iHELP: Personalised Health Monitoring and Decision Support Based on Artificial Intelligence and Holistic Health Records

George Manias[1], Harm op den Akker[2], Ainhoa Azqueta[3], Diego Burgos[3,9], Nikola Dino Capocchiano[4], Borja Llobell Crespo[5], Athanasios Dalianis[6], Andrea Damiani[4], Krasimir Filipov[7], Giorgos Giotis[6], Maritini Kalogerini[6], Rostislav Kostadinov[8], Pavlos Kranas[3,9], Dimosthenis Kyriazis[1], Artitaya Lophatananon[10], Shwetambara Malwade[11], George Marinos[1], Fabio Melillo[12], Vicent Moncho Mas[5], Kenneth Muir[10], Marzena Nieroda[10], Antonio De Nigro[12], Claudia Pandolfo[12], Marta Patiño-Martinez[3], Florin Picioroaga[13], Aristodemos Pnevmatikakis[2], Shabbir Syed-Abdul[11], Tanja Tomson[14], Dilyana Vicheva[8], and Usman Wajid[15]

[1] *University of Piraeus,* Piraeus, Greece
[2] *Innovation Sprint,* Brussels, Belgium
[3] *Universidad Politécnica de Madrid,* Madrid, Spain
[4] *Fondazione Policlinico Universitario Agostino Gemelli IRCCS,* Rome, Italy
[5] *Hospital de Dénia-MarinaSalud,* Alicante, Spain
[6] *Athens Technology Centre,* Athens, Greece
[7] *KODAR Systems,* Plovdiv, Bulgaria
[8] *Medical University Plovdiv,* Plovdiv, Bulgaria
[9] *LeanXcale,* Madrid, Spain
[10] *University of Manchester,* Manchester, United Kingdom
[11] *Taipei Medical University,* Taipei, Taiwan
[12] *Engineering Ingegneria Informatica SpA,* Rome, Italy
[13] *Siemens SRL,* Brasov, Romania
[14] *Karolinska Institutet,* Solna, Sweden
[15] *Information Catalyst for Enterprise,* Northwich, United Kingdom

gmanias@unipi.gr, hopdenakker@innovationsprint.eu, aazqueta@fi.upm.es, diego.burgos@leanxcale.com, nikoladino.capocchiano@policlinicogemelli.it, borja.llobell@marinasalud.es, T.Dalianis@atc.gr, andrea.damiani@policlinicogemelli.it, krasimir.filipov@kodar.net, G.Giotis@atc.gr, m.kalogerini@atc.gr, rostikosti@abv.bg, pavlos@leanxcale.com, dimos@unipi.gr, artitaya.lophatananon@manchester.ac.uk, sv14kekade@tmu.edu.tw, gmar@unipi.gr, fabio.melillo@eng.it, vicent.moncho@marinasalud.es, kenneth.muir@manchester.ac.uk, marzena.nieroda@manchester.ac.uk, antonio.denigro@eng.it, claudia.pandolfo@eng.it, mpatino@fi.upm.es, florin.picioroaga@siemens.com, apnevmatikakis@innovationsprint.eu, drshabbir@tmu.edu.tw, tanja.tomson@ki.se, dilyanav@yahoo.com, usman.wajid@informationcatalyst.com

Abstract—**Scientific and clinical research have advanced the ability of healthcare professionals to more precisely define diseases and classify patients into different groups based on their likelihood of responding to a given treatment, and on their future risks. However, a significant gap remains between the delivery of stratified healthcare and personalization. The latter implies solutions that seek to treat each citizen as a truly unique individual, as opposed to a member of a group with whom they share common risks or health-related characteristics. Personalisation also implies an approach that takes into account personal characteristics and conditions of individuals. This paper investigates how these desirable attributes can be developed and introduces a holistic environment, the iHELP, that incorporates big data management and Artificial Intelligence (AI) approaches to enable the realization of data-driven pathways where awareness, care and decision support is provided based on person-centric early risk prediction, prevention and intervention measures.**

Keywords—Holistic Health Records (HHRs), Pancreatic Cancer, Artificial Intelligence

I. INTRODUCTION

Cancer is one of the main causes of death in Europe and worldwide, after cardiovascular diseases. According to the World Health Organization (WHO), cancer represents the second most important cause of death and morbidity in Europe with more than 3.7 million new cases and 1.9 million deaths each year [1]. Among the different types of Cancer, Pancreatic Cancer (PC) is the seventh leading cause of cancer-related-deaths worldwide, with higher incidence in developed countries. As per the GLOBOCAN 2020 estimates, Pancreatic Cancer ranked the 14th most common Cancer in the world accounting to 495,773 new cases and 466,003 deaths (4.7% of all cancer deaths) in 2020 [2]. On top of this, Pancreatic Cancer has been identified across the European Union as the 4th most common cause of death from cancer in 2018 [3], while in the same year Western Europe recorded highest mortality rates (7.6 per 100,000 people), followed by Central and Eastern Europe (7.3), and further, Northern Europe and North America (equally: 6.5) [4]. More than half of deaths for Pancreatic Cancer were accounted by most developed countries (52.3%, 226,272 of deaths) [5]. Moreover, the incidence and mortality for Pancreatic Cancer is higher among men and is correlated with an increasing age [6].

To this end, a key rational to address this particular type of Cancer is the poor survival rate; worldwide for Pancreatic Cancer the approximate survival rate is 5%. Taking into consideration the high incidence and mortality rate, early risk assessment and early detection can be the key to reducing mortality. Coupled with that, targeted preventions and intervention measures, such as early screening programs for

high-risk individuals, can help save lives. Pancreatic Cancer usually lacks clear symptoms at its early stages. As the tumour progresses, it presents as a gradual onset of non-specific symptoms including jaundice, pruritus, weight loss, light-coloured stools, abdominal pain and fatigue. Furthermore, symptoms such as anorexia, early satiety, nausea, dyspepsia, etc. may also be presented, which can be differentiated into a broad number of diseases. Although, newer screening techniques are being evaluated, healthcare professionals still base the detection of Pancreatic Cancer mainly through blood markers such as CA19-9, CA-50, SPAN-1, DUPAN-2, cell surface-associated mucins (MUC), carcino-embryonic antigen and heat shock proteins. However, these tests have not been well evaluated yet [7]. Multidetector computed tomography (MDCT), magnetic resonance imaging (MRI) and endoscopic procedures are also performed for diagnosing Pancreatic Cancer, nonetheless with limited diagnostic ability.

In this respect, the screening tests until now have been focused to determine pre-invasive lesions, rather than early detection of Pancreatic Cancer. Clinical experts agree that the early identification of risks and pre-emptive mitigation can reduce the likelihood of Cancer. Once an invasive Pancreatic Cancer develops, it could rapidly spread beyond the Pancreas, limiting use of markers for invasive Pancreatic Cancer. These factors mark the importance of personalised and early prevention measures. Taking into consideration the cause and enlightening the risk factors that can help to identify high risk individuals are necessary for the primary prevention of the disease. Reducing the risk by acting on the modifiable risk factors that relate to lifestyle, behaviours, social interactions (e.g. reduction in smoking, alcohol, and red meat consumption, increasing intake of fruits and vegetables, and regular physical exercise), are among the best preventive measures against Pancreatic Cancer related risk reduction.

At the same time, vast and divergent amounts of health-related data are being produced, gathered and stored in various formats across multiple types of systems and data sources. In addition to typical medical systems at hospitals, medical centres and health organisations, new ICT services (e.g. that make use of Internet of Things (IoT) devices, sensors and mobile applications for health monitoring or decision support) also collect vast amount of health related data. Hence, the successful exploration and interpretation of all these data play a vital role in exploring opportunities of such data in mitigating disease risk factors [8]. Currently all these systems operate independently, thus the potential value from the exploitation of data in a combined and holistic way is limited. Due to lack of system/data integration and in many cases due to inadequate integration of the technology, it is either very difficult or generally impossible for healthcare professionals to provide a person and disease centred approach for health monitoring, risk predictions, preventions and personalised interventions.

Addressing these challenges requires a robust approach for data integration as well as advance tools for data exploitation that can support not only the development of decision support mechanisms (for personalised Early Risk Prediction, Prevention and Intervention), but also contribute towards advancing the proficiency of healthcare providers through increased knowledge about diseases and interpretation of symptoms and effects. The approach for robust data integration should consider additional health determinants (to typical medical history) at individual level, as highlighted by the World Health Organization, including age, physical, social, genetics and economic aspects, as well as relationships with friends and family. In parallel, advance data exploitation tools should leverage the potential of Artificial Intelligence (AI) to extract meaningful insight from complex integrated data. The application of AI-based machine learning, deep learning and systems neuroscience techniques can help build powerful decision support systems that can perform risk predictions based on health, age, gender, lifestyle, environmental, genomic, diet and other factors. AI is already being used to detect diseases e.g. the use of AI is enabling the analysis and translation of mammograms 30 times faster with 99% accuracy, reducing the need for unnecessary biopsies [9]. The latter results in timely risk assessment and more accurate interventions to be offered to the target persons. On top of this, the availability of high-tech wearables and mobile applications provide a user friendly delivery mechanism to offer AI-based personalised early risk identification and decision support (regarding prevention and intervention measures). The implementation of novel user-profiling techniques will facilitate the provision of not only personalised advice but also a personalised mode of delivering this advice to ensure the compliance with the advice and continuous use of the electronic device by the patient. The application of AI through wearable and mobile applications can support an increase in health literacy as well as better monitoring and detection of health and wellbeing issues at earlier and more treatable stages [10] – [12]. The exploitation of integrated data and application of AI not only benefits healthcare providers but also supports the creation of recommendations or guidelines for policy makers (e.g. public health authorities in Europe). Therefore the challenge is not only how to fuse and analyse information from various sources, but also how to provide different analytic views for the benefit of different stakeholders i.e. individuals (or citizens), healthcare providers (or Medics) and the key decision makers (or policy makers).

In this context, this paper presents the main research challenges and an architecture of an overall integrated environment for addressing the challenge of providing multi-dimensional analytic views on integrated data and extract value for different stakeholders in the healthcare domain. The proposed environment (namely iHELP) and approach for the convergence of Holistic Health Data and AI seeks to support the study of Pancreatic Cancer with particular focus on the early detection of associated risks and on the design of personalised prevention and intervention measures that can be delivered using personalised technological means. The analysis of the implications of specified preventative measures and interventions is another area for the convergence of holistic health data and AI. The latter will provide new insights into the psychosocial aspects of the disease. On top of this, the preventative measures and interventions developed in the contexts of this environment will aim at reducing the likelihood of developing Pancreatic Cancer in people that are at high risk.

II. RELATED WORK

The standardised integration of data and recalibration of learning models and analytics is one of the key aspects and challenges of the modern eHealth systems. Ambitious R&D initiatives have resulted in integrated eHealth, eCare and Ambient Assisted Living (AAL) systems, many of those with user support characteristics. It is worth to mention that

integrated approaches so far have failed to break the entry barriers in the healthcare sector and therefore the research community is seeking for the innovation ingredients that can create new experiences and services for the end-users. Two pioneering approaches of such solutions are being presented below, which set the baseline for the design and implementation of the iHELP platform and approach. The universAAL IoT platform [13] is an open-source platform that enables interoperability and rapid development of innovative IoT solutions. It empowers consumers to create a fully customized experience and lifestyle, accelerating and growing the potential for interconnectivity. The rapid evolution of universAAL IoT ensures that new features and applications are regularly added to integrated systems, bringing endless possibilities to the middleware platform. Intelligent products and services are able to instantly interact, enabling infinite possibilities for the IoT. The latter demonstrates the increasing need for the implementation of completely unique lifestyle environments and platforms in the modern healthcare domain, where users will be able to effortlessly share valuable data between devices and systems. The eWALL platform [14] resides at the intersection of the concepts of AAL, Enhanced Living Environments (ELEs) and Ambient Intelligence (AmI). eWALL, similarly as AAL, is utilising ICT technologies to facilitate independent living of elderly people by building ELEs to offer them support in everyday activities. The objective of eWALL is to support independent living, compensating for prevailing age-related physical impairments, leading to a significant prolongation of their functional capacity, a delay in institutionalisation, increased autonomy and improved QoL. eWALL is envisaged as a holistic infrastructure model and a wall mounted system, which includes a modular cloud-based platform capable of integrating various off-the-shelf and custom devices and a supporting Sensing Environment based in the primary user's home and responsible for explicit and implicit interaction with the primary user. The eWALL system comprises two distinctive logical entities, i.e. the eWALL Home and the eWALL Cloud Environments. The eWALL Home Environment resides in patients' homes and hosts the sensing capabilities of the entire system. Market available off-the-shelf devices provide all sensing capabilities within. The Device Gateway (GW) is responsible for seamless integration of the different devices and transparent integration with the eWALL Cloud Environment. The "Processing component" is responsible for quick interpretation of sensing data for critical purposes such as presence detection, fall detection etc. The "Home DB" keeps track of patients' data temporarily before it is transferred in the "Cloud DB" for permanent storage. The eWALL Cloud Environment is a centralized system responsible for hosting and analysing all patients' data coming from patients' homes. The system includes aspects such as gateways for interconnection, notification ability, reasoning ability and application incorporation.

Another key aspect is the harmonisation and integration of data in a holistic fashion; across multiple data stores and locations. To this end, electronic health records (EHRs) facilitate the management of patient data, for health care as well as other purposes, across any kind of institutional, regional or national border [15]. Hence, data can be stored, shared, processed and analysed in more effective ways for exploring trends more thoroughly, quality assurance, disease surveillance, public health monitoring and research. Nonetheless, what is required refers to standardisation of data

models to support interoperability between all the medical information systems that stored clinical information about patients. Although several standards are developed that enable structured clinical content for the purpose of exchange such as, HL7 FHIR [16], openEHR [17], and epSOS [18]; the use of different standards across different organisations is quite common. This leads to the needs for standardised security and data sharing models. Many efforts are concentrated into standard system architectures with the aim to provide standardised access and data sharing models to different stakeholders. However, consensus has not been reached in relation to security related standards, thus there is a need to develop more standardised data sharing models that allow better linking of RTD efforts across the health continuum to be able to use collective knowledge for the wider benefits of the involved stakeholders. To this end, a complete approach for capturing all health determinants in new structures was proposed by the CrowdHEALTH project, the HHRs [19]. Holistic Health Records (HHRs) are structured health records that may include several types of information that are relevant to a patient's health status, such as laboratory medical data; clinical data; lifestyle data collected by the patient or related people; social interactions data; physiological and behavioural data collected by medical devices and sensors. An HHR represents a single data object capturing individualised aggregated data from different sources such as healthcare records (from information systems in the healthcare domain) as well as citizen specific information such as age group, ethnicity, gender, lifestyle, fitness/wellbeing, nutrition, etc. and other health-related data.

III. PROPOSED APPROACH AND MAIN CHALLENGES

Artificial Intelligence offers huge advantage over traditional (arithmetic and statistical) analytic techniques and AI algorithms have already demonstrated the potential to perform healthcare tasks such as disease diagnosing faster and more accurately than humans [20] – [21]. Moreover, Machine learning techniques are already used for precision medication [22]. However, where the traditional supervised learning approaches have their advantages on providing more precise decision support, they also have limitations when it comes to early identification of risk or analysis of emerging conditions for which there are not enough training data available [23]. In iHELP the development of more frugal unsupervised learning algorithms will enable the early detection of risks and analysis of factors associated with Pancreatic Cancer based on the analysis of patients' clinical records, which are further mentioned asprimary data and which can provide information on the person's profile, diagnosis, comorbidity and medication. This analysis will be used to identify key risks, associated factors and also predictions (e.g. profile of people that are at high risk of developing Pancreatic Cancer). The outcomes of the analytic functions will be used to identify the individuals or candidates that have a high likelihood of identified risks and that show the trend towards factors that are associated with the risks of Pancreatic Cancer.

Moreover, iHELP seeks to capture, process, and analyse data about patients' lifestyle, which are further mentioned as secondary data. To this end, based on the study of known patients and their data that will be gathered continuously and ubiquitously from their everyday life through questionnaires, sensors, mobile and wearable applications, specific and personalised recommendations (concerning prevention measures and interventions) will be designed that can reduce

the risk for the identified candidates and/or prevent negative consequences on a person's quality of life and thus offer the potential to improve individual health. These recommendations and the mechanisms for their delivery (such as mobile and wearable apps) will be tuned to the needs of individual patients, clustered according to their psychological profiles to ensure compliance, uptake and continuous use of the advising and monitoring systems. The personalised recommendations will be delivered through technologies (such as IoT, mobile and wearable technologies) that offer the means to deliver specific interventions (such as behavioural nudges, lifestyle changes, monitoring and alerting) at individual level [24] – [28]. The use of these technologies is tuned to monitor the impact of prescribed interventions and capture patients' secondary data that are to be processed to yield higher-level information including relationships, interactions, experiences, habits and behaviours that can influence an individual's health, the design of preventative measures or the risk of developing (chronic) conditions. Hence, these data will also be used for the identification of emerging risks as well as the impact of prescribed recommendations (prevention and intervention measures).

To this end, the harmonisation and integration of these two types of collected data, i.e. primary and secondary data, results in HHRs that will be stored in the iHELP platform to support the continuous monitoring and analysis cycle. Within the iHELP platform, the research on advanced AI techniques is complemented with innovative approaches for data modelling, management and interoperability aspects. In this respect, iHELP will deliver an integral clinical-ICT framework that provides standardised data models and procedures to manage heterogeneous datasets and support the exploitation of collective knowledge for clinical decision making and also for improving the quality of life and awareness of health conditions at individual level. To this end, iHELP seeks to deliver an ICT platform with integrated tools that serve the needs of different stakeholders in the healthcare domain by providing the following innovations:

- *Tools to ingest and integrate health related data from heterogeneous sources*, including but not limited to hospitals, medical research organisations, wearable devices, mobile apps and IoT platforms. The integration mechanism considers a standardised and extensible approach, based on HL7 FHIR standard-compatible data structures, to compile the HHRs.

- *Tools that make use of advanced AI techniques to develop compressive models on top of HHRs.* AI-based analytic applications operate on these models to provide personalised decision support in terms of early risk predictions, probabilistic conditions, preventative measures, medical interventions, and behavioural nudges. The potential of advanced AI is leveraged to instantly analyse large number of scenarios and draw responses, decisions or advice based on a specific event or situation – where a situation is seen as a sequence of interrelated events that are identified in the HHR data, gathered and integrated from various sources.

- *Tools that provide Data & Analytics as a Service* towards different stakeholders in the healthcare domain, enabling them to utilise advanced AI data analytics and recommender system techniques –

combined with simulation tools – to aid the delivery of personalised decision support and development of new healthcare pathways and evaluations of strategies for addressing emerging trends in the healthcare domain.

Through the design and implementation of the above proposed platform several key challenges and objectives seek to be addressed. These challenges and the specific frameworks and approaches that will be utilised to address them are being presented and listed below.

A. Exploitation of heterogeneous data sources and knowledge acquisition through a holistic data management platform

The iHELP platform seeks to be implemented in an end-to-end big data management platform in which health analytics (e.g. predictions, risk assessment, causal analysis, health conditions' forecasting) will be performed over large amount of historic data, previously introduced as primary data, and real-time data, previously introduced as secondary data which also are gathered from heterogeneous sources such as sensors, mobile devices and wearables, without having to take care of scalability issues as they will be handled transparently by the platform. Unified analytics will be performed over all the heterogeneous streaming and stored data, offering the ability to share and reuse all this huge amount of data towards the provision of personalised actionable insights. The big data management framework will utilise state-of-the-art developments in master data management, SQL and noSQL technologies and ACID principles to support the realisation of personalised real-time risk prediction, health status monitoring, provision of actionable insights, and health maintenance. Hence, iHELP will enable automated adaptation of care by dynamically extracting knowledge through the combination of real-time facts (e.g. sensor readings), stored historical data for each patient, and analytics performed in a unified way on both these stored and streaming data.

B. Advanced AI-based Health Analytics

iHELP aims at the implementation and evaluation of two types of advanced AI techniques. These AI techniques are able to process large amount of data to unravel hidden insights, such as key risks that are associated with the development of Pancreatic Cancer, based on different profiles of individuals. In this context, neural networks and deep learning techniques are the suitable candidates for training and testing of supervised learning models based on the availability of large amounts of historical data [29]. The second type of AI relates to a new breed of more frugal AI algorithms that are able to perform reliable assessment and deliver early risk predictions based on unsupervised models that require limited or no training data. The latter, frugal AI algorithms will create an advanced sensing environment (based on data gathered from mobile, wearable and social media platforms) with intelligent reasoning, which will be able to preserve and interpret physical, cognitive, mental and social data in real-time.

C. Efficient prevention and intervention models for Pancreatic Cancer through the utilisation of IoT, mobile apps and wearable technologies

iHELP seeks to deliver a set of innovative multi-level decision support mechanisms [30] in order to facilitate the building of personalised risk detection, prevention and intervention models, based on the application of advanced AI

techniques (i.e. multi-criteria reasoning, big data analytics and machine learning) for adaptation to personal needs, cancer related factors, lifestyle, genetics, gender and other dispositional factors, conditions and preferences as well as the individuals' living environment and their social connections [28], [31] – [32]. On top of this, the use of AI mechanisms will allow clinical/medical partners to develop targeted prevention and intervention models against the identified key risks of Pancreatic Cancer. The targeted models will provide grounding for personalised recommendations, especially by the use of appropriate communication strategies. This is of growing importance, as recent data shows that many people stop using digital health aids after 2 weeks, which is a relatively short period of time to gather any feedback and analyse the impact of new solutions/services [24], [33]. Based on the availability of such models and their continuous evolution (based on the fusion of new data) the clinical/medical experts will be able to prescribe personalised recommendations (prevention and intervention measures) to specific individuals. In this respect, iHELP will provide advanced modelling and representation technologies through an explainable AI concept that will enable just-in-time decision support through modelling and selection of relevant recommendations against the detection of specific risks (as well as behavioural, lifestyle and social events) in specific profile of individuals.

D. User-engagement toolsets and personalised interfaces for early risk identification, prevention and intervention mechanisms

iHELP seeks to leverage the latest advances in mobile, IoT and wearable technologies to develop engaging and user-centric applications for personalised eHealth/eCare that will also help achieve intelligent personalised monitoring and consultancy [32], [34] – [36]. The latter will create an advanced sensing environment with intelligent reasoning, which will be able to preserve and interpret physical, cognitive, mental and social data in real-time. The iHELP applications will be designed based on end-user development techniques and usability principles; and they will be part of an end-to-end big data management framework in which health analytics (e.g. causal analysis, predictions, health conditions' forecasting, or real-time feedback) will be performed without having to take care of scalability and accessibility issues as they can operate across all individual profiles (e.g. age groups, or gender). Unified analytics will be performed over all the heterogeneous streaming (e.g. mobile app, wearable, or social media) and historic (e.g. biobank) data, offering the ability to share and reuse all this large amounts of data towards the provision of personalised actionable insights.

E. HHRs as a standardised base for data modelling and management

iHELP seeks to deliver standardised ways of modelling and storing heterogeneous health data through HHRs. This will facilitate the compilation of distributive knowledge towards the provision of efficient healthcare services and policies. The HHR structures in iHELP will enable the aggregation of data from different sources, sensors and online platforms to support seamless integration of multiple health dimensions. The development of HHRs will incorporate metadata structures capturing and enriching the main health record attributes, also extending to behavioural, lifestyle, nutrition, genomics and social properties. Approaches to extend HHR structures will be developed to exchange

knowledge, experiences, specific risks/events and also recommendations that influence the behaviour of both the individuals and the communities, analyse situations and their potential impact.

F. Secure multi-actor information exchange and enhanced quality, trustworthiness and reliability of data

iHELP aims at the implementation of a multi-actor approach and the development of a relevant governance mechanism to ensure that the data exchange, ethical and gender aspects, as well as privacy and security of data as well as stakeholders' interests is adequately preserved. By conducting digital trials, the platform will create personalised profiles of individuals through the exchanged data and used devices, while assuring the confidentiality of all the transferred data. The data governance mechanism implemented by iHELP will ensure address concerns related to data ownership, quality, sovereignty and reuse – among all relevant stakeholders in the personalised healthcare ecosystem. iHELP seeks also to provide a holistic privacy and security framework enabling different stakeholders to ingest, analyse and specify the legal, regulatory and ethical requirements that govern the health data information management within countries, while also incorporating mechanisms for hybrid access control (role-based and attribute-based) and data integrity to ensure adherence to legal, regulatory, security and privacy aspects, especially concerning the General Data Protection Regulation (GDPR).

IV. TECHNICAL ARCHITECTURE

The iHELP platform connects different stakeholders in the healthcare ecosystem, providing each stakeholder something of interest. As presented in "Fig. 1", the platform consists of five main subsystems while specific interfaces, tools and cross-subsystem elements address particular technical, business and operational requirements associated with the platform objectives.
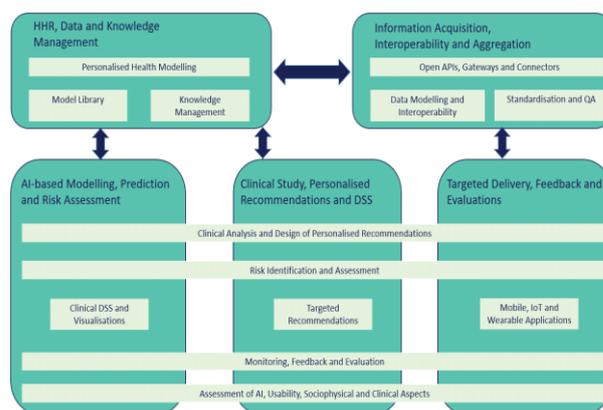

Fig. 1: High-level iHELP's architecture.

iHELP will be built on top of a *Scalable Big Data Platform* which will provide the necessary performance and features for the management and analysis of the heterogeneous data that will become available from various sources. *Trust, Security and Access Control* is the core security framework of the iHELP platform, which provides identification, authentication and authorization for business-to-business and business-to-consumer models. In addition, the framework ensures data confidentiality and integrity based on appropriate encryption mechanisms and methodologies. The framework will provide a holistic approach on the trust aspects considering the

participation of multiple administrative domains as well as the involvement of multiple stakeholders and sub-systems.

- **Information Acquisition, Interoperability and Aggregation** will enable the ingestion of heterogeneous historic and real-time datasets in the iHELP platform to realise the notion of HHRs. The open APIs will implement secure data pipelines that support the necessary transformation, interoperability and aggregation functions to be applied to the data before they are integrated into HHRs.

- **HHR, Data and Knowledge Management** will realise the concept of HHRs through implementation of big data and knowledge management techniques. The Health Record Modelling tools will support the integration and management of rich metadata and information within the iHELP environment. During data management and analysis of imported information, Reliability & Volatility Assessment services evaluate the quality of information, evaluate its uncertainty, while predictive algorithms and machine learning techniques will allow the adaptive selection of information sources based on their perceived reliability, provided information type and availability.

- **AI based Modelling, Predictions and Risk Assessment** focuses on extracting insights, enriching healthcare data and providing real-time decision support to different users of the iHELP platform. To this end, Deep Neural Networks (DNNs) and state-of-the-art models will be utilised with the scopes to identify rule associations between the patients' processed data and the Pancreatic Cancer, for anomaly detection and to simulate the evolution of the Pancreatic Cancer over time for a new patient. Hence, AI based analytic techniques will operate at several sub-systems in iHELP; from supporting data interoperability and aggregation mechanisms to facilitating the deep learning models in order to analyse the clinical records and develop personalised models that can be used for identification of impeding risks and generating predictions. On top of this, anomaly detection models must be trained, in order to distinguish healthy and diseased people. Anomaly detection models have seen a great development lately and more specifically Generative Adversarial Nets (GANs), Autoencoders and Variational Autoencoders (VAE) are the state-of-the-art models that learn to recognize normal behaviour and then calculate anomaly scores for the data in order to identity irregularities [37] – [39]. Moreover, in order for the iHELP platform to be able to simulate the evolution of the Pancreatic Cancer over time for a new patient, two different approaches can be followed. First, if the evolution of the disease of other patients is available, then traditional models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU) Neural Networks and Autoencoders, that look for similarities in the input data have proven to perform well and thus these models will be implemented and utilized [40] – [42]. Otherwise, novel and very promising models such as

EvoNet that try to predict the evolution of a dynamic graph by using Graph Neural Networks will be utilized [43]. Moreover, the advancements in AI are also seen in the development of frugal AI techniques that are able to perform real-time risk assessment and decision support on specific risks health conditions – based on the availability of personalised information from different sources. Frugal AI has met a considerable development lately, with many state-of-the-art models, tackling all different types of problems, especially in the healthcare domain. In particular, Active Learning intends to learn the characteristics of different classes by choosing which examples it will consider and, thus, reducing drastically the amount of data for training [44], in this sense models such as Matching Nets might be proven valuable [45]. Finally, the AI techniques will be bundled together in a Decision Support System (DSS) that also provides visualization support to the users.

- **Clinical Study, Personalised Recommendations and Decision Support** is facilitated by mechanism that extract meaningful information from the historic and real-time datasets. This sub-system consists of mechanisms and tools that support clinical studies and design of personalised Decision Support System functions with query building and visual analytic tools that support simulations, forecasting and causal analysis on specific symptoms/risks considering, among clinical parameters, the lifestyle and behavioural aspects as well as spatial and temporal factors. To this end, the DSS will support query building, analytic and visualization capabilities by means of different interfaces. The query builder interface will allow to create queries workflow and execute federated queries, while the dashboard interface will allow to show queries and analytic results in different chart types such as line, bar, pie, world maps or heatmaps and tables. The goal is to provide predictive analytic functions, addressing both clinical and patient needs, that are able to measure the risks for particular patients and groups and identify the factors that prevent the patients from making expect progress in medical and quality of life aspects.

- **Targeted Delivery, Feedback and Evaluation** is realised by integrated mobile and wearable solutions in the iHELP platform. The use of these technologies provides the necessary support to clinical partners for modelling, presentation, and delivery of targeted recommendations to individuals who have a high likelihood of developing Pancreatic Cancer – based on the identification and predictions of specific risks. The aim of this subsystem will be to observe, check and aggregate data coming from sensors, mobile and/or wearable devices and compare them with personalised targets issued by the clinician/medical experts. The solutions in this subsystem will enable the delivery of more effective and targeted healthcare and the ability to raise awareness (of disease, risks, associated factors, precautions and relevant interventions) among users. The feedback gathered from the targeted delivery of recommendations and their evaluations will allow the platform to provide

(clinical and policy making) recommendations to relevant stakeholders.

V. CONCLUSION

Despite the wealth of evidence that data analytic techniques (when applied on Pancreatic Cancer related datasets) have provided and the consequent advances in Pancreatic Cancer treatment that have been made in the recent years, many risks and associated conditions are still not being analysed and targeted recommendations remain to be explored. The iHELP platform aims at changing this by incorporating AI-driven actionable insights into the Pancreatic Cancer prevention strategies, screening programs and treatment processes. iHELP will implement innovative AI algorithms and techniques, ranging from multi-modal causal analysis and real-time anomaly detection to disease trajectories, real-time risks prediction, forecasting of risk trajectories and detection of behavioural swings. To this end, iHELP seeks to deliver a novel personalised healthcare framework that enables the collection, integration and management of health-related data from various sources (medical records, lifestyle, behaviours, social media interactions) in a standardised structure, the HHRs. The data in the HHRs will be analysed using advanced AI techniques to draw adaptive learning models that are used to provide decision support in the form of early risk predictions as well as personalised prevention & intervention measures (alerts, behavioural nudges, consultations medications, therapies, screening etc) that are delivered through user-centric mobile and wearable applications. Hence, the iHELP outcomes will not only enable the integration of data coming from multiple sources, but they will also provide ways of developing advanced AI techniques that can model the specific conditions of individual Pancreatic Cancer patients to gather reliable predictions that can be used in the development of effective care plans. Moreover, the availability of HHRs provide opportunities to validate iHELP outcomes (e.g. improvements in quality of life, reduced risks) through advanced analytic functions. Hence, iHELP solutions will facilitate the policy making procedures by providing decision support and social analysis on the design of new screening programs and new guidelines for bringing improvements in clinical, lifestyle and behavioural aspects of the fight against Pancreatic Cancer. The proposed iHELP approach and platform will be further evaluated with data obtained from multiple cohorts/regions in the EU and beyond to help validate data models, AI-based analytic techniques and also to facilitate knowledge exchange and collaboration at the clinician, research and policy making levels.

REFERENCES

[1] "Cancer data and statistics", World Health Organization Regional Office for Europe, 2021. [Online]. Available: https://www.euro.who.int/en/health-topics/noncommunicable-diseases/cancer/data-and-statistics [Accessed: 02-Jun-2021].

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

[3] J. Ferlay et al, "Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018", European Journal of cancer, vol. 103, pp. 356-387, 2018.

[4] M. E. Nieroda et al, "Online decision support tool for personalized cancer symptom checking in the community (REACT): acceptability, feasibility, and usability study", JMIR cancer, vol. 4, no. 2, p. e10073, 2018.

[5] J. G. Walker, S. Licqurish, P. P. C. Chiang, M. Pirotta, and J. D. Emery, "Cancer risk assessment tools in primary care: a systematic review of randomized controlled trials", The Annals of Family Medicine, vol. 13, no. 5, pp. 480-489, 2015.

[6] A. McGuigan et al, "Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes", World journal of gastroenterology, vol. 24, no. 43, pp. 4846, 2018.

[7] P. Rawla, T. Sunkara, and V. Gaduputi, " Epidemiology of pancreatic cancer: global trends, etiology and risk factors", World journal of oncology, vol. 10, no.1, p. 10, 2019.

[8] P. P. Jayaraman et al., "Healthcare 4.0: A review of frontiers in digital health", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 10, no. 2, e1350, 2020.

[9] S. M. McKinney et al., "International evaluation of an AI system for breast cancer screening", Nature, vol. 577, no. 7788, pp. 89-94, 2020.

[10] M. Nieroda, K. Keeling, and D. Keeling, "Acceptance of Mobile Apps for Health Self-management: Regulatory Fit Perspective", in UMAP Workshops, 2015.

[11] G. Villalobos-Zúñiga and M. Cherubini, "Apps That Motivate: a Taxonomy of App Features Based on Self-Determination Theory", International Journal of Human-Computer Studies, vol. 140, p. 102449, 2020.

[12] L. Rosemarijn, M. A. Neerincx, and F. Cnossen, "Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors.", International Journal of Human-Computer Studies, vol. 68, no. 6, pp. 386-397, 2010.

[13] S. Hanke et al, "universAAL–an open and consolidated AAL platform", in Ambient assisted living, pp. 127-140, Springer, Berlin, Heidelberg, 2011.

[14] S. Kyriazakos et al, "eWALL: An intelligent caring home environment offering personalized context-aware applications based on advanced sensing", Wireless Personal Communications, vol. 87, no. 3, pp. 1093-1111, 2016.

[15] P. Yadav et al., "Mining electronic health records (EHRs): a survey", ACM Computing Surveys (CSUR), vol. 50, no. 6, pp. 1-40, 2018.

[16] D. Bender and K. Sartipi, "HL7 FHIR: An Agile and RESTful approach to healthcare information exchange", in Proceedings of the 26th IEEE international symposium on computer-based medical systems, IEEE, pp. 326-331, June 2013.

[17] B. Christensen, and G. Ellingsen, "Evaluating model-driven development for large-scale EHRs through the openEHR approach", International journal of medical informatics, vol. 89, pp. 43-54, 2016.

[18] F. Linden, "epSOS local data providers", Acta Informatica Medica, vol. 3, pp. 142-145, 2009.

[19] S. kBioAssist et al, "Crowdhealth: Holistic health records and big data analytics for health policy making and personalized health", Informatics Empowers Healthcare Transformation, vol. 238, p. 19, 2017.

[20] J. A. Golden, "Deep learning algorithms for detection of lymph node metastases from breast cancer: helping artificial intelligence be seen", Jama, vol. 318, no. 22, pp. 2184-2186, 2017.

[21] K. H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare", Nature biomedical engineering, vol. 2, no. 10, pp. 719-731, 2018.

[22] G. Cammarota, et al, "Gut microbiome, big data and machine learning to promote precision medicine for cancer", Nature reviews gastroenterology & hepatology, vol. 17, no. 10, pp. 635-648, 2020.

[23] S. Hussein, P. Kandel, C. W. Bolan, M. B. Wallace, and U. Bagci, "Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches", IEEE transactions on medical imaging, vol. 38, no. 8, pp. 1777-1787, 2019.

[24] K. Gallagher and J. Updegraff, "Health Message Framing Effects on Attitudes, Intentions, and Behavior: A Meta-analytic Review", Annals of Behavioral Medicine, vol. 43, no. 1, pp. 101-116, 2012.

[25] G. E. Cole, P. A. Keller, J. Reynolds, M. Schaur, and D. Krause, "CDC MessageWorks: Designing and Validating a Social Marketing Tool to

Craft and Defend Effective Messages", Social Marketing Quarterly, vol. 22, no. 1, pp. 3-18, 2016.

[26] A. J. Rothman and P. Salovey, "Shaping perceptions to motivate healthy behavior: the role of message framing", Psychol Bull, vol. 121, no. 1, pp. 3-19, 1997.

[27] J. Covey, "The role of dispositional factors in moderating message framing effects.", Health Psychology, vol. 33, no. 1, pp. 52, 2014.

[28] M. Nieroda, K. Keeling, and D. Keeling, "Healthcare self-management tools: promotion or prevention regulatory focus? A scale (PR-PV) development and validation", Journal of Marketing Theory and Practice, vol. 23, no. 1, pp. 57-74, 2015.

[29] N. Shahid, T. Rappon, and W. Berta, "Applications of artificial neural networks in health care organizational decision-making: A scoping review", PloS one, vol. 14, no. 2, p. e0212356, 2019.

[30] K. L. Cheung, D. Durusu, X. Sui, and H. de Vries, "How recommender systems could support and enhance computer-tailored digital health programs: a scoping review", Digital health, vol. 5, p. 2055207618824727, 2019.

[31] J. Covey, "The role of dispositional factors in moderating message framing effects.", Health Psychology, vol. 33, no. 1, pp. 52, 2014.

[32] R. Best and N. Charness, "Age differences in the effect of framing on risky choice: A meta-analysis.", Psychology and aging, vol. 30, no. 3, pp. 688, 2015.

[33] J. M. Grisolía, A. Longo, G. Hutchinson, and F. Kee, "Comparing mortality risk reduction, life expectancy gains, and probability of achieving full life span, as alternatives for presenting CVD mortality risk reduction: A discrete choice study of framing risk and health behaviour change", Social Science & Medicine, vol. 211, pp. 164-174, 2018.

[34] M. Zanker, L. Rook, and D. Jannach, "Measuring the impact of online personalisation: Past, present and future", International Journal of Human-Computer Studies, vol. 131, pp. 160-168, 2019.

[35] M. E. Nieroda, M. Mrad, and M. R. Solomon, "How do consumers think about hybrid products? Computer wearables have an identity problem", Journal of Business Research, vol. 89, pp. 159-170, 2018.

[36] K. Lee, J. Choi, G. M. Marakas, and S. N. Singh, "Two distinct routes for inducing emotions in HCI design", International Journal of Human-Computer Studies, vol. 124, pp. 67-80, 2019.

[37] J. Pereira, and M. Silveira, "Learning representations from healthcare time series data for unsupervised anomaly detection", In 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE, pp. 1-7, 2019.

[38] D. Li et al, "MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks", In International Conference on Artificial Neural Networks, Springer, Cham, pp. 703-716), 2019.

[39] C. Baur, S. Denner, B. Wiestler, N. Navab, and S. Albarqouni, "Autoencoders for unsupervised anomaly segmentation in brain MR images: a comparative study", Medical Image Analysis, pp. 101952, 2021.

[40] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor ai: Predicting clinical events via recurrent neural networks", In Machine learning for healthcare conference, PMLR, pp. 301-318, 2016.

[41] M. Desai, and M. Shah, "An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN)", Clinical eHealth, 2020.

[42] L. Men, N. Ilk, X. Tang, and Y. Liu, "Multi-disease prediction using LSTM recurrent neural networks", Expert Systems with Applications, vol. 177, pp. 114905, 2021.

[43] C. Wu, G. Nikolentzos, and M. Vazirgiannis, "EvoNet: A neural network for predicting the evolution of dynamic graphs", In International Conference on Artificial Neural Networks, Springer, Cham, pp. 594-606, 2020.

[44] B. Settles, "Active Learning", Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 6, no. 1, pp. 1–114, 2012.

[45] O. Vinyals, C. Blundell, T. Lillicrap, and D. Wierstra, "Matching networks for one shot learning", Advances in neural information processing systems, vol. 29, pp. 3630-3638, 2016.