THE SANO CENTRE FOR COMPUTATIONAL MEDICINE: INTERNATIONAL RESEARCH AGENDA

Arkadiusz Sitek, Marian Bubak, Piotr Nowakowski

Sano Centre for Computational Medicine
International Research Foundation
Czarnowiejska 36
30 054 Kraków, Poland
e-mail: {a.sitek,m.bubak,p.nowakowski}@sanoscience.org

March 21, 2022

Abstract. The aim of this paper is to introduce the research agenda of the recently established Sano Centre for Computational Medicine - a research institution based in Kraków, Poland. Following an introduction to the Centre, the paper describes the strategy, details specific areas in which the Centre carries out research, and concludes with emphasizing the goals and impact.

Keywords: Sano, computational medicine, medical research

1 INTRODUCTION

This paper presents the long-term Research Agenda of Sano, developed in the context of IRAP (international Research Agendas Programme), the Polish funding mechanism which provides a major source of funding for the Centre. It defines the Research Agenda as a fundamental document for the Centre. Important steps which led to this stage of implementation are also presented,

The Sano Centre for Computational Medicine is a non-profit research institute dedicated to the advancement of computational medicine. The strategic objectives of the Centre are to become Central Eastern Europe's foremost academic institution for the identification, development, and implementation of techniques in Computational Medicine The Centre combines academic, industrial and clinical perspectives to inform innovative R&D processes, delivering competitive products and services to the marketplace and creating sustainability It also organizes internal and external training programmes to provide the next generation of healthcare technology innovators with the unique skills profile required by computational medicine.

The Sano Research Agenda, which is put forth by the Centre and subject to approval of its funding bodies, includes, among others:

- Description of the societal challenge which Sano addresses
- Overview of the proposed approach
- Discussion of risks related to the field and to the proposed approach itself
- Processes and activities which will support Sano's scientific teams
- Depiction of the value chain Sano is expected to contribute to, together with main external partners Sano will cooperate inside this value chain
- Presentation of six research teams which will constitute Sano, and how they interact with each other
- Objectives which this agenda sets out to achieve in short and long terms. The short term is related to the span of the current IRAP project, the longer term to a more distant vision of Sano and its impact on medicine and healthcare.

2 RESEARCH AGENDA OUTLINE

There is little doubt that the status quo in global healthcare systems cannot be sustained in the long run from the economic point of view. Poor efficiency, ageing populations and high medical error rates are only some of the other serious issues that plague healthcare. We elaborate on some of those problems in Section 2.1 Precision medicine (PM) briefly described in Section 2.2 with important molecular biology components described in Section 2.3 is a potential solution to this crisis. Although the ideas of PM are not new, its full implementation is difficult and the route to achieve it unclear. Computation will play an essential part in the implementation of PM concepts in clinics, which is argued in Section 2.4. The main goal of this RA is to create knowledge and solutions, which will bring computation to the clinics to assist in decision making, to gradually transform healthcare to the vision of the PM as outlined in Section 2.2. In the final section of the introduction, Section 2.6, we provide an outline of steps that will be taken to achieve our goals. We provide a general strategy to achieve our goals and a more detailed description of core research areas in Section 3 and 4 respectively. Section 6 summarizes the paper.

2.1 Global Challenges of Healthcare Systems

The past two decades witnessed tremendous progress in medical and computing technologies. Notwithstanding, the positive effects of technological growth seen in some important medical applications are mediocre. We still observe a preponderance of chronic medical conditions ravaging populations in developed and developing countries alike. The prospects for some of those conditions are dire. An example is dementia, caused largely by Alzheimer's disease, where the current number of people worldwide affected by the disease is expected to grow from 50 to 152 million by 2050[1]. There are other, even more alarming examples of chronic conditions already seen in children such as autism, diabetes, learning disabilities and attention deficit hyperactivity disorder (ADHD), whose incidences are on the rise. Despite decades of research and generous funding by governments and industry, only moderate success can be claimed in the fight against cardiovascular and cancer related morbidities, which are the two most frequent causes of death in Europe and the US. For example, we observe a steady decrease in cancer deaths since the nineties, mostly because of the reduction in cigarette smoking. However, this positive trend is slowing[2] and it seems unlikely that substantial positive change will take place in the next decades without some dramatic breakthrough. Similarly, for cardiovascular disease (CVD), the number one killer in Europe and the US, a similar trend exists whereby years of research and preventive measures have led to only minor improvements in CVD incidence and related deaths[3]. Direct and indirect costs of CVD are staggering and reach half a trillion Euro per annum in Europe and the US. The economic cost of CVD, along with other diseases such as cancer, neurodegenerative diseases, mental disorders, and many others, represent existential threats capable of bankrupting modern societies.

2.2 Precision Medicine

Precision medicine has been heralded as the future of healthcare. Following the visionary works and writings of Leroy Hood et al.[4], we summarize Precision Medicine (PM) by referring to "4Ps". They stand for Predictive, Preventive, Personalized and Participatory which shift focus of medicine from controlling disease to maintaining well-being. Precision medicine as defined by 4Ps reasserts the teachings of Hippocrates, the father of modern medicine, who saw patients as unique human beings and saw physicians as those who predict and prevent. "Predict" means that any medical intervention or lack thereof should be associated with prediction. Ideally, we would like to have the capability to predict a conditional future. For example, one can predict medical outcomes given different medical/non-medical interventions. "Prevent" applies only when we know how to predict. If physicians and patients alike become aware, based on predictive models, of a given path (lifestyle, social situation, diet, medical treatment) likely leading to unwanted outcomes (disease, serious side effects of medical treatment, death), then preventive steps can be implemented. "Personalized" in 4Ps means that medicine treats each patient as a unique entity. Each human has a unique genome (except for identical twins) and progress in genome sequencing has made this information readily available to PMs (see also Section 2.3). The genome is only part of what makes humans unique. Another very important factor which shapes our phenotype and who we are as humans is the history of our interaction with the environment. Genome and genome-derived characteristics of humans, coupled with epigenetics, proteomics and other molecular biology disciplines and environmental/historical factors, define our unique characteristics. We can use those characteristics to predict and prevent. To implement the three Ps already described, we need patients to "participate" (fourth P) to enrich the data pool used to build predictive models and actively involve patients in health-related decision making. Self-medication, self-monitoring and self-education are integral parts of their participation. How to achieve 4P medicine is far from obvious, but computational decision support may represent an integral part of the solution (see Section 2.4).

2.3 Molecular Biology

To adapt healthcare to the needs of different individuals, which is one of the premises of PM, one must harness the power of modern molecular biology (MB). Applications of MB are increasingly entering clinical practice thanks to, among others, the cost of genome sequencing falling well below €1000. Yet, these multi-omics techniques have driven important discoveries - new driver mutations in oncology, the relationship between the gut microbiome and type-2 diabetes as well as other diseases, and new therapeutic targets for rheumatoid arthritis - and are regarded by many experts as the real key to personalized medicine. With further improvements about to become mainstream, the generation of multi-omics datasets, even from single-cells, will become even faster, easier, cheaper, and more precise. Sano has a strong link to MB through two members of Sano Clinical Advisory Committee leading top MB laboratories in Kraków. Furthermore, in its preparatory research, Sano is already investigating, through simulation, the impact of clustered SNPs on protein function. Further links to MB are presented in Section 2.6.

2.4 Need for Artificial Intelligence (AI) and Computational Models

Doctors have traditionally customized their medical advice to individual factors such as age, gender, patient preferences, mobility levels, community resources, pre-existing conditions, and other mitigating circumstances for many years. However, throughout the millennia of medicine's history, we observe an increase in the knowledge about human biology, along with a steady growth in the number of medical treatments, devices, and drugs developed each year. This expanding knowledge contributes to substantial increase in the complexity of healthcare delivery, and makes it very challenging for the human brain to take all relevant factors into account when making healthcare-related decisions. When faced with a complex health issue, we no longer visit just one physician. Instead, we likely undergo a series of tests (such as laboratory tests or radiology) with results examined and interpreted by other physicians whom we have never met and who know little about our particular case. Computers do not have such limitations and can access nearly unlimited amounts of relevant information specific to the patient in a relatively short time. This breadth of information might provide invaluable support for healthcare workers and patients to make optimal health-related decisions if efficiently communicated, interwoven in clinical workflows and analyzed by computational models (artificial intelligence). We use the term AI rather loosely and in contrast to its formal definition. We associate AI with computer algorithms providing "intelligent" assessment or advice on the problem at hand using data and modeling. Some examples of AI's uses in medicine include: (1) advice whether to discharge or admit patients in the ER, (2) predictions of efficacy of medical procedures, (3) design of optimal medical interventions, (4) guidance for the surgeon during surgery, (5) interpretation of radiological data, (6) assessment of the relevance of medical information and efficient display of this information, (7) prediction of disease spread or identification of hot spots during pandemics, (8) automatic completion of medical forms, charts, and many others. The computing technology which underwrites AI varies from machine learning and deep learning to optimization, modeling and simulation. Unquestionably, AI has provoked a fundamental reassessment of the utility of clinical data in identifying effective approaches to healthcare. Yet, particularly in medicine, there is an unease that the underlying basis for the success of AI techniques needs to be more firmly established in medical practice. Sano recognizes the compelling need to establish such a foundation and, with expertise across all the required disciplines, is ideally positioned to drive this pursuit of the fundamental science critical for successful implementations and acceptance of AI. By establishing this goal as one of the specific targets within the computational modelling group (see Section 4.1, we position Sano as the Centre of choice for collaborators seeking to use AI for otherwise intractable clinical challenges (competitive advantage). Data science algorithms are typically built based on large amounts of data and machine learning. Sano also concerns itself with building predictive models from the ground up, using modelling and simulation, without heavy reliance on the availability of data. This approach forms the basis of the pan-European Virtual Physiological Human (VPH) initiative[5], which began in the nineties and continues today. In the modeling and simulation approach, we use computational models of the human body at various scales to diagnose, measure or predict. We note one important difference: data science approaches look at the prediction problem from a different perspective than modeling and simulation approaches. Rather than trying to explain and model the phenomena using basic principles (physics, biology, etc.), extensive amounts of data are collected first and exploited by machine learning (ML) algorithms. The strengths of both approaches rely on different assumptions. The modeling and simulation approach is conditioned on the correctness of the assumed model. Data science relies on data availability, but both approaches are largely complementary, which is our competitive advantage as we build on the rich experience of our partners (Sheffield and Jülich) in modeling and simulation, and extend the methodology toolbox with a wide variety of data science approaches. Unique combination of expertise in these two methods constitutes the original approach of Sano to precision medicine and its impact on future healthcare.

2.5 Risk and Risk Mitigation

We briefly touched upon a few important areas where computation can be a major factor assisting transformation to better medicine. There are many others, such as healthcare delivery logistics, telemedicine, medical errors and side effects, regulatory effectiveness, health disparities or ethics. Despite its clear potential in those areas, the use of computation in medicine carries risks. The most prominent risk is lack of acceptance by the medical community. To account for it, methods need to be seamlessly interwoven into clinical workflows, accurate and useful. They must also be as transparent as possible. If a difference of opinion between human and machine occurs, coming to a common conclusion is difficult without transparency. Regulatory risk is another very important aspect of applications of computing in medicine. Data science requires access to large amounts of patient data, which is highly regulated and appropriate steps must be taken to ensure protection of patient privacy along with regulatory compliance. In recent years, the number of applications for regulatory approvals of artificial intelligence has increased dramatically. The regulatory landscape for those applications is still forming and, as such, poses risks to the commercialization process and widespread use. Sano develops its own expertise in regulatory pathways. Another important risk involves poor algorithm performance and lack of generalizability. AI algorithms which rely on deep neural networks show amazing performance, often matching or exceeding humans in many medical and non-medical applications, all of which is documented in scientific literature. These findings are relatively recent and the generalizability of those reports remains unclear. This risk can be mitigated by using diverse data sources when creating AI algorithms, but that, of course, runs into regulatory issues described previously, along with the risk of limited data access. The latter risk is directly addressed in this proposal by development of a new pipeline for data harvesting and AI model training.

2.6 Outline of the Research Agenda

The dominant idea of Sano's Research Agenda (RA) is to harness innovations in high-performance computing, artificial intelligence, and clinical workflows to create practical solutions which will tangibly improve healthcare in Poland and the rest of the world. Sano executes this agenda through clinical problem-driven research activities schematically presented in Figure 1. The Centre comprises six research groups (see Figure 2) which were chosen keeping in mind the diversity of technical and application expertise, along with the potential for synergistic collaboration (see Section 3.1) between these research teams.

Section 4 provides an overview of the application areas for the six groups. Through these areas, Sano pursues its mission of developing and testing - in clinical environments - novel approaches to data visualization and remote interactions/communications between patients, healthcare professionals, AI, and telemedicine (Figure 5, health informatics group).

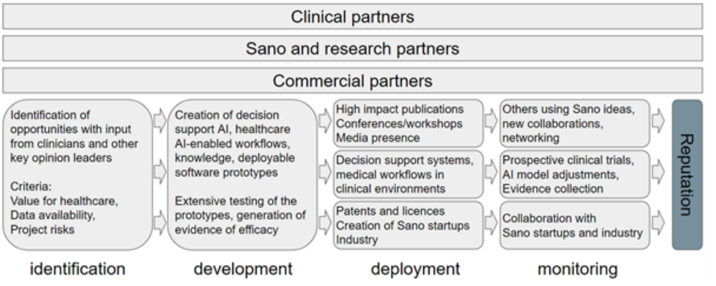

Fig. 1. Diagram illustrating the value chain of Sano.

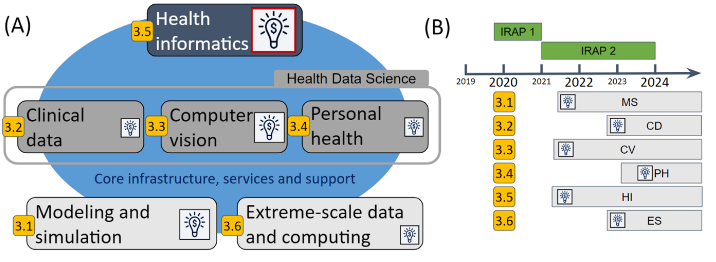

Fig. 2. (A) Schematics of the six scientific pillars which correspond to research teams of Sano Centre. Core (blue ellipse) is an underlying computing infrastructure with many elements shared by the pillars. There is the modeling and simulation (MS) group which uses basic science and models of biological systems (see section 3.1). There are three data science groups (CD, CV, PH) with efforts concentrated on three different types of health-related data and described in sections 3.2-3.4. Health informatics (HI, section 3.5) group is an interface to the clinical world. Extreme-scale data and computing group (ES) addresses fundamental computer science problems driven by Sano research (section 3.6). Bulb icon next to each group indicates that all groups have entrepreneurial potential and will interact with entrepreneurs at some point of the project. (B) Timeline for expected time periods when groups will interact with entrepreneurs.

We aim at the development of computer support in decision making based on patient-specific data. Our goal is to test those systems in many disciplines including, but not limited to, emergency medicine, oncology, cardiology and neurology. In particular, we harness recent progress in computer vision to develop and test clinically relevant use cases in radiology, digital pathology, surgery, and other medical specialties which rely on visual information (Figure 1; computer vision group). We investigate socioeconomic region- and nationwide medical event data, Internet-of-Things (IoT), self-health monitoring approaches, health risk models and the promotion of health (Figure 1; personal health group). We research bottom-up approaches based on computing and simulation of in-vivo processes at all scales with applications to medicine (Figure 1; modelling and simulation group). The extreme-scale data and computing group in Figure 1 carries out work on computer science solutions in data science, modeling and simulation. As each group produces commercializable output, all remain in contact with entrepreneurs from the outset. Note that core infrastructure services and IT and security support is a non-research concern, and will be addressed by a specialized team.

The Sano research agenda is being executed in close collaborations with research and clinical partners in Poland, Europe, and the USA. We work to expand existing clinical collaborations in Poland and the EU to include top medical academic sites in the USA, with whom Sano representatives have previously forged academic and business ties. We also work towards extending our clinical collaborations to China either through bilateral programs or directly with clinical/academic sites. For more details on planned collaborations please refer to Section 4.2.

3 STRATEGY FOR SUCCESS

We build the strategy with short (IRAP funding horizon) and long-term goals in mind. Here we present how we see this agenda as a foundation to carry on research during and beyond the IRAP funding.

We design areas of expertise of six Sano teams with synergy in mind Section 3.1 and pay close attention to mentorship and education (Section 3.2). We differentiate between clinical and research collaborations. For clinical collaborations (Section 4.2), we seek partners who can either share clinical data or test our clinical decision models at their institutions (real-world testing). In research collaborations, we expect scientists from collaborating institutions working together with Sano faculty and trainees on data science, health informatics, or modeling and simulation projects. These institutions are ready for research exchange programs with Sano at faculty, post-doc, and student levels, and their researchers serve as co-mentors for Sano's master and doctoral students. Section 3.5 and 3.6 touch upon our commercialization strategy, while Section 3.7 describes the computing infrastructure supported both by IRAP and its complementary funding stream: the Teaming for Excellence project (which began in August 2019). Allocation is close to 9 million Euro, and 15 million Euro to the entire Sano partnership network. These processes, strengthened by direct involvement of Sano's Advanced Partners over the next six years, ensure a supportive environment for research.

3.1 Interactions and Synergies between Sano Teams

We regard medical care as a decision science. Medical professionals, patients, and relatives decide about prevention, diagnosis, and therapy. The goal of Sano is to provide support in those decisions based on data and computational models to have decision makers choose the most informed decisions possible. We stress here that we see the methods developed as an aid in health-related decision making and always (if possible) consider humans in the decision chain. As visualized in Figure 3, three research groups (personal health, clinical AI, and computer vision) are data science groups. They differ by the type of data that they will focus on. However, there is much synergy between those groups as they will all predominantly use machine learning. Although each deals with a different clinical specialty and type of data (structured, unstructured, image, video), the resulting expertise and its diversity will benefit other groups which also apply machine learning.

The Modeling and Simulation group will work with all types of data. As mentioned in Section 2.4 and elaborated in Section 4.1, the group uses approaches complementary to methods utilized by the data science groups. Information obtained from the data (either through data science or modeling/simulation) is meant to be applied synergistically when medical decisions are to be recommended. Occasionally, medical decisions can be based on a single type of data, say imaging. This is either because of the specific nature of the decision or because other types of data may not be available, but more often than not processing accounts for ancillary information and a better decision can be reached when such information is taken into account. Sano dedicates only a single group to modeling and simulation and relies on strong ties to the University of Sheffield with its considerable strength in this area. The Health Informatics group (Figure 1) works on novel approaches to deliver insight from computer algorithms to clinical users and optimize human/human and human/machine decision making. This obviously fits well with other groups, and is complementary to data science and simulation groups. All of the listed groups will collaborate with the Extreme Scale Computing group which will develop novel methods for easier and more secure access to healthcare data and new sophisticated state-of-the-art pipelines to process them.

3.2 Mentorship and Education

Any flourishing research institute relies on exceptional faculty members, postdocs, and students. It is therefore critical to ensure steady recruitment of talented personnel. To attract talent, we will build a reputation as a world-class research institution by producing high-visibility research output (RO) published in top journals, presented at global scientific meetings, and featured in the mainstream media. Through collaborations - again, some already forged in my prior academic engagements - with top world-renowned research institutions (Harvard, MIT, Penn U, Stanford) and Industry (Compugroup, Salesforce, IBM) we work to establish student and postdoc exchange programs and internships. These exchanges directly affect the IRAP project by creating joint RO, and indirectly by widening the expertise of Sano employees. We strive to have Sano graduates well-rounded and adequately prepared for their future careers in either academia or industry, ultimately creating long-lasting positive influence on Polish science and economy.

Good understanding of technology, trust and technical expertise of potential users and cooperation partners is essential to increase the impact of Sano's R&D activities and to successfully implement new technologies. Sano works to provide nonacademic training sessions targeted at potential users and collaboration partners, including clinicians and other healthcare staff, medical students, business representatives and researchers. The rationale behind these sessions is that computational medicine technologies, their opportunities and existing limitations must be well understood by existing and potential cooperation partners, since only then will they be able to make the best possible contribution to joint R&D projects. Understanding is also a prerequisite of trust, and trust in new healthcare applications is of crucial importance for the acceptance of Sano's technology. Therefore, Sano seeks to address future users and researchers from different fields by offering customized training. Seminars, workshops and internships not only raise interest in Sano's activities among future user groups and familiarize them with relevant technologies, but also act as a platform for continuous dialogue on relevant topics between themselves and Sano. Such ongoing dialogue helps Sano keep track of emerging needs and concerns. In addition, training activities create opportunities to address the critical need of interdisciplinary expertise by informing and attracting specialists from relevant disciplines, including generating interest from potential future employees, which in turn, helps nurture the interdisciplinary environment at Sano.

3.3 Clinical and Research Collaborations

Although Sano concentrates on developing computing solutions, we must strongly emphasize that in order to achieve success and meaningful innovation, the expertise and guidance from clinicians and clinical scientists is of paramount importance. We see the Sano Institute implementing its Research Agenda by using a top-down approach whereby we identify relevant problems together with clinicians, researchers and scientists (see also Figure 3.3 - Sano partners with Fraunhofer ISI on analysing the computational medicine ecosystem and eliciting user needs) and then develop and deploy those solutions in close collaboration with clinical experts. Sano is in the process of establishing a Clinical Advisory Committee (CAC), composed of renowned clinicians - indeed, fifteen senior clinical investigators have already expressed interest in CAC membership (specific names can be provided on request). Once formally established, CAC will be an important source of research topics for Sano groups and PhD theses. Clinical collaborators, as well as patient organisations, will be directly involved in Sano projects.

The emphasis on the clinical aspect of the projects will contribute to cross-fertilization and better understanding of the clinical and scientific context. We stress that the number one risk identified in Section 2.5 is the lack of acceptance of new methods by the medical community. Therefore, close collaboration with and guidance from clinicians at all stages of projects is of extreme importance and absolutely necessary for Sano's success. To this effect, we extend the current CAC base by seeking additional strong relations and trust between Sano and collaborating medical centers in Poland, the EU, USA, and beyond (Figure 3.3).

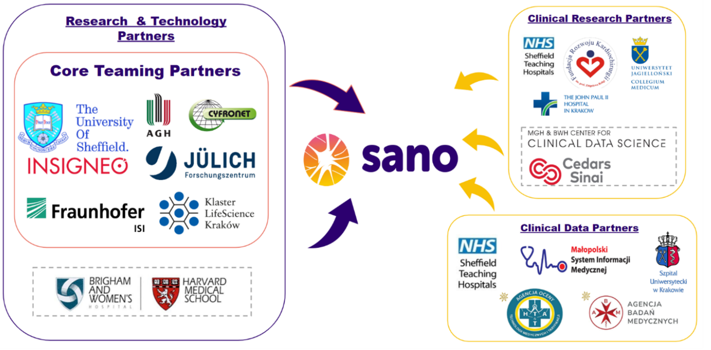

Fig. 3. Research partners (left) and clinical/research partners right which are in collaboration with Sano.

3.4 Clinical Data Acquisition Strategy

Sano would not be able to pursue its ambitious technological advancement agenda without adequate sources of high quality clinical data, access to which is not trivial. Large amounts of patient data are necessary to build high-impact AI, simulate operations of the human body using in silico methods, or combine both approaches. We have devised and are enacting a multi-faceted strategy to ensure acquisition of the data needed for the projects and mitigate any risks related to e.g. relying on a single data source.

In order to be able to build an adequate critical mass of statistically significant cohorts, Sano seeks access to registries of meta-information on EHR in Poland, either in the framework of NFZ (nationwide - AOTMiT, ABM) or the Małopolska region (MSIM), with whom joint research proposals have already been submitted to funding agencies. Having agreements with these agencies in place will allow Sano to rapidly obtain information on clinics treating given diseases or performing specific medical procedures, along with statistical information. Once relevant hospitals are identified, Sano's business development unit can negotiate agreements, fully respecting all ethical privacy concerns according to the current European and Polish regulations of access to data, simultaneously ensuring collaboration of clinicians involved in treatment of patients in question.

Sano is also securing framework agreements with several large Polish teaching hospitals (this includes Poland's largest University Hospital in Kraków). The scale of these institutions, along with their strong involvement in research, should guarantee an adequate supply of research ideas and data for Sano researchers. Involvement of clinical personnel will be stimulated with co-authorship of publications, co-supervision of PhD theses, joint technology development, early access to prototypes, joint applications for research grants, and similar non-monetary opportunities.

Successful application of computational methods in medicine calls for data from many institutions to create robust generalizable models. Heterogeneity of data is also important when the developed solutions are tested for accuracy. Even if a data-intensive solution (e.g. based on AI) works on data harvested at one institution, it may fail when supplied with data from another institution, due to differences in recording, procedures, personnel expertise, hardware etc. Moreover, lack of generality might stem from differences in patient demographics used for training models compared to the demographics upon which the model is clinically deployed or used for research studies. It is therefore desirable to use data from as many institutions as possible, and once Sano establishes working contacts and operating procedures with Polish clinical data, it will expand its data acquisition operations to other countries.

Guided by its advanced partners, Sano develops mechanisms to ensure protection of patient rights. In particular, General Data Protection Regulation (GDPR) or other regulations adequate for the countries of data origin are duly enforced, with regular training sessions organized for Sano employees and collaborators. Full compliance when dealing with large amounts of data requires a significant amount of innovation (acknowledging recent developments in distributed computing, big data analytics and federated learning) and effort, but it is a cornerstone of a robust and successful data science institute. Moreover, such innovative mechanisms constitute an important asset that can be disseminated in the community through licensing or using open-source mechanisms advancing Sano's reputation.

We are working to establish collaboration with European initiatives such as EC's European Health Data Space (EHDS), which fosters exchange and sharing of various types of health data (EHR, genomics, registries, etc.) in Europe. The goal of the initiative is to support delivery of primary care, development of new treatments, drugs, medical devices and services, whilst simultaneously protecting citizens' data. EHDS is currently at the planning stage and we will take steps to be actively involved in this initiative, closely collaborating with the European Federation of Pharmaceutical Industries (EFPIA). One of such steps is Sano's current involvement in formulation of a joint declaration on the steps to create EHDS.

3.5 Industry Relations

In addition to fostering academic collaboration, Sano strives to establish strong co-operation with the industry. This is necessary not only because of potential revenue streams, but also because it provides a different-than-academic perspective needed to balance the research goals of the Centre. With this in mind, Sano has already forged some bonds with commercial institutions and received a number of letters of interest. This list includes representatives of large corporations, like AstraZeneca, GE Healthcare and Siemens, as well as smaller "rising stars" of Polish biotech (e.g. Captor Therapeutics, OncoArendi and Medicalgorithmics). In analogy to CAC, Sano is working to create an Industrial Advisory Committee (IAC), which will engage industrial partners in joint research projects that yield clinically exploitable results. Collaborative creation of joint IP with the industry may happen across multiple domains, where integration of new Clinical Decision Support Systems (CDSS) with existing solutions (Hospital Information Systems, diagnostic equipment), awareness of international healthcare delivery standards, and provisioning of training as well as a trained workforce, are all key elements of success. Industrial engagement is feasible at many stages throughout the Sano R&D cycle. To this end, Sano has also established relationships with international clusters and alliances, e.g.: Life Science Nord, Biosaxony, EFPIA and Associations or Avicenna Alliance for Predictive Medicine.

The overriding goal of Sano is to create better medicine through technological advancements in computing. There is no better reward for a medical researcher than to see their inventions making a difference in human lives. However, the road from an idea to implementation, especially in the strongly regulated medical sector, is arduous and fraught with uncertainty. We therefore attach great value to input from industrial partners in this regard. While building lasting collaborations between the industry and research institutions is challenging, as it is necessary to find common factors that reward both sides, results provide more efficient pathways to the market, which eventually benefits all stakeholders. Our long-term strategy is to attract commercial partners by building a solid reputation as an excellent medical computing institute capable of performing, among others, contractual research in this highly competitive field. Moreover, depending on the needs of the commercial partners, Sano will offer research output, access to data (GDPR-compliant), joint development projects, expertise, training, access to intellectual property (IP), specialized services. Examples of such services may include testing industrial AI prototypes against clinical and regulatory constraints, or providing assistance in obtaining CE marking. Since the global market for AI solutions is already enormous and rapidly expanding, many industries will need to go through FDA or EMA scrutiny with their products, which will create demand for such services. This strategy may provide revenue to Sano and, in the long run, strengthen it as a European leader in applications of computing in medicine.

3.6 Entrepreneurial Activity

Research groups will work on solutions with commercialization potentials. A timeline of interaction between the groups and entrepreneurs is presented in Figure 2B. They will be incentivized and assisted by Sano support and development teams to carry out market research and protect potential IP at the identification stage (Figure 1) of the project. Throughout the entire cycle of project development, group leaders and members are encouraged to interact with entrepreneurs and investors to seek commercialization opportunities using private and public capital through spinoff mechanisms.

We build an environment and culture which encourage and nurture entrepreneurship in Sano's students and postdocs. The business development team identifies commercial-value opportunities and supports development (business-oriented, entrepreneurship or IP protection courses) of team members to ensure appropriate inventive processes. This team, along with Sano's legal advisors, further assists Sano scientists with an adequate IP protection approach, and by attending seedfunding opportunities organized either by Sano or by outside entities. Such events (boot camps, startup scene etc.) operate in two ways: through networking with innovation hubs, Sano scientists are able to gain valuable knowledge from experienced entrepreneurs, and discuss engagement opportunities with them, but such events will also allow some Sano graduates to take entrepreneurial pathways in their further careers, building new companies based on Sano's IP, with continuing help and participation from Sano in finding capital and fostering growth. To achieve these goals, Sano sets up procedures for formal establishment of new spinoffs with third-party investment, and helps them achieve quicker growth, among others by addressing regulatory, legal, recruitment and collaboration issues. An important goal of Sano is to produce graduates with a strong entrepreneurial spirit who would carry Sano's ambitions to improve healthcare and people's health and well-being.

3.7 Computing Infrastructure and Services

Without doubt, computational medicine requires considerable computing and data storage capabilities. Access to adequate computing infrastructures is provided to Sano mainly via collaboration with the Academic Computer Centre Cyfronet AGH and Juelich Supercomputing Centre (JSC), both of which are its Teaming project partners. Cyfronet operates Prometheus, the largest supercomputer in Poland, which Sano uses for large-scale simulations, data analysis and AI workloads. This infrastructure is available to all research institutions in Poland, including Sano, while special demands will require separate agreements. For even larger experiments, we rely on European infrastructures, such as PRACE and EuroHPC, in which both JSC and Cyfronet are active. Finally, the Centre may obtain additional computing and storage capacity from external providers, such as Amazon, Nvidia or Microsoft. Partnerships with industry will be sought to complement publicly available academic computing capabilities, if needed.

Sano's IRAP researchers are assisted by experienced technical personnel: research software engineers, UX programmers, data managers/analysts, etc. This cross-group team, following in the footsteps of advanced institutions such as the Netherlands eScience Centre or Juelich Simulation Labs, focuses on high-level support for scientific software development, creation of research databases, producing interfaces for users and external systems, supporting processing pipelines and managing access to computing, networking and storage infrastructure the Centre is using. This multifaceted approach leads to more efficient translation of research outputs into industrial and clinical prototypes.

4 CORE RESEARCH AREAS AT SANO

In this section, we provide a general description of the six research pillars of the Research Agenda and a description of the interactions between these pillars. Specific focus of each group will be determined based on guidance from prospective group leaders and clinical collaborators.

4.1 Modeling and Simulation

With close collaborations with our research partners, we aim to expand research from the Virtual Physiological Human initiative. A rich assortment of complementary modelling approaches has already been applied in healthcare, including 3D (e.g. Navier-Stokes solvers, Lattice Boltzmann approaches), 1D and 0D models of fluid mechanics for cardiovascular and respiratory applications, Finite Element Analysis, Growth and Remodelling frameworks and Agent-Based Models of structural mechanics for cardiovascular, musculoskeletal, respiratory, and oncology applications. This research area is concerned with the development of fundamental modelling tools and workflows. We simulate physiology and predict multi-morbidity through integration with the healthcare data and data science technologies.

We also look into applications of in-silico clinical trials which concerns modelling and simulation of devices, drugs and interventions to be evaluated over trial populations covering complete target treatment populations, hugely reducing the time and cost of traditional clinical trials through the use of computing. Significant research questions remain around these approaches, particularly over missing and inaccurate data. We look for solutions using the Data Science and Healthcare Informatics for input data distribution and accuracy, and with the Modelling and Simulation (MS) for output uncertainty determination.

Some machine learning (ML) approaches such as deep learning suffer from an inherent problem of the lack of explainability. ML may provide accurate recommendations, but it rarely explains why it has made particular decisions. Models based on accurate multiscale biology physics can be developed in the attempts to explain AI decisions. On the other hand, ML can also complement MS as it can be used to reduce the computational burden associated with simulation of complex biological systems. ML can be considered non-linear adaptation of data to target. The latter can be represented, for example, by the likelihood of morbidity or recommendations concerning therapy. The rule of thumb is that more abundant and more accurate data results in more accurate estimation of the target. We see MS as a tool which uses knowledge of biology and physics to generate relevant data that can be used either as features or/and targets in ML. Given that MS approaches are sensitive to initial parameters, which are often missing, machine learning can be used to estimate the values and uncertainty of missing parameters.

4.2 Clinical Data Science

Sano conducts research into improvements in decision making at all stages of patient interaction with the healthcare systems and in particular: (1) Primary care physician (PCP); diagnosis - analysis of data to provide differential diagnoses with predictions of uncertainty, (2) Primary care physician (PCP); risk - analysis of data to estimate risks of future morbidity (e.g. cardiovascular events, breast cancer, prostate cancer, lung cancer), (3) Selection of medical intervention (PCP and specialist, e.g. oncologist) - patient-specific data used to determine the optimal path of treatment (value-care medicine), involving predictions of drug/treatment/medical procedure side effects, risk/benefit of hospitalizations, etc., (4) Treatment monitoring - analysis of therapy effectiveness indicators or outcomes in treatment management (value-care medicine), involving time-series analysis of data related to treatment to provide recommendations about its effectiveness and the likelihood of success based on patient-specific data. The decision support AI relies on data science machine learning and deep learning models based on a combination of one or more of the following patient-specific data types: patient presentation, medical history (electronic health records EHR), laboratory results, clinical notes, medical imaging data (current and past), molecular biology data, socioeconomic factors, patient-specific simulation data, and other available relevant data.

There is a trade-off between the accuracy of AI and interpretability. Although accurate, AI decisions produced by deep learning models are opaque to interpretation and often referred to as black-box decisions. This may create serious problems with the acceptance of such methods in clinical practice[6]. We investigate machine learning models based on optimized scoring results that are fully and easily interpretable by humans. These types of approaches require substantial computational power and advanced algorithms because linear and non-linear mixed-integer problems are to be solved[7].

4.3 Computer Vision Data Science

Interpretation and analysis of visual information in healthcare is among the most attractive and impactful applications of AI which can dramatically and positively impact medical practice. In recent years, convolutional neural networks and deep learning have revolutionized computer vision (CV) applications. The impact of this revolution is felt in many industries, including healthcare. AI used to interpret medically relevant visual data will bring medical practice closer to the ideals of 4P medicine.

Many healthcare-related applications of CV exist, some of which are schematically presented in Figure 4.3. One of the most obvious applications is automatic interpretation of radiological images which may improve patient-specific accuracy and reduce radiologist burnout (Figure 4.3A). These technologies will provide support to radiologists to reach accurate and consistent expert-level decisions or provide a second opinion about a potential pathology in longitudinal imaging data. There are extensive amounts of longitudinal radiology data accumulated for patients and it is an impossibility for radiologists to thoroughly examine them without AI, thus creating a high-impact application area for AI technology. Applications in automatic image interpretation during procedures such as endoscopy or AI assistance during surgery (Figure 4.3B) and for automatic interpretation of digital pathology (Figure 4.3C) provide additional examples of the applicability of CV algorithms. AI systems can analyze thousands of hours of surgical videos to develop patient-specific surgical procedures. For example, we can detect anatomical structures such as major nerves and provide information to surgeons to prevent nerve damage. Feeds from classical/infrared cameras can be used in conjunction with CV machine learning for patient safety, logistical optimization, monitoring the spread of infections, health assessment, mobility, and so on (Figure 4.3D).

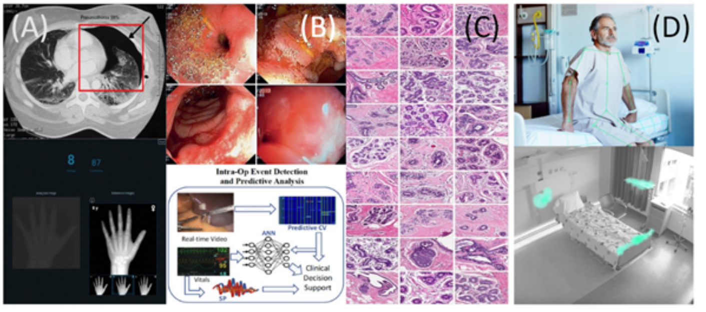

Fig. 4. Examples of visual information obtained in healthcare which can be processed with AI. (A) Medical imaging, (B) endoscopy, colonoscopy, surgery, robotic surgery, (C) digital pathology, (D) Optical/infrared camera and other sensor feeds. Image credits: Choy et al, Radiology, 288, 2018, Lourenço et al, Port. Journal of Gastroenterology, 24, 2017, Hashimoto et al, Annals of surgery, 268 2018, Ouva.com.

It is important to note that data science image/video-based information can also be combined with other types of medically relevant information to provide synergistic decision support based on visual and non-visual cues (e.g. clinical history, molecular biology, etc). The interrelations between the phenotype, determined on the basis of medical imaging, and molecular biology are also of high relevance to precision medicine and will be investigated.

4.4 Personal Health Data Science

In this pillar, we concentrate on research aimed at shifting healthcare philosophy from reactive to proactive. This is very much consistent with the 4P medicine approach outlined in the introduction, and it is an integral part of our vision of Healthcare 2050. Over the past thirty years there has been a fundamental change in how we interact with information through the rise of the Internet and popularity of mobile devices such as cell phones. Most people are now constantly "online", consuming and providing enormous amounts of information through the Internet. We have witnessed spectacular growth in the popularity of social media such as Facebook, Twitter, etc. This trajectory will continue and although it is always risky to predict the relatively distant future, humans will probably be even more strongly connected to larger amounts of information, and will generate more of it by the year 2050. Interestingly, as we witness our social behaviours transform, healthcare delivery remains relatively unchanged compared to thirty years ago. We therefore perceive strong potential for creating a positive change in how we manage and influence our health.

The long-term goal is to create a personal health tracker that will collect all available information about individuals and provide prediction of our health. Information that the tracker will use includes longitudinal data from medical records, genetics, microbiome genetics, behavioral information (exercise level, sleep quality, physical activity, etc.), stress levels, dietary habits, food quality, environmental factors (e.g. pollution, allergen levels) and other socioeconomic factors. The tracker will not only compute existing risks, but will also predict them if, hypothetically, the input changes. Therefore, it will empower individuals to affect their own health by providing knowledge on how to improve it. It will give the power of preventing disease and enable people to live longer and in better health.

Through this RA we will work towards the long-term goals exploiting Poland's unique position as a single-payer healthcare system with centralized historical data about interaction of a vast number of patients with the healthcare system (collaboration with the Agency for Medical Research, Ministry of Health), census, death records, environmental data which will be used to develop machine learning for analytics and to create predictive models of personal health. We also work with medical institutions (existing and future collaborations) to obtain detailed medical data (EHR imaging, genetics) of patients. Given a subset of patients from nationwide records, if chosen strategically using carefully selected institutions around Poland, we will be able to develop accurate machine learning models despite the potentially large number of missing values for the entire population.

4.5 Health Informatics

Health informatics is information engineering applied to healthcare. It manages the use of patient healthcare information and is a vast multidisciplinary field that spans information science, computer science, social science, behavioral science, management science, and others. It deals with the resources, devices, and methods required to optimize acquisition, storage, retrieval, and use of information in medicine. Within this Research Agenda we concentrate on a new generation of medical communication and incorporation of AI in medical workflows. We define communication in medicine as exchange of information between decision agents (DAs), i.e. entities which can influence health-related decisions about individual patients. We identify four types of DAs, as shown in Figure 5. The patient and their family are obvious DAs, and the most important ones as they often make the final decisions. From healthcare professions, we identify primary care physicians (PCPs) who in many healthcare systems are the first points of contact for patients. Typically, PCP will also have the most interactions and spend the most time with the patient. Care teams are another DA, comprising other healthcare professionals involved in patient care. This includes medical specialists, nurses, technologists, etc. Finally, in the new paradigm we identify data and AI as a distinct DA because data is not merely a passive source of information for other DAs but in fact enables AI to make active decisions. In some implementations, we may use the AI only to improve data presentation, but even in this application we consider it as a DA because the way it presents the data may affect decisions.

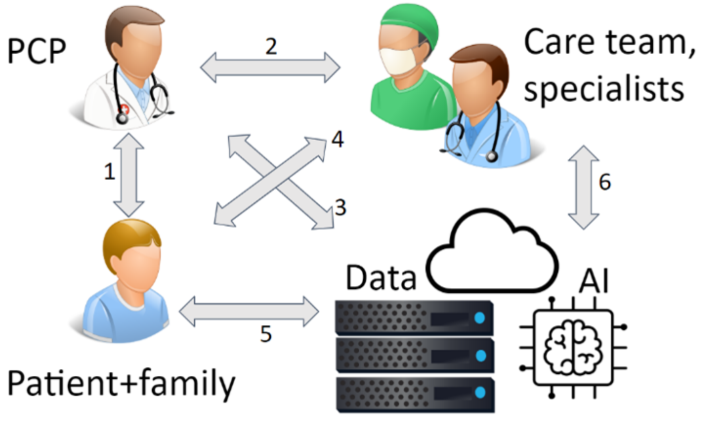

Fig. 5. Different types of communications in healthcare between four DAs, primary care physician (PCP), other members of the care team, such specialists (cardiologists, radiologist, oncologist, etc.)., Data and AI engines, and patient/patient family. There 6 types of communications between them indicated by numbered arrows. More than 2 DAs can also participate (e.g. patient, PCP, care team, etc) not drawn for clarity.

In particular, our goal is to investigate augmented reality computer interfaces for enhanced perception of multi-modality, multi-source medical data: (1) efficient AI-driven visualization of relevant patient-specific data in 3D virtual interactive environments (3DVIE); (2) exploring approaches to remote interactive communication in 3DVIE without the need for physical specialist-PCP or specialist-specialist co-presence (Figure 4.3). Similarly, we focus on novel approaches to facilitate consultations between specialists. Those approaches will not require physical presence at the same location of two or more parties.

In situations when medical institutions do not employ specialists (small medical centers), 3D virtual environments allow for remote virtual interaction between patients and specialists. Remote cameras will create virtual environments. The HI pillar will also investigate optimal decision-making when machine recommendations, expert human knowledge, and decision loss function have to be factored in during the decision-making process. We note here that combining heterogeneous information is a very complex problem because it must account for human-machine collaboration as well as patient preferences.

4.6 Extreme Scale Data and Computing

The unique combination of Modeling and Simulation with Health Data Science and Health Informatics poses new fundamental challenges that require basic computer science and engineering research. The computational and data processing needs of the Centre push the boundaries of current state-of-the-art infrastructures for AI, HPC, big data and cloud computing. This includes alignment of traditional HPC systems with big data analytics and ML/AI workloads, using CPU, GPU, many-core, hybrid, virtualized and containerized environments with the computational needs of systems required to deliver patient-specific in silico care on clinically-appropriate timescales. Moreover, health data science will benefit from novel approaches in distributed computing and security research, such as Federated Learning, Differential Privacy or Encrypted Computation, which can be applied to medical data in a secure and privacy-conscious manner. On the other hand, new developments in computer hardware, programming models, cloud computing and emerging services influence development, deployment and execution of computational and AI models at extreme scale. Close integration with health data science should match computational resources to the processing needs of simulation and machine learning - currently one of the most demanding and rapidly developing areas in computer science and supercomputing applications. All of these require constant evaluation of new technologies and platforms, experimenting with novel approaches, and prototyping new solutions, which can be further developed into production tools and services at Sano.

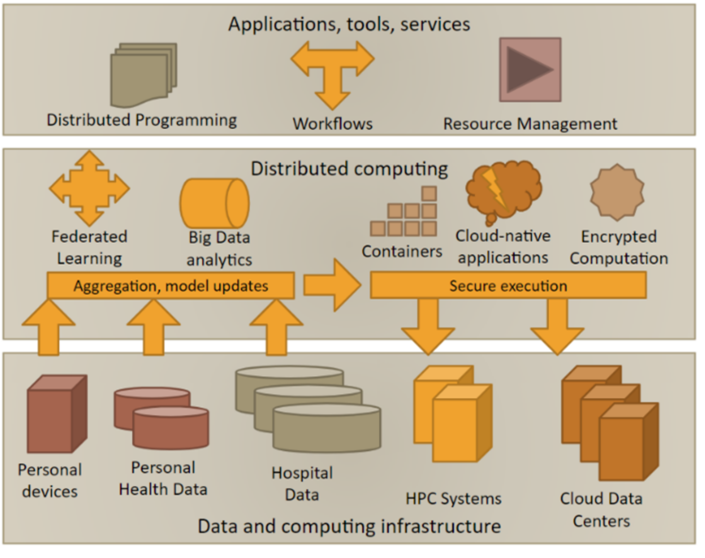

Fig. 6. The key layers of extreme-scale data and computing research in Sano. Data acquired from the multiple sources is securely processed on cloud and HPC infrastructure using recent developments in distributed computing.

Sano's core extreme-scale data and computing addresses the following research areas: (1) Convergence of HPC with AI/big data infrastructures and workloads, resource management on HPC and cloud platforms; (2) Distributed programming, workflows, containers, cloud-native applications, federated learning, encrypted and privacy-preserving computation (see Figure 4.6. Of particular interest is a scenario where medical institutions are not allowed to release primary data to the outside world. Sano therefore embraces an approach referred to as federated learning, initially postulated by Google, which is especially attractive in sensitive-data situations such as medical or financial applications. In short, in federated learning the data never leaves the hospital - only updates to machine learning model parameters are sent to the central aggregator. This approach is successfully being tested and applied in the academic and commercial world (e.g. Tensorflow). Development of systems operating in clinical, research, HPC, Big Data/AI environments require novel, transparent techniques, and the delivery of state-of-the-art in health data science and in-silico techniques will require exascale computing resources, meaning the Centre can also become a driver for developments within EU-level HPC and cloud initiatives, including PRACE and EuroHPC.

5 GOALS AND IMPACT

In summary, we expect to create a vibrant medical research institute and establish:

- a wide net of collaborations with top clinical and academic sites in Poland and abroad,
- substantial scientific output concerning the use of computing and AI in a wide range of medical domains in peer-reviewed publications, conference presentations/papers, and media presence,
- at least several of our innovations to form part of clinical workflows in real-world scenarios (clinical and usability trials) in collaboration with our clinical partners; continuing to support the deployed innovations in the long run to generate evidence of their efficacy,
- Sano's technology transfer mechanism and expertise in IP management, regulatory affairs, and industry relations,
- relations with venture capital and other sources of capital, with at least one startup based on Sano IP and at least partially operated by Sano graduates,
- partnerships with industry for joint development of projects initiated at Sano or in the industry.

Overall, we look to achieve excellence in research and build Sano's reputation as a leading center of computational precision medicine in Central Europe, and then on the global level. Finally and importantly, we hope to mentor and train graduate students and postdocs who will carry on the message of medical innovation and precision medicine in their further careers outside Sano.

The long-term goal of Sano is to substantially contribute to bringing computing and artificial intelligence to healthcare on a global scale by the year 2050 (H50 goal). We want to imbue medical practice with the paradigm of 4P medicine, enabling people to live longer, healthier, and more fulfilling lives.

At the *regional* level, Małopolska (Lesser Poland Region) is one of the most rapidly developing economic European regions as confirmed by its high status in rankings such as fDi European Cities and Regions of the Future 2020/21 (fDi Intelligence). In the 2014-2020 EU Financial Framework, Małopolska has nearly 3 billion Euro allocated to the development of infrastructure, support of entrepreneurship, new technologies and improvement in the quality of life. The Centre contributes to the Małopolska smart specialization, particularly with regard to Life Science, Information and Communication Technologies and Electrical Engineering and Machinery Industry, through impact in the life science sector, building a regional competitive advantage in research and development (R&D) and introducing and retaining new skills in the technological areas addressed by the IRAP program. The Centre aims to achieve significant lasting change in these specializations, by providing a critical mass of research activity, to speed up the implementation of regional S3 priorities.

At the *national* level, the Centre will introduce much needed coherence in the currently fragmented Polish computational medicine community by integrating common R&D projects with the strongest national partners. The Polish government supports AI topics, among others by engaging in the consultation process for the European white paper on artificial intelligence: European approach to excellence and trust, and supporting the directions charted by the European Commission under the Digital Europe Programme 2021-2027. Moreover, Sano's activities are in line with Poland's strategic policy mission in the field of artificial intelligence (Policy for the Development of Artificial Intelligence in Poland for 2019-2027). The policy will support science, research and AI development for the growth of innovation and productivity of the knowledge-based economy. It will also support citizens and businesses in the processes of transformation of the work model towards ethical AI adoption (adequately protecting human dignity and fair competition). It will positively influence economic growth by creating entrepreneurial activity of high revenue potential into the national and global clinical market. An important aspect of such activities is to counteract attrition of Polish researchers, clinicians, and skilled practitioners by offering attractive and scientifically challenging employment options, both directly through creating PhD and postdoc positions, and indirectly by instigating research projects with Polish research partners.

At the *international* level, the RA will generate significant impact through participation of the Centre's researchers in international research programmes and direct engagement with academia and industry from outside Poland. Promotion of the Centre within international consortia will provide opportunities for exploitation of research by both international companies and small to mid-size enterprises (SMEs) based in other countries, including licensing of individual outputs (e.g. software, algorithms etc.) and partnerships to deliver services to SMEs, healthcare industry, and healthcare providers (e.g. HPC, software engineering services, infrastructures to test AI products, etc.)

6 CONCLUSIONS

The original proposal for the Sano Centre was fundamentally based on the realization that in order to harness powerful computational techniques in the fight to tackle complex medical challenges, there is a need for a radically new cross-disciplinary type of institution to unite the necessary skills. Sano is the embodiment of this ambition. Sano's Research Agenda thus defines a concrete set of internal structures into which we will breathe the daily life of scientific development. Sano, now formally established as a foundation under Polish law, will generate impact through translation of excellence in computational science in the domains of healthcare data science, artificial intelligence techniques, high performance computing and modelling and simulation, and into applications guided by clinicians and targeted at the needs of patients and sustained through industrial engagement and collaboration. At the heart of the International Research Agendas Programme is an interdisciplinary design that we believe is the only workable approach to creating the momentum of collaboration needed to build integrated systems and bridge the widely-acknowledged gap in Technology Readiness Levels that elsewhere separates academic innovation from sustainable exploitation and product delivery. Whilst the Sano Centre benefits from mature relationships with more traditional academic partner institutes (AGH, USFD), clinical centres in Kraków and beyond, and industrial partners, the Centre's independence facilitates an agile approach to mapping the fundamental development of new and emerging technologies to address the most pressing and significant clinical needs. Crucially, the co-location of staff with a diverse range of backgrounds within a dedicated unit will achieve the broadest of sustainability goals – the continuing success of Sano itself – through direct training of the next generation of researchers and entrepreneurs with the gamut of expertise required to drive the evolution of computationally-enabled clinical decision support systems and precision medicine and, in the long run, to transform healthcare closer to the ideals of 4P medicine.

REFERENCES

[1] Patterson, C., 2018. World Alzheimer report 2018: the state of the art of dementia research: new frontiers. Alzheimer's Disease International (ADI): London, UK.

[2] Torre, L.A., Siegel, R.L., Ward, E.M. and Jemal, A., 2016. Global cancer incidence and mortality rates and trends—an update. Cancer Epidemiology and Prevention Biomarkers, 25(1), pp.16-27.

[3] Wilkins, E., Wilson, L., Wickramasinghe, K., Bhatnagar, P., Leal, J., Luengo-Fernandez, R., Burns, R., Rayner, M. and Townsend, N., 2017. European cardiovascular disease statistics 2017.

[4] Hood, L. and Friend, S.H., 2011. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nature reviews Clinical oncology, 8(3), pp.184-187.

[5] Hunter, P., Coveney, P.V., De Bono, B., Diaz, V., Fenner, J., Frangi, A.F., Harris, P., Hose, R., Kohl, P., Lawford, P. and McCormack, K., 2010. A vision and strategy for the virtual physiological human in 2010 and beyond. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 368(1920), pp.2595-2614.

[6] Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), pp.206-215. Sitek, A. and Wolfe, J.M., 2019. Assessing cancer risk from mammograms: deep learning is superior to conventional risk models. Radiology, 292(1), pp.67-68.

[7] Rudin, C. and Ustun, B., 2018. Optimized scoring systems: Toward trust in machine learning for healthcare and criminal justice. Interfaces, 48(5), pp.449-466.