HORIZON 2020
The EU Framework Programme for Research and Innovation

CAPABLE
CAncer PAtients Better Life Experience

Grant Agreement No. 875052
Start Date: 01/01/2020 (48 Months)

Deliverable No. 5.2

Framework Defined (Including Patients' Needs) Based on Available Data and Modelling Approaches

Due Date: [30/06/2021]
Submitted On: [02/07/2021]

Coordinator	University of Pavia (UNIPV)
Deliverable Lead Partner	IBM ISRAEL
Contributing Partners	
Contact	Prof. Silvana Quaglini
Email	silvana.quaglini@unipv.it
Website	www.capable-project.eu

Deliverable Type		
R	Document, report	**X**
DEM	Demonstrator, pilot, prototype	
DEC	Websites, patent filings, videos etc.	
OTHER		
Dissemination Level		
PU	Public	**X**
CO	Confidential (Consortium members including the Commission Services)	
CI	Classified Information (Commission Decision 2015/444/EC)	

This deliverable is a part of a project receiving funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875052.

Table of Contents

List of Figures

List of Tables

1. Versions History

Version	Date	Author	Comments
1.0	15/05/2021	Ella Barkan	Initial document
1.01	31/05/2021	Flora Gilboa-Solomon	Executive Summary and some modifications
1.02	06/06/2021	Ronald Cornet	Internal Review Comments
1.03	15/06/2021	Flora Gilboa-Solomon, Ella Barkan	Updates after review
1.04	22/06/2021	Szymon Wilk	Revision of Section 5.2 (figures and text)
1.05	24/06/2021	Ella Barkan, Flora Gilboa-Solomon, Szymon Wilk	Final Version

2. Executive Summary

The aim of this deliverable is to describe a framework which is part of the CAPABLE system architecture. A **framework** is defined as a set of concepts, libraries, tools, practices, and analytic methodologies. In this document we focus on the artificial intelligence (AI) framework, while we consider AI in a broad notion that covers formal knowledge representation and logic-based reasoning, as well as machine learning techniques (Russel, 2016). This document describes the means in which the AI framework supports the requirements of melanoma and renal cancer patients, their caregivers and physicians.

Section 3 provides a general introduction to the CAPABLE overall concept, along with the different functionalities and the interfaces with its users.

Section 4 presents an overview of CAPABLE's users requirements gathered through interviews of patients and physicians from the Netherlands Cancer Institute (NKI), the Instituti Clinici Scientifici Maugeri hospital (ICSM) and the Italian Cancer Patient Organization "Associazione Italiana Malati di Cancro, Parenti e Amici" (AIMAC). The interviews revealed that patients are interested in information about their diagnosis, treatments, side-effects, nutrition, physical activity and sleep. They expect evidence-based feedback and advice on symptom management following their symptom reports. They see the value in early monitoring of their well-being and would like to receive information on their supportive care options (nutritional, psychological and rehabilitation support) provided by their hospital as well as interventions such as mindfulness, creative therapy, relaxation and physical activities. Physicians are interested in remote patient monitoring, the progress of disease, the occurrence of acute moderate/severe symptoms and side-effects, and the overall wellbeing of patients.

Section 5 provides the AI framework components overview which are developed based on two main sources:

1. Real World Data (RWD): Patient data coming from Electronic Health Records (EHR), sensors, and questionnaires, including patient-reported outcomes (PROs).
2. Knowledge captured in scientific publications, clinical guidelines, standard terminologies and educational materials.

Based on these types of data sources, the AI framework encompasses two groups of approaches to provide either **knowledge-driven** or **data-driven** models. For example, data-driven models include risk predictive models based on retrospective data while knowledge-driven models include logical guideline-based reasoning.

Section 6 details the RWD and knowledge sources used to develop the framework components.

Section 7 describes each one of the framework components. These can be grouped into two categories:

1. **Low-level (back-end)** components include: **knowledge-driven** models such as plan models, ontological models and rule-based/algorithmic models and **data-driven** models such as population-based risk predictive models and personalized predictive models, as well as statistical summaries on data.
2. **High-level (closer to front-end)** components: they deliver support to the physician and patient, and offer hybrid support that combines knowledge and data-driven techniques.

In addition, this section describes the **Auxiliary components** that support both high-and low level-components in inter-component communication, managing patients' data and processing flow as well as interaction with framework users.

3. CAPABLE Concept

CAPABLE provides means of remotely acquiring measurements from the patients or caregivers, either actively using questionnaires or allowing self-reporting symptoms or passively, using wearable or environmental sensors. Figure 3.1 presents the CAPABLE concept. The data sent by the patients will be collected and merged with the data available in the Electronic Health Record (EHR) of the treating institutions.

CAPABLE includes AI-based analytic methodologies. Some of them are based on real world patient data, like the predictive models and some are based on knowledge sources as physicians' guidelines. CAPABLE will continuously and automatically scan all that information looking for criticalities and providing interpretations for clinicians as well as coaching and motivation to patients in adopting life-style changes and improving adherence to treatment.

Specifically, for patients, the CAPABLE mobile application aims to support and improve physical, mental, and social well-being. The system will consist of symptom reporting functionalities including immediate feedback to patients, evidence-based patient information, and evidence-based non-pharmacological interventions. For physicians, the CAPABLE web-based dashboard will support care by presenting the patient-reported symptom and well-being outcomes. It will be possible to manage a list of patients informative of patients' current health status. The dashboard will provide decision support based on those patient-reported symptoms for symptom management, supported by clinical practice guidelines.

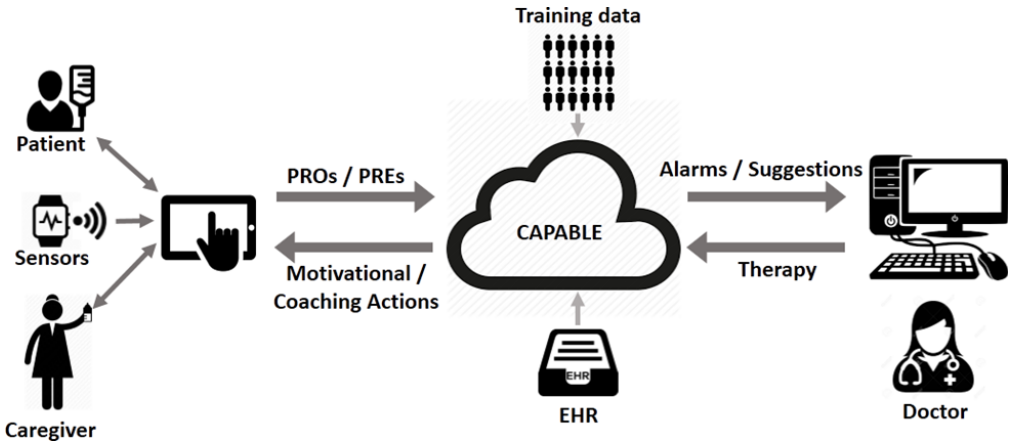

Figure 3.1. The CAPABLE Concept

4. Users' Needs

The framework aims at providing support for both patients and physicians. The following sections provide details of these users' needs.

4.1. Patients' Needs

Patients' needs from CAPABLE patient app were collected previously via interviews of patients from NKI, ICSM and AIMAC. These patient needs (PaNs) can be summarized as follows:

- **Pa-N1** Patients indicated varying levels of satisfaction with the information provision currently provided by their respective healthcare professionals. Patients are interested in a wide range of topics, including their diagnosis, treatments, side-effects, nutrition, physical activity and sleep. Furthermore, patients are interested in interventions such as mindfulness, creative therapy, relaxation and physical activities.
    - The CAPABLE system should support patients by presenting information directly to patients or by referring to existing information sources.
- **Pa-N2** Patients experience barriers in reporting symptoms to their care team, due to perceiving their call as intrusive to their clinician, or because they are not sure if the symptom should be a cause of concern. Patients see opportunities in reporting symptoms remotely via a mobile application. The primary benefit of using such an application would be to receive feedback and advice regarding symptom management. Patients indicated that both computer-based feedback or feedback from their clinician would be satisfactory. Patients expected their motivation to report symptoms to rapidly decline without receiving any feedback.
    - The CAPABLE mobile application must allow for patients to report their symptoms remotely.
    - The CAPABLE mobile application must provide feedback and advice on symptom management to patients. Patients expect evidence-based feedback and advice on symptom management following their symptom reports.
- **Pa-N3** Patients indicated not being informed or not being informed in a timely manner of the supportive care options (nutritional, psychological and rehabilitation support) provided by their hospital. Patients see the value in early monitoring of their well-being through questionnaires via the CAPABLE app, and would like to receive general and personalized information on their supportive care needs and options.
    - The CAPABLE mobile application should monitor the physical and mental well-being of patients through questionnaires and inform both patient and care team when necessary.
    - The CAPABLE mobile application should provide information on supportive care to patients.

For a complete overview of the interview results and the subsequently concluded patient needs, see section 4 of Deliverable D2.1 (Peleg, 2020).

4.2. Physicians' Needs

The needs of healthcare professionals (HCPs) for the CAPABLE clinician dashboard were collected previously via interviews at NKI and via interviews and questionnaires at ICSM. The needs of HCPs can be summarized as follows:

- **HCP-N1** HCPs have a need for remote patient monitoring to have insight in the patient's well-being at home. HCPs want to be notified by such a remote patient monitoring system when the patient's well-being deteriorates and if the patient reports acute moderate/severe symptoms.
    - The CAPABLE dashboard must allow for monitoring of the progress of disease, the occurrence of symptoms and side-effects, and the overall quality of life and wellbeing of patients.
    - The CAPABLE dashboard must be able to process these monitoring data and provide a flexible clinical decision support system to identify critical situations, following the recommendations from clinical practice guidelines.
- **HCP-N2** HCPs are often involved in complex decision making regarding treatment for patients. There is a need for supportive tools to facilitate this complex decision making.
    - The CAPABLE dashboard should provide patient-specific and/or general evidence-based tools to support these complex decisions.
- **HCP-N3** HCPs have a need for general information such as summaries of survival statistics for different patient groups to facilitate discussion with patients regarding survival and prognosis.
    - The CAPABLE dashboard should provide such summaries of survival statistics.
- **HCP-N4** HCPs indicated the need for the CAPABLE dashboard to complement their current work processes and not cause an increase in administrative burden.
    - The CAPABLE dashboard should extract and reuse data items available from the hospital infrastructures to prevent manual redundant recording by HCPs. This requirement is handled outside the AI framework

For a complete overview of the interview and questionnaire results and subsequently concluded healthcare professional needs see section 5 of D2.1 in (Peleg, 2020).

5. AI Framework

5.1. CAPABLE Architecture

Figure 5.1.1 captures the overall architecture which underlies the CAPABLE project (Lanzola, 2021). It serves the different actors involved in the CAPABLE use cases. The main principles guiding the design of the CAPABLE system are modularity and decoupling: different components will be developed by different project's partners as separate web services. Most of the interaction between the components will happen for data exchange and events.

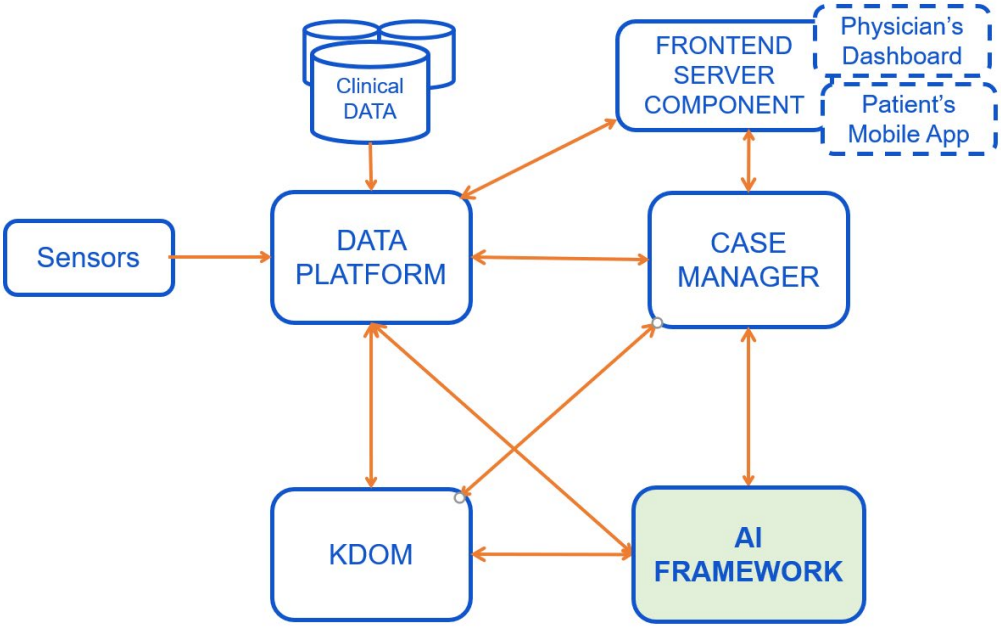

Figure 5.1.1. The CAPABLE architecture including the AI framework

5.2. AI Framework Overview

The framework aims at providing support to physicians, patients and their home caregivers. Since the support offered to home caregivers is a subset of support provided to patients, in the following text we will refer to the latter. The different modules in the AI framework are being developed based on two main categories of sources: real world data (RWD) and knowledge.

RWD includes retrospective EHR data, prospective EHR data, prospective sensor data, patient questionnaires and data gathered from a question-answering forum.

Knowledge includes knowledge captured in publications, clinical guidelines, and educational materials, interpreted with the help of the CAPABLE clinical partners.

Figure 5.2.1. shows an overview of the framework and presents the different models based on this separation of data-driven support (Figure bottom box) and knowledge-driven support (upper box).

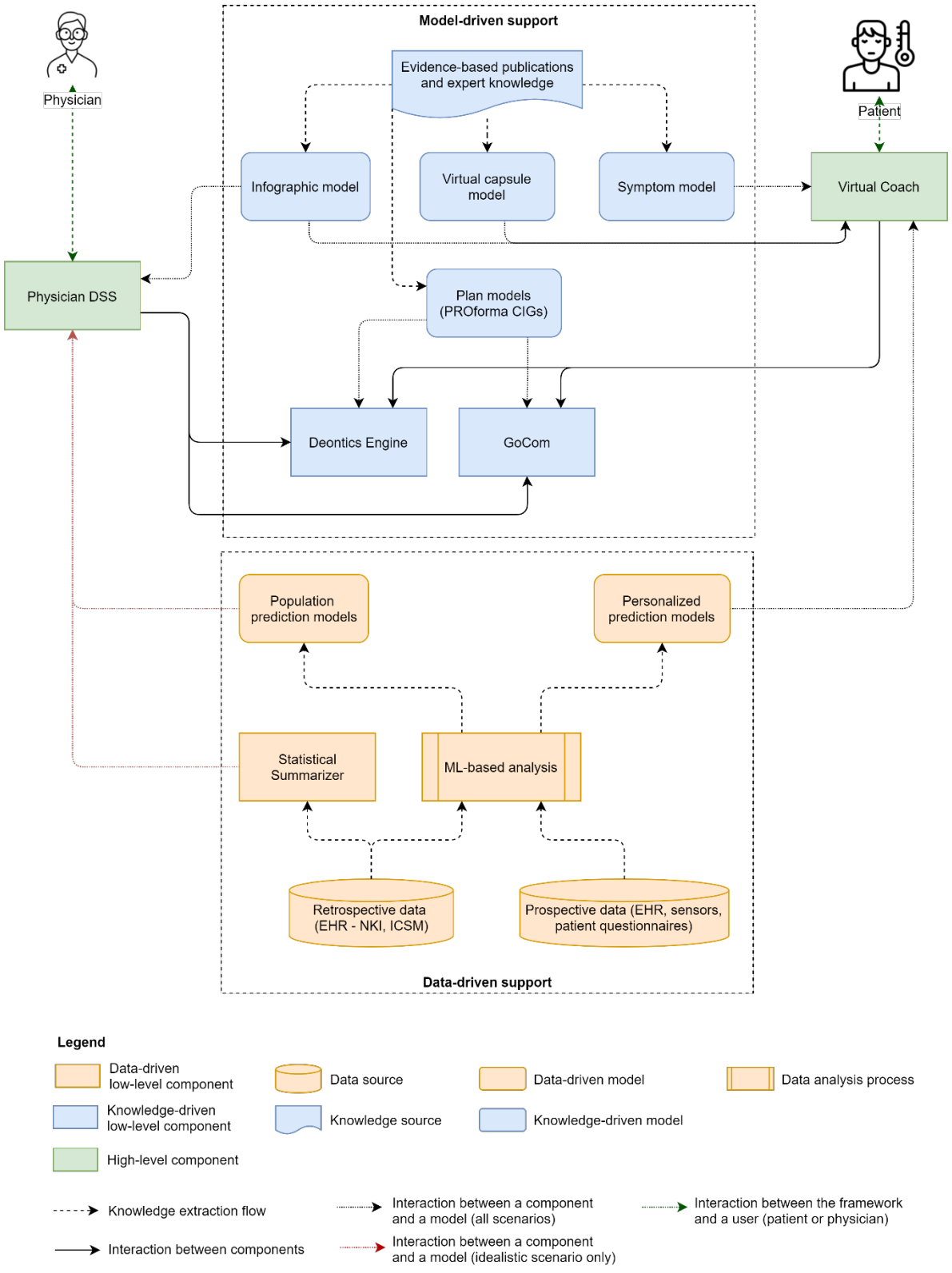

Figure 5.2.1. Overview of the AI framework

The provided support addresses a range of needs related to the care process and it offers the following high-level functions typically considered in the context of clinical decision support (Musen, 2014):

- Providing access to relevant data, information, and knowledge

- Alerting and reminding about relevant events or actions
- Providing patient-specific recommendations and feedback

Specific forms of the above support modalities depend on whether they are provided to the physician or the patient. Examples of modalities are given in Table 5.2.1.

Table 5.2.1. Examples of support modalities

Support	**Physician**	**Patient**
Access to data, information and knowledge	• statistical summaries for groups of patients • recent symptoms reported by the patient	• current vital signs collected from wearable sensors • educational materials related to immunotherapy
Alerts and reminders	• notification about symptoms requiring additional investigation during the next follow-up visit • alert about a prolonged period when the patient has not used the patient app	• notification about an upcoming follow-up visit • request to report on an on-going symptom • reminder to take a prescribed drug
Patient-specific recommendation and feedback	• a CIG*-based recommendation to prescribe a specific as-needed treatment • recommendation to revise a proposed treatment due to adverse interactions	• recommendation to start a virtual capsule with Tai Chi exercise program • recommendation to perform a breathing exercise after detecting stress • recommendation to start a prescribed as-needed drug after detecting a triggering condition

*CIG= computer-interpretable clinical practice guideline

The AI framework components can be separated into high-level (closer to the *front-end)* and low level (*back-end)* components. The high-level components provide support directly to patients' application and physicians' dashboard. They rely on the low-level components of the framework. As described above, the back-end components encompass two groups of techniques to provide the support: **knowledge-driven** and **data-driven**.

Knowledge-driven techniques rely on formal models that represent domain knowledge (e.g., coming from publications) and are constructed in a non-automatic way with the help of domain experts. Examples of such models include *plan models*, *virtual capsule model, infographic model* and *symptom model*.

- *Plan models* are used to represent physician- and patient-related clinical plans (see Section 6.1.1). These models capture CIGs and other auxiliary plans, e.g., controlling the process of symptom reporting or filling questionnaires. For the sake of simplicity, in the following text we refer to CIGs.
- *Virtual capsule model* defines concepts that are necessary to characterize virtual capsules, i.e. non-pharmacological interventions for improving wellbeing. It also

introduces instances of these concepts to represent specific virtual capsules that may be offered to patients (see Section 6.1.2).

- *Infographic model* defines concepts that specify relevant properties of infographics (e.g., condition or symptom that are related to) and their content. In this way available materials may be automatically filtered or prioritized for a given patient. This model also contains instances of introduced concepts that represent specific infographics (see Section 7.1.4.1).
- *Symptom model* defines concepts related to symptoms that may be reported by the patient and possible relations between symptoms. It also contains instances of the concepts capturing all symptoms considered in the CAPABLE system (see Section 7.2.2)

In CAPABLE the plan models are represented using the PROforma language (see Section 7.1.1). We also use the Deontics Engine (see Section 7.1.1) to execute CIGs captured by these models. Moreover, we employ the GoCom component (see Section 7.1.3) to identify and mitigate conflicts between actions recommended by several CIGs applied to the same patient, and/or actions recommended by a CIG and already being executed. The remaining models are in fact *ontological models* (they define concepts and their instances related to a specific body of domain knowledge) and they are represented using the OWL (Web Ontology Language) (W3C, 2013).

We should note that some of the support functions rely on several models. For example, managing symptom reporting and providing relevant feedback to the patient calls for a plan model that describes the general reporting process and the symptom model that provides detailed information (e.g., desired reporting frequency) for a given symptom.

Data-driven support techniques rely on knowledge discovered from data with diversified data mining methods, such as exploratory analysis, statistical analysis or machine learning. Within the AI framework we apply such methods to collected clinical data in order to develop *statistical summaries* and *prediction models*.

- Statistical summaries are static findings from the analysis that provide concise characteristics of the processed data, however, they do not allow for deriving patient-specific recommendations.
- Prediction models are executable findings from the analysis that can be applied to new patients in order to provide diversified recommendations and insights, related to clinical and non-clinical treatment, such as response to treatment, survival or appropriate timing to start exercise from a virtual capsule.

Within the AI framework we consider two groups of data sets that are employed by the data-driven support techniques:

- *Retrospective data* providing structured clinical characteristics with melanoma and renal cancer together with the information about the course and effects of clinical treatment (see Section 7.1.4.2), as well as unstructured textual data from patient forum (see Annex 1).
- *Prospective data* of patients participating in the clinical trial of the CAPABLE system. This data shares the structure with the retrospective data from EHR and in addition it includes symptoms self-reported by patients, entered by physicians during follow-up visits, imported from EHR and collected automatically using sensors embedded in a consumer-grade smartwatch (see Section 6.3.1).

Retrospective data is used to develop population-based prediction models and summaries. We will run the developed (retrospective-based) prediction models on the common

structured data from EHR, which is part of the prospective data and summarize the results. We anticipate that the actual outcome results will be available for part of the patients, depending on when they were enrolled to the study as there is need for at least 1 year follow up time.

The prospective data is employed to refine population-based models and to construct *personal* prediction models for individual patients: (1) When using the retrospective data, we predict overall survival rate, response to treatment and toxicity, while when using prospective data, we can further tune these models. (2) Using the prospective data, we work on developing models for evaluating the stress level and cognitive load of patients to further customize offered support, e.g., by identifying most appropriate times during the day to provide reminders related to virtual capsules.

As already mentioned, the front-end of the AI framework is composed of two components – Physician DSS (Section 7.2.1) and Virtual Coach (Section 7.2.2), respectively. These components offer **hybrid support** that combines knowledge- and data-driven techniques. For example, the Physician DSS employs plan models (CIGs) to control patient management and when consulting with the physician decision options (e.g., treatments), it augments them with outcomes of population prediction models, as well as relevant infographics and statistical summaries. Moreover, the Virtual Coach employs ontological and algorithmic models for initial recommendation of virtual capsules, and then employs personalized prediction models derived from prospective data to further customize these capsules.

The proposed framework allows for satisfying the needs of patients and clinicians mentioned in Sections 4.1 and 4.2 respectively. On the one hand, patients are offered a wide selection of education materials (stand-alone or associated with virtual capsules). They are supported by manual and automatic supporting facilities (e.g., forms and reminders for collecting symptoms and questionnaires, and wearable sensors for collecting vital signs). Finally, based on the provided information they are given feedback (recommendations, instructions, or explanations) that is based on evidence-based models (CIGs, virtual capsules) and on personalized data-driven prediction models. On the other hand, physicians have instant access to data collected by their patients and are alerted about situations requiring their attention. Moreover, as explained above, they are provided with diagnostic and therapeutic recommendations driven by evidence-based models (CIGs) enhanced with infographics and outcomes of population-based prediction models, so they can make informed decisions. In Table 5.2.2 we show how specific platform components and models address specific physician and patient needs (x indicates that a given component or model is required to address a specific need).

Table 5.2.2. Components and models in the AI platform required to address specific patient and physician needs

	Pa-N1	Pa-N2	Pa-N3	HCP-N1	HCP-N2	HCP-N3
Physician DSS				x	x	x
Virtual Coach	x	x	x			
GoCom		x	x	x		
Deontics Engine		x	x	x		
Plan models		x	x	x		
Infographics model	x		x		x	
Virtual capsule model			x			
Symptom model		x	x			
Statistical summarizer						x
Population prediction models					x	
Personalized prediction models			x			

Finally, we note that knowledge-driven techniques are based on validated research that physicians are using in clinical practice. Therefore, giving advice in the prospective stage based on the knowledge-driven support will be well accepted. In contrast, data-driven support, and in particular population-based prediction models that are based on the retrospective smaller data set, although validated on a test set, have no FDA/CE approval to be used in clinical practice and will not be part of the prospective stage (for those legal constraints). The conclusions of these prediction models will be provided in a report to the physicians in a statistical manner, with no personal advice per patient.

To account for the above limitations, during the development and clinical validation of CAPABLE, we consider two classes of scenarios: *idealistic*, where the decision support is based also on the prediction model's results (as if they had FDA/CE approval) and *realistic* in which the prediction model is not providing advice at the point of care. Idealistic scenarios would be followed for demonstrating the CAPABLE system on synthetic (artificial) patients, while realistic scenarios will be used during the clinical trial.

6. Data & Knowledge Sources

6.1. Sources for Knowledge-driven Models

6.1.1. Guidelines Encoded in PROforma Language

PROforma (see section 7.1.1) is a formal language for specifying clinical knowledge. It aims to capture the framing, structure and logic of clinical decisions - when a decision can and should be made, what options are to be decided between, what information is relevant for making the decision, and the effect of available data on the different options. PROforma may be used to capture clinical knowledge from any source, but in current use it is typically used to formalise existing published clinical guidelines. In the CAPABLE project it is used to formalise two bodies of existing clinical knowledge, relating to published clinical guidance for patients and physicians.

Patient-oriented plans

A number of guidelines will be modeled and used in order to produce recommendations for the patient:

- Immunotherapy toxicity
- Diarrhea
- Fatigue
- Mucosal Injury
- Skin toxicity

Physician-oriented plans

A number of guidelines will be modeled and used in order to produce recommendations for the physician:

- Immunotherapy toxicity
- Diarrhea
- Fatigue
- Mucosal Injury
- Skin toxicity
- Drug-induced hypertension
- Cutaneous Melanoma
- Renal Cell Carcinoma

6.1.2. Publications Encoded as Virtual Capsules

Virtual capsules (capsules in short) are non-pharmacological based interventions aimed at improving mental, physical and social well-being of the patients -- see Section 4.5.3 in D2.1 (Peleg, 2020) for details. Similarly to clinical guidelines, the proposed capsules are based on evidence, and they complement CIGs by addressing non-clinical aspects of care. Virtual capsules combine educational materials with actionable recommendations, such as physical or breathing exercises. Moreover, educational materials may be static, where the patient just reads or watches them without further interactions (examples of such materials include infographics, textual descriptions or videos), or dynamic, where the patient can use their knowledge to solve simple problems, such as quizzes.

Virtual capsules are represented using an ontological model -- see Section 4.5.4 and 4.6 in D2.1 (Peleg 2020) where this model is referred to as the Patient Model. This model introduces concepts representing capsules and their detailed characteristics, including dimensions of wellbeing they address, their duration and intensity, and selected patient characteristics that are relevant for selecting and personalizing the capsules (e.g., hobbies or literacy level). The ontological model also allows for specifying awards for following recommendations from capsules and for reporting on the experience associated with the capsules. This allows Virtual Coach to apply gamification techniques, such as scoring, social competition and comparison to peers (Cheng, 2018) to make the capsules more appealing to the patients. Finally, the ontological model defines instances of specific capsules. Examples include 30x30 nature walk, tai chi or garden bowl that allow to improve various forms of well-being. A simplified version of the ontological model for virtual capsules is given in Fig. 6.1.2.1.

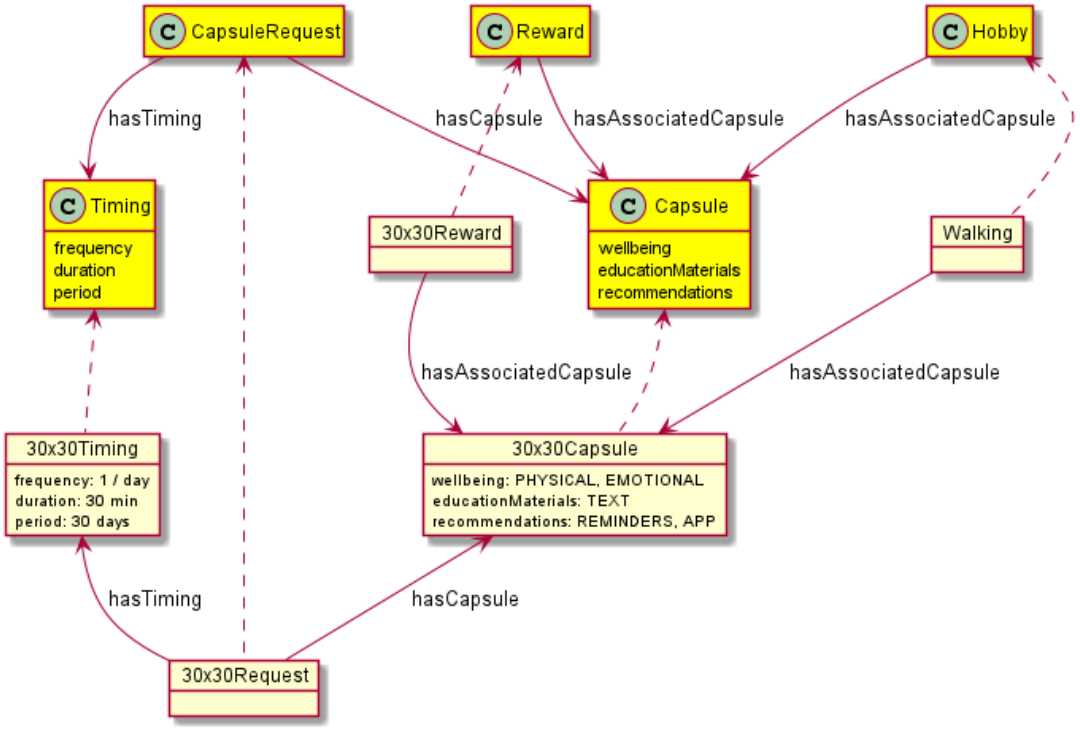

Figure 6.1.2.1. Part of the ontological model describing virtual capsules

6.2. Sources for Data-driven Models

6.2.1. Retrospective Data

The retrospective data is EHR based structured clinical data of the patients collected for two types of cancers (Melanoma, Renal Cancer) in two hospitals in the Netherlands and Italy.

6.2.1.1. Melanoma Data Set (NKI)

This data set comprises 500 Melanoma patients (men - 44%, women – 56%, median age - 65 IQR [55,75]) from NKI in Amsterdam. The patients are of stages III and IV of Melanoma (having metastases) and treated with Immune Checkpoint Inhibitors (ICI): anti-PD1 antibodies (Nivolumab, Pembrolizumab) alone (78%) or combined with anti-CTLA4 antibodies (Ipilimumab) (22%) as presented in 560 treatment lines. Each subject is presented with demographic information, diagnostic melanoma details and detailed information about treatment lines. Each treatment line includes metastases status, blood tests, patient

condition, treatment details, other clinical information and the outcomes: response, toxicities, reason for termination. In addition, timing information is available for diagnosis date, treatment start/end, follow-up status update dates during the treatment, as well as last follow up date or moment of death. In 46% of lines - patients have complete or partial response to treatment, in 20% remain stable and in 30% have only progression. Toxicity of grade 3-4 reported in 23% of treatment lines and caused treatment discontinuation in %17 of treatment lines.

Detailed descriptive statistics of data can be viewed in Section 8.4, Figures 8.4.1-8.4.3 of D2.1 (Peleg, 2020). Additional information related to this data can be viewed in Section 3.2 of D5.1 (Barkan, 2020)

6.2.1.2. Renal Cancer Data Set (ICSM)

This data set comprises 343 metastatic renal cancer patients (men - 77%, women - 23%, median age 56 IQR[47,63]) from ICSM in Italy. The data set includes 917 treatment lines with up to 7 treatment lines available per patient, where 89% of treatment lines are targeted therapy (antiangiogenic), 6% are immunotherapy and 5% are chemotherapy. Each subject is presented with demographics information, diagnostic renal cancer details and detailed information about treatment lines. Each treatment line includes metastases status, blood tests, patient condition, treatment details, other clinical information and the outcomes (best response, toxicities, reason for termination and survival). In addition, timing information is available for diagnosis, treatment start/end for each line, dose reduction, brain metastasis diagnosis and treatment dates as well as last follow-up date or moment of death. In 18% of lines - patients have complete or partial response to treatment, in 52% remain stable and in 29% have only progression. Toxicity of grade 3-4 reported in 19% of treatment lines and caused treatment discontinuation in %7 of treatment lines. Descriptive statistics of data can be viewed in Section 8.4, Figures 8.4.4-8.4.5 in D2.1 (Peleg, 2020). Additional information related to this data can be viewed in Section 3.1 in D5.1 (Barkan, 2020)

6.2.1.3. AIMAC

AIMAC, Italian Association for Cancer patients, relatives and friends, provides three kinds of data:

1. Data from its discussion forum. The forum contains unstructured data, i.e., texts (forum posts) that patients or their relatives share within their community;
2. Data from questionnaires (structured data) that are filled-in when patients contact AIMAC volunteers either by telephone or by accessing the AIMAC headquarters;
3. Data from a questionnaire (structured data) that has been recently put online during the COVID-19 pandemic.

Additional information related to this data can be viewed in Section 3.1 in D5.1 (Barkan, 2020).

Initial exploratory work that was done with AIMAC data is described in Annex 1. We will explore if additional usage of the forum data can be of benefit for CAPABLE users.

6.2.1.4. Other Public Data Sources

In order to validate and strengthen predictive models we are exploring additional data sources. After exploration, we will report if and how we used it to support the AI framework.

IBM MarketScan Research Databases (https://www.ibm.com/il-en/products/marketscan-research-databases)

- This is a set of databases of insurance claims compiled from health insurers, large employers, and government programs in the USA.
- Initial estimation (based on 8 months information) shows
    - Availability of ~1800 Melanoma patients with treatments similar to NKI data
    - Availability of ~2400 patients of mRCC treated with targeted and immuno therapy.

The Surveillance, Epidemiology, and End Results (SEER) (https://seer.cancer.gov/). This is a national cancer registry that is used by the National Cancer Institute (NCI) for surveilling cancer in the United States. SEER collects information about the demographic characteristics of patients diagnosed with cancer.
Although these data sets have patients with similar diagnosis and treatment they are missing essential clinical information about diagnostic details, lab results, outcomes, etc. that are present in NKI and ICSM data sets. We are currently researching how to leverage the available data although it is not complete using some learning approaches, e.g transfer learning.

6.3. Data Sources for Patient Specific Support

6.3.1. Sensors

To enhance the decision support processes, data collected from the sensors are being used in the CAPABLE Project. Each patient is going to be provided with a smartwatch (ASUS VivoWatch) equipped with the following built-in sensors:

- *Photoplethysmography* (PPG) sensor detects changes in blood volume using an optical light source (typically red or green diode) and a photodetector. This provides insight into the volumetric variations of blood circulation. Heart rate or blood oxygen saturation can be extracted from such measurement. Many PPG sensors are embedded in wrist-worn devices.
- *Electrocardiography* (ECG or EKG) sensors record the electrical signals in a patient's heart. It is a noninvasive way of detecting common heart problems. The electrocardiography waveform consists of five peaks and valleys. Heart rate can be easily extracted by measuring the R-peaks of the ECG waveform.
- *G sensor* (accelerometer) is responsible for measuring the acceleration (or vibration) of a device. Data collected by accelerometers embedded in wearables are mainly used to detect the number of steps taken by the users. Switching the smartphone's orientation when flipping it is also achieved thanks to the G sensor.

These devices will provide vital signs data, which will be collected and saved in the Data Platform (section 7.3.1). Table 6.3.1 lists the built-in sensors together with the collected vital signs.

Table 6.3.1.1. Data collected using wearable sensors

Vital Sign	Sensor	Unit
Daily Steps	G sensor	Steps
Systolic blood pressure	PPG sensor + ECG sensor	Millimeter mercury column (mm[Hg])
Diastolic blood pressure	PPG sensor + ECG sensor	Millimeter mercury column (mm[Hg])
Heart rate	PPG sensor + ECG sensor	Beats per minute (bpm)
Light sleep duration	PPG + G sensor	Hours (h)
Deep sleep duration	PPG + G sensor	Hours (h)

Vital signs listed in Table 6.3.1.1 are synchronized automatically with OmniCare -- a cloud platform provided by ASUS for healthcare applications (ASUS Life Corporation, 2021). Specifically, a dedicated app running on a patient's smartphone connects with the smartwatch and uploads the raw data onto the cloud every 15 minutes. Then, the Sensors module in the CAPABLE system retrieves the data, preprocesses them if necessary, and then stores them in the Data Platform. In this way vital signs become visible for all other components of the system. Typical methods of preprocessing include aggregation of raw values into larger "chunks". For example, information about the type of sleep is collected at 1-minute intervals and these readings need to be translated into the duration of deep and light sleep at night.

6.3.2. EHR and Questionnaires data

In D2.1 (Peleg,2020), section 5.7, the clinical data to be collected within the CAPABLE system throughout several data sources was described. Table 6.3.2.1 reports the source for each category of data.

Table 6.3.2.1. Data type descriptions and data sources (extracted from (Peleg,2020))

Data type description	EHR	Patient app	Physician dashboard	Sensors
Symptoms	x	x	x	
Tumour characteristics	x		x	
Risk factors	x		x	
Clinical history	x		x	
Cancer treatment	x		x	
Other treatments	x		x	
Exams and visits	x		x	
Laboratory tests	x		x	
Physiological and lifestyle data		x	x	x
Environmental data				x
Questionnaires		x	x	

CAPABLE EHR data available for NKI and ICSM will be extracted from the hospital infrastructure through dedicated, automatic ETL procedures and loaded into the Data Platform. The scheduling of this task is to be defined. When data is not available in a structured format (or not available at the time that the user is interacting with CAPABLE) then the user will be asked to enter the data.

In the following sections, we will describe the NKI EHR data source and ICSM EHR data source in more detail. Additionally, Section 4.1.1. in (Cornet, 2021) describes both data sources.

6.3.2.1. NKI EHR Data

The prospective data source available for NKI is an existing reporting/research extract that can be used. Items that are not currently in this database can be added, provided that they exist in the patient record as structured data items. The database is currently updated nightly, but there are plans to update it with near-live data (every few minutes) in the future. Potential delays due to COVID-19 might mean that the near-live data will not be available yet at the time of the trial. We are developing the system to function well in either situation.

This database will provide data about demographics information, medication prescribed in NKI, procedures performed in NKI, blood tests (if performed in NKI), scheduled visits in NKI and possible hospitalizations. Table 6.3.2.1.1 shows the current feasibility assessment of extraction of data from the NKI database, as well as a summary of the data items to be entered by users. All data can be entered via the clinician user interface, so the clinician can enter any data items that are missing from the EHR. Data entered by the patient and collected from sensors will also be accessible to the patient, and the clinician via the clinician dashboard.

Table 6.3.2.1.1. Feasibility assessment of extraction of data from NKI EHR

Data type description		**From EHR / Uncertain**
Patient profile		
Patient demographics	name, birthdate, height, education, etc.	From EHR
Caregiver information	Name, contact information	From EHR
Treatments and tests		
Medication	name, code, start date, end date, frequency, dose, route, indication	From EHR
Home medications	name, code, start date, end date, frequency, dose, route, indication	Uncertain
Procedures (in hospital)	name, code, start date, end date	From EHR
Past procedures	name, code, start date, end date	Uncertain
Confirmed diagnoses list / problem list	name, code, date confirmed	Uncertain
Laboratory tests (blood/fluid tests)	name, code, date ordered/performed, +/-result	From EHR
Other laboratory tests/exams	name, code, date ordered/performed, +/-result	Uncertain
Referrals within hospital	name, code, date	From EHR
Questionnaires		
Questionnaires administered within NKI and by CAPABLE	EORTC QLQ-Q30, EQ-5D-5L, FACT-M	Uncertain

6.3.2.2. ICSM EHR Data

The prospective data sources available for ICSM are the Hospital Information System (HIS) and a dedicated pathology registry (managed by REDCap, a software for electronic data capture). The HIS will provide data about demographics information, blood tests (if performed in ICSM),

scheduled visits and possible hospitalizations. Scheduled visits for patients currently enrolled in pharmaceutical clinical trials are not managed by the HIS, so for this subset the information is not available. Data related to the next scheduled appointments will be inserted by the clinicians in the CAPABLE interface.

The pathology registry is hand-curated by the clinical staff of ICSM and it contains detailed information about treatment lines and outcomes. The data set details are available in the Section 6.2.1.2 of this document. The pathology registry does not contain information about drugs scheduled administrations, this data will be inserted by the clinicians in the CAPABLE interface.

Table 6.3.2.2.1. Feasibility assessment of extraction of data from ICSM EHR

Data type description		**From EHR / Uncertain**
Patient profile		
Patient demographics	name, birthdate, height, education, etc.	from EHR
Caregiver information	Name, contact information	Uncertain
Treatments and tests		
Medication	name, code, start date, end date, frequency, dose, indication	from EHR
Home medications	name, code, start date, end date, frequency, dose, indication	Not available
Procedures (in hospital)	name, code, start date, end date	from EHR (only if hospitalized patient)
Past procedures	name, code, start date, end date	from EHR (only if hospitalized patient)
Confirmed diagnoses list / problem list	name, code, date confirmed	from EHR (only if hospitalized patient)
Laboratory tests (blood/fluid tests)	name, code, date ordered/performed, +/-result	from EHR
Other laboratory tests/exams	name, code, date ordered/performed, +/-result	Uncertain
Referrals within hospital	name, code, date	from EHR
Questionnaires		
Questionnaires administered within ICSM and by CAPABLE	EORTC QLQ-Q30, EQ-5D-5L, FACT-G, FACT-R, MST, PHQ9, GAD7, EORTC QLQ-INFO25	from EHR

6.3.2.3. Questionnaires

As described in section 5.2 of D7.1 (Sacchi, 2020), the clinical pilot study protocols foresee the administration of several questionnaires in order to generate evidence on the effect of 'a systematic web-based collection of patient-reported symptoms and mobile coaching system (CAPABLE)'. The clinical pilot trial will be a prospective experimental cohort study with patients receiving the CAPABLE app throughout their systemic treatment, and one prospective cohort study with patients receiving standard care (without CAPABLE app). Both groups of patients will fill-in questionnaires starting in Q1, 2021 (ICSM May/June).

The details about questionnaire scoring can be found in Annex 2.

7. AI Framework Components

AI Framework components can be grouped into two categories according to their task:

- Low-level components - provide model-driven support by application of plan models, ontological models and algorithmic models by Deontics Engine and GoCom and data driven support by application of population-base and personalized predictive models, as well as statistical summaries on data (Section 7.1);
- High-level components - deliver support to the physician and patient, offer hybrid support that combines knowledge- and data-driven techniques. (Section 7.2).

In addition, there are components that support both high- and low-level components that we call **Auxiliary components** as they are out of the AI framework. These auxiliary components assist in inter-component communication, managing patients' data and processing flow as well an interaction with framework users (section 7.3).

7.1. Low-Level Components

7.1.1. Deontics Engine and the PROforma language

The Deontics engine is an execution engine for Computer-Interpretable Guidelines (CIGs) written in the PROforma language. Within CAPABLE, PROforma CIGs are used to model the guideline knowledge used by the two higher level components discussed above, the Physician DSS and the Virtual Coach. The Deontics engine provides a CIG execution service for both components.

7.1.1.1. PROforma

PROforma is an executable process modelling language, designed primarily for modelling clinical processes and decisions. The syntax and operational semantics of the language are public (Sutton and Fox 2003). The PROforma process model is based on a small set of task classes that can be composed into networks representing arbitrarily complex plans or procedures.

There are four main task types:

- *Actions* represent procedures to be carried out in the external world (e.g. administer a drug).
- *Enquiries* are carried out to acquire information from some person or external system.
- *Decisions* define choices about what to believe or what to do.
- *Plans* group tasks (including other plans) together into a single unit.

The execution model is based on a simple notion of task state. Each task may be in one of four states:

- *Dormant*: task has not yet been started.
- *In progress:* task has been started and is being executed.
- *Completed:* task has been completely executed.
- *Discarded:* task is unable to execute because it has preconditions which are not fulfilled or is a plan that is unable to complete because it has component tasks that are discarded.

A plan can therefore effectively be cancelled part-way through if it becomes unable to proceed. Individual tasks however are considered to be atomic and correspond to operations like sending a medication request which cannot be cancelled once they have been triggered.

Cancelling such operations in general will require positive actions to be taken, which may be complex or non-trivial (e.g. it may not be certain whether the patient has yet received a requested medication). This is best handled by a plan that can include contingency procedures to follow if things don't go as expected.

CIG execution can be controlled by simple scheduling constraints, where a task can only be performed if another task has been completed, and by more flexible preconditions and postconditions. A precondition is a boolean condition that must be satisfied for the task to start execution, and a postcondition allows internal variables to be written to if the task completes successfully.

PROforma CIGs define an internal data model that logical conditions can operate against. Typical data types are supported including integer and real numbers (e.g., test results, vitals etc), text, dates, and sets (e.g. lists of current diagnoses or medications).

PROforma defines a particular model for Decisions. Rather than simply choice-points in a process (which can be handled by logical preconditions as in any programming language), in PROforma a Decision represents a point where a human user - a clinician or patient - needs to make a decision for which they require support. Decisions are modelled as a set of *Candidates*, which represent the valid options that can be decided between. Each Candidate may have one or more *Arguments* attached to it. Arguments represent logical "reasons" for or against that Candidate and can in turn be based on evidence (such as published guidance, clinical trials etc) that the user can refer to if they need further justification for a decision. Decisions may have a logical rule defined (a *recommendation rule*) that will recommend a particular Candidate to the user, for example if it has more arguments in favor than others. However, this is not required, and typically the user is in any case allowed to see the various arguments for and against each candidate before making a decision.

When a PROforma plan is started all component, tasks are in the dormant state. The execution engine repeatedly examines the scheduling constraints, pre- and post-conditions of all tasks in order to determine what state changes and other actions should occur.

7.1.1.2. Deontics Engine

Deontics provides tooling for writing and testing PROforma CIGs, and also an engine for executing them. The engine typically runs on a server and exposes a web-based API to allow PROforma CIGs to be uploaded and executed, to allow data values to be set, and to allow the results of CIG execution (e.g. recommended decisions and their justification) to be retrieved.

7.1.2. Predictive Models

7.1.2.1. Predictive Models from Retrospective Data

Data Sets

Predictive models will be developed using two retrospective data sets that are described in 5.4.1.

Objectives for Prediction

After consulting with clinicians from NKI and ICSM we defined a scope of work for development predictive modes. The work will be done in two stages:

1. Developing models for predicting of: Overall Survival, Response to treatment (Eisenhauer, 2009), Toxicity and Reason for treatment stop;

2. Exploring additional models for prediction Progression Free Survival (PFS) during the treatment (only for Melanoma data set), Time from diagnosis to start of systemic treatment.

Models Research and Development

First step of building the predictive models is to generate descriptive statistics for the data sets in use, in order to verify the distribution of the desired questions/predictions we want to explore as well as distribution of patients parameters like age, gender, treatment type, etc. Descriptive statistics for Melanoma and Kidney data were calculated and can be viewed in Section 8.4, Figures 8.4.1-8.4.5 in D2.1 (Peleg, 2020).

The next step will be to identify the existing biases in the cohort either due to unbalanced data or to presence of confounders in the data. In case if biases are found, we are going to apply corrections methods like Inverse Probability of Treatment Weighting (IPTW) using propensity score matching (Rosenbaum,1983) and SMOTE (Chawla,2002).

During the feature generation process we will manage with (1) generation features in high-dimensional data of patients (up to several hundred of fields in raw data) using expert knowledge, (2) aggregation information by line and patient; (3) grouping features according to medical episodes: baseline (diagnostic information), staging, treatment line related. On the generated features we will apply Data Scaling and Missing data imputation using standard imputation approaches, e.g. mean or more sophisticated approaches like MICE (Van Buuren, 2011).

For categorical outcomes, such as Response Type or Toxicity, we will use machine learning classifiers such as XGBoost (Chen, 2016), Logistic Regression (Wright, 1995), Random Forest (Breiman, 2001). In order to get better performance, we will use an ensemble approach, as it is known that an ensemble of several (even weak) classifiers generates a better performing classifier. We will use AUC-ROC (Hanley, 1982) as a metric for performance evaluation of classification models.

For time to event prediction like Overall Survival (OS) and Progression Free Survival we will apply statistical methods such as CoxPH (Cox, 1972) and Weibull (Weibull,1951) as well as Machine Learning methods such as Random Survival Forest (Ishwaran, 2008) We will use Concordance Index(CI) (Harrell, 1984), log-likelihood and Mean Absolute Error(MAE) as performance evaluation metrics.

We will train and validate our models using five-fold cross validation, one-leave out and split to train/validation/test sets.

In order to understand which features influence most the prediction model generation, we use several methods that visualize or output the significant features like univariate analysis with statistical significance estimation, SHAP analysis (Lundberg, 2017) of each model (see an example in Figure 7.1.2.1.1), plotting of coefficients for CoxPH model (see example in Figure 7.1.2.1.2). As well we will apply correlation analysis (using Chi-Square, t-test, etc.) in order to detect correlated features in our features set.

In addition, we will perform sub-group analysis to identify the model's performance across different criteria (age, etc.).

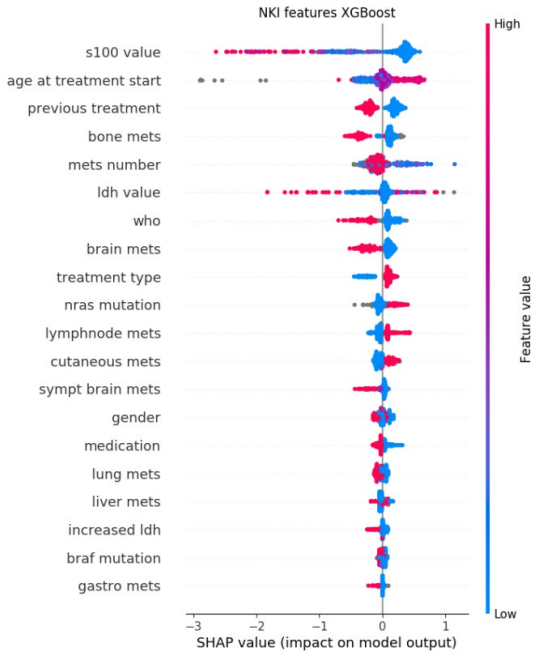

Figure 7.1.2.1.1. SHAP analysis for feature importance for predicting treatment Response for Melanoma Data Set

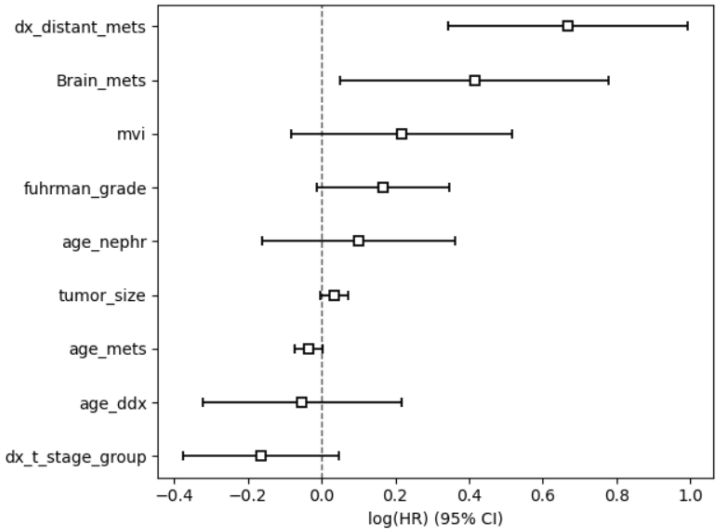

Figure 7.1.2.1.2. Plotting coefficients of CohPH model for prediction of overall survival for renal cancer data set

7.1.2.2. Personalized Predictive Models

Objectives

The Virtual Coach aims to help cancer patients form positive health habits that will improve their quality of life and wellbeing. It will suggest simple short activities from the domain of mindfulness and positive psychology (virtual capsules), shown in the literature to have a positive impact on one's well-being. Each activity is meant to become a habit aligned with patients' physical and psychological wellbeing goals.

Fogg (Fogg, 2019) proposes that habits formation (i.e., performing a target activity) depends on three factors: motivation, ability to perform the task and the presence of a trigger reminding the person to perform the target behavior (See Figure 7.1.2.2.1)

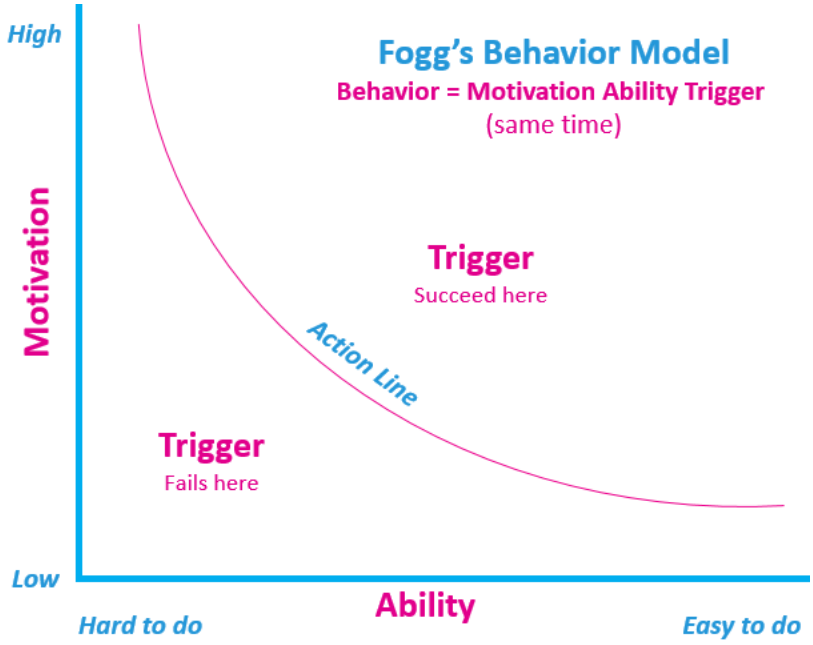

Figure 7.1.2.2.1. Fogg's behavior model

The personalised predictive models have a dual role (See Figure 7.1.2.2.2):

1. Predict where the patient is on the Fogg's action line curve, so that the reminders to perform the target activities are send only at the right time.
2. Detect if the activity was performed or not (note this is only possible for some activities. e.g., rhythmic breathing, which can influence one's heart rate variability).

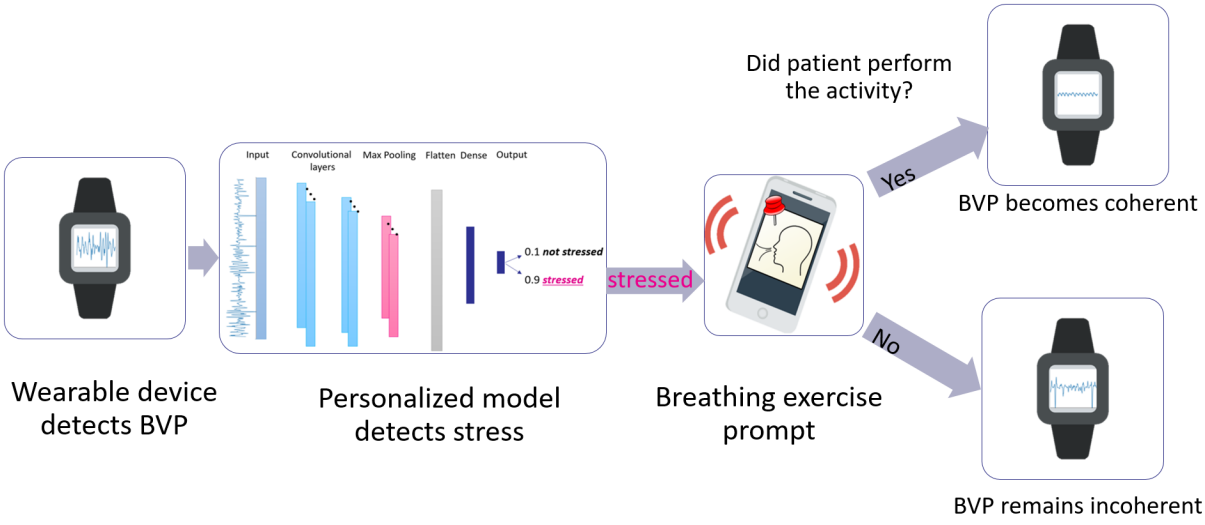

Figure 7.1.2.2.2. Application of personalized stress prediction models

Multiple factors might affect patients' motivation, ability to perform the target task and effectiveness of the prompt. For example, the motivation might depend on the quality of sleep (Dolsen, 2017) and patients' positive valence (Jowsey, 2014). The ability is influenced by

whether the patient is currently occupied by other tasks (Chan, 2020) and whether they have performed the target activity successfully before. The effectiveness of the prompt depends on the patient's arousal state (Goyal, 2017) and time of the day (Bidargaddi, 2018). Note that some of these variables are captured directly by smartwatch and a mobile application, e.g., hours of sleep or activity rating (respectively), others need to be inferred from the raw Blood Volume Pulse (BVP) signal captured by smart-watch e.g cognitive load, valence and arousal. Therefore, as the first step we develop personalised models that take as an input raw BVP and output predicted mental state. In the second step we take a predicted mental state, alongside other information gathered by the app and predict whether to prompt the patient or not.

Methods

1. Machine learning models for emotion and cognitive load classification.

   In two feasibility studies we use annotated BVP signal data from publicly available datasets WESAD (Schmidt, 2018) and CLAS (Markova, 2019) to develop stress and cognitive load detectors respectively. In both cases we train simple 1D convolution neural network models directly on the snippets of the BVP signal. In the first feasibility study we aimed to obtain a model which distinguishes between stress, amusement and neutral emotion (Lisowska, 2021a). Both amusement and stress are characterised by high physiological arousal but differ in the valence and the distinction between the two might be important for determining whether the activity should be suggested to the patient or not at the given time. In the second feasibility study we were interested in detection of the high cognitive load (Lisowska, 2021b) to avoid prompting patients at the times they are already overloaded.

   The next step will be to collect data using the ASUS smartwatch from a small population of volunteers to check how transferable/ adaptable are the models developed on clean data to data gathered outside of experimental conditions (real life setting).

2. Models for personalized prompt timing.

   To personalize timing of the prompts we train a reinforcement learning agent that takes information about the patient state and predicts action (prompt or not). Currently we use Fogg's behavior model to simulate the patient's response to the prompt. Deep Q-learning (Mnih, 2015) agent receives observation of the patient environment state that include the following variables: time of the day (morning, midday, evening, night), week day (weekend or not), time since the last response to the prompt, the last activity score, location (home or other), patients motion (stationary or walking), whether patient is sleeping or awake and physiological state of valence, arousal and cognitive load. The aim of the agent is to learn to prompt the patient only at the time the patient responds to the prompt.

7.1.3. GoCom

GoCom provides higher-level components (e.g., the Physician DSS and the Virtual Coach) with information about potential or existing interactions among the patients' diseases, drugs and other types of treatments. Clinical Practice Guidelines (CPGs) and the CIGs that are based on them often focus on one morbidity. The risk of an interaction occurring due to guideline-based clinical recommendations is greater for a multimorbid patient, since multiple guidelines are involved. Also, the patient may be receiving treatments that are not guideline-based or taking over-the-counter medications that cause a severe adverse reaction when combined with the patient's other treatments or diseases.

For example, a patient may be taking St. John's wort (an over-the-counter medication) for their mood. If the patient must be treated with loperamide for diarrhea, GoCom will provide a warning, since St. John's wort and loperamide may negatively interact.

The notification should be provided both to the physician, before loperamide is prescribed, and to the patient to make sure that they are informed about the possible interaction and will stop taking St. John's wort.

GoCom uses the Data Platform and the Case Manager in order to receive requests and notifications from the Physician DSS and Virtual Coach, as well as send back information about possible interactions if relevant.

When the Physician DSS or the Virtual Coach need to run a guideline due to new patient information entering the Data Platform, GoCom is notified and retrieves the relevant recommendations from the guideline that was run in order to check them for interactions. GoCom is goal-oriented and generates a hierarchical structure of goals for the patient's new recommendations and other data in order to detect interactions.

GoCom utilizes several reasoning patterns in order to detect interactions, and uses a number of terminologies and data sources: RxNorm (National Library of Medicine, 2021), openFDA (U.S. Drug and Food Administration, 2020), and NDF-RT (National Library of Medicine, 2017). GoCom retrieves data automatically from these sources and reasons upon it in order to detect the following types of interactions:

- Temporal interactions
- Drug-drug interactions
- Drug-disease interactions; Treatments causing adverse events
- General treatment interactions (virtual capsules, procedures).

If interactions are found, GoCom provides different combinations of valid clinical solutions with explanations for the user and notifies (via the Case Manager) the component that requested the interaction check. If interactions are not found, GoCom also provides a notification. Figure 7.1.3.1 is a high-level sequence diagram showing the workflow between GoCom and the other components.

More information is available in (Lanzola, 2021).

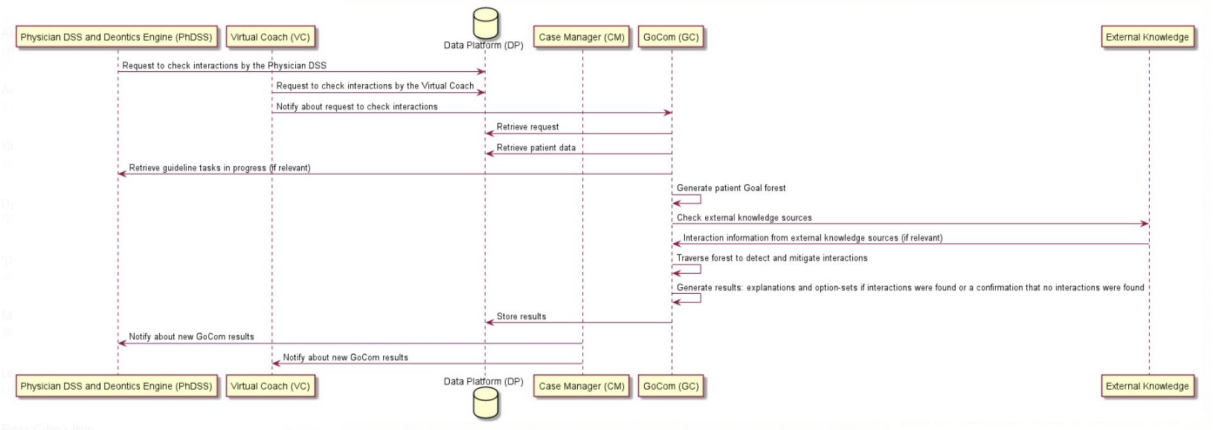

Figure 7.1.3.1. A high-level sequence diagram showing the workflow between GoCom and the other components.

7.1.4. Statistical Summarizer

7.1.4.1. Summaries on Data – Infographics

Infographics for Healthcare Professionals

We have identified relevant information needs (see Section 4 above and Deliverable 2.1) which could be supported by the CAPABLE clinician dashboard. HCPs have a need for general information such as summaries of survival statistics for different patient groups as clinical decision aids and to facilitate discussion with patients regarding survival and prognosis. Infographics are planned to be incorporated in the CAPABLE clinician dashboard as a clinician decision aid. An example of a complex decision that could be supported by infographics based on evidence-based sources is the clinical decision to treat patients with irresectable stage 3 or 4 melanoma with mono immunotherapy (nivolumab) or combination immunotherapy (nivolumab + ipilimumab). Currently, we identified the ESMO Cutaneous Melanoma guideline and the ESMO Management of Immunotoxicity guideline as evidence-based sources for these infographics. During infographic development, clinicians from the respective hospitals will be consulted to check whether the proposed infographics fulfill their information needs. Additionally, we are monitoring for emerging information gaps/clinical dilemmas as the guidelines are formalized.

Infographics for Patients

We have identified relevant information needs (see Section 4 above and (Peleg,2020)) which could be supported by the CAPABLE patient app. Topics include diagnoses, treatments and side-effects. If available and feasible, information from evidence-based sources will be integrated in the CAPABLE patient app, or referred to (linked to) from the CAPABLE patient app. Additionally, we will support patients with information acquisition by presenting patient-friendly infographics in the CAPABLE patient app. Specifically, infographics displaying risks of immune-related adverse events will be developed.

We identified relevant sources for these infographics. Table 7.4.1.1.1 presents an overview of immune-related adverse events and corresponding relevant publications. While developing these infographics, recommendations will be taken into consideration from Trevena et al. (2013) and Bonna et al. (2021), describing the current best practice for presenting probabilities and quantitative information in patient decision aids. We will review the resulting information with a nurse practitioner and support consultant, who are responsible for providing educational information to patients at NKI.

If a patient reports a symptom in the CAPABLE patient app that is associated with educational material, the app will inform them that there is educational material about the symptom in the library. If possible, this will also include a link to the educational material, for example "See library for more info". This addresses the patients' need for more detailed information on side effects that was identified in Section 4 in (Peleg, 2020)

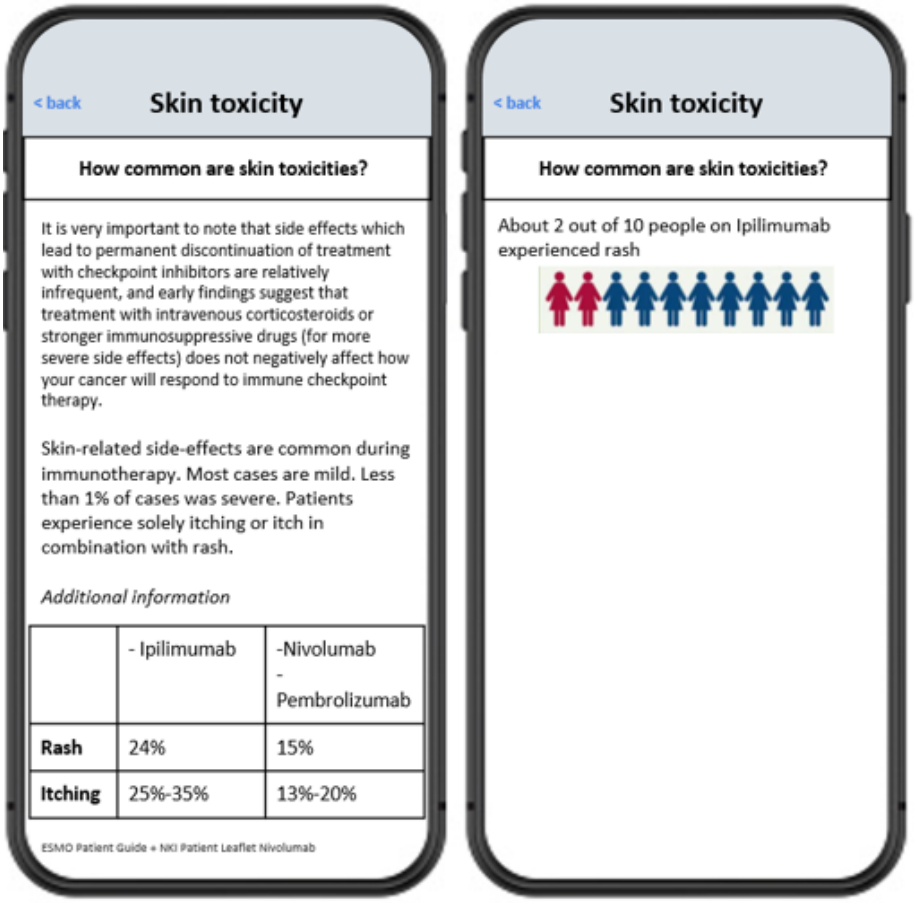

Figure 7.4.1.1.1. Mock-ups of patient education and infographics displaying risks of immune-related adverse events.

Table 7.4.1.1.1 Publications to be used for development infographics immune-related adverse events

Type of immune-related toxicity	Author (Year)	Title
Immune-related skin toxicity	Hodi et al. (2010)	Improved Survival with Ipilimumab in Patients with Metastatic Melanoma
	Wolchock et al. (2010)	Ipilimumab monotherapy in patients with pretreated advanced melanoma: a randomised, double-blind, multicentre, phase 2, dose-ranging study
	Lacouture et al. (2014)	Ipilimumab in patients with cancer and the management of dermatologic adverse events
	Larkin et al. (2015)	Combined Nivolumab and Ipilimumab or Monotherapy in Previously Untreated Melanoma
	Boutros et al. (2017)	Safety profiles of anti-CTLA-4 and anti-PD-1 antibodies alone and in combination
Immune-related endocrinopathies	Hodi et al. (2010)	Improved Survival with Ipilimumab in Patients with Metastatic Melanoma
	Larkin et al. (2015)	Combined Nivolumab and Ipilimumab or Monotherapy in Previously Untreated Melanoma
	Eggermont et al. (2016)	Prolonged Survival in Stage III Melanoma with Ipilimumab Adjuvant Therapy
	Robert et al. (2015)	Nivolumab in previously untreated melanoma without BRAF mutation
	Motzer et al. (2015)	Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma
	Robert et al. (2015)	Pembrolizumab versus Ipilimumab in Advanced Melanoma
Immune-related hepatotoxicity	Larkin et al. (2015)	Combined Nivolumab and Ipilimumab or Monotherapy in Previously Untreated Melanoma
	Robert et al. (2015	Pembrolizumab versus Ipilimumab in Advanced Melanoma
Gastrointestinal toxicity	Gupta et al. (2015)	Systematic review: colitis associated with anti-CTLA-4 therapy
	Horvat et al. (2015)	Immune-Related Adverse Events, Need for Systemic Immunosuppression, and Effects on Survival and Time to Treatment Failure in Patients With Melanoma Treated With Ipilimumab at Memorial Sloan Kettering Cancer Center
Pneumonitis	Rizvi et al. (2015)	Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial

	Topalian et al. (2012)	Safety, Activity, and Immune Correlates of Anti–PD-1 Antibody in Cancer
	Topalian et al. (2014)	Survival, Durable Tumor Remission, and Long-Term Safety in Patients With Advanced Melanoma Receiving Nivolumab
	Robert et. al (2015)	Nivolumab in previously untreated melanoma without BRAF mutation
	Boutros et al. (2017)	Safety Profile of Nivolumab Monotherapy: A Pooled Analysis of Patients With Advanced Melanoma
	Garon et al. (2015)	Pembrolizumab for the Treatment of Non–Small-Cell Lung Cancer
	Hazarika et al. (2017)	U.S. FDA Approval Summary: Nivolumab for Treatment of Unresectable or Metastatic Melanoma Following Progression on Ipilimumab

7.1.4.2. Retrospective Data Statistics

Statistical summaries for Retrospective Data will include:

Descriptive statistics for baseline characteristics and outcomes in the data. Examples can be seen in Figure 7.4.2.1.1 (a-d)

- Pie-Charts/Bar-Plots for categorical variables
- Histograms for continuous variables with Kernel Density Estimation(KDE)
- Box-and-Whisker plots for presenting quartiles of continuous variable allowing comparison between variables

Variables' Correlation

- Scatter Plots for pairs of variables (e.g. age at treatment start vs. comorbidities index)
- Correlation Matrix with heatmap

Non-parametric Survival Analysis Methods for different time-to-event outcomes: Overall Survival, Progression Free Survival, Time-to-treatment and their comparison for different variables (treatment type, age, etc)

- Kaplan-Meir (Kaplan,1958) (see examples in Figure 7.4.2.1.2)
- Nelson-Aalen (Nelson, 1972 ;Aalen 1978)

Clustering methods like Principle Components Analysis (PCA), K-Means and Hierarchical clustering presented with cluster visualization techniques ,like tSNE (Van der Maaten,2008), and descriptive statistics for each cluster.

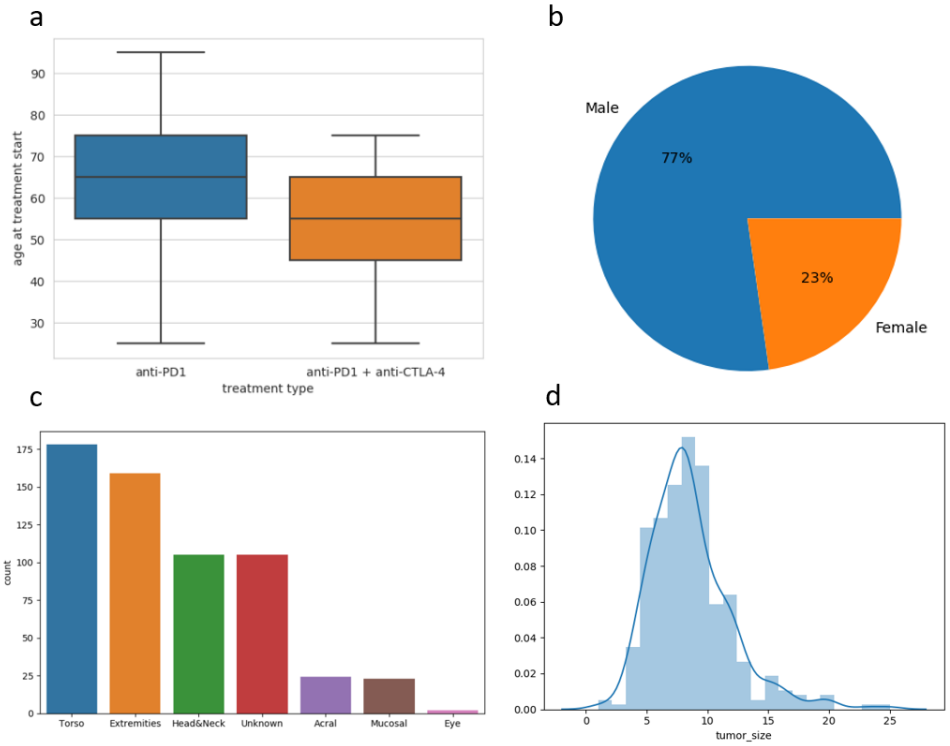

Figure 7.4.2.1.1. (a) Box-and-Whisker plot for comparing age at treatment start for two treatment types for Melanoma data (b) Pie-chart for Male and Female patients distribution in Renal cancer data set (c) Bar-chart for Melanoma primary location distribution (d) Histogram + KDE for tumor size in Renal cancer data set

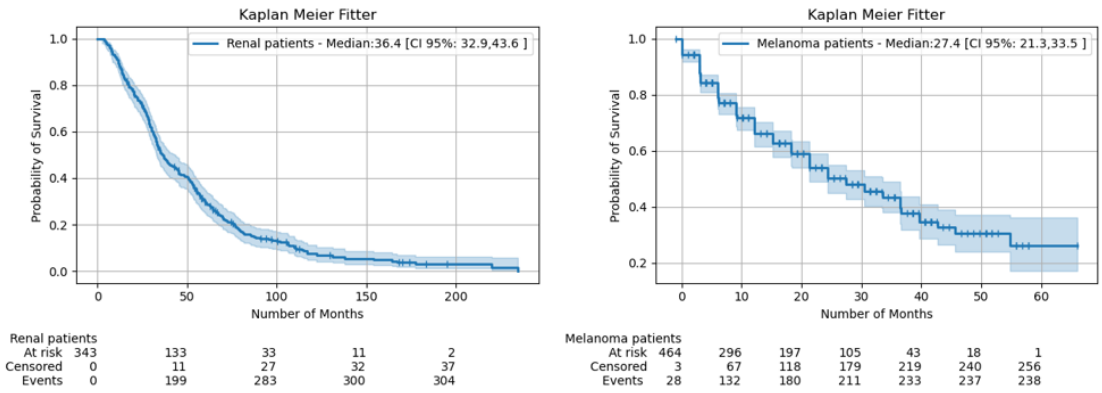

Figure 7.4.2.1.2. Kaplan Meier Fitter calculated for Melanoma and Renal data sets

7.2. High-Level Components

7.2.1. Physician DSS

The Physician DSS component provides decision support for Physician users of the CAPABLE system. It uses the Deontics Engine (see Fig. 7.2.1.1) to provide decision support based on executable clinical logic in the form of PROforma CIGs. The Physician DSS interfaces between the rest of the CAPABLE architecture (in particular the Case Manager and Data Platform) and the Deontics engine, and manages event handling, data synchronization and CIG execution.

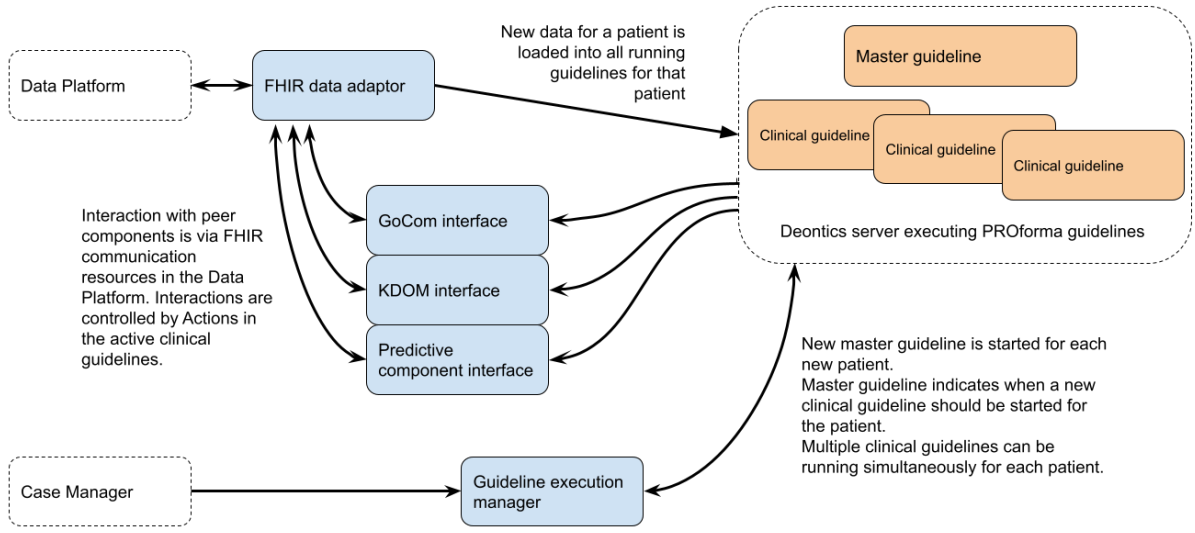

Figure 7.2.1.1. Overall view of Physician DSS

The Physician DSS uses the Case Manager to receive notifications of new patients entering the CAPABLE system, and of any changes in live clinical data for those patients.

A master pathway guideline is used to determine what clinical guidelines (clinical knowledge encoded as PROforma guideline files) should apply to any patient in the system. When a new patient is registered in the Data Platform, a new instance of the master pathway guideline is started on the Deontics server, dedicated to that patient. As any new clinical data for the patient becomes available in the Data Platform the master pathway guideline is updated, and any recommendations from the guideline are pulled from the Deontics server.

The primary purpose of the master pathway is to make recommendations for appropriate individual clinical guidelines. When a clinical guideline is recommended an instance of that guideline is started on the Deontics server for the patient.

When the Case Manager notifies the Physician DSS of any new clinical data available for a patient that is enrolled on a clinical guideline, all available patient data is pulled from the Data Platform and the relevant clinical pathways are updated on the Deontics server. Any new guideline recommendations are retrieved.

Clinical guidelines may recommend actions (such as tests or procedures) or changes in medication for a patient. These are translated to appropriate FHIR resources for storage in the Data Platform, however whether any of these recommendations should be carried out will be subject to analysis for potential conflicts or interactions by the GoCom component, and in any case the final decision will always rest with the clinician. These FHIR resources are therefore marked as "provisional", and GoCom is notified of new recommendations to be analysed.

7.2.2. Virtual Coach

Virtual Coach aims at providing a comprehensive support (coaching) to patients. The support covers both *clinical* and *non-clinical* aspects of care. The clinical support relies on plan models representing patient-oriented CIGs together with the symptom model, while the non-clinical support is concerned with patient well-being and education, and typically it employs the virtual capsule model, the infographic model, and personalized prediction models. In some

situations we also consider plan models for providing non-clinical support, e.g., to control the process of collecting data related to the quality of life through standardized questionnaires.

Virtual Coach has been already presented in Section 6.11 in (Peleg, 2020) where we focused on plan models, their extended definition using meta-properties and execution in cooperation with the Deontics Engine. Below we summarize this part, and then focus on the new functionality, such as invoking personalized prediction models or specialized algorithms (e.g., set cover aimed at optimization) embedded into the Virtual Coach.

In Fig. 7.2.2.1 We show a high-level scheme of Virtual Coach operations related to execution of tasks from a patient-related CIG. In this scheme we also refer to the Patient App to clearly indicate interactions with the patient. Patient App and Physician Dashboard (see Section 7.3.4) are not included in the AI framework, however, they act as a view or presentation layer that delivers outputs from the high-level support components to the users. Handling of a task depends on its type (enquiry, action or decision -- see Section 7.1.1 for details). Most complex processing is required by enquiry tasks that are concerned with obtaining data from various sources. Note that for the brevity of presentation we assume a single data item is associated with an enquiry task, while in practice there are usually multiple items. Depending on the indicated data source, Virtual Coach may invoke a specific algorithmic model or a prediction model to obtain a value of the data item (information about a specific model is provided as part of meta-properties associated with the data item).

Virtual Coach is designed and implemented following the actor model (Charousset, 2013) where functionality is divided into multiple independent entities that act in parallel. Following this paradigm, specialized computational algorithms are implemented as dedicated actors (we refer to these actors as *computational actors*) that are instantiated and run when necessary. Moreover, we also introduce actors that act as wrappers around personalized prediction models and apply these models to patient data available in the Data Platform.

One of the computational actors frequently used in practice is an optimization actor that applies the set cover technique to establish a recommended set of virtual capsules based on the well-being goals assigned to the patient and their hobbies. Specifically, it uses the virtual capsule model as one of its inputs. However, ontological models can be also used separately to provide values of a specific data item, e.g., an ontological model defining patient-reported symptoms may be queries to obtain a reporting frequency for a given symptom (see Fig 7.2.2.1).

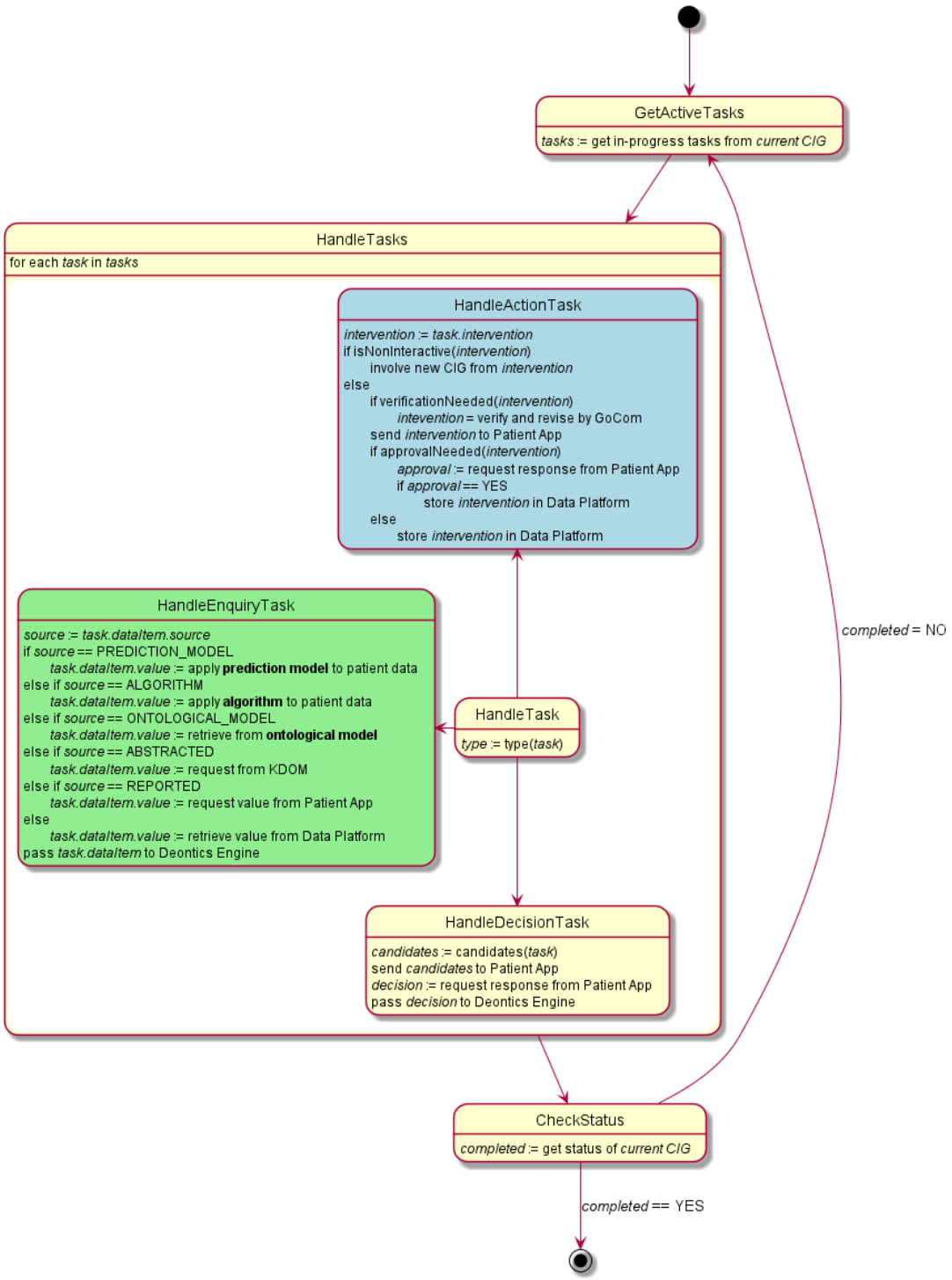

Figure 7.2.2.1. A high-level scheme of Virtual Coach operations

Values of data items may also be obtained from KDOM (see Section 7.3.3), if they need to be abstracted from available information, requested from a patient through reporting facilities offered by the Patient App, or retrieved from the Data Platform. Here we should note Virtual Coach supports both *non-solicited* (spontaneous) reporting (the patient reports a symptom without being explicitly asked), as well as *solicited* reporting.

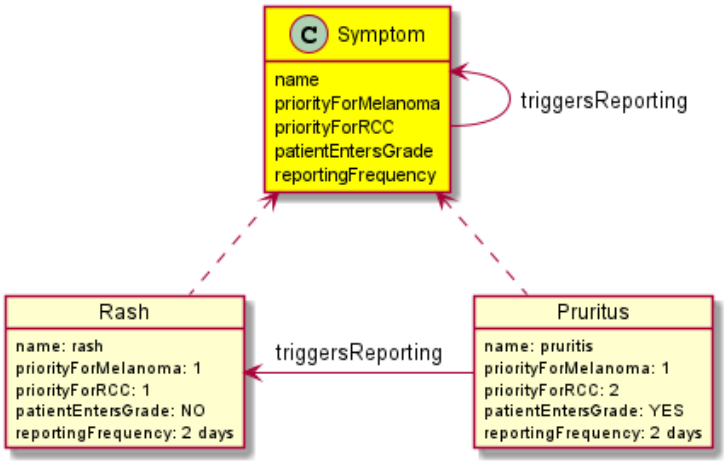

Figure 7.2.2.1 Part of the symptom model

Processing of an action task starts with checking whether the associated intervention is interactive or not. Non-interactive intervention translates into starting another CIG -- this allows for implementing master CIGs, mentioned also in the context of Physician DSS. In case of interactive intervention (requiring interaction with the patient), Virtual Coach checks whether the intervention needs to be processed by GoCom. If this is the case, GoCom looks for possible interactions caused by the intervention and revises it if necessary (the intervention reported back by GoCom may be different than the initial one). Then, the intervention is delivered to the patient via the Patient App. If it requires explicit approval, Virtua Coach requests it from the patient. Finally, if the intervention has been approved, it is stored in the Data Platform. Approval is required only for selected interventions, such as pharmacological or capsule recommendations, where the CAPABLE system needs to know the approval for proper further processing (e.g., displaying related reminders). Simple interventions associated with displaying hints do not need to be approved.

Handling decision tasks is similar to handling solicited reporting. The patient is presented with a set of decision candidates (together with supporting and opposing arguments) via the Patient App, and the selected candidate is reported back. We should note that in order to properly divide responsibilities related to decision making, relevant decisions that may affect a patient's health (e.g., prescription of a specific as-needed drug) are included in physician-related CIGs and are made "manually" by physicians. On the contrary, patient-related CIGs include minor decisions (e.g., starting a prescribed as-needed drug after observing specific symptoms) that are made automatically, thus relieving the patient from the decision burden. In such cases the patient is only asked to confirm or reject the recommended intervention.

In addition to executing CIGs as described above, Virtual Coach can also run pro-actively some algorithmic models. These models may represent strategies for generating reminders or notifications -- due to their characteristic it is difficult to represent them as plan models in PROforma. Such algorithmic models also have access to ontological and personal prediction models and to data available in Data Platform.

7.2.3. Auxiliary Components

This section presents an overview of Auxiliary components. They support both high- and low-level components in providing and delivering support. More details about these components can be found in (Lanzola, 2021).

7.2.3.1. Data Platform

The main objective of CAPABLE Data Platform is to provide a persistent layer where to store and fetch all patients-related data. To guarantee a state-of-the art level component we chose Observational Medical Outcomes Partnership (OMOP) as a Common Data Model (CDM) and HL7 Fast Healthcare Interoperability Resources (FHIR) to manage data exchange (both input and output). All project's data will be exchanged (between data producers and Data Platform and between Data Platform and data consumers) as FHIR Resources using REST APIs. The development of such an API layer is one of the developed products of the project: it is aimed at receiving requests (both to read and write) for standard FHIR resources, fetch/write the OMOP CDM accordingly and create FHIR-compliant responses. Another custom functionality of the CAPABLE Data Platform is establishing a private notification channel towards Case Manage - another core component of the system.

7.2.3.2. Case Manager

To prevent components from continuously scanning the blackboard and facilitating their activation, CAPABLE provides an additional one called Case Manager, specifically designed to coordinate the component efforts and activate the right ones only when needed. The Case Manager guides the reasoning process through an event triggering strategy, where events are combinations of facts regarding the patient's clinical status. The Case Manager monitors the Data Platform to detect the occurrence of such events, so that the interested component can be promptly notified, possibly resuming a reasoning process that has previously been "paused" waiting for the arrival of new data. Thus, the Case Manager is the only component allowed to send messages directly to any other one, promoting a clear modular design and a uniform communication protocol.

7.2.3.3. Knowledge Data Ontology Mapper (KDOM)

Clinical Practice Guidelines (CPGs) contain evidence-based recommendations intended to improve patient care. Specifying CPGs formally as Computer-Interpretable Guidelines (CIGs) enables their execution over a patient's Electronic Health Record (EHR) to deliver patient-specific recommendations, supporting clinicians. Knowledge Data Ontology Mapper (KDOM) allows mapping of abstract medical concepts from CIGs to EHR, that will allow querying EHR data in terms of clinical concepts. Hence, KDOM helps the physician to get better and accurate diagnose results by executing abstractions (abstraction is similar to query which get the patient medical data and execute it on CIG, e.g. "Diarrhea for the last 3 days" abstraction) and return the abstraction result back to the EHR.

7.2.3.4. UI Components

UI components include: Front-End Server, Clinician's Dashboard and Patient's Mobile Application

Front-End Server Component provides users of the system (patients and doctors) with tools to access the CAPABLE system according to their needs and roles. The main two objectives of the component are:

- To manage the connection of the external clients (mobile and web) to the CAPABLE system functionality of Data Platform, Case Manager
- To manage the users (patients and physicians) notification functionality.

The main objective of the web application of **Clinician's Dashboard** is to provide physicians to enroll/import lists of new patients, to access the list of existing patients from the web browser, to access each patient profile and treatment reports for its analysis and modification.

The main objective of the **Patient's Mobile Application** is to provide patients with the tool where they can:

- Manage their profile
- Report symptoms
- Receive treatment recommendations, so that they could manage them at home
- Track their condition using Bluetooth Low Energy (BLE) connected devices
- Receive educational tips on side effect's management.

8. Glossary

BLE	Bluetooth Low Energy
BoK	Bag of Keywords
CIG	Computer-interpretable Clinical Practice Guideline
CPG	Clinical Practice Guideline
DSS	Decision Support System
EHR	Electronic Health Record
FHIR	Fast Healthcare Interoperability Resources
HCP	Healthcare Professional
ICSM	Instituto Clinico Scientifico Maugeri hospital
NKI	Netherlands Cancer Institute
OWL	Web Ontology Language
OS	Overall Survival
PFS	Progression Free Survival
PRO	Patient-reported Outcome
PWD	Real World Data

9. References

Aalen, O. (1978): Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pp.701-726.

ASUS Life Corporation (2021): OmniCare Medical IoT Platform. https://omnicare.asus.life/.

Barkan, E., Gilboa-Solomon, F., Quaglini, S. (2020): CAPABLE D5.1: Data Ready for Modelling and Reasoning Development. Zenodo. https://doi.org/10.5281/zenodo.4540570

Bidargaddi, N., Almirall, D., Murphy, S., Nahum-Shani, I., Kovalcik, M., Pituch, T., Maaieh,T. & Strecher, V. (2018): To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. *JMIR mHealth and uHealth*, 6(11), e10123.

Breiman, L., (2001): Random forests. *Machine learning*, *45*(1), pp.5-32.

Chan, S. W., Sapkota, S., Mathews, R., Zhang, H., and Nanayakkara, S. (2020): Prompto: Investigating Receptivity to Prompts Based on Cognitive Load from Memory Training Conversational Agent. Proceedings of the ACM on Interactive, Mobile, Wearable and *Ubiquitous Technologies*, 4(4), 1-23.

Charousset, D., Schmidt, T.C., Hiesgen, R. Wählisch, M. (2013): Native Actors: A Scalable Software Platform for Distributed, Heterogeneous Environments. Proceedings of the 2013 Workshop on Programming Based on Actors, Agents, and Decentralized Control, ACM, New York, pp. 87-96.

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., (2002): SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, pp.321-357.

Chen, T., Guestrin, C., (2016): Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Cheng, V., Davenport, T., Johnson, D., Vella, K., Hickie I.B. (2019): Gamification in Apps and Technologies for Improving MentalHealth and Well-Being: Systematic Review. *JMIR Mental Health* 6 (6), e13717.

Cox, D.R. (1972): Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B*. **34** (2): 187–220. JSTOR 2985181. MR 0341758

Devlin, J., Chang, M.W., Lee, K., Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Vol. 1 Long Short Pap., Association for Computational Linguistics, Minneapolis, Minnesota, 2019: pp. 4171–4186

Dolsen, M.R., Soehner, A.M., Morin, C.M., Bélanger, L., Walker, M., Harvey, A.G. (2017): Sleep the night before and after a treatment session: A critical ingredient for treatment adherence?. Journal of consulting and clinical psychology, 85(6), 647.

Eisenhauer, E.A., Therasse, P., Bogaerts, J., Schwartz, L.H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M. and Rubinstein, L., (2009): New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European journal of cancer*, *45*(2), pp.228-247.

Fogg, B.J. (2019): Tiny habits: The small changes that change everything. Eamon Dolan Books.

Goyal, N., Fussell, S.R. (2017): Intelligent interruption management using electro dermal activity based physiological sensor for collaborative sensemaking. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 1(3), 1-21.

Hanley, J.A., McNeil, B.J. (1982): The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), pp.29-36.

Harrell Jr, F.E., Lee, K.L., Califf, R.M., Pryor, D.B., Rosati, R.A. (1984): Regression modelling strategies for improved prognostic prediction. *Statistics in medicine*, *3*(2), pp.143-152

Hulth, A., Megyesi, B.B. A Study on Automatically Extracted Keywords in Text Categorization, in: Proc. 21st Int. Conf. Comput. Linguist. 44th Ann. Meet. Assoc. Comput. Linguist., Association for Computational Linguistics, Sydney, Australia, 2006: pp. 537–544

Ishwaran, H., Kogalur, U.B., Blackstone, E.H. and Lauer, M.S. (2008): Random survival forests. *Annals of Applied Statistics*, *2*(3), pp.841-860.

Jowsey, T., Pearce-Brown, C., Douglas, K. A., and Yen, L. (2014): What motivates Australian health service users with chronic illness to engage in self-management behaviour? *Health Expectations*, 17(2), 267-277.

Kaplan, E.L., Meier, P. (1958): Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, *53*(282), pp.457-481.

Khattak, F.K. , Jeblee, S., Pou-Prom, C. , Abdalla, M. , Meaney, C., Rudzicz, F. A survey of word embeddings for clinical text, J. Biomed. Inform. X. 4 (2019)

Lanzola, G., Parimbelli, E., Polce, F., Quaglini, S., Kogan, A., Leizer, R., … Glasspool, D.. (2021): CAPABLE D1.3: Deployment Architecture and Roadmap of the System Iterations [confidential deliverable, accessible to consortium members and Commission Services].

Le, Q., Mikolov, T. Distributed representations of sentences and documents, in: Proc. 31st Int. Conf. Int. Conf. Mach. Learn. - Vol. 32, JMLR.org, Beijing, China, 2014: p. II-1188-II–1196.

Lisowska, A., Wilk, Sz., Peleg, M. (2021a): Catching patient's attention at the right time to help them undergo behavioural change: Stress classification experiment from blood volume pulse. *Proceedings of the 19th International Conference on Artificial Intelligence in Medicine (AIME)*, vol. 12721.

Lisowska, A., Wilk, Sz., Peleg, M. (2021b): Is it a good time to survey you? Cognitive load classification from blood volume pulse. *Proceedings of the IEEE 34th International Symposium on Computer Based Medical Systems (CBMS)* (to be published).

Lundberg, S.M., Lee, S.-I. (2017):A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst*. 30, 4765–4774

Markova, V., Ganchev, T. , Kalinkov, K. (2019): CLAS: A database for cognitive load, affect and stress recognition. *International Conference on Biomedical Innovations and Applications (BIA)*. IEEE, pp. 1–4.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., … Hassabis, D. (2015): Human-level control through deep reinforcement learning. *Nature*, 518 (7540), 529–533.

Musen, M.A., Middleton, B., Greenes, R.A. (2014): Clinical Decision-Support Systems. In: Shortliffe, E., Cimino, J.: *Biomedical Informatics*, Springer

National Library of Medicine (2017): National Drug File - Reference Terminology Source Information. https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/.

National Library of Medicine (2021): RxNav Home Page. https://rxnav.nlm.nih.gov/.

Nelson, W. (1972): Theory and applications of hazard plotting for censored failure data. *Technometrics*, *14*(4), pp.945-966.

Peleg, M., Lanzola, L., Panzarasa, S., Parimbelli, E., Polce, F., Quaglini, S., … Vicente, V. (2020): CAPABLE D2.1: Requirements Table and Use Case Description. Zenodo. https://doi.org/10.5281/zenodo.4540452.

Qiao, Y., Xiong, C., Liu, Z., Liu, Z. Understanding the Behaviors of BERT in Ranking, *ArXiv190407531 Cs*. (2019). http://arxiv.org/abs/1904.07531 (accessed February 3, 2021).

Rosenbaum, P.R.; Rubin, D.B. (1983): The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. **70** (1): 41–55. DOI:10.1093/biomet/70.1.41.

Russel, S., Norvig, P. (2016): Artificial Intelligence: A Modern Approach (3rd Ed). Pearson.

Sacchi, L., Girani, E., Panzarasa, S., Parimbelli, E., Quaglini, S., Veggiotti, N., … Ottaviano, M. (2020): CAPABLE D7.1: Study Plan, Protocols Definition, and Informed Consent/Assent Drafts. Zenodo. https://doi.org/10.5281/zenodo.4540503

Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., Van Laerhoven, K. (2018): Introducing wesad, a multimodal dataset for wearable stress and affect detection. *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pp. 400–408

Sutton D, Fox J. (2003): The Syntax and Semantics of the PROforma Guideline Modeling Language. *Journal of the American Medical Informatics Association*, 10(5), pp. 433–443.

U.S. Food and Drug Administration (2020): openFDA: Open-source APIs. https://open.fda.gov/.

Van Buuren, S., Groothuis-Oudshoorn, K. (2011): mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, **45**(3), 1-67. https://www.jstatsoft.org/v45/i03/.

Van der Maaten, L. and Hinton, G., 2008. Visualizing data using t-SNE. *Journal of machine learning research*, *9*(11).

Wang (2014) Wang, F., Wang, Z., Li, Z., Wen, J.R. Concept-based Short Text Classification and Ranking, in: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* 2014: pp. 1069–1078.

Weibull, W. (1951): A statistical distribution function of wide applicability. *Journal of Applied Mechanics, Transactions of the American Society of Mechanical Engineers*, 18(3), 293–297.

Wright, R.E. (1995): Logistic regression.

W3C (2013): Web Ontology Language (OWL). https://www.w3.org/OWL/.

10. Annexes

Annex 1: AIMAC Forum Analysis - Exploratory Research

In this section we describe our initial effort on applying AI techniques to a big repository of information represented by a cancer patients forum. As a partner of the CAPABLE project, the *Italian Association of Cancer patients, relatives, and friends* (Aimac) is making the data collected in its discussion forum available for several analytics tasks, including the identification of patient needs. This forum offers a virtual place where people facing directly or indirectly cancer can meet, share their experiences and discuss.

We worked on text classification (including topic classification), which refers to the Natural Language Processing (NLP) task of assigning a sentence or document an appropriate category by learning associations and patterns between pieces of the text. In order to be understood by machine learning models, documents, in our case the *posts*, must be transformed into a numerical representation (vectors) before being classified.

Count-based approaches like Bag-of-Words (BoW) represent posts based on word distributions that are estimated by counting the frequency of occurrence of each word of the vocabulary in a post (Wang, 2014), while prediction-based approaches provide representations where the semantic similarities between the words are well-preserved (Khattak, 2019). One of the most advanced ways to produce these vectors, commonly referred to as embeddings (Khattak, 2019), is BERT, which stands for Bidirectional Encoder Representations from Transformers, a language representation model based on transformers that also addresses Italian in its multilingual version (Devlin, 2019).

By combining these two approaches (BoW and embedding), we can build models that can be useful for developing ancillary tools for the AIMAC forum, such as facilitated moderation, search for posts on specific cancer within the entire forum, and search for specific cancer patients' issues (e.g., nutrition, sleep problems, depression, etc.).

Data Preparation and Pre-processing

We analysed 74930 posts, written in Italian by 3955 unique users. Discussions are distributed in 35 predefined subforums based on their main subject. Each of these subforums hosts multiple threads. The average number of threads in a forum is 117, while the average number of posts is 2140, varying from 6 to 35095.

We considered the subforums with the highest participation in terms of opened threads and published posts, excluding the general-purpose ones (e.g., "Introduce Yourself", "Scattered Thoughts", "Staff News"). This leaves us with 8 hot-topics: "Prostatic Cancer", "Pancreatic Cancer", "Lung Cancer", "Colorectal Cancer", "Head and Neck Cancer", "Brain Tumor", "Breast Cancer" and "Liver Cancer".

Forum posts have been filtered considering only the first M messages of each thread. Each post has been truncated to the first W tokens and annotated leveraging the inner structure of the forum itself with a One-vs-Rest approach: positive (on-topic) if the message belongs to the subforum of interest, negative (off-topic) otherwise. The M and W parameters are manually defined, and different configurations have been tested.

Punctuation and special tokens (e.g. HTML tags, escape sequences, emoticons) have been removed from the text. No stemming nor lemmatization has been deployed, as well as stop

words removal, because non-trivial words may still provide useful contextual information for Transformer-based models like BERT (Qiao, 2019).

We split the data randomly with stratification into an 80% training set and a 20% test set, using 20% of the training set as the validation set for BERT fine-tuning.

The NLP pipeline and the results achieved

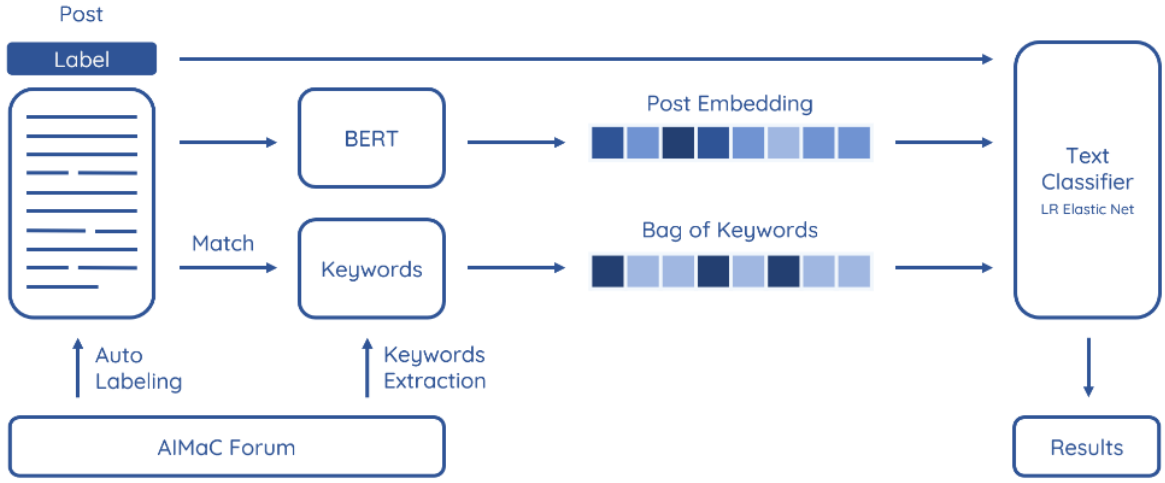

Figure 8.1.1. The pipeline for the topic classification of the messages in the AIMAC forum

Figure 8.1.1 shows our NLP pipeline, which has been implemented with Python 3 on a consumer machine's CPU.

Figure 8.1.2 shows an example of an automatically extracted keywords set for pancreatic cancer and prostate cancer posts. These include a wide variety of medical concepts. We can find treatments and medications like "Abraxane", a prescription medicine used to treat advanced cancer; generic cancer-related terminology like "chemotherapy", "tumor" or "metastasis"; medical procedures like "PSA", a screening test for prostate cancer, and "Gleason", a popular grading system used for prostate cancer prognosis; conditions like "incontinence", a common side-effect in treated prostate cancer patients; target body parts like "prostate" and "pancreas", as well as indirectly involved organs like "liver", which is strictly related to pancreatic adenocarcinomas, as more than 50% of patients with pancreatic cancer have liver metastases at the time of diagnosis.

Figure 8.1.2. Italian keyword clouds for Pancreatic Cancer (top) and Prostatic Cancer (bottom)

Word embeddings for each word of each post have been extracted averaging the last 4 layers of BERT, as described in (Devlin, 2019). For each post, word embeddings have been merged to obtain post embeddings, a document-level representation (Le, 2014). In this work we used the uncased multilingual BERT-Base configuration (https://huggingface.co/bert-base-multilingual-uncased), which consists of 12 encoding layers, 12 attention heads and a hidden size of 768, adding a single linear layer at the end for sequence classification.

Each post has been labeled for binary topic classification with no additional human supervision, leveraging the forum structure itself and the previous, implicit, annotation work done by moderators and users when they decide on which discussion subforum to post their messages. The representation methods described above have been combined in four ways for classification:

1. BERTFT. No keywords, posts are represented using embeddings only. Pretrained representations have been fine-tuned (FT) on a validation set.
2. Bag-of-Keywords. Keywords match representation, posts are mapped using the BoK approach alone.
3. BERTFT+BoK. Post embeddings are concatenated with BoK, resulting in a vector sized 768+K and made of both continuous and binary variables.
4. BERTPT+BoK. Pre-trained (PT) post embeddings are concatenated with BoK without further fine-tuning.

Post representations in turn become the input of a logistic regression, regularized through Elastic Net to control the high number of variables. For BERTFT, we directly used the output of the final layer instead.

Models have been evaluated in multiple configurations, varying the size of the keyword set K. For each configuration, the pipeline has been run 5 times with different seeds to monitor variability and stability of results. The truncation threshold W has been set to 380 tokens, obtained by using the WordPiece sub-word tokenization. This is equal to the 95th percentile of the MMB message length distribution after preprocessing and allows us to speed up computation by cutting only the tail of the distribution.

We evaluated models' performance with respect to their AUC-ROC, accuracy, F1-score and recall. Standard evaluation metrics have been complemented with recall to emphasize the prediction of actual positives.

Experiments have been carried out on multiple topics to check the robustness of the models, changing the subforum of interest and the binary labeling accordingly.

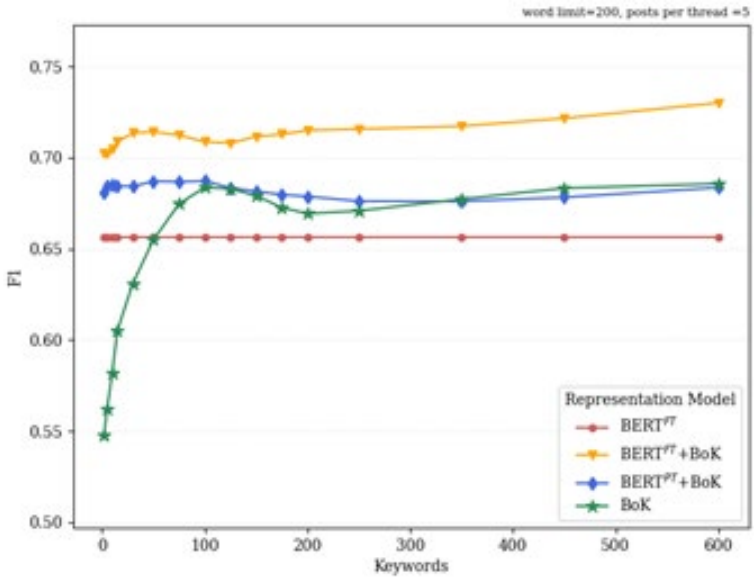

Figure 8.1.3. F1 score trend for the "Breast Cancer" topic

Figure 8.1.3 shows the average F1-score trend for breast cancer" over all the tested values of K. Similar curves have been found for other topics of interest.

In conclusion, we demonstrated that combining Bag of Words strategies like BoK with contextual embeddings like BERT seems to improve the quality of predictions for every topic we tried, providing a relatively inexpensive way to overcome BoW limitations and charting a course towards new feasible approaches for real-world situations involving languages other than English. We plan to use this approach for further analyses on the AIMAC forum.

Annex 2: Questionnaires with scoring details

<table>
<tr><th>QUESTIONNAIRE</th><th>NUMBER OF QUESTIONS</th><th>OUTPUT</th></tr>
<tr><td>EORTC QLQ-C30</td><td>30</td><td>To measure patient's quality of life.

* Item range is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving range = 3.
† (revised) scales are those that have been changed since version 1.0, and their short names are indicated in this manual by a suffix "2" – for example, PF2.

For all scales, the RawScore, RS, is the mean of the component items:
$RawScore = RS = (I_1 + I_2 + ... + I_n)/n$

Then for Functional scales:
$Score = \left\{1 - \frac{(RS-1)}{range}\right\} \times 100$

and for Symptom scales / items and Global health status / QoL:
$Score = \{(RS-1)/range\} \times 100$

Examples:
Emotional functioning
$RawScore = (Q_{21} + Q_{22} + Q_{23} + Q_{24})/4$
$EF\ Score = \{1 - (RawScore - 1)/3\} \times 100$

Fatigue
$RawScore = (Q_{10} + Q_{12} + Q_{18})/3$
$FA\ Score = \{(RawScore - 1)/3\} \times 100$</td></tr>
<tr><td>EORTC QLQ INFO-25</td><td>25</td><td>To provide information such as receiving disease information, medical testing, treatment and other support services.

* "Item range" is the difference between the possible maximum and the minimum response to individual items. Most items are scored 1 to 4, giving range = 3.
[a] Items 50, 51, 53, and 54 should be coded yes = 1, no = 2
[b] Items 53 and 54 have an extra question for patients who have replied yes, in which they are invited to specify on which topics they wish to have received more or less information.

Principle for scoring
1) Raw score
Items 50, 51, 53 and 54 should be coded yes = 1, no = 2.
Take into account that the scoring of questions 50 and 51 must be reversed prior to statistical analysis.

For each multi-item scale, calculate the average of the corresponding items:
$Raw\ Score = RS = \left\{\frac{(I_1 + I_2 + ... + I_n)}{n}\right\}$
For each single-item measure, the score of the concerning item corresponds to the raw score.

2) Linear Transformation
To obtain the Score S, standardize the raw score to a 0 – 100 range using the following transformation:
$Score: S = \left\{\frac{(RS-1)}{range}\right\} \times 100$

3) Global score
To obtain a global score of the questionnaire, all scales and single-items must first be converted into a 0 – 100 range as explained above. Then, calculate the average of all the scales / single-item measures:
$Global\ score = \frac{INFODIS+INFOMEDT+INFOTREAT+INFOTHSE+INFODIFP+INFOHELP+INFOWRIN+INFOCD+SATINFO+RECMORE+RECLESS+OVERHELP}{12}$</td></tr>
</table>

<table>
<tr><td>EuroQoL 5D-5L</td><td>5 + scale</td><td>It measures mobility, self-care, habitual activities, pain, discomfort, anxiety and depression.
SCORE NOT AVAILABLE</td></tr>
<tr><td>FACT-G</td><td>27</td><td>General aspects about tumor.
Instructions:* 1. Record answers in "item response" column. If missing, mark with an X
2. Perform reversals as indicated, and sum individual items to obtain a score.
3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the subscale score.
4. Add subscale scores to derive total FACT-G score. The higher the score, the better the QOL.

<table>
<tr><th>Subscale</th><th>Item Code</th><th>Reverse item?</th><th></th><th>Item response</th><th>Item Score</th></tr>
<tr><td rowspan="8">PHYSICAL WELL-BEING (PWB)

Score range: 0-28</td><td>GP1</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td>GP2</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td>GP3</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td>GP4</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td>GP5</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td>GP6</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td>GP7</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td colspan="5">Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____=PWB subscale score</td></tr>
<tr><td rowspan="8">SOCIAL/FAMILY WELL-BEING (SWB)

Score range: 0-28</td><td>GS1</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GS2</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GS3</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GS4</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GS5</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GS6</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GS7</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td colspan="5">Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____=SWB subscale score</td></tr>
<tr><td rowspan="7">EMOTIONAL WELL-BEING (EWB)

Score range: 0-24</td><td>GE1</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td>GE2</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GE3</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td>GE4</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td>GE5</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td>GE6</td><td>4</td><td>-</td><td>_____</td><td>=_____</td></tr>
<tr><td colspan="5">Sum individual item scores: _____
Multiply by 6: _____
Divide by number of items answered: _____=EWB subscale score</td></tr>
<tr><td rowspan="8">FUNCTIONAL WELL-BEING (FWB)

Score range: 0-28</td><td>GF1</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GF2</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GF3</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GF4</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GF5</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GF6</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td>GF7</td><td>0</td><td>+</td><td>_____</td><td>=_____</td></tr>
<tr><td colspan="5">Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____=FWB subscale score</td></tr>
<tr><td colspan="5">TOTAL SCORE:
Score range: 0-108
_____ (PWB score) + _____ (SWB score) + _____ (EWB score) + _____ (FWB score) = _____=FACT-G Total score</td></tr>
</table>
</td></tr>
<tr><td>FACT-M</td><td>51</td><td>See FACT-G for:
• Physical Well-Being (PWB)
• Social/Family well-being (SWB)
• Emotional well-being (EWB) and
• Functional Well-being (FWB)
Additional:</td></tr>
</table>

<table>
<tr>
<td></td>
<td></td>
<td>
Subscale | Item Code | Reverse item? | Item response | Item Score
MELANOMA SUBSCALE (MS)
Score range: 0-64
M1 4 - _____ =_____
M2 4 - _____ =_____
M3 4 - _____ =_____
B1 4 - _____ =_____
ITU4 4 - _____ =_____
An10 4 - _____ =_____
Hep3 4 - _____ =_____
C1 4 - _____ =_____
C6 0 + _____ =_____
M5 4 - _____ =_____
M6 4 - _____ =_____
ITU3 4 - _____ =_____
MS8 4 - _____ =_____
M8 4 - _____ =_____
M9 4 - _____ =_____
HI7 4 - _____ =_____
Sum individual item scores:_____
Multiply by 16: _____
Divide by number of items answered: _____=MS score

Subscale | Item Code | Reverse item? | Item response | Item Score
MELANOMA SURGERY SCALE (MSS)
Score range: 0-32
M10 4 - _____ =_____
M11 4 - _____ =_____
M12 4 - _____ =_____
M13 4 - _____ =_____
M14 4 - _____ =_____
M15 4 - _____ =_____
M16 4 - _____ =_____
M17 0 + _____ =_____
Sum individual item scores:_____
Multiply by 8: _____
Divide by number of items answered: _____=MSS score+

To Derive a FACT-Melanoma total score (this excludes the surgery scale):
Score range: 0-172
_____+_____+_____+_____+_____+ =_____ FACT- M Total score
(PWB score) (SWB score) (EWB score) (FWB score) (MS score)

+ The Melanoma Surgery Scale (MSS) is an independent scale, and therefore not included in the summary TOI and total FACT-Melanoma scores.
</td>
</tr>
<tr><td>NCCN-FACT FKSI-19</td><td>19</td><td>Specific for kidney tumor.
SCORE NOT AVAILABLE (possibly the same as FACT-G)</td></tr>
<tr><td>MST</td><td>3 (2+1)</td><td>To screen patient malnutrition.
The score is the sum of the values of the answered questions.</td></tr>
<tr><td>PHQ9</td><td>9</td><td>It is used for the diagnosis, monitoring and determination of the severity of depression.
The score is the sum of the values of the answered questions.</td></tr>
<tr><td>GAD7</td><td>7</td><td>To evaluate generalized anxiety disorder.
The score is the sum of the values of the answered questions.</td></tr>
</table>