtained by automatic speech recognition (ASR) which is mentioned in detail at Deliverable D4.3.

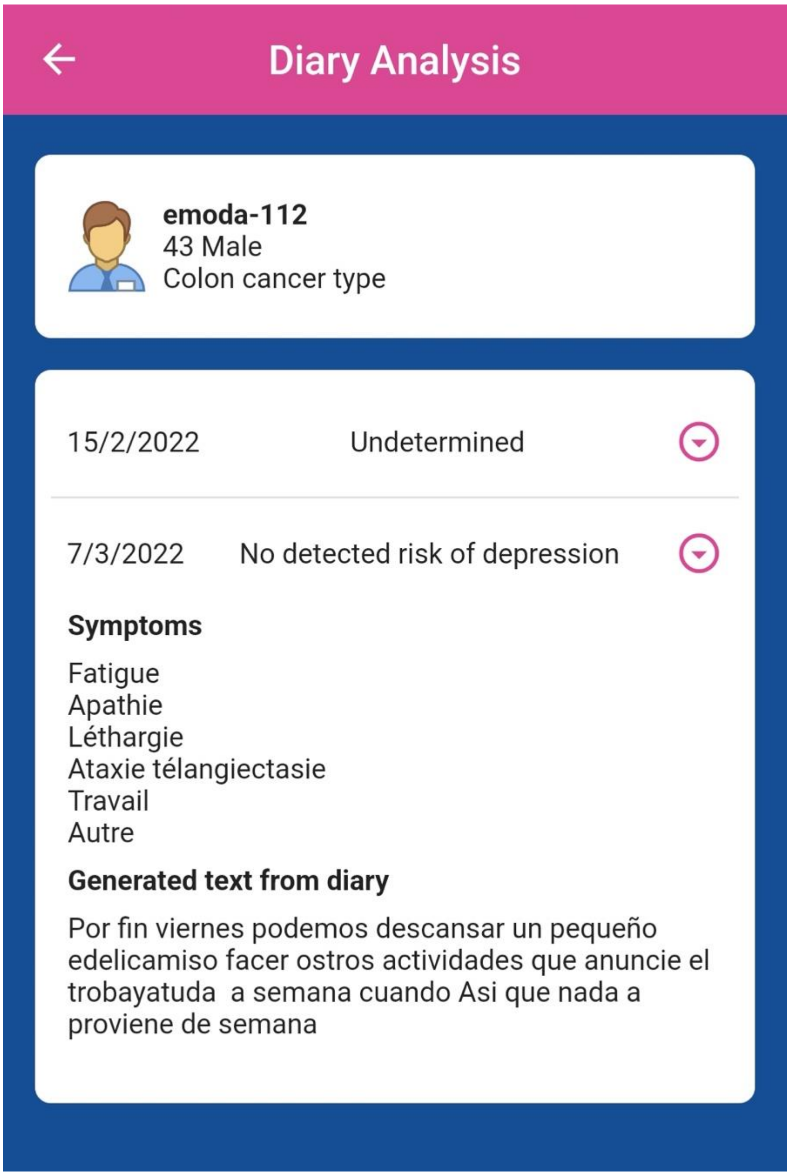

Figure 10 Diary analysis result displayed in mClinician application

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Evaluation of the CDSS

1. Overview of the evaluation process

Evaluation of CDSS is the process of determining the value of the system to improve quality of care and measure its effectiveness. Since different CDSSs serve different purposes, there is no general measure for all these systems; however, features such as consistency can be determined through top-down approaches. The evaluation criterion for a CDSS depends on the purpose of the system, for example alert mechanism can be evaluated based on the consistency and precision, but for recurrence prediction and trajectory services cannot be evaluated with the actual patient data collected in PERSIST due to time needed for the conclusion of their period.

Specific test plans including cases and expected behaviour were created and executed to evaluate the outputs of the CDSS. After the launch of mClinician UI, the users, in this case clinical partners in PERSIST, begin using and providing feedback for the system. Some changes regarding these feedbacks have also been implemented.

2. Evaluation of Risk Assessment Services

To select most appropriate test cases to evaluate the performance and accuracy of the risk assessment services of CDSS, the first step was to understand the criteria and factors that the solution uses to assess the risk of breast cancer recurrence, colon cancer recurrence and cardiovascular disease. This includes the information about the patient's medical history, current health status, and other factors that could affect their risk of cancer recurrence. Once there was a clear understanding of these factors, we selected a set of test cases that represent a range of different scenarios and conditions, to evaluate the performance of the solution under a variety of different circumstances. This might include cases where the patient has a high risk of cancer recurrence or cardiovascular disease, as well as cases where the risk is lower. By carefully selecting and evaluating a range of test cases we were able to determine the effectiveness of the solution in assessing the risks.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

2.1. Breast Cancer Recurrence Risk Assessment Service

When selecting test cases to evaluate the solution that assesses breast cancer recurrence risk, it was important to choose cases that are representative of the various factors that can affect the risk of recurrence.

Our approach was to focus on hypothetical patients specifically at higher or lower risk of breast cancer recurrence. For example, test cases that include patients who are known to have a BRCA1 mutation, as this is a known risk factor for breast cancer. Also, we included test cases that include patients with specific cancer subtypes. By including test cases that are representative of these subgroups, we were able to evaluate how well the solution performs for patients who are known to be at higher or lower risk of recurrence.

- *Test Patient I – emoda-124*
  A 50-year-old woman with a BMI of 25, who consumes alcohol regularly, has a CA 15-3 level of 19, and is a current smoker. She has family history of breat cancer. She has been diagnosed with stage 3 breast cancer and has a tumor size of 2 cm and has tested positive for estrogen and progesterone receptors. This case would represent a high-risk scenario, with several factors that are known to increase the risk of breast cancer recurrence.

- *Test Patient II – emoda-125*
  A 65-year-old woman with a BMI of 22, who does not consume alcohol or smoke, and has a CA 15-3 level of 13. She has been diagnosed with stage 1 breast cancer and has a tumor size of 1 cm and has tested negative for estrogen and progesterone receptors. This case would represent a low-risk scenario, with several factors that are known to decrease the risk of breast cancer recurrence.

- *Test Patient III – emoda-126*
  A 40-year-old woman with a BMI of 28, who consumes alcohol occasionally, has a CA 15-3 level of 17, and is a former smoker. She has been diagnosed with stage 2 breast cancer and has a tumor size of 1.5 cm and has tested positive for estrogen and progesterone receptors. Patient also has BRCA1 positive. This case would represent a moderate-risk scenario, with some factors that increase the risk of breast cancer recurrence and others that decrease the risk.

- *Test Patient IV – emoda-127*
  A 35-year-old woman with a BMI of 21, who does not consume alcohol or smoke, and has a CA 15-3 level of 15. She has been diagnosed with stage 3 breast cancer and has a tumor size of 3 cm and has tested negative for estrogen and progesterone receptors. This case would represent a high-risk scenario, with the advanced stage of cancer and the large tumor size increasing the risk of recurrence.

- *Test Patient V – emoda-128*
  A 60-year-old man with a BMI of 25, who consumes alcohol regularly, has a CA 15-3 level of 16, and is a current smoker. He has been diagnosed with stage 2 breast cancer and has a tumor size of 1.5 cm, and has tested positive for estrogen and progesterone receptors. This case would represent a moderate-risk scenario, with some factors that increase the risk of breast cancer recurrence and others that decrease the risk.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Test Patient	Expected Risk	Calculated Risk	Improvements	Calculated Risk After Improvements
I	High	High	Rearranged threshold levels for CA 15-3 and BMI level effects	High
II	Low	Medium		Low
III	High	High		High
IV	High	High		High
V	Medium	Medium		Medium

Table 1 Test case results for breast cancer recurrence risk assessment service

The solution resulted with most of the expected results in test results, after inspecting the test results we found only one test with minor deflection due to effects of levels of BMI and CA 15-3 slightly reworked to provide more accurate recommendation.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

2.2. Colon Cancer Recurrence Risk Assessment Service

Having the parameters age, CEA, smoking, alcohol, physical activity, diabetes mellitus, BMI, sleeping, and C-reactive protein level and certain medical tests and biomarkers (such as CEA and CA 19-9) can provide important information about a patient's current health status and their risk of colon cancer recurrence.

It was important to include a mix of patients with different risk factors for colon cancer recurrence, as well as patients who have undergone different treatment methods. Additionally, with detailed medical records and information collected during the project, by using a diverse set of test cases and carefully analysing the patient data, we were able to evaluate the solution's performance in determining colon cancer recurrence risk.

- *Test Patient I*
  A 65-year-old non-smoker with a family history of colon cancer who consumes alcohol regularly, has a high C-reactive protein level and was diagnosed with stage III colon cancer.
  Expected result is High risk.

- *Test Patient II*
  A 45-year-old female patient with no family history of colon cancer and a healthy lifestyle, who was diagnosed with stage I colon cancer, has a normal BMI and no history of diabetes mellitus.
  Expected result is Low risk.

- *Test Patient III*
  A 55-year-old smoker with a high BMI and a history of diabetes mellitus, a family history of colon cancer and a history of smoking, who was diagnosed with stage IV colon cancer.
  Expected result is High risk

- *Test Patient VI*
  A 32-year-old female patient with no family history of colon cancer and BMI value 25 kg/m^2, who was diagnosed with stage II colon cancer and has a high level of CA 19-9 value.
  Expected result is Medium risk

- *Test Patient V*
  A 75-year-old smoker with a low BMI and a high level of CA 19-9 who was diagnosed with stage II colon cancer
  Expected result is High risk

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

Test Patient	Expected Risk	Calculated Risk
I	High	High
II	Low	Low
III	High	High
IV	Medium	Medium
V	High	High

Table 2 Test case results for colon cancer recurrence risk assessment service

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

2.3. Cardiovascular Disease Risk Assessment Service

The numerous elements that may affect a patient's risk of cardiovascular disease must be represented in the test cases that are considered, and this is essential. We focused on subgroups of hypothetical patients with known higher or lower cardiovascular disease risk. Choosing test patients with a history of cardiovascular disease or stroke, for instance, as these conditions are recognized risk factors for cardiovascular disease. Also including test cases who had certain treatments, such as chemotherapy or radiation therapy, which can raise the risk of cardiovascular disease. It is assessed how effectively the solution works for patients who are known to be at higher or lower risk by incorporating test cases that are typical of various subgroups.

- *Test Patient I*
  A 45-year-old female breast cancer survivor with a high BMI, a history of hypertension, and a low physical activity tolerance, who is taking multiple medications to manage her conditions.
  This test case expected to output high risk.

- *Test Patient II*
  A 60-year-old male colon cancer survivor with a history of high blood pressure and heart rate, who is taking medication to manage his conditions but has a high physical activity tolerance.
  This test case expected to output high risk.

- *Test Patient III*
  A 50-year-old female breast cancer survivor with a normal BMI, no history of hypertension or heart problems, and a high physical activity tolerance, who is not taking any medications.
  This test case expected to output low risk.

- *Test Patient IV*
  A 70-year-old male colon cancer survivor with a high BMI, a history of high blood pressure and heart rate, and a low physical activity tolerance, who is taking multiple medications to manage his conditions.
  This test case expected to output high risk

- *Test Patient V*
  A 40-year-old female breast cancer survivor with a normal BMI, no history of hypertension or heart problems, and a high physical activity tolerance, who is taking medication to manage her condition.
  This test case expected to output medium risk

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Test Patient	Expected Risk	Calculated Risk
I	High	High
II	Low	Low
III	High	High
IV	Medium	Medium
V	High	High

;

Table 3 Test case results for cardiovascular disease risk assessment service

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

3. Evaluation of Alert Mechanism

The Alert Mechanism consists of several services concerned about responses to different questionnaires each. The evaluation of this solutions was about to evaluate triggering and delivering alerts, rather than specifically testing if the questionnaires can be responded.

One part of the evaluation was selecting test cases that include a variety of different scenarios, such as answers that change the output of the inference engine for that scenario. For each test case, input the answers to the questionnaire and verify that the solution produces the expected output which in this case was triggering or not triggering the alert.

The second part of the evaluation was to conduct user testing, where a group of representative users are asked to use the solution and provide feedback on receiving notifications on alert situations and their overall experience about the feature. This was also aiming to provide valuable insights into the solution's usability, user experience, and overall effectiveness.

3.1. Test Scenarios for alert trigger

No	Test Case	Expected Result	Actual Output
	Response to Rectal bleeding part of Colorectal cancer interval questionnaire		
1	Question d with “No”	Triggers Alert	Alert triggered containing the question and response
2	Question e with “No”	Recommendation	Text with recommendation
3	Question e with “Yes”	Triggers Alert	Alert triggered containing the question and response
	Response to Changes in feces coloration part of Colorectal cancer interval questionnaire		
4	Question c with “Yes”	Triggers Alert	Alert triggered containing the question and response
5	Question d with “Yes”	Triggers Alert	Alert triggered containing the question and response
6	Question e with “Yes”	Triggers Alert	Alert triggered containing the question and response
7	Question e with “No”	Recommendation	Text with recommendation
	Response to Change in bowel habits part of Colorectal cancer interval questionnaire		
8	Question a with “No”	Recommendation	Text with recommendation

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

9	Question b with “Yes”	Recommendation	Text with recommendation
10	Question c with “Yes”	Triggers Alert	Alert triggered containing the question and response
11	Question d with “Yes”	Triggers Alert	Alert triggered containing the question and response
12	Question d with “No”	Recommendation	Text with recommendation
	Response to Breast cancer interval questionnaire		
13	Question Breast lumps with “Yes”	Triggers Alert	Alert triggered containing the question and response
14	Question Skin retraction with “Yes”	Triggers Alert	Alert triggered containing the question and response
15	Question Nipple discharge with “Yes”	Triggers Alert	Alert triggered containing the question and response
	Response to Cardiovascular risk questionnaire		
16	2nd Question with answer “YES”	Triggers Alert	Alert triggered containing the question and response
17	3rd Question with answer “YES”	Triggers Alert	Alert triggered containing the question and response
	2nd or 3rd Question with “YES” and		
18	Chemotherapy treatment based on specific drugs.	Adds risk context to alert	Triggered alert contains related information
19	Diabetes or high blood pressure or hyperlipidemia.	Adds risk context to alert	Triggered alert contains related information
20	BMI > 30	Adds risk context to alert	Triggered alert contains related information
21	Smoking	Adds risk context to alert	Triggered alert contains related information
	Lymphedema questionnaire with “Yes” response to any of the questions 1 to 4		
22	and “Yes” to the lymphangitis questions 5 or 6	Triggers Alert	Alert triggered containing the question and response
23	Once and “No” to the lymphangitis questions 5 or 6.	No Alert	None
24	Second time	Triggers Alert	Alert triggered containing the question and response
	Response to Pain questionnaire		

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

25	With "Yes" to 1st question and rating greater than 4 to the 2nd question	Triggers Alert	Alert triggered containing the question and response
	Hormonal disbalances and when patient is female		
26	With "Yes" response to any question	Triggers Alert	Alert triggered containing the question and response
	Response to Sexual dysfunction questionnaire		
27	With "No" to 1st question or "Yes" to 2nd question	Triggers Alert	Alert triggered containing the question and response
	Response to Fatigue questionnaire		
28	With "Yes" to 1st or 2nd , or rating greater than 6 to 3rd question	Triggers Alert	Alert triggered containing the question and response
	Psychosocial status: anxiety questionnaire		
29	With GAD-7 scoring > 8	Triggers Alert	Alert triggered containing the question and response
	Response to PHQ-2		
30	With Score > 2	Triggers Alert	Alert triggered containing the question and response
	Response to Cognitive function questionnaire		
30	With "Yes" to any question	Triggers Alert	Alert triggered containing the question and response
	Weekly response to Malnutrition questionnaire with "No" to intake of nutrition / healthy nutrition / Fluid intake		
30	Two weeks in a row	No Action	None
31	Three subsequent weeks	Triggers Alert	Alert triggered containing the question and response
32	4 out of up to 9 weeks	No Action	None
33	5 out of 10 weeks	Triggers Alert	Alert triggered containing the question and response
	Weight evolution in 3 months		
34	is more than 5%	Triggers Alert	Alert triggered containing the question and response
	Gastrointestinal conditions questionnaire		

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

35	"Yes" to Swallowing issues	Triggers Alert	Alert triggered containing the question and response
	WHO questionnaire: 12-item Instrument Scoring Sheet "at least severe" for questions S1 to S12		
36	Two weeks in a row	No Action	None
37	Three subsequent weeks	Triggers Alert	Alert triggered containing the question and response
38	4 out of up to 9 weeks	No Action	None
39	5 out of 10 weeks	Triggers Alert	Alert triggered containing the question and response
	WHO questionnaire: 12-item Instrument Scoring Sheet "more than 10 days" for questions H1 to H3		
36	Two weeks in a row	No Action	None
37	Three subsequent weeks	Triggers Alert	Alert triggered containing the question and response
39	8 out of 16 weeks	Triggers Alert	Alert triggered containing the question and response

Table 4 Test cases for alert mechanism

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

3.2. User Testing

As part of the general user interface tests, users were also able to check notifications arriving to their mobile application upon a triggered alert. The feature was tested for the metrics accessibility, accuracy and arriving duration of the notifications. Accuracy measures the percentage of the alerts received as notifications through MQTT. Accessibility measures the percentage of the alerts accessed through the mobile device’s notifications interface and redirected to the correct patient inside the application and arriving duration is the average duration between the alert trigger and arrival of the notification to the user.

Test Device/User	Total number of triggered alerts tested	Accuracy	Accessibility	Time
1	60	96,66%	98,27%	2,3 mins
2	44	97,77%	100%	2,1 mins
3	24	100%	100%	1,9 mins
4	65	95,38%	98,38%	2,7 mins
5	52	100%	98,07%	2,1 mins
6	30	96,66%	100%	2,5 mins
Average	45,83	97,74%	99,12%	2,26 mins

Table 5 Statistical results of user tests for the alert mechanism

The results show that triggered alerts are sent as notifications and on the receiving and notification can be engaged with high ratios, the decrease in accuracy may be related to listening the MQTT server or due to some malfunctioning while receiving or sending the message and should be improved by introducing counter mechanisms such as message queues and acknowledgements to resent failed messages. For accessibility, most of irregular cases are because of the device not handling the user interaction, in all of the cases the app did not open at all but it always redirected the user to the patient related to alert successfully. Time between the alert trigger condition happening and notification arrival is due to the interval time of checking user reported measures to be saved and sent to inference engine and the inference engine response leading to triggering the alert through MQTT.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

4. Evaluation of integrated services

4.1. Evaluation of Recurrence Prediction

A hold-out strategy is a common approach to evaluating the performance of a machine learning model. In this approach, the dataset is split into two subsets: a training set, which is used to train the model, and a test set, which is used to evaluate the performance of the trained model. In the case of the AI service mentioned in the paragraph, 90% of the data is used for training and 10% is used for final testing. This means that the model is trained using 90% of the available data, and then its performance is evaluated using the remaining 10%.

Hold-out validation can be a useful approach for evaluating the performance of a machine learning model, but it has some limitations. One limitation is that it only uses a single split of the data into training and test sets, which means that the evaluation may not be representative of the model's performance on other data. To address this limitation, the AI service mentioned in the paragraph also uses k-fold cross-validation on the training dataset.

K-fold cross-validation is a method for evaluating the performance of a machine learning model that involves dividing the dataset into k equal-sized subsets (or "folds"), training the model on k-1 subsets, and then evaluating the performance of the model on the remaining subset. This process is repeated k times, with each fold serving as the test set in turn. In the case of the AI service mentioned in the paragraph, k=6, meaning that the data is split into six equal-sized subsets and the model is trained and evaluated six times. Each time, 75% of the data is used for training and 15% is used for validation.

Hyperparameter tuning is the process of adjusting the values of the hyperparameters of a machine learning model to optimize its performance. In the case of the AI service mentioned in the paragraph, grid-search is used for hyperparameter tuning. Grid-search is an exhaustive search algorithm that considers all possible combinations of hyperparameter values within a specified range. This allows the algorithm to identify the combination of values that produces the best performance on the training data.

As a result, the AI service described in the paragraph uses a combination of hold-out validation, k-fold cross-validation, and grid-search to evaluate and optimize the performance of its prediction models for breast and colon cancer survivors. This approach allows the service to provide accurate predictions of the likelihood of relapse recurrence in the next five years for cancer survivors.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Model	F1-weighted	accuracy	precision-weighted	recall-weighted	roc_auc
Deep Neural Network	0.708	0.647	0.822	0.647	0.567
XGBoosting	0.731	0.676	0.827	0.676	0.525
Gradient Boosting	**0.774**	**0.735**	0.838	**0.735**	0.533
Decision Tree	0.730	0.706	0.756	0.706	0.433
Logistic Regression	0.660	0.588	**0.855**	0.588	**0.683**

Table 6 Evaluation outputs of colon cancer model of Recurrence Prediction

Model	F1-weighted	accuracy	precision-weighted	recall-weighted	roc_auc
Deep Neural Network	0.869	0.855	0.893	0.855	0.800
XGBoosting	**0.920**	**0.928**	**0.923**	**0.928**	**0.844**
Gradient Boosting	0.907	0.904	0.913	0.904	0.796
Decision Tree	0.861	0.855	0.868	0.855	0.703
Logistic Regression	0.820	0.795	0.863	0.795	0.718

Table 7 Evaluation results of breast cancer model of Recurrence Prediction

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

5. Evaluation of integrated services

5.1. Evaluation of DCD integration

To evaluate the DCD, the functionality of the integration has been considered. Since the DCD starts on demand of the clinician user, for further investigation upon an alert, the expected behaviours can be listed.

No	Given	When	Then
1	An alert on a patient	Clinician user clicks "Start DCD" button	Patient gets a notification to start the DCD Questionnaire
2	An alert with a related DCD process completed	Clinician user clicks "Show DCD Result" button	UI displays the questions answered by the patient in
			UI displays the possible causes to that alert

Table 8 Test cases for the DCD integration

As a result integrated service tested and works as expected.

5.2. Evaluation of Diary Analysis integration

To evaluate the diary analysis, the functionality of the integration has been considered. Diaries are analysed automatically in a queued fashion, if there is an analysis result it is already displayed on the Diary Analysis page of the patient detail view. If the diary has not been analysed yet, clinician can request the diary to be analysed. When this request arrives to the PERSIST Depression Analysis Framework the diary video begins to be analysed asynchronously.

No	Given	When	Then
1	An already analysed diaries on the list	Clinician taps on the diary item on the list	Symptoms extracted by the service and generated text is displayed
2	A not yet analysed diary on the list	Clinician taps on the diary item on the list	A "Request Analysis" button reveals
3	A not yet analysed diary on the list with revealed "Request Analysis" button	Clinician taps on the button	Request to asynchronously analyse the diary recording is sent to PERSIST Depression Analysis Framework

Table 9 Test case for the diary analysis integration

As a result integrated service tested and works as expected.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

6. Evaluation of User Interface

The user interface is tested for functionality and integration. For functionality, the interface tested to see if the features and tools are working as expected, and for integration we tested to see if any services integrated through CDS Hooks can be accessed without the need for updating the application. The tests are handled by users and the bugs and misfunctions pointed out by the testing users are issues and worked during the evolution.

No	Feedback / Issue	Category	Solution
1	Trends view chart font size issue on devices using larger font-size as default	Visual	Enforce device to use given font-size and increase the intervals display on chart for both axis
2	Displaying long names for the abbreviations create crowd it should be simplified by using only abbreviations	Visual	Removed long names when possible
3	Patient trajectory chart labels are duplicated and hard to understand	Functional	The labels were used directly from the response of the trajectory service. Service updated to fix the issue.
4	Recurrence risk services is displaying both cancer types should be related to patient cancer type	Functional	Filtered the services according to patients' cancer type
5	Text misspellings in CDSS output	Visual	Corrected spellings
6	Patient 65 years old, in his "Negative Effects" appears 30<Age<35 in CDSS output	Functional	Inference engine misinterprets the FHIR resource, the issue has been fixed.
7	Some laboratory test value units are missing, and the displayed text has a missing space	Visual	Corrected missing values and
8	Patient list is not displaying any patients	Functional	It was due to an error on server side and resolved.

Table 10 Reported issues and feedbacks from user functionality tests between version releases

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Conclusions

In this deliverable we presented the summary of the integration into the CDSS in PERSIST and how the clinical decision support is evaluated with its components, and the integrated services.

The evaluation showed how patient data is crucial for the effective operation of a CDSS. The data collection in PERSIST shown its strength by being able to collect data from various sources as well as further analyse the data recorded with AI and ML techniques to create additional information.

The importance of patient data for a CDSS is twofold. First, the data provides the basis for the recommendations generated by the CDSS. Without accurate and complete patient data, the CDSS may not be able to provide accurate or useful recommendations.

Second, the data can be used to improve the performance of the CDSS over time. By constantly updating and refining the patient data in its database, a CDSS can improve its ability to generate accurate and relevant recommendations. This can help healthcare professionals make more informed and evidence-based decisions about their patients' care.

In conclusion, the CDSS system has demonstrated strong performance in our evaluation process. The system was able to accurately process and analyse large amounts of clinical data, and generate appropriate recommendations based on established guidelines and protocols. The user interface was intuitive and easy to use, allowing clinicians to quickly access and incorporate the system's recommendations into their decision-making process. By leveraging the latest advances in AI and machine learning, the system is able to extract relevant features from patient records and video data, generate accurate and timely recommendations.

Additionally, the system showed a good level of scalability, with no significant degradation in performance even when handling large volumes of data. Overall, the CDSS evaluation has shown the system to be a valuable tool for improving clinical decision making and patient care and also shown the value of incorporating AI and machine learning technology into our healthcare setting.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

References

- Flutter for Mobile Development, Web Page: https://flutter.dev/
- HL7 FHIR Observation, Web Page: https://www.hl7.org/fhir/observation.html
- HL7 FHIR Appointment, Web Page: https://www.hl7.org/fhir/appointment.html
- HL7 FHIR Questionnaire Response, Web Page: https://www.hl7.org/fhir/questionnaireresponse.html
- HL7 FHIR Diagnostic Report, Web Page: https://www.hl7.org/fhir/diagnosticreport.html
- HL7 FHIR Patient, Web Page: http://www.hl7.org/fhir/patient.html
- HL7 FHIR Document Reference, Web Page: https://www.hl7.org/fhir/documentreference.html
- MQTT Protocol, Web Page: https://mqtt.org/
- MQTT Introduction, Web Page: https://www.hivemq.com/blog/mqtt-essentials-part-1-introducing-mqtt/
- PERSIST Project Deliverable, D2.5 PERSIST overall Big Data architecture
- PERSIST Project Deliverable, D4.4 Beta version of Multimodal Sensing Network
- PERSIST Project Deliverable, D4.5 Full version of PERSIST mHealth Application
- PERSIST Project Deliverable, D5.4 High risk markers for patient stratification
- PERSIST Project Deliverable, D5.5 Clinical decision support system
- PERSIST Project Deliverable, D5.1. CDSS requirements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406