AI IN HEALTHCARE WHITEPAPER

BDVA Task Force 7 - Sub-group Healthcare

November 2020

BIG DATA VALUE ASSOCIATION

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

FOREWORD

Healthcare is facing enormous challenges especially now when the world is hit by a pandemic. First, healthcare quality and patient outcomes needs to be improved. Secondly, healthcare costs need to be reduced. We are currently spending in EU around 10% of our GDP on healthcare. With this trend in 30 years from now, we will be spending more than 50% of GDP. Thirdly, patient experience and the work life of healthcare providers needs to be improved. In short, the need for healthcare transformation is eminent.

Another important ongoing trend is digitalization of healthcare. Data has been increasingly created in healthcare (with annual growth of 50%). Healthcare providers should be using this data to provide evidence-based treatment. However, the question is whether they are equipped to make sense out of large amount of data. Artificial intelligence (AI) and data technologies are the key enablers of the digital transformation of healthcare. The connected medical devices will be everywhere, from hospital to home, providing a rich variety of health and context specific data. AI will be instrumental to turn this data into actionable insights across the continuum of care. It will augment healthcare providers and help them to treat more patients for the same amount of time. These technologies can help to better diagnose and treat patients, based on evidence.

HealthTech companies are working on such solutions. AI is already integrated in some products to improve their primary function and make them more intelligent. AI is also applied to make the use of healthcare products and solutions seamless by improving user experience. Moreover, AI and data technologies are used to gain new knowledge, understand new diseases, pandemic outbreaks and to improve patient outcomes (improve quality of healthcare).

This is a great start but imagine if we would be able to create a patient digital twin that can be used to check our health status at any time. This would also help us to predict our health conditions and give us time to react, make interventions and avoid sickness. AI will facilitate the transition from healthcare to health!

Prof. Dr. Milan Petković

BDVA Vice President

Department Head Data Science at Philips

ABSTRACT

Artificial Intelligence is already part of our daily lives. Many applications that are currently in our smartphones claim to have an AI algorithm behind them. However, when it comes to Healthcare the trend changes dramatically: a powerful tool that may serve for prevention, diagnosis, treatment and monitoring is many times delayed due to regulations, ethical constraints or lack of standardization. This whitepaper aims to feed policy makers with the information that is needed for realizing the full potential of AI in healthcare.

We have leveraged the network of the Big Data Value Association (BDVA) to define towards European policy makers a unified viewpoint of care organizations, academia and industry. We have conducted an online survey and an experts' workshop to gather our community knowledge and experience in respect to seven topics: (1) what is the importance of AI in healthcare, (2) what are the key areas in which AI has the most potential, (3) what are the major technical challenges, (4) what are key challenges of non-technical nature, (5) what are the enablers/facilitators for further uptake, (6) which use cases can serve as success stories and (7) what are specific tasks for which AI is expected to be successful in the future.

We have compared the findings from our survey and workshop to results from the literature to analyze which key message had to be emphasized to policy makers. Our collective efforts delivered the following results. Artificial intelligence is already adding value in rather small but promising areas of healthcare. The uptake is still far below its potential, even though the COVID19 pandemic has led to an acceleration in the adoption of digital tools. Legal, social and regulatory challenges are still major barriers and policy makers should unblock them. Most pressingly, technical and regulatory actions should be taken at the level of unified data access across Europe.

1 INTRODUCTION

Technical breakthroughs in Artificial Intelligence (AI) enable unprecedented automation and precision levels for prevention, diagnosis and treatment. Numerous publications demonstrate potential gains in self-management, as well as in the timeliness, dynamism and resources effectiveness of care [1]. In specific areas of radiology, studies are emerging that demonstrate how AI systems can outperform human experts, or at least aid significantly in reducing their workload [2]. At the same time, we face an inconvenient truth [3]: the pace at which AI techniques are emerging is out of balance with the adoption rate in practice. Both US and EU studies have documented practical issues surrounding the implementation of AI [4, 5]. Therefore, European policy makers need from their European ecosystem coordinated guidance for creating the resources that are needed. Only then can we realize the full potential of AI within the health continuum (ranging from prevention and cure to rehabilitation and long-term care).

Before moving to health and care specific matters, it is critical to define what is the scope of aforementioned “artificially intelligent” techniques: Fetzer for example starts from a plain dictionary definition to derive properties that are commonly associated with intelligence [6], including:

- mental abilities: the ability to learn or understand from experience;
- ability to acquire and retain knowledge reasoning capabilities in solving problems: the ability to respond quickly and successfully to a new situation.

AI is then defined as machine intelligence, which can be further decomposed based on differences in the underlying computational techniques. One common approach is to decompose it in two dimensions: first, techniques can be divided based on the types of the learning algorithm, and second, techniques can be divided based on the representation style of the knowledge.

In the first dimension, one can distinguish between supervised learning versus unsupervised learning. While general data mining algorithms techniques (e.g., for finding “interesting” anomalies, or “interesting” relations) do not fall easily in either of these categories, they are surely also accepted as within the scope of contemporary AI literature. Furthermore, some AI techniques do not rely on learning, instead they are tailored to searching efficiently through gigantic spaces [7]. In the second dimension, popular knowledge representation schemes are artificial neural networks (ANNs), versus decision trees, support vector machines and Bayesian networks. In recent years, especially the literature on deep learning (DL) has received increased attention. DL models (networks) are ANNs that are composed of many layers that transform input data (e.g., images) to outputs (e.g., disease present/absent) while learning increasingly higher level features for the subsequent layers [8].

Unfortunately, these technical classifications provide little guidance to scholars or practitioners who wish to apply the most suitable AI methods meaningfully in health and care settings. Furthermore, policy makers have little guidance on where to put the priorities for future funding schemes dedicated to AI in health(care). A recent industry report confirms the relevance of “future research efforts to provide a comprehensive, quantitative view, potentially surveying practitioners across all EU Member States and across all disciplines and specialties.” [5].

Innovation policy is also strongly continent-specific. Within Europe, the new policies tend to be developed according to the triple Helix model of Innovation, that is by means of structured collaborations between academia, industry and government. Focus is achieved by organizing European policy development in about 10 areas, where policy on Artificial Intelligence, big data, high performance computing and cloud computing is developed via the Big Data Value Association (BDVA). The BDVA is a not–for-profit organisation with 200 members all over Europe and a well-balanced composition of large, small, and medium-sized industries as well as research centres, Universities and user organizations. From within a healthcare-specific task force, we have reached out to the BDVA members and their partners to provide a comprehensive overview of the European viewpoint on AI in healthcare. This whitepaper documents the outcome of a literature study, an online survey and a physical workshop with 63 experts.

This whitepaper takes into account previous studies which were conducted nationally within the EU ecosystem (e.g., [9], where stakeholders from universities like Aalto and Helsinki, and from companies like IBM, VTT, GE HC and Nokia were interviewed) and where also multiple national workshops were conducted. We also take into account prior European policy reports on topics like Big Data in health care [10, 11, 12]. We also consider AI-based solutions that deal with the recent COVID-19 pandemic. The overall whitepaper is surely broader, but the pandemic serves well to illustrate specific AI opportunities.

The Healthcare Subgroup in the Application Task Force of the BDVA has organized multiple AI related workshops in recent years. While some workshops were primarily driven by expert lectures [13], one workshop (i.e., the BigMedilytics workshop in Valencia [14]) was organised exclusively for collecting input and feedback of its large audience. Furthermore, various discussion meetings took place in Brussels (e.g., in October 2019 and February 2020) with the aim to define a set of core European AI and robotics enabling technologies. Within those meetings, experts from multiple task forces have discussed cross-sectoral facilitators and barriers, as also reflected in [15]. In this whitepaper, we focus on the Valencia workshop for which the results have not been published before [14].

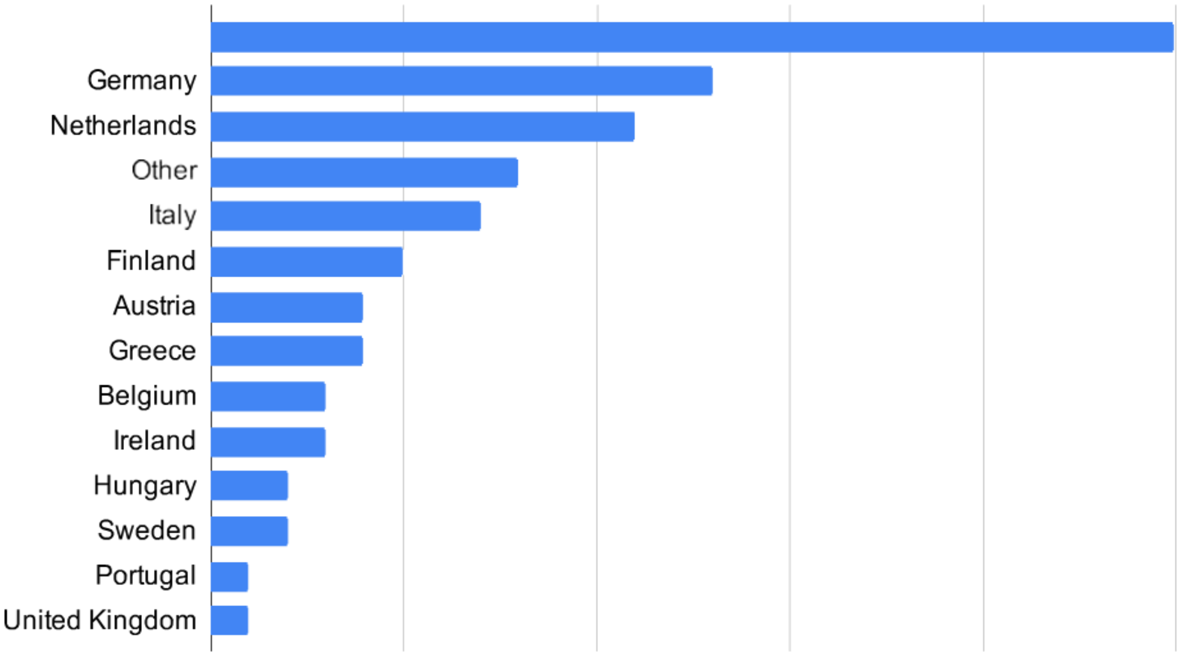

Figure 1 - Survey respondents organised by country

The workshop was organized to consolidate the results from a survey among BDVA member and their network. That survey had yielded 131 responses from organizations operating across more than 15 European member states (see Figure 1).

For the survey respondents, 52% had an academic affiliation, 15% were affiliated with a small or medium enterprise, 11% were affiliated with a larger enterprise. Others were from an industrial association or a standardization organization. For the workshop, 40% of the participants were linked to a research institute, 24% indicated a hospital affiliation and 36% represented industry.

The remainder of this report is structured as follows: Sections 2 to 8 share results based on our literature review, survey and workshops. Specifically, Section 2 covers results demonstrating the importance of AI in healthcare and Section 3 covers the key areas of AI in healthcare. Sections 4 and 5 are devoted to the challenges faced when applying AI in healthcare. Section 6 addresses enablers and instruments for introducing AI in healthcare, Section 7 presents some successful use cases and Section 8 covers the European vision on the future of AI in healthcare. Finally, Section 9 concludes.

2 IMPORTANCE OF AI IN HEALTHCARE

2.1 Literature review

AI affects every aspect of our lives. Especially its impact on healthcare domain is truly significant. AI is rapidly bringing a new era to different areas around healthcare. Moreover, it is offering benefits to specializations such as radiology, pathology, ophthalmology, and cardiology [16]. It is making substantial contribution to economic growth and well-being. It is worth noting that by 2021, the growth in the AI health market is expected to reach $6.6 billion [17]. In the United States, healthcare expenditures reach to $ 3.5 trillion, while in Europe it is $0.4 trillion [18].

Starting with the different opportunities AI brings to healthcare domain and going further to its impact on human entity, AI changes how healthcare professionals work and act in this ecosystem. It also changes significantly the care outcomes, patient experience as well as the access to healthcare services. Overall, AI changes and improves the quality of our lives [19].

Many studies have been conducted to investigating the role of AI in healthcare. The result emphasizes its importance and the impact on patients, clinicians, and the pharmaceuticals industry. Carefully observing their findings, there is no doubt that AI draws strategies, as an integral part of healthcare today and in the future [20]. Especially these days, AI has proven its usefulness in predicting, explaining and managing different issues caused by the health crisis due to COVID-19 pandemic [21].

AI is being used for a wide variety of healthcare applications, such as high-quality clinical decision support systems [1], virtual assistants for healthcare professionals [22], automated processes in image diagnosis [23], disinfection [24], care and socially assistive robots in hospitals [25] and many others. Indicatively, the implementation of AI-based clinical decision support systems has become more and more attractive for healthcare professionals (clinicians, caregivers, health authorities, etc.). Knowledge-based and learning-based techniques have been widely used to strengthen medical decision support systems. AI provides better understanding of medical processes where humans are unable to handle due to the volume, nature heterogeneity and/or complexity of data. It also leads to the derivation of valuable insights and knowledge [1]. Furthermore, automated dispensing technology based on AI dominates in pharmacy sector. This procedure is used by 30-40% of community pharmacies in Europe today. Automated distribution of packs using robots, central filling systems and automated daily dosing systems are also included in AI applications. Dispensing safety and the save working time on dispensing are obvious advantages that reinforce the growing need of pharmacies to use AI and robots [26]. In addition, faster implementation of algorithms and better predictive models can improve efficiencies for operational management of health care operations. Their evolution and their acceptance lead to accurate diagnosis and treatment in personal medicine and increase insights to enhance cohort treatment [27].

2.2 Survey and workshop results

In the survey, participants were asked: “What is the expected impact and value of AI in Healthcare? and what are the benefits and outcomes generated from it?”.

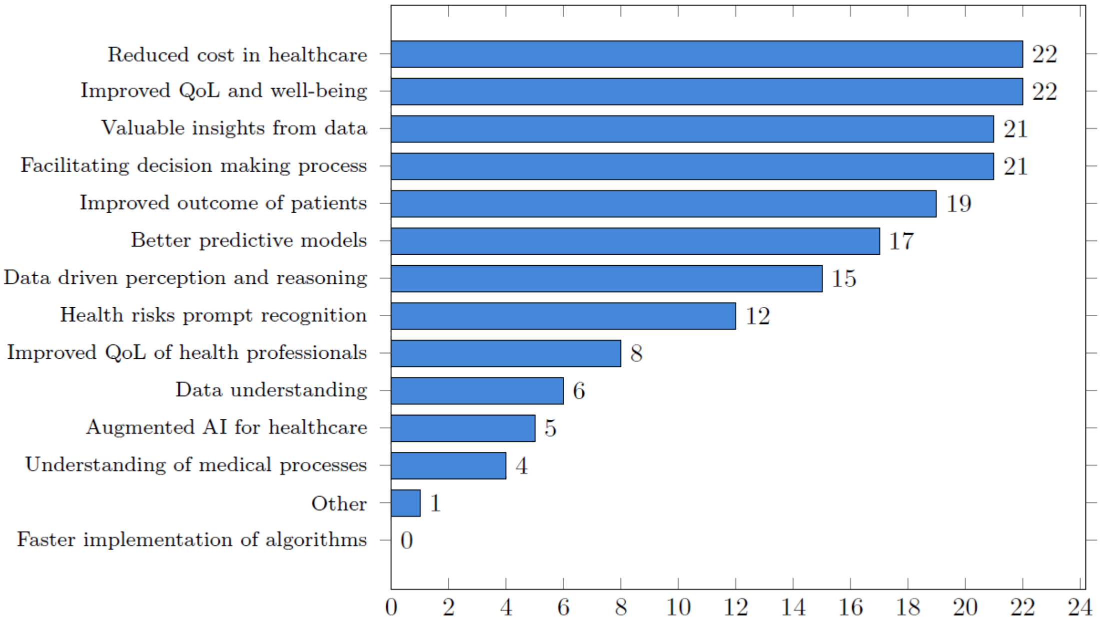

Figure 2 - Expected impact and Value of AI in Healthcare

Figure 2 depicts the answer to the question reflecting the importance of AI. The participants' interest showed reduced cost in healthcare as well as the improvement of quality of life (QoL) and wellbeing of patients as the most important factors. The derivation of valuable insights and knowledge from data, facilitating decision making process and improved outcome of the patients were also popular responses.

Apart from the fact that AI reduces hospital admissions and its potential to transform how care is delivered [5], we can also learn from its findings for the future. A good example is the current pandemic. Although we do not have final answer yet, using AI, the progress that is slowly being made and the results that are constantly delivered, we can better understand what and how it happened [28]. This also means that we may be prepared for other similar situations in the future [29].

3 KEY AREAS FOR AI IN HEALTHCARE

3.1 Overview of key areas in health

As introduced in the previous section, healthcare can benefit broadly from the advances in AI, going beyond the commonly known classical data intensive clinical health domains like medical imaging and (electronic) health records. The amount of health-related data generated spans the whole life span “from cradle to grave”, covering besides clinical health data increasingly also self-monitoring data from wearables and non-medical sources. In a wider perspective, any data affecting humans, like environmental data or work-related information, can be regarded health data and the term “exposome” has been coined to signify its importance for health. The data is not only used by medical professionals, but increasingly also by the patients – or rather; individuals themselves for health awareness and self-management, and by policy makers involved in decision making for population health, care process improvement or insurance policy development.

For this reason, it is important to first gain an understanding of the key areas and its characterising dimensions. Figure 3 shows the three dimensions that define the key areas, which will be further explained below. Enablers are additionally listed in the grey circle.

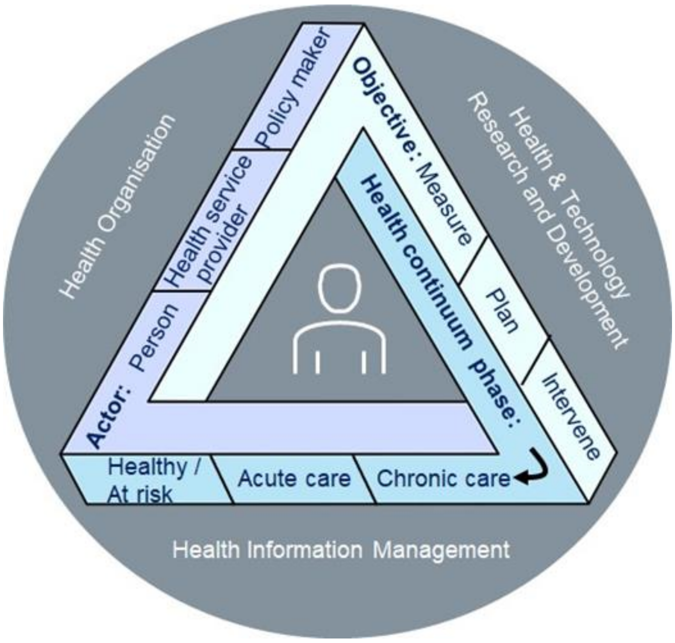

Figure 3 - Health related key areas categorisation

The first dimension is the actor, i.e. the person or office using the AI-driven health related solution. Clear actors are:

Person The individual using the well-being or health application or service based on data from themselves. In case of clear symptoms or after diagnosis, the person will be a patient. However, with the focus increasingly being on health coaching, we include any person interested in their health or well-being as well.

Health service provider Traditionally, this would be the physician or medical professional treating the patient (and is thus looking at data of other persons, typically at individual level). More generally any medical service provider from diagnostics to treatment and care belongs to this actor. Focus on preventive health makes it imperative also to include personnel or companies delivering services to maintain or improve well-being. Informal carers, even when not paid, are important for long term care, and also associations may play a role.

Policy maker This actor is motivated to improve the healthcare processes either on a (sub) national, institutional, or company level. They monitor national health, aggregate data, collect performance indicators, allocate resources, define strategies, etc. Besides politicians and population health services, it includes management of large health organisations and hospitals, insurance companies and investors.

Health data, tools, medicine and medical devices (and their developers) are in this categorisation viewed as enablers instead of actors, and are listed separately. The second dimension specifies the phase in the health continuum where the application or service is used. There are a number of options in the second dimension. From the patient's point of view, one could define a scale from completely healthy to diseased with levels of e.g. minor discomfort, curable disease, chronic disease to terminally ill. This can be extended on the healthy side with various levels of wellness like proposed in the Illness-Wellness continuum by Travis. The perspective of a service or solution provider may be slightly different. For example, Philips includes the steps of healthy living, prevention, diagnosis, treatment and home care in their health continuum view. A policy maker in turn will have more eye for obtaining insight in the public health situations, factors that may influence it, resources to maintain it, and responsibilities of the various stakeholders. The following three phases capture the essential ones, where needs and objectives clearly differ:

Healthy at risk phase This phase covers healthy people as well as those with identified risks, but without diagnosis. It focuses on primary and secondary prevention, providing lifestyle coaching to maintain or improve people's health and manage risks. Government programs will aim at prevention, where people are encouraged to take responsibility for their lifestyle choices. Service providers will focus mostly on wellness and health coaching, especially focusing on identified risks in health check-ups. Early detection (or, if possible, prediction) of the onset of diseases will be beneficial at this stage to make healthcare more proactive.

Acute care phase When symptoms are indicating a likely disease, the person will be enrolled into the healthcare system and turns into a patient. The patient will be tested by clinical means in order to come to a diagnosis. A treatment plan will be made, and the patient treated accordingly. The acute care phase is typically led by trained medical professionals and the patient is following a care path designed for the identified disease. For most, primary care will be sufficient, but some will get a referral to secondary care, which may include hospitalisation. Treatment may involve medical interventions like surgery and use of prescription drugs. The treatment may be given at the hospital or elsewhere. The acute care phase ends when the disease is cured, and the patient returns to the Healthy & at-risk phase. However, when the diseases are not curable, chronic, or requires long term rehabilitation or regularly recurring treatments, the patient (for the purpose of this categorization) will move to the *Chronic care phase*.

Chronic care phase A person needing longer term regular medical attention due to a diagnosed (chronic) disease, recovery from a disease, or general deteriorating health will enter the *Chronic care phase*. That person may be called a "caree", although that term is not regularly used. Rehabilitation is care intended to improve health aiming to ideally fully recover from the disease. Health improvement may also be needed to improve the quality of life of a patient, or to prepare for surgery, which will be beneficial for outcome expectations and patient well-being. Chronic diseases require continuous care aiming at controlling the disease and improving patient well-being, but without expectations of cure. Palliative care focuses only on alleviating pain and improving the quality of life. Elderly often receive care to cope with their ailments and daily living challenges. While many elderlies are suffering of chronic diseases and even multimorbidities, ageing is not a disease and care may include elements focusing on general age-related fading functional ability. This care is increasingly provided at home, instead of elderly care facilities. Some diseases (e.g., mental) and disabilities require long term care in specialized facilities.

The health continuum covers the typical phases of care looking from the person's perspective; The person is healthy, a patient in acute care, or a caree. In case of multimorbidity, the person may be in a different phase for each of the illnesses, making management of the disease(s) even more complex. Within each of the basic phases, there are objectives that need to be supported by the available tools. These objectives form the third dimension:

Measure The first objective is to get insight in the person's health condition by "measuring", i.e. collecting information by means of medical examination, clinical tests, imaging etc. This initial step of “data collection” is inherent to classical clinical diagnosis, but there are increasingly methods available to get insight in one’s own health by means of apps and wearables. Measuring can be done continuously, when it becomes monitoring. Population health analytics has become possible by aggregation of health-related data.

Plan The next objective is to gain an understanding of the health state or problem and plan for steps to remedy or maintain it. This includes the traditional steps of triage, stratification, diagnosis, as well as treatment planning. In general, it combines available information (historical, present, medical and other) with existing (medical) knowledge to arrive at a diagnosis (or state assessment) and allows for the selection of treatment steps (or interventions) with the highest possibility for a positive outcome. Planning includes decision making, but also managing of the process.

Intervene The implementation of the planned actions, or interventions, is the final objective. It constitutes the actual treatment and care provided to the patient. But on a macro-scale, it will involve actions, such as launching preventive health programmes for the population and managing resources in a hospital.

Typically, the process will involve an iteration of these objectives, e.g. measuring during intervention to ensure efficiency and update the plans when needed. In addition to these three dimensions categorizing the key areas, there are several enablers providing a base to all health-related activities:

Health information is a crucial element for health care provision today and the increasing amount of data available will allow to work on improvements of care. Systems to manage health information in a privacy respecting secure manner are more and more in place.

Health organization takes care that all stakeholders cooperate to provide the best achievable health care with the available resources.

Health and technology research and development is an ongoing endeavor to bring innovations to healthcare, providing new drugs and treatment methods, insight in human health issues and diseases, as well as methods for measurement, cure and decision making. Research has increasingly set its own demands on available technologies.

Finally, modern healthcare puts the person (citizen, individual) at the center, which has given rise to changes in care organization and care delivery. The paradigm continues to transform healthcare and promises personalized (medicine) solutions to improve both medical outcomes and patient experience. Detailed key areas can be found by examining the 27 intersections of health continuum phase, actor, and objective. For clarification, some major key areas were identified by the authors in Table 1. It is worth noting that besides these areas, also the enablers *Health information management, Health organization and Health & technology research and development* are key areas by themselves. Key areas voted for in the survey are indicated in the table in bold.

3.2 Survey and Workshop results

In the survey, participants were asked: “Which are the key areas and biggest opportunities for AI in Healthcare?”. The answers were given as free text, so that clustering was done by hand. The graph in Figure 4 depicts the number of times a certain key area was suggested.

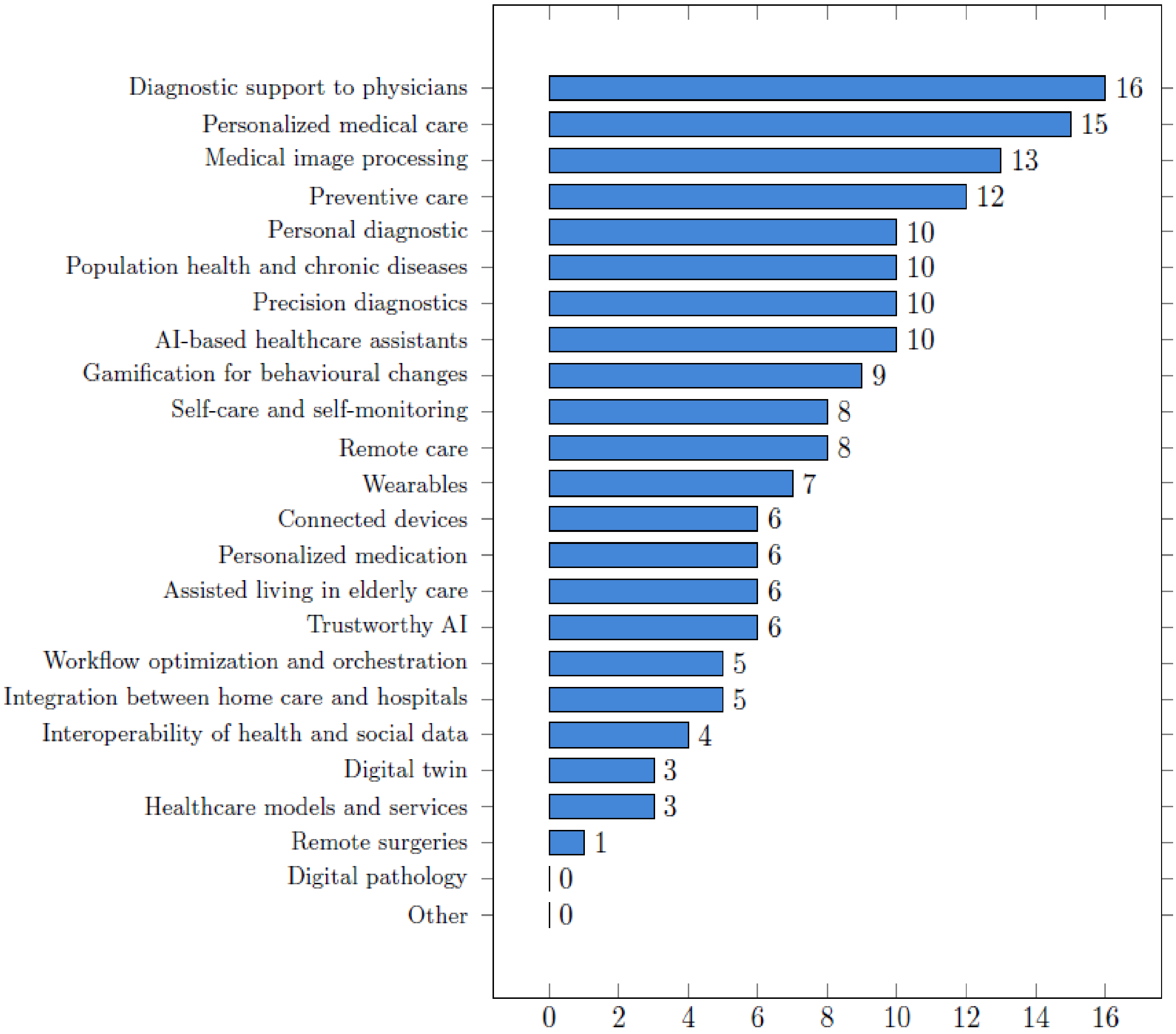

Figure 4 - Key areas mentioned in the survey with the number of times mentioned

When we map the mentioned areas and their votes in our 3-dimensional key area matrix, we can see some patterns evolving. However, care must be taken when interpreting the results, as the mapping is not always perfect (e.g. diagnostic support to physicians may include both the measurement and planning part) and this categorisation was not known to the participants beforehand (i.e. they may easily miss out on some less obvious areas).

- The majority of the votes concentrates on the support of the physician in diagnostics and treatment planning (*Acute care / Health service provider / Plan*). This includes also personalised medicine.
- The second highest vote concentration is on support for personal health in the form of preventive care, personal diagnostics or health coaching using gamification or AI-based assistants (*Healthy & At risk / Person / Plan*)
- Health care optimisation, both on population level and workflows in a hospital, is also mentioned as an important opportunity (*Acute care / Policy maker / Plan*)
- In measurements, medical image processing (Acute care / Health service provider / Measurement) is a clear focus point, but also self-monitoring (e.g. with wearables) received some votes (*Health & At Risk + Chronic care / Person / Measure*)

Table 1 - Identified key areas with examples. Bold items (with votes) were suggested in the survey

Healthy & At risk	Measure	**Self-monitoring (8)**, **Wearables (7)**, Risk assessment (genome, lifestyle), early detection	Health check, screening, symptom tracking	Population health tracking & risk assessment
	Plan	**Preventive care (12)**, **Personal diagnostic (10)**, **AI-based healthcare assistants (10)**, **Gamification for behavioural change (9)**, Lifestyle management, risk management	Health coach, proactive services, primary prevention	Primary and secondary prevention programmes
	Intervene	Wellness activities, Health coaching	Wellness providers, lifestyle coaching	Health awareness, health management support, Education
Acute care	Measure	Self-monitoring of physiological parameters, Wellness data use in diagnosis, Self-screening	**Medical image processing (13)**, Triage, Stratification, Laboratory tests, Examination	Population disease tracking. Treatment & care efficiency tracking (PROM, PREM). Intervention risk and outcome prediction, Data aggregation
	Plan	**Digital twin (3)** Self-diagnosis, Risk awareness	**Diagnostic support to physicians(16)**, **Personalised medical care(15)**, **Precision diagnostics (10)**, **Personalised medication (6)**, Diagnosis (primary & secondary care), Treatment & care planning and optimisation, Intensive care, Clinical decision making, Drug prescription	**Population health and chronic diseases (10)**, **Workflow optimization and orchestration (5)**, Healthcare system improvement, Planning for disaster and pandemic; Diagnostics and treatment strategy development; Insight from big health data; Resource planning
	Intervene	Self-care (limited)	**Remote surgery (1)**, Treatment; medical intervention, care provision, care management	Implement healthcare strategies; resource allocation; Education
Chronic care	Measure	Continuous health monitoring (also for elderly and chronic patients), Symptom and outcome monitoring, Adherence monitoring	Regular disease specific health measurements; Drug efficiency tracking; Progress monitoring	Rehabilitation efficiency tracking, Care service efficiency tracking, Drug usage and efficiency tracking
	Plan	Patient centric care planning, self-care	Rehabilitation planning, disease management, Patient centric care	Outcome prediction, care and drug efficiency assessment, Care recommendations
	Intervene	**Self-care (8)**, **Remote care (8)** Rehabilitation, controlled exercise, medicine use, chronic disease self-management	**Assisted living in elderly care (6)**, Rehabilitation services, exercises, medicines, Elderly home care, Chronic disease management services	Resource planning for rehabilitation, elderly and chronic care provision
Enablers	Health Information Management	**Connected devices (6)**, **Trustworthy AI (6)**, **Integration between home care and hospitals (5)**, **Interoperability of health and social data (4)**, Secure and privacy aware health data access		
	Health Organisation	**Healthcare models and services (3)**, systemic organisation improvement, efficiency and impact assessment, logistics, cost efficiency		
	Health & Technology Research	Medicine development, Large scale pilots, Secondary use of data, Big data.		

- On the intervention side, particularly elderly care related remote care and assisted living were mentioned (Chronic care / Person + Health service provider / Intervene)
- Not surprisingly, the enabler related to *Health information management* was also mentioned frequently. Topics like integration of data between home and hospital, health and social care, and devices prevail, while also trustworthiness (of AI) was deemed important.

In the BigMedilytics workshop, the results were confirmed (see Table 2). Physician diagnostic support and decision making tools allowing for personalised medicine and treatment are clearly leading topics. Prevention, home care and self-monitoring also receive some attention. Finally, the workshop identified other minor focus areas, such as population health and chronic diseases.

Table 2 - Results from the BigMedilytics workshop

Areas	**No. of voters**
Diagnostic support to physicians	11
Personalized medical care	9
Medical image processing	9
AI-based healthcare assistants	6
Population health and chronic diseases	5
Personal diagnostic	4
Precision diagnostic	4
Self-care and self-monitoring	4
Preventive care	4
Remote care	2
Gamification for behavioural change	1
Wearables	1

4 TECHNICAL AND RESEARCH CHALLENGES

AI can be useful for the society, industry and research organizations. However, there are several challenges that AI tools and techniques must face to be able to provide reasonable and accurate results in health. To furnish consistent insights, a survey has been conducted that covers the most significant challenges in application of AI techniques for the healthcare panorama. The result of the conducted review can be seen in Figure 5.

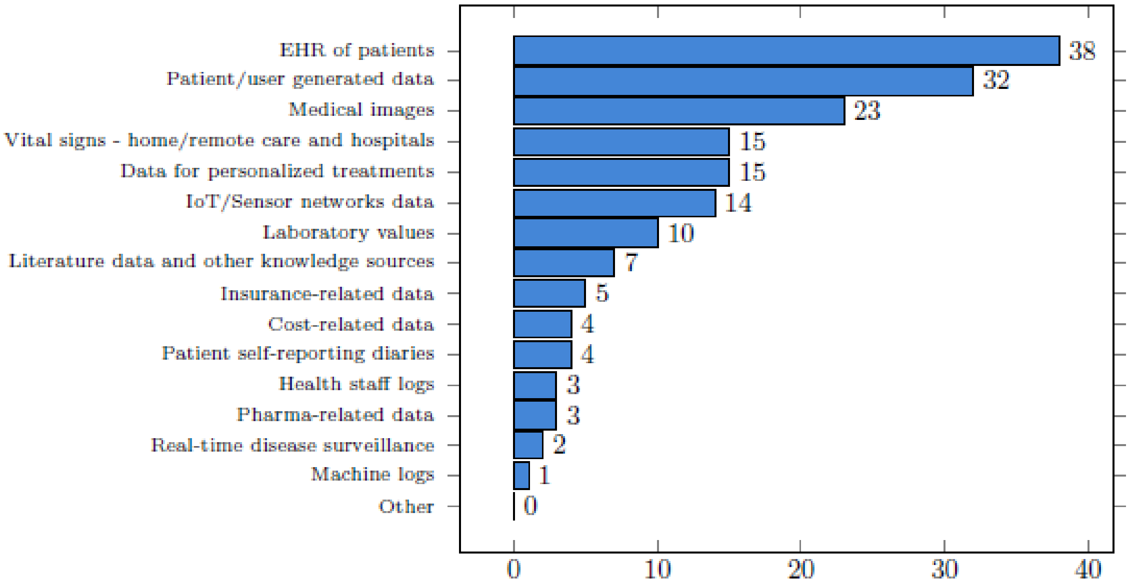

Figure 5 - Critical data resources

The results show how the respondents recognize that most of the technical and research challenges that AI experts face when dealing with the healthcare resources are data related. This is because when the information arrives from heterogeneous sources of all kinds: health records, context information, genomics, and other data formats, the quality of the data becomes the number one priority. During data collection, tools and techniques must be able to deal with noise as well as with missing and inaccurate values to ensure the quality of the dataset. This is especially important, as decisions or suggestions provided by machine learning models, mostly depend on the availability of accurate information, particularly in context of unstructured data, because it is highly variable and, in many cases, incorrect or incomplete. The integration of the curation techniques into a common analysis framework will allow clinicians and other main stakeholders to make better decisions.

Still, compiling and curating large-scale datasets is an intensive and time-consuming task. It requires an interactive approach granting human-in-the-loop scenarios where for instance medical experts manually inspect, clean, and tag the data. Dealing with the complexity of storing and accessing heterogeneous data is one of the main research areas when using AI models in the medical domain. Hospitals and care centers might have different storage systems that raise challenges in terms of interoperability. The need for efficient and secure interoperability layers to allow the communication among health players is an open research question in the data management community.

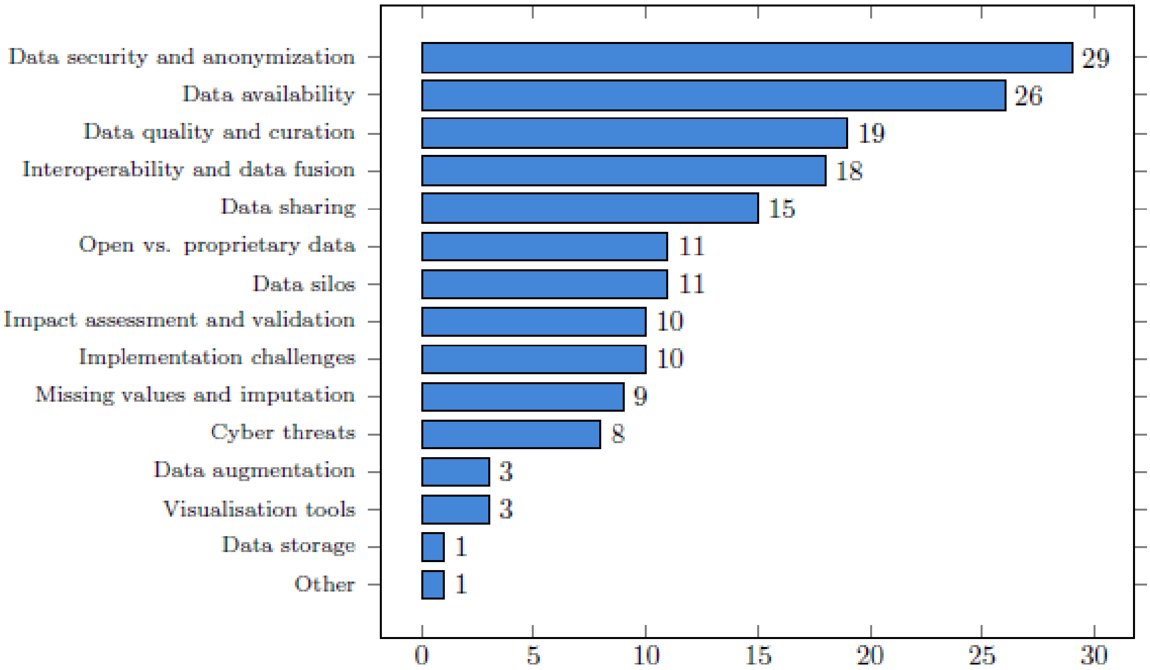

Figure 6 - Technical challenges

Also, although an efficient interoperability layer would provide direct benefits, it is important to take into account the regulatory challenges at each stage of the process of managing healthcare data. In fact, data security and anonymization are two of the most popular technical challenges that respondents have found so far (Figure 6). In relation to data cybersecurity, the development of infrastructures and technologies that provide the necessary protection to host and store health data is required. A good example of these efforts is H2020 projects such as CUREX [30], SPHINX[31], SAFECARE [32] or ASCLEPIOS [33], among others. These projects aim at improving the security of hospitals and care centres by reducing the risks to protect privacy, data infrastructures.

Moreover, as the availability of data is normally crucial for efficiency of machine learning models, one of the cutting-edges approaches for AI in healthcare is data sharing among different hospitals. Here, cyber security measures do not apply, because legal barriers restrict free flow of data among hospitals and countries [34]. For these reasons, there are initiatives based on Federated Machine Learning [35], which allow hospitals to collaboratively learn from a shared medical predictive model, while maintaining all data in each hospital, decoupling the ability to do machine learning from the need to store the data in a single place. An example of this approach in H2020 projects is the MUSKETEER project.

We also identify another promising area related to early diagnosis, prevention and intervention of mental disorders' diseases, such as Alzheimer Disease (AD). This and other related dementia disorders faced by elderly people are associated with chronic diseases and several comorbidities [36]. Traditionally, most of the techniques used for AD diagnosis (and related dementia disorders) are focused on highly invasive and costly techniques, such as lumbar puncture, PET (Positron Emission Tomography) or MRI (Magnetic Resonance Image). Even more, these techniques are usually considered in an isolated manner, thereby losing the capacity of a more concrete diagnosis and prevention. Therefore, not only the use of non-invasive techniques (e.g. Electroencephalogram, Near-Infrared Spectroscopy, or Internet of Things devices), but also the proper treatment of the data gathered from them will significantly improve not only the diagnosis but also the prevention of these diseases. Big Data and AI techniques will play a key role in this area.

First, Big Data processing platforms should be further developed and tailored to properly process and store the real time data underneath these novel non-invasive techniques.

Last but not least, the development of novel machine learning algorithms and techniques applicable for integrated repositories of health data will allow us to predict not only specific care events but also to diagnose and predict a vast range of diseases considering heterogeneous sources of Big Data. One crucial side effect of these challenges is that they will grant us to provide more user-centered solutions based on the continuous data that is being gathered and monitored. On the other hand, it will lead to the challenge of implementing efficient interfaces that allow the adequate use on the side of patients, physicians and the different set of stakeholders involved in the diagnosis, prevention and intervention of these diseases. Obviously, these user-centered solutions will have to implement general and particular Key Performance Indicators and Key Results Indicators as claimed in [37].

5 NON-TECHNICAL CHALLENGES

5.1 Literature review

In this section, a review on the non-technical challenges that AI poses in the Healthcare domain is provided. This domain has been traditionally very different from many others in terms of non-technical challenges as the final subject of application is the human being, the patient or the care provider, the person as such. It is easy to understand that when we are addressing healthcare, the consequences of something going wrong, can be dramatic.

AI may play a role in decision making, but the right decisions need to be ensured, as a mistake can lead to a human's life loss. Therefore, as in many other occasions in the health domain, AI presents a set of challenges that may represent a real NO GO factor at the end of the day. Even being one of the sectors which can obtain more benefits, it could be one of the most reluctant ones to introduce it.

The next sections discuss the most critical aspects to realize a massive adoption of AI in Healthcare.

Proposed blocks are "frequent suspects" in this field: Regulatory, Legal, Standard and Ethical aspects. These aspects are brand new for this particular case, but they are our day to day barriers when trying to make healthcare embracing digital transformation path at full performance.Due to the current health situation we are all living at the time this document is being prepared, a mention to the possible implications of using AI for supporting managing COVID-19 crisis is also addressed.

5.1.1 Regulatory challenges

Every time a new technology enters the scene, regulatory aspects and legislations come into play. In the particular case of AI there is a major point to be taken into account: patients' safety first.

The first major challenge is well known: data protection. The information sources as well as derived information resulting from AI analyses should be protected. This protection has to be done at two levels. First, at the level of the infrastructure or the devices that are involved. This applies to the bigger ones (like a major repository at a hospital) as well as to the smaller ones (a personal device like a smart watch). All the hardware should be protected to avoid data theft or system breaches via their weakest links (for instance edge processing machines at certain hospitals set-up). The second level to be protected is that of the aggregated result. Indeed, sharing data in the analysis phase is not always as secure as it looks like.

Related to the data protection, there are also challenges related to data sovereignty. Data sovereignty is a country-specific requirement that data is subject to the laws of the country in which it is collected or processed and must remain within its borders. Now that patients and citizens are generating more and more personal data, regulation should be put in place to regulate who is owner of the data. Additionally, it should be regulated how such data can flow to health institutions. With the enactment of the EU's GDPR [38] and the huge fines that go along with it, organizations have begun to take a much more serious look at their data sovereignty requirements and capabilities. The GDPR requires that all data collected on citizens must be either stored in the EU, so it is subject to European privacy laws, or within a jurisdiction that has similar levels of protection. Additionally, it applies to both data controllers and data processors so, an organization is affected regardless of whether it uses or provides a cloud service that processes EU resident data.

A brand-new regulatory element comes with the massive adoption of AI is linked with the ethical aspects of this technology. High level regulatory bodies [39] like the Food and Drug Administration (FDA) have started regulating AI/ML and prescribe best practices to be followed, for example in relation to continuously changing models. An additional major aspect is related to how models can be used, especially considering when further permission should be asked to the patients who are linked to the source data.

Of course, a final aspect that should be regulated is the potential consequences of the use of AI as a tool for diagnosis, treatment, risk management or process administration. For example, what happens if a decision is not the right one and has very serious consequences (e.g., a person dying)? Who is the civil responsible for that? The doctor? The medical center?

Some of these aspects are starting to be regulated but others will not be soon, as they may look like far in time. The lack of these regulations across the globe (or even at national, regional or local levels) is a very serious barrier to include AI into daily practice.

5.1.2 Standardization challenges

When designing the future regulatory framework for AI, it will be necessary to decide on the types of mandatory legal requirements to be imposed on the relevant actors. These requirements may be further specified through standards to ensure data portability, completeness and accessibility.

Standards help promoting technology transfer as enable interoperability between different systems while safeguarding security. Standards can help managing highly complex processes in a coordinated manner across multiple sectors and manufacturers [40]. The use of IHE integration profiles (https://www.ihe.net/) and RESTful implementations that conform to the HL7 FHIR standard are recommended in this setting [41].

Both integration profiles and HL7 FHIR implementation guides follow the concepts of FAIRness. HL7 FHIR is providing interoperability resources that can be mapped into any health care solution and is widely assessed currently for use in artificial intelligence and machine learning scenarios [4]. Interoperability activities should also be linked with the EU-specific policies, as defined in the EU Rolling plan for ICT standardisation [42]. Due to high relevance of standardisation in e-health, the Steering Committee number 42 (SC42) of the ISO (International Standard Organisation) have published 5 standards and is currently working on 14 standards of Big Data and AI Intelligence (ISO/IEC JTC 1/SC 42). Basic standards on data processing, quality and security (ISO 8000, 25012 and 27000) should also be considered when appropriate when dealing with data, and even more, with personal data.

5.1.3 Ethical and social challenges

One of the major concerns that arise for the use of AI in general is the potential replacement of people in their daily work. This becomes a major situation in healthcare where professionals are a key element in the full process. It is near impossible that AI may replace doctors, but AI can be a real support in their activities. A key success factor is the therefore user acceptance and recognition as help in the routine.

At the beginning of the technology introduction, the first move is to show the full potential of the tools as assistants and facilitators. This will show the full value and how the technology can facilitate professionals' work. Tools that should be provided need to be flexible, personalised and be available on demand. It is not a good option to put something in place that was not considered relevant by the professional. Once value is recognised, the synergy starts: humans and tools start to collaborate in an effective manner and the wheel moves in a smoother and faster manner: the magic takes place and technologies and humans work towards the same direction. The technology can speed up some processes like diagnosis and analysis and can perform some monotonous work, leaving more time to the healthcare practitioners to look after the patients, but it will never fully replace practitioners. It will then improve the productivity and efficiency of the care delivery.

For being accepted by the healthcare professionals, the concept of "Explainable AI" is very important. This recent term refers to methods and techniques in the application of AI making the results of the solution understood by human experts. It contrasts with the "black box" concept, that can lead to a lack of acceptance by the experts. The AI systems are currently being designed to adopt this concept, explaining "what is behind" the result.

One should also consider that machines are not humans and therefore complex legal and moral considerations should be made. Various authors call for "Ethics by Design" [43], and a paradigm shift from "are we compliant" to "are we doing the right thing".

Several solutions arose to better manage the COVID-19 pandemic, which could help the management of the situation, tracking and tracing all the contacts of COVID-19 infected people.

They could help identifying quarantine violations, trace routes of infected persons, identify bystanders who may be affected, and detect clusters of affected people. However, these tools are based on geolocation and that raises questions about the privacy and security mechanisms.

There are cultural factors that set apart the choices that individuals in the East and West might make. The more sensitive the data is from a privacy protection standpoint, the more useful it is from an epidemiological point of view. This implies that citizens may be faced with having to choose between anonymity and convenience, which is something we all have to do in any case on a daily basis when using digital services. In Eastern countries, some apps were launched by the government to collect information on the user's health, location and places they visited, based on the device's GPS tracking information. Most citizens are consenting the use of their data because they see the advantage to identify COVID-19 hotspots early to isolate them and keep them from spreading. In contrast, in Europe solutions that use anonymised and aggregated data are potentially "less powerful". Solutions could not offer fully individualised advice and overall uptake has been lower. Still, these solutions measure useful information like the effectiveness of different preventive measures (e.g. correlating the decrease in mobility metrics with the flattening of the epidemic curve). This privacy debate is open in society.

5.2 Survey and workshop results

As already mentioned in previous sections, BDVA conducted a survey and a discussion workshop where different experts coming from academia, industry and policy makers were represented. As per Figure 7, it is possible to see that main non-technical challenges identified by this experts' group can be grouped in the blocks also identified in the literature review.

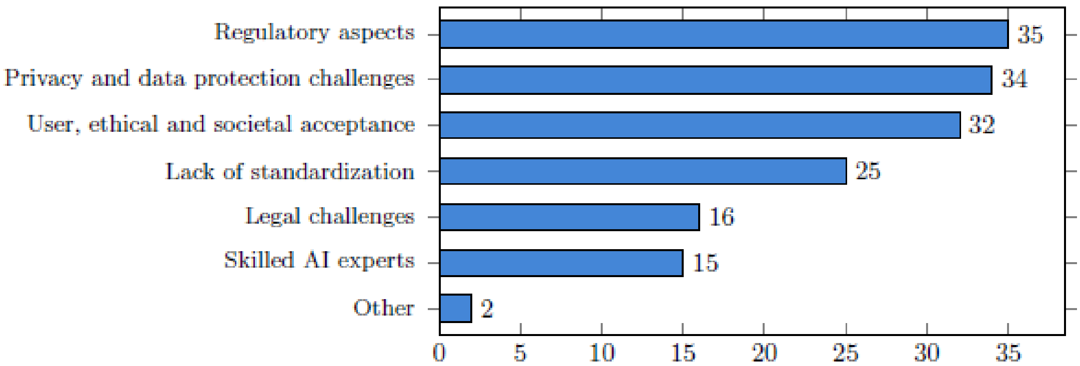

Figure 7 - Survey result on the top non-technical challenges of AI

For the survey, 159 answers were obtained in the section on non-technical challenges.

- 32% of responders identified the regulatory aspects as the most relevant ones to be considered as the non-technical challenges posed by AI technology applied to the Health domain. These included both the more general regulatory aspects and the lack of them including the different countries regulations and legislations. These responses confirm what the literature extensively shows in terms of the different paces: technology has a faster track and it does not stop in its development while regulatory and legal aspects are slower and therefore, in some occasions may be late or not covering the solution but the problems that already happened. The need of having a faster and more flexible regulation that can be adopted in a fast manner is a must in the technological regulatory space.
- 20% of survey responders selected Ethical and Social Challenges. As part of the massive adoption of a technology, the final acceptance as well as the respect of ethical and social principles is a must. In the case of AI for Healthcare, as the literature shows, there are many different aspects that can be seen as a threat. Medical doctors are the major driving force for massive adoption and a clear vision about AI as an "assistant" rather than a "competitor" will help importantly in this acceptance. As the literature shows, recent movements towards this final acceptance in several sectorial activities, are already taken place. A major

discussion point here would be if these efforts will suffice in the Healthcare industry where digitalization is lagging behind.

- The third major opinion group (21%) was focused on standardization challenges. This problem is well known into the Healthcare sector as it is one of the key barriers and difficulties that we have been facing for a long time. Although AI poses new concrete needs in this regard, maybe trying to solve old problems with new momentum is what we really need.
- Lastly, within our survey, a final group (9%) of people, identified the lack of professionals as well as a real challenge in the sector. This problem somehow looks like transversal to any application sector but combined with the intrinsic problems of the healthcare sector (late or poor digitalization, poor information standards, etc.) is the perfect storm.

Summarizing, the non-technical challenges contribute significantly to the barriers for the massive adoption of AI in the Healthcare domain. This represents a kind of a paradox as the social and ethical threats of the technology are delaying the great social value that AI may bring to the sector (for example in creating pandemic management tools).

6 ENABLERS/INSTRUMENTS

6.1 Literature review

Healthcare can benefit from the introduction of AI, yet to do so it is important to analyse the enablers and instruments needed to foster its development and adoption.

The need of greater access to data and data sharing is a common struggle within the Healthcare domain. The EC is supporting the sharing of data with initiatives like the European Open Science Cloud (EOSC, https://www.eosc-portal.eu/) and Digital Health Europe (https://digitalhealtheurope.eu/), promoting FAIR (Findable, Accessible, Interoperable and Reusable) principles. The former allows easier replicability of results and limits data wastage of e.g. clinical trial data (research integrity); the latter, aims, among others, at empowering AI development for the health data ecosystem, by earmarking more resources to promote the accessibility and interoperability of health data and addressing its provenance as well as curation [44]. Nonetheless, while sharing data some important privacy issues can be raised and to face them techniques are put into practice, namely: anonymisation or de-identification [45].

Most medical data lack of interoperability as they are in isolated databases, incompatible systems and/or proprietary software, making it difficult to exchange, analyse, and interpret them [46]. Indeed, as stated also in [5], healthcare data can be notoriously complex. Data are often gathered by proprietary software and compiled in siloed databases that are part of largely incompatible systems. In that sense standardization, interoperability, and scale of data aggregation and transfer are not always achieved in practice [47]. Interoperability, be it technical, structural or semantical, is strongly dependent on quality, accuracy and completeness of data as such indicator will impact on the performance and the capabilities of the AI tools. In that sense several initiatives have been ongoing in the healthcare sector to standardize data format like Health Level Seven International (HL7) or Integrating the Healthcare Enterprise (IHE), which specify health IT standards and their use across systems.

Additionally, the co-existence of private and public healthcare services reveals data ownership issues, complicating the achievement of unified and complete health information of each individual. The mass adoption of Personal Health Records (PHRs) can be a step towards this goal. When integrated with other systems, PHRs can enable patients and/or their caregivers with rights to supplement, validate and enrich data from sources such as Hospital Information Systems but also lifestyle related data sources [48]. Blockchain technologies have recently been proposed as a solution to issues regarding the distributed management of PHR data sets [49].

This EIT Health 2020 report [5] highlights that “there is a need to define the critical use cases that can deliver the biggest impact in healthcare through AI”. Despite the increasing levels of attention on AI in healthcare [44] and increased funding ([5], page 34), the widespread adoption of AI solutions is still lacking [50].

The benefits of AI should remain accessible to all [44]. For that to happen, IT, Health organizations, patients and public administration should maintain a regular dialogue. This is in alignment with the Finnish SRIA [9] that stresses the value of having strategic and long- lasting collaboration between stakeholders in healthcare, academia, business and experts in areas like legal and regulatory. These, among others, should lead to the development of sound business models that take into account the value chain of social and health care services together with the changing (data-driven) relationships between regulators, payers, healthcare organizations, industry, technology providers and citizens.

Besides, some reports have stressed the need to generate trust to prevent resistance and/or accelerate the take up of AI solutions. There are concerns about the safety and effectiveness of decisions or recommendations induced by the technology. Without trust from different stakeholders on a wide geographical scale, it is hard to envision how solution providers, specially SMEs, will grow outside their initial implementations. Successful engagement of healthcare professionals and patients, especially in the co-creation of solutions [51], seems instrumental to speed up the acceptance of AI in the health space.

The trust towards some AI ”black box” [52] models is sometimes reduced as the results they produce, while highly accurate, can be difficult to interpret being too complex or opaque. To overcome this issue, Explainable AI (xAI) looks at why a decision was made. xAI models can be more interpretable as they potentially make it possible to open up the black box and reveal the full decision-making process in a way that is easily comprehensible to humans [53], enabling them to understand why the system arrived at a specific decision or performed a specific action. This can help, for instance, to explain a seemingly anomalous result [54].

There is a range of reasons why some form of interpretability in AI systems might be desirable: (i) giving users confidence in the system; (ii) safeguarding against bias; (iii) meeting regulatory standards or policy requirements; (iv) improving system design; (v) assessing risk, robustness, and vulnerability; (vi) understanding and verifying the outputs from a system; and (vii) autonomy, agency, and meeting social values [55]. Transparency is important also for regulators who “may require clinical AI tools to be explainable to clinicians to whose decision making they are coupled; to quality assurance officers and IT staff in a health provider organization who acquire the clinical AI and have risk-management/legal responsibility for their operation; to developers; to regulators; or to other humans” [47].

6.2 Survey and workshop results

Participants were asked about the necessary enablers / instruments for AI in Healthcare and, in doing so, out of the over 130 involved participants, 58 prioritized the suggested enablers and instruments as follows:

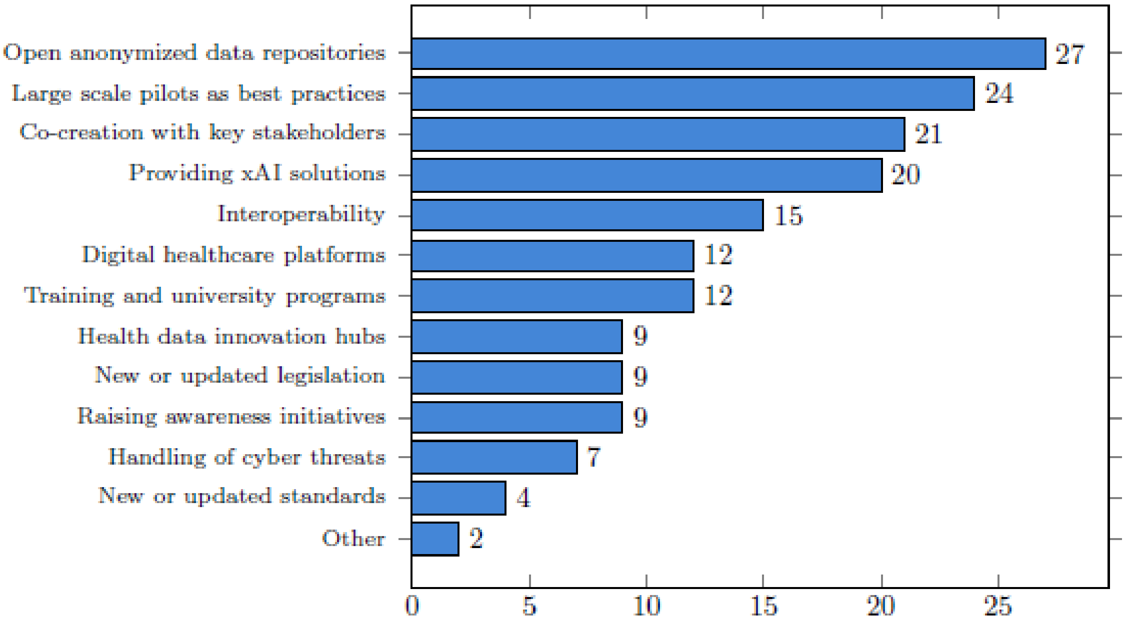

Figure 8 - Survey result on AI enablers and instruments

Over 47% of the respondents identified in the availability of *“open, anonymized medical data repositories”* the most important enabler for AI. Moreover, 41% of the participants reported the relevance of *“large-scale pilots as the best practice examples”* and 34% both the *provision of transparent and explainable AI-solutions* and the *inclusion of key stakeholders. Interoperability* and the rest of the instruments complete the list with a score below 26%.

In line with the survey, the *BigMedilytics* workshop, held in Valencia in November 2019, provided the following qualitative *insights* about the enablers.

Regarding *open, anonymized medical data repositories*, one of the main spotted issues is the fact that many researchers find unclear the approaches that are required to collect anonymized data to ensure final users’ privacy. Indeed, some also claim difficulties in collecting data at all. In any case, AI can support here. In particular, Generative Adversarial Networks (GANs) can generate meaningful synthetic data which do not suffer from the confidentiality constraints of the source data [56]. When it comes to open data, the importance to ensure their quality and "correctness" is highlighted.

As for the *inclusion of key stakeholders* in solution development as an instrument to prevent resistance due to changes that new technologies bring, a mistrust of healthcare organizations on AI companies on the use of the data can be noticed. Besides, they have the impression that the potential return of their data sharing is limited, having to purchase the solution they have helped to create. Those are the main reasons why they are reluctant to share it. Involving the clinicians at early co-creation stages is seen as a best practice to overcome these reluctancies. Some healthcare professionals are willing to be part of the development process, becoming bridges between the AI companies and patients.

Healthcare professionals also stress out that algorithms need to be *transparent* in order to be trusted and reliable in current practice, which makes evident the need to use Explainable AI solutions.

Finally, *interoperability* issues were mentioned. On the one hand, the importance of semantic interoperability is highlighted, as this is necessary to integrate data produced in different contexts overcoming the limitation of simple structural interoperability. On the other hand, General Data Protection Regulation (GDPR) differences are an additional obstacle worth to be mentioned, in particular, when it comes to the realization of data interoperability between different countries.

6.3 Discussion and proposal

After analysing the potential AI enablers for healthcare along with their relevance in previous sections, herein, the main conclusions are summarized, and some proposals made.

1. **Promote open, anonymized medical data repositories.** In particular by easing the responsibilities towards the GDPR for research purposes as to facilitate the creation of open repositories and stimulating their fair exploitation and enrichment. Besides, owners of existing repositories could collaborate to securely interconnect their data and/or EU-funded projects are encouraged to make available synthetic data sets that sufficiently resemble source data while avoiding privacy issues. Generative Adversarial Networks (GANs) have demonstrated potential, but their general applicability has not been established yet.

2. **Foster large scale pilots as the best practical example,** in order to lower the barriers for AI adoption within healthcare organizations. For example, by funding and facilitating first experiences with AI with own and/or open data repositories. Besides, the dissemination of success stories, best practices and lessons learnt should be supported, both from the healthcare and the technological sector. It is also critical that successful pilots are adopted and scaled, surviving the 'valley of death'.

3. **Inclusion of key stakeholders in solution development.** By fostering demand-driven co-creation initiatives where healthcare professionals (and patients) take an active role in the conceptualization, design and validation of the AI solution. Public funding should keep on supporting demand-driven initiatives (e.g. PCP or PPI for big projects or the inDemand model [51] for smaller ones) to encourage the early participation of healthcare professionals. Also, the learning-by-doing initiatives where healthcare professionals and managers get hands-on knowledge on AI and its possibilities should be encouraged.

4. **Generate trust among different stakeholders** by promoting joint initiatives with fair returns for all those involved. Efforts should be put to investigate and share alternative win-win approaches (beyond economic exchanges) to motivate and make projects feasible from their conception. For example, by openly making available templates of contracts or agreements to be re-used and adapted in other projects. This should generate consciousness about the need to state clear returns on all participating, in particular the healthcare organizations and its participating professionals, during and after the collaborative project.

5. **Promote more explainable AI-solutions** and implementations in the healthcare sector to increase trust and acceptance in AI across the professionals and decision makers.

7 SUCCESSFUL USE CASES

This section aims to present a variety of success use cases developed in the framework of European funded projects focused on AI applied in the Health domain. A selection has been made among the wide variety of existing projects, to present some recent use-cases coming from projects from the BDVA community. They cover a wide range of the key areas introduced in this White paper, presenting solutions addressing from the prevention to care stages, and different kind of actors: the user, the patient, the health professional and even the healthcare solution provider and Health data scientists and researchers. As such, they also include the topics that have the first and second largest number of votes as an answer to the question about the key areas and biggest opportunities for AI in Healthcare: diagnosis support for physicians and personalized medical care. The use- cases selected belong to three projects at different maturity stage: one successfully finished (2008), one in an advanced status (to be finished in August 2021) and one in its halfway point.

7.1 EuResist

In 2006-2008 the EuResist EU project (https://www.euresist.org/) developed one of the first AI-based solution to tackle the challenge of optimal selection of antiretroviral drugs for HIV patients while considering the potential development of resistance given the mutations in the HIV genome of patients.

The idea was to use multiple machine learning approaches and to combine them into one prediction engine. The project was composed of several steps. As a first step, the variables to be used for learning were defined: the fundamental ones, also called "treatment change episode" (TCE, a baseline viral load and genotype, a new treatment and a follow-up viral load obtained while still on that treatment) and the supplementary parameters (such as: indicators of treatment history, previously identified resistance mutations, route of infection, ethnicity, age, CD4 counts, consensus B local similarity, viral subtype). A fairly large amount of this data coming from real clinical practice in various centres in Europe was collected. The EuResist Integrated Database (EIDB) was thus created, which started with about 18.000 patients in 2007 and counts now more than 100.000 patients from Europe, Russia and Africa (http://engine.euresist.org/database/). Using this data as training and (separate) test sets, three engines were developed and found to perform similarly in validation tests. After extensive exploration of sophisticated machine learning techniques such as Support Vector Machines, Fuzzy logic, Case Based Reasoning, and Random Forests, all of the three engines ended up with using more popular Logistic Regression models for the classification of the therapies as successes or failures. However, the three engines use different approaches to derive from the EIDB useful extra information not directly contained in the TCE. A Generative-Discriminative engine uses a Bayesian network to derive the probability of therapy success on the basis of clinical markers only (without genotype). An Evolutionary engine focuses on a model of viral evolution under the selective pressure exerted by a specific drug, deriving a measure of the genetic barrier to resistance. A Mixed Effects engine considers a number of second and third order interactions among variables (drug x drug, drug x drug x drug, drug x mutation), thus accounting for composite effects [57]. The three individual engines were combined into one single prediction engine by a simple mean combiner methods [58] which predicts the correct outcome in 76% of cases.

The EuResist engine has been compared with other state-of-the-art tools available for assisting the choice of antiretroviral therapy in clinical practice and also with humans: a selected panel of HIV resistance experts [59]. In both cases results favour the use of the EuResist engine as an improved treatment decision support tool, with performance similar to or better than human experts, in the limited study panel. The EuResist engine has been made available on the web in July 2008. The user could input just the HIV genotype or add other attributes summarizing the patient history and baseline status. The output is a list of the top ten best treatment regimens each with its probability of success.

7.2 BigMedilytics

BigMedilytics is an EU Horizon2020 project that uses state-of-the-art Big Data and AI technologies to improve the productivity of the Healthcare sector. The project implements twelve diverse pilots covering all major disease groups in the EU, which address disease prevention as well as diagnosis. Furthermore, the pilots implement secure and privacy preserving architectures to support different national privacy regulations. One of the core outcomes of the project is a blueprint of Big Data and AI and best practices in Healthcare, together with an analysis and comparison of the different technologies and solutions. Table 1 for instance shows a simplified overview of the variety of data sources used across all pilots. This data sources diversity was the source of many of the challenges faced by BigMedilytics. In the following, we briefly explain how three representative pilots use Big Data and AI technologies and challenges they faced in BigMedilytics.

Data / Pilot	EMR					APP									Devices	Other								
	EMR	Socio-demographi	Structured	Image	Text	APP	Food intake	Medication intake	Vitals	Exercise	Gestational weight	CAT scores	Symptom scores	location data	Glucometer	Pollution	Weather	Insurance	Publications	Knowledge Bases	phone logs	RTLS	Machine Logs	Staffing
1	X	X	X																					
2	X	X	X		X	X		X	X															
3	X	X				X	X			X	X				X									
4		X										X	X	X		X	X							
5	X	X	X															X						
6	X	X	X	X	X																			
7					X														X	X	X			
8	X	X	X	X	X																			
9			X																			X	X	X
10			X																			X	X	
11																						X	X	
12	X			X	X																			

Figure 9 - Overview about data variety

Radiology Workflows. This pilot provides an efficient search engine for radiological data to improve the response time and the quality of the diagnoses. Given a marked region of interest in the image, the pilot searches in a large database of cases, ranks the results, and shows a summarization with scores of the findings. Radiologists can access comparable cases for differential diagnosis, based on visual queries in the imaging data they are reading. As the manually labeled dataset is small, the pilot uses NLP techniques to (weakly) label images with their corresponding findings. Then, the pilot performs de-identification on images, extracts feature vectors, and indexes the vectors to be used in the search engine. The two major challenges in this pilot are efficient search on a large image database and integration of the search engine in the existing radiology workflow.

Kidney Disease. This pilot supports clinical staff to treat patients before and after a kidney transplant. To this end, the pilot monitors adherence of outpatients, outliers, and possible risk prediction (e.g., rejection in the next 90 days) by using complex event processing and machine learning on integrated data from electronic health records and a patient app. As many real- world datasets, the data contains noise, errors, missing values, and duplicates and thus cannot be directly fed into the ML models. Therefore, the preprocessing is challenging because it includes interdisciplinary interaction with technicians, medics, and ML experts. Moreover, the quality of the trained model might not be easily generalized to other datasets.

Stroke Management. The pilot monitors and characterizes workflows within a hospital to reduce unnecessary delays and bottlenecks in the time-critical and hyper-acute stages of the therapy workflow. A workflow consists of all the processes, such as triage, tomography scan, and blood tests. The pilot uses multiple data streams to accurately characterize a patient's care pathway by integrating a diverse set of data streams: real-time location of patients, assets and staff, clinical and laboratory data. The characterization of the workflows helps determine and predict bottlenecks thus allowing care personnel to optimize their daily routines. The pilot also uses several ML techniques, such as data cleaning and feature selection, model training and prediction. Two main challenges of the pilot are data quality (i.e., data includes missing and noisy values in data streams) and data privacy issues.

These three representative pilots, together with the remaining ones of the BigMedilytics project, tackle a spectrum of different challenges, such as data quality, privacy, availability of large datasets, data representation and preparation, and model evaluation when applying Big Data and AI techniques. The pilots provide various solutions for the challenges.

7.3 DeepHealth

The DeepHealth - Deep-Learning and HPC to Boost Biomedical Applications for Health - project (https://deephealth-project.eu/) is funded by the EC under the topic ICT-11-2018-2019 "HPC and Big Data enabled Large-scale Test-beds and Applications". DeepHealth is a 3-year project, kicked-off in January 2019 and is expected to conclude its work in December 2021.

DeepHealth aims to push the use of technology in the healthcare sector, reducing the current gap between the availability of mature enough AI-medical imaging solutions and their deployment in real scenarios. The main goal of the DeepHealth project is to put High-Performance Computing (HPC) power at the service of biomedical applications that require the analysis of large and complex biomedical datasets and apply Deep Learning and Computer Vision techniques to support new and more efficient ways of diagnosis, monitoring and treatment of diseases.

DeepHealth provides two different kind of success use-cases: on the one hand, one of its main results is an European unified framework completely adapted to exploit underlying heterogeneous HPC, Big Data and cloud architectures assembled with Deep Learning and Computer Vision capabilities, ready to be integrated into any software platform that allows an easy and fast development and deployment of new applications for specific problems. On the other hand, to ensure the approach can be easily adopted by the industry and reach the clinicians and health researchers and the patients, the framework will be integrated and used in 7 industrial and research software platforms building specific solutions that cover different use-cases and health purposes.

DeepHealth results present several benefits that address "the implementation challenges" identified in this paper, reducing the complexity derived from the current toolkits to create easily AI-based applications and promoting HPC and distributed computing for reducing the needed time to train models and port them into specific applications. Results increase the productivity of IT staff working in the health sector by optimising the training of predictive models nurturing current and future AI-based medical software platforms.

It also provides several appointed enablers, in particular it relaxes the need of having "available skilled AI experts in all necessary domains", facilitating health developers' work allowing them to design, train and test predictive models without a profound knowledge in Deep Learning, HPC, Big Data, distributed computing or Cloud computing; Furthermore, it provides "large scale pilots as examples" and form of validation of the proposed framework, both in terms of time-related KPIs as well as clinical KPIs depending on each particular use-case, to promote its usage.

The DeepHealth Toolkit: a key European open-source asset for health AI-based solutions

A key result of the project is the DeepHealth toolkit, a free open source software providing European AI technology in a current ecosystem of tools led by USA and Asia. The toolkit is composed of two core libraries, the European Distributed Deep Learning Library (EDDLL) and the European Computer Vision Library (ECVL), plus a back-end offering a RESTful API to other applications to facilitate the use of the libraries, and a dedicated front-end that interacts with the back-end for facilitating the use of the libraries to computer and data scientists without the need of writing code. Coupled with the toolkit, DeepHealth is also developing HPC infrastructure support for an efficient execution of the libraries, with a focus on usability and portability. Thus, the libraries are being adapted to exploit the performance of advanced hardware accelerators (CPUs, GPUs, FPGAs, etc.), and on the other hand, the procedure for training predictive models will be efficiently distributed on Hybrid and Heterogeneous HPC, Big Data and Cloud architectures.

At the time of publishing this paper, first version of the libraries and the toolkit are already publicly available in GitHub and can be accessed for building up successful use-cases.

Specific applications and use-cases

The DeepHealth innovations will be validated in 14 use-cases setting up different test-beds through the use of seven biomedical and AI software platforms that integrate and exploit the libraries. The use-cases cover the main diseases and pathologies in the fields of (1) neurological diseases, (2) tumour detection and early cancer prediction, and (3) digital analysis of heart and brain pathologies and automated image annotation. Among them, proposed solutions include the main key areas highlighted in previous sections and represented by some success use-cases presented above such as decision support systems for clinicians, physician diagnostic support tools, preventive care and self-monitoring systems as well as population studies. For the sake of variety, in the following we present two use-cases that differ substantially from the rest of cases already presented, addressing patient's lifestyle and health researchers.

Migraine prediction: This use-case is selected for being a good example of how AI improves daily living. Within this use-case it is being developed a user-friendly personalized healthcare solution for non-hospitalized epileptic patients who would like to better monitor their seizures and stay safe. It is based on the MigraineNet platform, a cloud-based system that continuously collect data from mobile app and wearable devices in order to forecast when seizures are more likely to happen. More specifically, the system will use advanced machine learning to a) detect unusual patterns before the episode that may be associated with seizures, and b) immediately notifies caregivers triggering an alarm. Additional benefits from the personalized disease management tool for people who suffer from depression symptoms will be also considered. The benefits of this cloud-based service which will be developed are (a) the prediction of the progression of depressive symptoms, (b) the detection of the events that lead to specific type of progression of depressive symptoms and (c) the support to the doctors for better recommendations on what the user should do or avoid. The piloting activities for the validation of the enhanced capabilities of the MigraneNet application started on July 1st and will last for three months, while prediction validation tests will follow. In addition, pilot rounds are also discussed inviting non-hospitalised patients who will be constantly monitored by wearable devices (EEG, HR, etc.) for testing and validation purposes.

Predictive and Population Model for Alzheimer's Disease using Structural Neuroimaging: This use-case is a good example of exploiting the rich information in large datasets of biomedical images for investigating new avenues for advanced public health services. Specifically, this use case focuses on exploiting MRI images from more than 10,000 individuals that provide detailed anatomical and morphological data of the brain to study which areas of the brain produce the possible variability between healthy and pathological subjects, specifically in relation to Alzheimer's disease. The use case will design and develop population models of the volumetric degeneration of the grey matter due to Alzheimer's disease, through the correlation of three measures (volume, area and cortical thickness) of the brain surface areas (identified and segmented through predictive models) according to several criteria (age, gender, pathology and geographical area). The obtained models will be used to extract and establish the indicators of degeneration that could be used at the end to assess the health status of an individual. This use-case will be supported by different platforms in the project for the efficient training of the predictive models and for the visualization of results.

8 VISION AND FUTURE POTENTIAL OF AI IN HEALTHCARE

8.1 Literature review

A large number of publications from the past decade are devoted to the important question of how AI will impact healthcare. According to many thinkers AI and Machine Learning are expected to be one of the most transformative technologies in the 21st century in general and in the medical world in particular (see e.g. [60] and [61]). The two medical areas highlighted in the literature as mostly affected are radiology and pathology (see e.g. [62]). While the pace of change of these professions is slower than expected, and actual AI-based products are not yet abundantly used, a number of recent publications have shown that the quality of the solutions developed within the research community have achieved human levels and in specific cases surpassed the medical experts' performance. An example is breast cancer detection in screening mammography where a number of studies show that the Deep Learning technology can achieve human level ([63], [64]). Indeed, the research community activities in the area of medical imaging increased significantly in the past couple of years, from 2563 AI papers in medical imaging indexed in pubmed or appearing in archives in all of 2019 to 2292 only in the first half of 2020 [65]. Yet, it is not clear to what extent and when will the potential of the technologies will be utilized. Much is written about the slow rate of AI adoption in healthcare (see e.g. Miller and Brown [66]). The explanation provided in the literature is that while healthcare systems continuously explore ways to provide better care in sustainable costs, the complex unique structure of such systems present significant challenges that prevent the extensive usage of AI for this purpose. The list of related barriers discussed in the literature is aligned with the discussion in the previous sections. It includes non-technical challenges such as regulatory challenges and other, as discussed in Section 6, as well as technical challenges, mainly related to data availability as discussed in Section 4. The scarcity of data by itself is attributed to insufficient quality measures of health care, lack of advanced IT technologies and the complex incentive systems (see e.g. [67]).

8.2 Survey results

As part of the online survey discussed above, a diverse group of 69 leaders shared their perspective on this topic. Specifically, the question was phrased as follow:" Where do you see the highest future potential of AI in Healthcare? What are the specific tasks that AI could support in the future?". Most of the answers in the survey refer to the person as the main actor that has the highest future potential to benefit from the technologies. The next most prominent actor mentioned by the experts is the health service provider; only very few refer to policy makers.

As discussed in Section 3.1, another dimension that can be used for dividing the many health related tasks into categories has to do with the patient journey. The person is healthy, a patient in acute care, or a caree. Across the different answers provided by the experts, the tasks that AI could support in the future are across all three phases, from preventive medicine through better diagnosis and improved prognosis, better treatment in the acute setting and of chronic conditions. In particular, many of the answers refer to tasks discussed vastly in the literature such as improved diagnosis and medical imaging related tasks as areas that will most benefit from AI.

Finally, as previously discussed, within each of the above three phases there are objectives that need to be supported by the available tools. The answers in the survey addressed all three objectives, with the following examples of answers for each of the categories:

- **Measure**: "Most benefit we could get by very simple and user friendly gadgets that can collect lot of data about person" and "I see the highest potential of AI in enabling precision medicine based not only on -omics data integration and analytics, but also based on patient-generated data and IoT data integration, including complex event processing and other analytics that capture data on the flight".

- **Plan**: “Patient-specific intensive care unit prognosis: standard patient risk scores such as SOFA, APACHE, SAPS, and MPM, consider only a relatively small number of factors” and “... would give more accurate and affordable estimations of the patient's condition as well as estimations of the development about the condition.”
- **Intervene**: “AI-assisted robotic surgery, virtual nursing assistants” and “Helping with treatments from data gathered from patients.”

8.3 The Pandemic

In 2020, with the outbreak of the global COVID-19 pandemic, the world faced the biggest public health challenge of a generation. Everyday life has drastically changed in many countries as non-pharmaceutical interventions (NPIs) were imposed by many regions. Almost all European countries and regions as well as many countries outside the European Union announced confinement. A similar set of regions closed schools in the second quarter of 2020 and many more NPIs were applied [68]. The burden on the health system in many countries was huge, given the large number of patients in general and severe patients in particular. While in some health systems the crisis might have slowed the non- critical investments in collecting data and developing AI tools, the crisis might have accelerated other activities. The development of AI solutions for chest radiology, for instance, have accelerated, where the number of related studies increased [69]. An even more significant acceleration that health systems see is in the adoption of technologies that support remote care including AI based technologies that support chatbots, remote diagnosis, automatic risk assessment and more [70]. Note that these shifts are not part of the data collected in this paper and did not affect the vision of participants as the data were collected prior to the pandemic.

9 CONCLUSIONS AND OUTLOOK

It is widely accepted that health systems are the cornerstone of the welfare systems and states. However, even though governments are spending a considerable amount of budget for many years, there is a wide consensus in that health systems should be more efficient and effective as they are attending more and more population year after year due to the continuously increasing of life expectancy. In this context, Artificial Intelligence (AI) is being applied in different areas and sectors to help decision makers to improve their decisions and anticipate problems and drawbacks. In this paper, we, as representative members of the Big Data Value Association (BDVA) have provided our vision on how AI can be used in health systems to improve their services effectively. Our vision is based on an analysis of the literature, a survey and a workshop. Different stakeholders (e.g., users, caregivers, clinicians and technology providers) provided their opinion on topics like the importance of AI, the challenges faced in the healthcare context, and enablers for moving forward. These stakeholders report significant potential of AI in healthcare, both in terms of cost reductions and in terms of quality improvement. Our results also demonstrate that AI is not restricted to purely clinical settings. Surely, AI powered precision diagnostics for clinicians (i.e., personalized medicine) does score highest as a perceived key area but AI-based health coaching for citizens/patients scored almost equally high.

Our whitepaper also illustrates key areas, running solutions and projects (from the H2020 Program) where AI techniques already play a main role to improve health systems. Even though several solutions were already implemented in the health area, our results suggest a future where AI techniques play an even bigger role. We envision more personalized health systems with preventive AI-based assistants for patients and timelier and more precise diagnoses for doctors. AI will also deliver more actionable information to the health authorities to improve their decisions.

Based on our analysis, one important policy action related to enabling data access stood out. This action needs to be addressed from two perspectives:

Technical aspects of data availability Even though EHRs already provide abundant data within individual care institutes, and even though patients generate an increasingly large stream of data by themselves, data should be better integrated and labelled with proper quality assurance for large scale AI applications. Especially since clinics are still applying rather basic techniques regarding privacy and security, there is a reluctance to share data. Our whitepaper described AI-based approaches to data anonymization which could alleviate such technical concerns. Additionally, we demonstrated how federated learning-based architectures enable consortia to train models collectively, without disclosing to each other their privately-owned data sets.

Regulatory aspects of data availability Data is often also held back when key clinical stakeholders perceive that they will receive insufficient personal returns from innovation trajectories. Policy makers should provide clear and fair rules to all stakeholders to overcome this barrier. We have described some pragmatic mechanisms in this context. Besides sharing template copies of successful collaboration contracts, Europe should keep investing in large scale pilots that bring together public and private partners.

The results of this whitepaper have natural limitations: even though we collected as many good survey responses and workshop participants as possible, our ultimate sample is debatable. We should continue our efforts and report status on a regular basis. For the next iteration, we will focus on involving even more the healthcare professionals, which were only represented by one fourth of our workshop participants.

REFERENCES

[1] Thomas Davenport and Ravi Kalakota. The potential for artificial intelligence in healthcare. *Future healthcare journal*, 6(2):94–98, 06 2019.

[2] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mary Chesus, Greg S. Corrado, Ara Darzi, Mozziyar Etemadi, Florencia Garcia-Vicente, Fiona J. Gilbert, Mark Halling-Brown, Demis Hassabis, Sunny Jansen, Alan Karthikesalingam, Christopher J. Kelly, Dominic King, Joseph R. Ledsam, David Melnick, Hormuz Mostofi, Lily Peng, Joshua Jay Reicher, Bernardino Romera-Paredes, Richard Sidebottom, Mustafa Suleyman, Daniel Tse, Kenneth C. Young, Jeffrey De Fauw, and Shravya Shetty. International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788):89–94, 2020.

[3] Trishan Panch, Heather Mattie, and Leo Anthony Celi. The “inconvenient truth “about AI in healthcare. *NPJ Digital Medicine*, 2(1):77, 2019.

[4] Jianxing He, Sally L. Baxter, Jie Xu, Jiming Xu, Xingtao Zhou, and Kang Zhang. The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, 25(1):30–36, 2019.

[5] Jorge Fernandez Garcıa, Angela Spatharou, Jan-Philipp Beck, and Jonathan Jenkins. Transforming healthcare with AI the impact on the workforce and organisations, March 2020.

[6] James H. Fetzer. *What is Artificial Intelligence?*, pages 3–27. Springer Netherlands, Dordrecht, 1990.

[7] Bruno Bouzy and Tristan Cazenave. Computer go: An ai oriented survey. *Artificial Intelligence*, 132(1):39 – 103, 2001.

[8] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sanchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60 – 88, 2017.

[9] Ilkka Korhonen, Miikka Ermes, and Jari Ahola. *Strategic Research Agenda (SRA) on “Finnish Innovation Hub for Artificial Intelligence for Health (AI for Health)”*. Number 304 in VTT Technology. VTT Technical Research Centre of Finland, Finland, 2017.

[10] BDVA Task Force 7 – Healthcare Subgroup. BDVA position paper: BDVA’s response to the European commission’s public consultation on health and care in the digital single market, October 2017.

[11] BDVA Task Force 7 – Healthcare Subgroup. Big data technologies in healthcare: Needs, opportunities and challenges, December 2016.

[12] Pharmaceutical Group of the European Union. Position paper on big data and artificial intelligence in healthcare, February 2019.

[13] Milan Petkovic, Mark van Gils, Marco Marsella, and Mika Ylihautala. Transforming healthcare with artificial intelligence, October 2019.

[14] BDVA Task for Health. Big Data: Fueling the transformation of Europe’s healthcare sector. https://www.big-data-value.eu/the-bigmedilytics-organizes-the-event-big-data-fueling-the-transformation-of-europes-healthcare-sector/, September 2019.

[15] Strategic research, innovation and deployment agenda for an AI PPP: A focal point for collaboration on artificial intelligence, data and robotics. second consultation release, September 2019.

[16] Abhimanyu S. Ahuja. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ.*, (e7702), 2019.

[17] Accenture. Position paper on big data and artificial intelligence in healthcare, February 2020.

[18] Justus Wolff, Josch Pauling, Andreas Keck, and Jan Baumbach. The economic impact of artificial intelligence in health care: Systematic review. *JOURNAL OF MEDICAL INTERNET RESEARCH*, 22(2):1, 2020.

[19] Artificial Intelligence in Healthcare: the future is amazing. https://healthcareweekly.com/artificial-intelligence-in-healthcare/, 2020.

[20] Economic World Forum. Here are 3 ways AI will change healthcare by 2030. https://www.weforum.org/agenda/2020/01/future-of-artificial-intelligence- healthcare-delivery/, January 2020.

[21] Council of Europe. AI and control of COVID-19 coronavirus. https://www.coe.int/en/web/artificial-intelligence/ai-and-control-of-covid-19-coronavirus, July 2020.

[22] Ossisto. What do medical virtual assistants do? and how are medical virtual assistants helpful for healthcare professionals? https://ossisto.com/blog/medical-virtual-assistant- services, May 2020.

[23] Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz, and Hugo J. W. L. Aerts. Artificial intelligence in radiology. *Nature Reviews Cancer*, 18(8):500–510, 2018.

[24] Christian Oxborough, Chris; Cameron, Euan; Rao, Dr. Anand; Westermann. Danish disinfection robots save lives in the fight against the Corona virus. https://ec.europa.eu/digital-single-market/en/news/danish-disinfection-robots-save-lives-fight-against-corona-virus, 2020.

[25] howpublished = http://www.statewatch.org/news/2019/may/ep-analysisrobots- health.pdf. year = 2019 Z Dolic, R Castro, A Moarcas, title = Robots in healthcare: a solution or a problem? Policy Department for Economic, Scientific and Quality of Life Policies, Directorate General for Internal Policies, European Parliament.

[26] John A Doucette, Atif Khan, Robin Cohen, Dan Lizotte, and Hooman M Moghaddam. A framework for ai-based clinical decision support that is patient-centric and evidence-based. In *International Workshop on Artificial Intelligence and NetMedicine*, page 26, 2012.

[27] Deloitte Insights. Predictive analytics in health care, emerging value and risks. https://www2.deloitte.com/us/en/insights/topics/analytics/predictive- analytics-health-care-value-risks.html, July 2019.

[28] How AI and machine learning are helping to fight COVID-19. https://www.weforum.org/agenda/2020/05/how-ai-and-machine-learning-are-helping-to-fight-covid-19/, 2020.

[29] Raju Vaishya, Mohd Javaid, Ibrahim Haleem Khan, Abid Haleem. Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4):337–339, 2020.

[30] secure and private health data exchange, grant agreement id: 826404 topic: Su-tds-02-2018. call: H2020-sc1-fa-dts-2018-1. funding scheme: Ria - research and innovation action. funded under: H2020-eu.3.1.5.1. https://curex-project.eu, 2020.

[31] Toolkit for assessing and reducing cyber risks in hospitals and care centres to protect privacy/data/infrastructures, grant agreement id: 826183 call:su-tds-02-2018. funding schema: Ria – research and innovation action. https://sphinx-project.eu, 2020.

[32] Safeguard of critical health infrastructure, grant agreement id: 787002 funding schema: H2020-eu.3.7.4.

[33] Advanced secure cloud encrypted platform for internationally orchestrated solutions in healthcare, grant agreement id: 826093 call:h2020-eu.3.1.5.1. topic: Su-tds-02-2018. https://www.asclepios-project.eu/.

[34] BDVA. Data protection in the era of artificial intelligence. trends, existing solutions and recommendations for privacy-preserving technologies, October 2019.

[35] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.

[36] Monica Machon, Maider Mateo-Abad, Mercedes Clerencia-Sierra, Carolina Guell, Beatriz Poblador-Pou, Kalliopi Vrotsou, Antonio Gimeno-Miguel, Alexandra Prados-Torres, and Itziar Vergara. Multimorbidity and functional status in older people: a cluster analysis. *European Geriatric Medicine*, 11(2):321–332, 2020.

[37] Alejandro Mate, Juan Trujillo, and John Mylopoulos. Specification and derivation of key performance indicators for business analytics: A semantic approach. *Data Knowl. Eng.*, 108:30–49, 2017.

[38] Council of Europe. Gdpr. https://gdpr.eu/.

[39] MissingLink.ai. FDA artificial intelligence: Regulating the future of healthcare. https://missinglink.ai/guides/deep-learning-healthcare/fda-artificial- intelligence-regulating-future-healthcare/, 2020.

[40] EOSC. Eosc interoperability framework (v1.0). https://www.eoscsecretariat.eu/, 2020.

[41] HL7. Hl7 FHIR release 4.0.1, 2019.

[42] European Commission (EC). The rolling plan on ict standardisation, September 2018.

[43] Celestino Guemes. Ethics by design: Taking responsibility in the age of machines. https://atos.net/en/blog/ethics-design-taking-responsibility-age-machines, July 2019.

[44] Digitaleurope 2020 - harnessing the power of AI in health applications. https://www.digitaleurope.org/, 2020.

[45] Vicki Xafis and Markus K Labude. Openness in Big Data and Data Repositories. *Asian Bioethics Review*, 11(3):255–273, 2019.

[46] Moritz Lehne, Julian Sass, Andrea Essenwanger, Josef Schepers, and Sylvia Thun. Why digital medicine depends on interoperability. *npj Digital Medicine*, 2(1):79, 2019.

[47] M. Matheny, S. T. Israni, M. Ahmed, and D. Whicher, editors. *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. National Academy of Medicine, Washington, DC, 2019.

[48] David C. Kaelber, Ashish K. Jha, Douglas Johnston, Blackford Middleton, and David W. Bates. A Research Agenda for Personal Health Records (PHRs). *Journal of the American Medical Informatics Association*, 15(6):729–736, 11 2008.

[49] Alex Roehrs, Cristiano Andr´e da Costa, Rodrigo da Rosa Righi, Valter Ferreira da Silva, Jos´e Roberto Goldim, and Douglas C. Schmidt. Analyzing the performance of a blockchain- based personal health record implementation. *Journal of Biomedical Informatics*, 92:103140, 2019.

[50] Marina Turea. 4 Barriers to Adopting Artificial Intelligence in Healthcare and How to Over-come Them, 2020.

[51] Timo Hyrkas, Elina; Pikkarainen, Minna; Alalaakkola. Demand-driven open innovation orchestration at the hospitals. In *Proceedings of the ISPIM Innovation Conference - Celebrating Innovation - 500 Years Since Da Vinci*, pages 1–15, Florence, Italy, 2019. Lappeenrannan teknillinen yliopisto.

[52] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

[53] Christian Oxborough, Chris; Cameron, Euan; Rao, Dr. Anand; Westermann. Explainable AI: Driving business value through greater understanding. http://www.pwc.co.uk/xai, 2018.

[54] Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21st ACM SIGKDD International Conference on*

Knowledge Discovery and Data Mining, KDD ’15, page 1721–1730, New York, NY, USA, 2015. Association for Computing Machinery.

[55] The Royal Society. *Explainable AI: The Basics*. The Royal Society, November 2019.

[56] Brett K. Beaulieu-Jones, Zhiwei Steven Wu, Chris Williams, Ran Lee, Sanjeev P. Bhavnani, James Brian Byrd, and Casey S. Greene. Privacy-Preserving Generative Deep Neural Networks Support Clinical Data Sharing. *Circulation: Cardiovascular Quality and Outcomes*, 12(7), jul 2019.

[57] Michal Rosen-Zvi, Andre Altmann, Mattia Prosperi, Ehud Aharoni, Hani Neuvirth, Anders Sonnerborg, Eugen Schulter, Daniel Struck, Yardena Peres, Francesca Incardona, et al. Selecting anti-hiv therapies based on a variety of genomic and clinical factors. *Bioinformatics*, 24(13): i399–i406, 2008.

[58] André Altmann, Michal Rosen-Zvi, Mattia Prosperi, Ehud Aharoni, Hani Neuvirth, Eugen Schulter, Joachim Buch, Daniel Struck, Yardena Peres, Francesca Incardona, et al. Comparison of classifier fusion methods for predicting response to anti--1 therapy. *PloS one*, 3(10): e3470, 2008.

[59] Maurizio Zazzi, Rolf Kaiser, A S¨onnerborg, Daniel Struck, Andre Altmann, Mattia Prosperi, M Rosen-Zvi, A Petroczi, Y Peres, E Schu¨lter, et al. Prediction of response to antiretroviral therapy by human experts and by the euresist data-driven expert system (the eve study). *HIV medicine*, 12(4):211–218, 2011.

[60] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[61] Alvin Rajkomar, Jeffrey Dean, and Isaac Kohane. Machine learning in medicine. *New England Journal of Medicine*, 380(14):1347–1358, 2019.

[62] Ziad Obermeyer and Ezekiel J Emanuel. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, 375(13):1216, 2016.

[63] Ayelet Akselrod-Ballin, Michal Chorev, Yoel Shoshan, Adam Spiro, Alon Hazan, Roie Melamed, Ella Barkan, Esma Herzel, Shaked Naor, Ehud Karavani, et al. Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology*, 292(2):331–342, 2019.

[64] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mary Chesus, Greg C Corrado, Ara Darzi, et al. International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788):89–94, 2020.

[65] Jannis Born, Nina Wiedemann, Gabriel Brandle, Charlotte Buhre, Bastian Rieck, and Karsten Borgwardt. Accelerating covid-19 differential diagnosis with explainable ultrasound image analysis. *arXiv preprint arXiv:2009.06116*, 2020.

[66] D Douglas Miller and Eric W Brown. Artificial intelligence in medical practice: the question to the answer? *The American journal of medicine*, 131(2):129–133, 2018.

[67] Inger Ekman, Reinhard Busse, Ewout van Ginneken, Chris Van Hoof, Linde van Ittersum, Ab Klink, Jan A Kremer, Marisa Miraldo, Anders Olauson, Walter De Raedt, et al. Health- care improvements in a financially constrained environment. *The Lancet*, 387(10019):646–647, 2016.

[68] Parthasarathy Suryanarayanan, Ching-Huei Tsou, Ananya Poddar, Diwakar Mahajan, Bharath Dandala, Piyush Madan, Anshul Agrawal, Charles Wachira, Osebe Mogaka Samuel, Osnat Bar-Shira, Clifton Kipchirchir, Sharon Okwako, William Ogallo, Fred Otieno, Timothy Nyota, Fiona Matu, Vesna Resende Barros, Daniel Shats, Oren Kagan, Sekou Remy, Oliver Bent, Shilpa Mahatma, Aisha Walcott-Bryant, Divya Pathak, and Michal Rosen-Zvi. Wn- trac: Artificial intelligence assisted tracking of non-pharmaceutical interventions implemented worldwide for COVID-19, 2020.

[69] Jannis Born, David Beymer, Deepta Rajan, Adam Coy, Vandana V Mukherjee, Matteo Manica, Prasanth Prasanna, Deddeh Ballah, Pallav L Shah, Emmanouil Karteris, et al. On the role of artificial intelligence in medical imaging of COVID-19. *medRxiv*, 2020.

[70] Thanh Thi Nguyen. Artificial intelligence in the battle against coronavirus (covid-19): a survey and future research directions. *Preprint, DOI*, 10, 2020.

CONTRIBUTORS

List of contributors

Ane Alberdi Aramendi Mondragon Unibertsitatea
Faculty of Engineering, Biomedical Sciences
e-mail: aalberdiar@mondragon.edu
Timon Brussaard TNO
Business Developer
e-mail: timon.brussaard@tno.nl
Monica Caballero Everis
Supported Smart Innovation
e-mail: monica.caballero.galeote@everis.com
Davide Dalle Carbonare Engineering Ingegneria Informatica spa
e-mail: davide.dallecarbonare@eng.it
Tiblets Demewez Philips
Research Data Scientist
e-mail: tiblets.demewez 1@philips.com
Marija Despenic Philips Research
Research Scientist
e-mail: marija.despenic@philips.com
Roberto Diaz Tree Technology
Head of Artificial Intelligence Unit
e-mail: roberto.diaz@treelogic.com
Jorge Gonzalez Ticbiomed
Director of Ticbiomed
e-mail: jorge.gonzalez@ticbiomed.net
Blanca Jordan Atos Spain
Healthcare and Life Sciences Consulting Manager for Southern Europe
e-mail: blanca.jordan@atos.new
Matteo Antonio Melideo Engineering Ingegneria Informatica spa
Head of the Heathcare R&D unit
e-mail: Matteo.Melideo@eng.it
Ernestina Menasalvas Universidad Politecnica de Madrid
Centro de Tecnologia Biomedica
e-mail: ernestina.menasalvas@upm.es
Amalia Ntemou WINGS ICT Solutions
Solution Architect
e-mail: antemou@wings-ict-solutions.eu

Johan Plomp VTT Technical Research Centre of Finland
Senior Researcher, Project Manager
e-mail: johan.plomp@vtt.fi
Roberto Pratola Engineering Ingegneria Informatica spa
Researcher
e-mail: roberto.pratola@eng.it
Alejandro Rodriguez Universidad Politecnica de Madrid
Centro de Tecnologia Biomedica
e-mail: alejandro.rg@upm.es
Michal Rosen-Zvi IBM Research - Haifa
e-mail: rosen@il.ibm.com
Tatiana Silva Tree Technology
R&D Project Manager
e-mail: tatiana.silva@treetk.com
Juan C. Trujillo University of Alicante
Lucentia Research Group, Lucentia Lab
e-mail: jtrujillo@dlsi.ua.es
Henk-Jan Vink TNO
Information & Communication Technology
E-mail: henk-jan.vink@tno.nl
Pieter Van Gorp Eindhoven University of Technology
Program Manager of Health within EAISI
e-mail: p.m.e.v.gorp@tue.nl

Contributors to use cases

Francesca Incardona IPRO - InformaPRO S.r.l. EuResist Network
e-mail: f.incardona@informa.pro
Jorge-Arnulfo Quiané-Ruiz German Research Center for Artificial Intelligence (DFKI)
Intelligent Analytics for Massive Data
e-mail: jorge.quiane@dfki.de
Alireza Rezaei Mahdiraji German Research Center for Artificial Intelligence (DFKI)
Intelligent Analytics for Massive Data
e-mail: alireza.rm@dfki.de
Roland Roller German Research Center for Artificial Intelligence (DFKI)
Speech and Language Technology
e-mail: roland.roller@dfki.de
Jack Thoms German Research Center for Artificial Intelligence (DFKI)
Intelligent Analytics for Massive Data

e-mail: jack.thoms@dfki.de
Salvador Vigo German Research Center for Artificial Intelligence (DFKI)
Intelligent Analytics for Massive Data
e-mail: salvador.vigo@dfki.de

www.bdva.eu