WAY AHEAD
PROJECT RESULTS AND FUTURE PERSPECTIVES

Dear Readers,

welcome to the final issue of the MD-Paedigree newsletter, where we have taken the opportunity to present you some of the project accomplishments we are most proud of. **When the MD-Paedigree project had been initially conceived, the development of computational models and advanced analytical tools to support decision making in clinical care – the so called clinical decision support systems, or CDSSs – was still in its embryonic stage.** In fact, as computer-based predictive modelling has been extensively adopted in other branches of knowledge, medicine had accumulated a decades long delay in this regard. As a result, the routine work of physicians remained largely based on the single clinician's experience, all to the detriment of patients. Clinical decisions could not rely on quantitative assessment parameters, the vast record of clinical cases available on the broad scale or all the new scientific knowledge produced at a constantly increasing rate, impeding optimal diagnosis, risk assessment and therapy assignment, particularly for challenging or uncommon clinical cases.

In this context, the last decade has seen huge investments into advanced information technology and simulation systems to create realistic representations of human physiology incorporating specific patients' information, and able to provide insights into prognosis to support more effective and efficient patient management. Conspicuous investments from the European Commission fed key initiatives such as the Virtual Physiological Human, whose major milestone has been the Avicenna Coordination and Support Action for in silico Clinical Trials, and under which several grants within the 6th and 7th EU Framework Programmes, such as Health-e-Child and Sim-e-Child, have been funded. All these efforts have led to significant results, with the development of a variety of CDSSs in various disease areas.

MD-Paedigree grand challenge has been of integrating patients' data with the most advanced physio-pathological models and data analytics solutions in one single framework for paediatric care, to provide clinicians with critical clinical decision support directly at the point of care. As a starting point, the project has taken into account clinical predictive models developed in previous European projects (i.e. Health-e-Child and Sim-e-Child) and extended them to encompass new disease areas in paediatrics, namely paediatric cardiomyopathies, cardiovascular disease risk in obese children and adolescents, juvenile idiopathic arthritis and neurological and neuromuscular diseases.

While having reached high technological maturity, however, these models still lacked a comprehensive validation on large cohorts of patients. To bridge this gap, MD-Paedigree put a specific effort for achieving an ultimate assessment of accuracy, efficacy and prognostic value in the clinical setting. This goal has been pursued validating the resulting mechanistic models and statistical predictors against actual patient outcomes at participating clinical centres, for an overall of 638 clinical cases. These studies will be the object of the forthcoming MD-Paedigree Final Conference, where project results will be presented at the presence of clinical experts of the field from all over Europe. **Despite several criticalities, this validation effort has ultimately demonstrated statistically significant validity and applicability for the vast majority of tools in a wide range of clinical scenarios, playing a pivotal role for demonstrating the strategic value of CDSSs in clinical care.** This critical effort contributed to lead this research area to a latter phase, where investments in basic research are going to be replaced with focused efforts on applied research for further implementation of mature technologies.

INDEX

PROJECT NEWS

From Algorithms to Clinical Decisions: the MD-Paedigree Final Conference

How to advance paediatric care improving diagnosis and treatments outcomes through the adoption of model-based clinical decision support systems: this will be the key topic of the forthcoming **MD-Paedigree Final Conference "From Algorithms to Clinical Decisions"** during which, at the end of its four-year endeavour, MD-Paedigree will present its final results and future perspectives.

The conference, taking place on **May 22-23 at Ospedale Pediatrico Bambino Gesù in Rome, will see experts from all around Europe convening to discuss project results and their potential impact on clinical practice**, in the broader framework of the latest news, developments and impending innovations in the field of Virtual Physiological Human modelling, advanced big (healthcare) data analytics and artificial intelligence tools, and relevant clinical validation.

During the two-day conference, **all MD-Paedigree partners will contribute to the presentation of project results, showcasing the implemented tools and discussing relevant outcomes involving a panel of external experts in the different project disease areas**, from cardiovascular disease and cardiovascular risk in obese children, to inflammatory disease (such as juvenile idiopathic arthritis) and neurological and neuromuscular diseases, passing through the innovative studies on the putative role of microbiome in disease progression. Also, specific sessions will be devoted to discussing **the MD-Paedigree IT Infostructure, as well as the relevant data management, curation and analytics tools.**

The so called "validation challenge" will also represent one of the major focuses of the conference programme, going hand in hand with the final adoption of the new tools in the everyday clinical practice, meanwhile investigating ethical and security aspects relevant to safety of their use in children's care.

The conference, which will be launched by the President of Ospedale Pediatrico Bambino Gesù, **Mariella Enoc**, and chaired by the Project Coordinator, **Prof. Bruno Dallapiccola**, will also encompass the participation of national and international speakers and decision makers, such as **Walter Ricciardi** (President of the Italian Istituto Superiore di Sanità) and **Terje Peetso** (Head of Policy Sector at Unit eHealth, Wellbeing & Ageing/H3 DG Connect, European Commission).

Fostering big (health) data value: from the MD-Paedigree Infostructure to the MyHealthMyData project

From data to information
With its 150 exabytes of data produced worldwide each year, **healthcare is surely a bright example of the "Data Explosion Phenomenon", expected to grow even further in the forthcoming years thanks to decreasing costs in high-throughput medical tests** such as genome sequencing, high definition diagnostic imaging and new biomolecular disease markers, not to mention the huge amount of data which will be coming from mobile and wearable devices, as by 2020 40% of all IoT technologies are expected to be healthcare-related.

It's indeed undeniable that healthcare is currently experiencing a post-(data) scarcity condition. Still, as data is the new oil, what we actually need is petrol, i.e. useful information extracted from the newly-available data. For this reason, **specific focus should be given to appropriate tools for managing datasets and making them interoperable, while spotting inconsistencies and extracting value through appropriate analytics tools.** MD-Paedigree addressed this issue by implementing an integrated data infrastructure, which aims at making the data usable, sharable among stakeholders, and useful for clinical purposes.

Extracting clinical value from data: the MD-Paedigree Infostructure

The MD-Paedigree Infostructure (Figure 1) has **been built on previous EU-funded projects' architectures (namely Health-e-Child and Sim-e-Child),** and its federated structure enabled clinical centre to share only copies of the datasets, while keeping the original ones within the internal IT system's firewall. This solution allowed for a full data anonymization (achieved thanks to a tool developed by our project partner gnúbila), before sharing it with other clinical centres and researchers, all at advantage of the patients' privacy. Also, the cloud-based solution adopted allowed to leverage on distributed computational resources, without requiring specific on-site hardware for running advanced decision support algorithm in the clinical setting.

The Infostructure integrates the software solutions implemented throughout the project in one unified framework, enabling the clinician to access a full range of resources from a single access point: MD-Paedigree clinical partners can easily access the Infostructure (logging in with a username and password), browsing a repository of pseudonymised patients' datasets, and can directly run a set of analytics and curation tools. Among

Figure 1. Outline of the MD-Paedigree Infostructure.
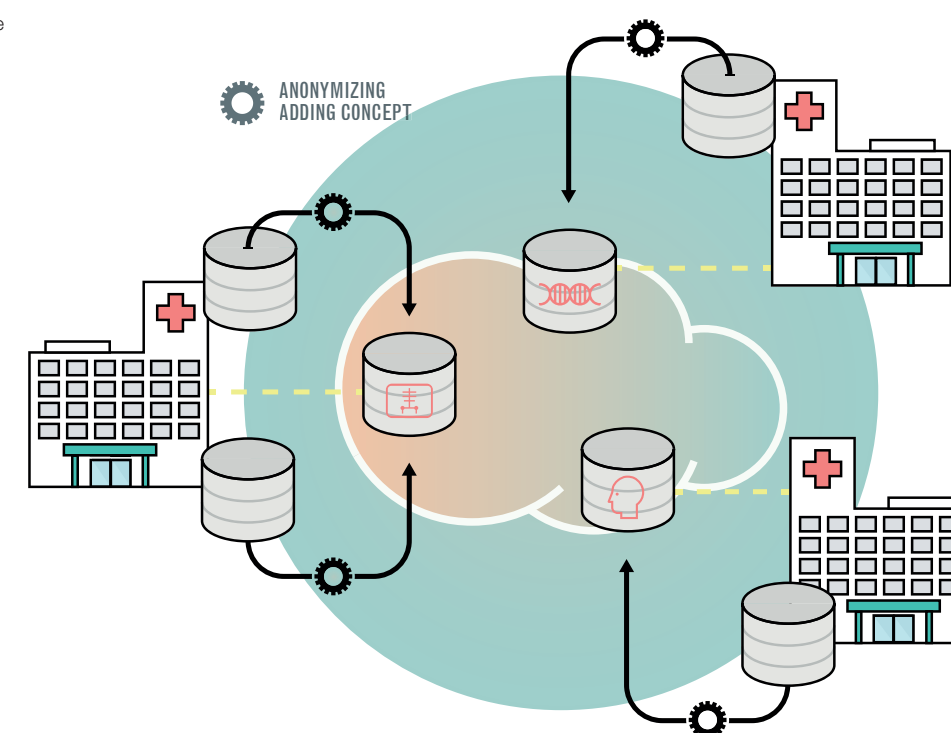
these, the data curation and validation tool (pages 27-28) allows the clinicians to review the datasets, spotting out errors or inconsistencies. Also, the case-based retrieval service (page 29) allows the clinicians to search for patients similar to the ones they are currently taking care of, on the basis of free-text queries, which will be then analysed on the basis of the patients' discharge summaries available in the repository. Other analytics tools, such as the Case Reasoner implemented by Siemens Healthineers, allow clinicians to find "patients-like-mine", comparing several parameters extracted from imaging data.

New trends and perspectives: the MyHealthMyData project

The Infostructure has been conceived at the time of the project submission to the European Commission (EC), in 2013. In the meantime, **new important trends in healthcare data management and usage** have emerged. In particular, **a new era of patients' engagement** and empowerment has been advocated by several stakeholders, including the EC, and **the new General Data Protection Regulation** has been released, with its acknowledgment of new patients' rights to data access and data portability. Ultimately, the recognition of **a new civil right** to health data ownership has been outlined.

To keep the pace with these emerging perspectives, the MD-Paedigree Infostructure has been re-used as the basis of a new EU-funded project, namely **MyHealthMyData (MHMD), which will extend the MD-Paedigree Infostructure, making it the basis of a new blockchain-based in-**frastructure to serve as a European health data platform for healthcare, research and business purposes (Figure 2)**, leveraging on new decentralised and distributed approaches towards individual patients' empowerment.

Following this perspective, the project aims at **securely storing patients' data**, using the MD-Paedigree Infostructure as a basis, at the same time granting **patients full control over their data,** allowing them to gather their health history into a single cloud-based personal data account. MHMD will also develop **solutions for managing consent** through an innovative dynamic consent interface.

Finally, taking into account the increasing concerns about data security and hospitals' IT systems vulnerabilities, the project will explore **the application of privacy-preservation and security technologies** (anonymization, pseudonymisation, homomorphic encryption, secure multi-party computation, etc.), **and their usage in combination with advanced analytics** to leverage the value of large datasets of biomedical de-identified data.

The ultimate goal of MHMD is to create **a EU-based information marketplace**, laying the foundation of **new mechanisms of trust and direct, value-based relationships between citizens, hospitals, research centres and businesses.** Thanks to MHMD, the MD-Paedigree Infostructure will be brought to the next level, entering into the future healthcare and contributing to it by laying the basis for innovation in patient-centric data management systems.
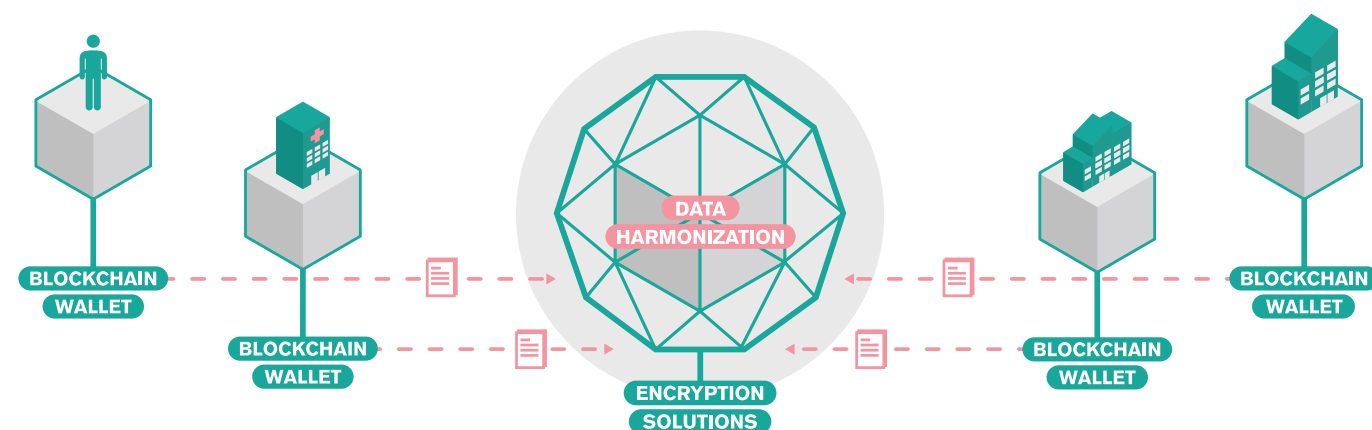

Figure 2. Outline of the MyHealthMyData (MHMD) data platform.

Health Data Interest Group and Research Data Alliance: a public arena for health data sharing issues

MD-Paedigree can be seen as a bright example of how the integrated use of patient health data from different sources, integrated with proper modelling and analytics tools, can lead to the development of personalised therapeutic approaches in all branches of clinical care. The exploitation of the full potential of these innovations is hindered, though, by increasing concerns about data privacy and security issues associated with the extensive employment of personal health data. What emerges is **the urgency of providing adequate solutions for protecting patients' rights without slowing the pace of data driven innovation.**

In this context, in 2013 **the Research Data Alliance (RDA) emerged as a new entity committed to support the open sharing of data and data-driven research, also through the provision of adequate social and technical infrastructures,** with the support of the European Commission, the National Science Foundation and the Australian Government. In 2016, RDA gave rise to the **Health Data Interest Group (HDIG),** a more specialised interest group focused on the **privacy and security issues underlying the use of health data,** co-chaired by **Yannis Ioannidis** (President and General Director of the ATHENA Research and Innovation Center and MD-Paedigree partner), **Edwin Morley-Fletcher** (Lynkeus President and MD-Paedigree Project Manager) and **Anthony Chang** (Chief Intelligence and Innovation Officer at Children's Hospital of Orange County).

"Data Infostructures for Open Science" has been the theme of the **9th RDA Plenary Assembly (Barcelona, April 5-7, 2017)**, organised by the Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS), where HDIG has been responsible for the **"Meaningful Health Data for Research and for Industry"** and **"Making Use of Blockchain in Dealing with Health Data"** open sessions. While the former was dedicated to discussing, with a panel of experts, basic but controversial concepts such as health data anonymisation and pseudonymisation, requirements for Open Science, and data provenance, comparability and traceability in biotechnology, the latter explored potentialities of the blockchain system application for the management of health data. This topic will be the object of a further dedicated session within the forthcoming **10th RDA Plenary Assembly (Montreal, September 19-21, 2017)** possibly leading to the establishment of a dedicated Working Group on Blockchain and Health.

PAEDIATRIC CARDIOMYOPATHIES

C

ardiomyopathies affect heart muscle fibers impairing normal pumping function, leading to arrhythmias and cardiac failure. They can result from gene mutations, but can also occur as consequence of other diseases, such as infections, high blood pressure and ischemic syndromes. According to clinical presentations, they are classified as dilated, hypertrophic, restrictive, arrhythmogenic right ventricular, and "miscellaneous" or unclassified cardiomyopathies. **While predominantly affecting adults, they are estimated to affect at least 100,000 children under the age of 18, worldwide, and at least one in every 100,000 in the USA, with the highest rate in children under 1 year of age.** Morbidity and mortality are high: nearly 40% of children with symptomatic disease undergo heart transplantation or die within 2 years since diagnosis. Due to small patient numbers and limited outcome data, it is still difficult to define the course of the disease, as well as to predict patients' response to treatments (open surgery, catheter interventions or drugs) and the optimal timing for intervention.

In MD-Paedigree, clinical predictive models developed in Health-e-Child and Sim-e-Child projects have been extended to cardiomyopathies to capture the patho-physiology of the heart, valves, arteries and peripheral circulation, integrating imaging data, pressure monitoring and clinical observations. Among these studies, Siemens Healthineers (SH) has developed a pipeline to construct **personalised anatomical and structural heart modelling by extracting anatomy of heart ventricles and valves from dynamic MRI and ultrasound data (next page).** Besides, researchers at Inria have implemented **3D cardiac models able to compute quantitative parameters describing individual heart properties and predicting physiological responses under challenging conditions (page 10).** Also, SH has realized **hemodynamical models able to simulate physical properties of the blood flow in the patient's heart and body (page 11).**

By merging scattered information in comprehensive virtualization and statistical predictions of heart function, these tools represent valid support tools for cardiologists, capable of bringing substantial improvement in patient care. To further assess the added value of clinical decision support systems (CDSSs), Empirica Communication & Technology Research conducted **a socio-economic assessment of the potential impact of the translation of bio-computational modelling into the clinical flow (page 12).** By taking into account the costs associated with the adoption of CDSSs and possible benefits in the treatment of specific diseases, preliminary results confirmed how the translation of bio-computational modelling will counterbalance a slight increase in initial costs with substantial savings in the long run due to significant improvements in patient management.

Personalised anatomical and structural heart modelling

Tobias Heimann
Siemens Healthineers
Technology Center

An accurate representation of the patient-specific anatomy is of paramount importance for the personalised heart models of the MD-Paedigree project. To process data from different input sources, Siemens Healthineers developed a pipeline consisting of three steps: segmentation of ventricles (i) and valves (ii), and fusion into a common model (iii).

Ventricles from dynamic MRI are extracted by a machine learning-based approach. The learned model contains mean anatomy, statistical shape information, and corresponding classifiers to adapt it to new image data. If required, results can be corrected interactively before they are passed on to other time steps by a temporal tracking approach.

This workflow is implemented in an easy-to-use prototype (Figure 1), which is able to process a complete case in less than 30 minutes.

Valve segmentation from ultrasound data is also based on a learned anatomical model, but relies heavily on interactive elements for semi-automatic initialization and model fitting. The semi-automatic modelling system supports editing of the valve models where needed, particularly to cope with complex valve geometry in pediatric cases. The models for aortic and mitral valves are composed of multiple surface landmark structures including aortic root, three aortic leaflets, commissures and hinge points as well as mitral anterior and posterior leaflets with trigone and commissure landmarks (Figure 2).

To fuse the extracted ventricles and valves into a common model, the valves are registered into MRI coordinates using correspondences from the left-ventricular base orifice and the valve models. The alignment is then rigidly refined by manual translation and rotation of the model to fit the MRI data as accurately as possible.

The team has been able to segment the heart chambers for all MRI data in the MD-Paedigree repository, and the resulting ventricle models have been validated by comparing the computed volumes with actual clinical data, yielding an excellent agreement. For the future, the team is aiming at further automating valve segmentation and fusion, as this would open up more opportunities to apply hemodynamic modelling in the clinical workflow, providing more information to the treating physician.
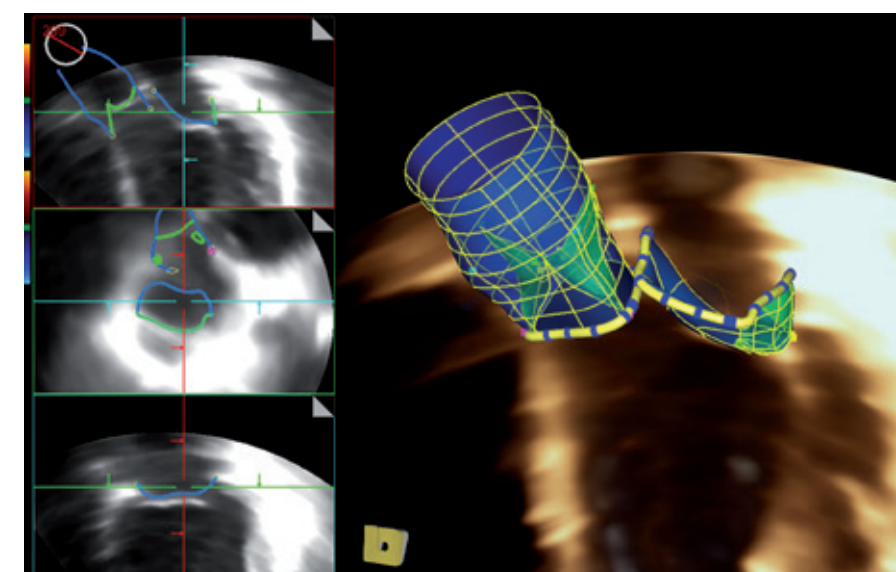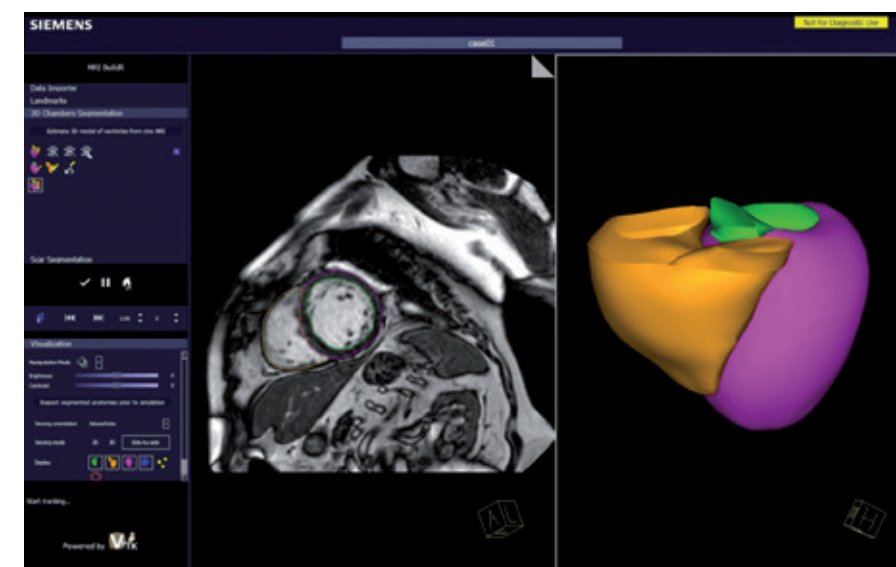
Figure 1. Prototype software to segment heart ventricles from MRI data.

Figure 2. Prototype software to segment heart valves from ultrasound data.

# Multifidelity 0D/3D modelling for fast personalisation of 3D cardiac models

Roch Mollero
Université Côte d'Azur, Inria

The main focus of electromechanical modelling in the area of paediatric cardiomyopathies has been **the estimation of clinical parameters able to capture intrinsic properties of the heart and predict its possible responses under challenging conditions, such as exercise or drug treatment, by means of personalised 3D cardiac models**. Devising such studies poses a series of challenges: firstly, they must rely on large homogeneous cardiac database, where same information should be available for all cases; secondly, the personalised parameter calculation implies a high computational complexity to perform the personalization (parameter estimation), particularly because of the time required to compute 3D model simulations, ranging from a few minutes up to several days for the most complex models.

To bypass the computational burden, researchers at Inria developed a "multi-fidelity personalisation method" that speeds up the personalization of a 3D model by order of magnitudes (Figure 1). As first step, a reduced "0D" version of the 3D model [1] is built by making simplifying assumptions, such as the spherical symmetry of the geometry and myocardial forces. The resulting 0D model is extremely fast (15 beats per seconds), in comparison to the 3D models (around 15 minutes per heartbeat). Then, a "multi-fidelity coupling" mechanism is employed to build a mapping which converts the parameters of the 3D model into parameters of the 0D model for which simulation outputs, such as the ejected blood volume or the arterial pressure, are the same in both models. This enables a very fast approximation of the 3D simulation outputs with the 0D model which can be applied to parameter estimation, resulting in a rapid and efficient personalization method for the 3D model [2]. Later on, prior probabilities have been assigned to the parameters in order to reduce the estimation variance. This resulted in a very efficient personalization: the estimation of 6 parameters from 3 clinical measurements over 121 cases was performed in around 2.5 days only.

Ultimately, the team investigated the relation between the estimated parameter values and the actual clinical condition, using follow up data of cardiomyopathy patients. In this way, they could demonstrate how **the evolution of parameters was naturally related with the improvement of the heart conditions under therapy.** The model parameters also revealed **a diagnosis power, proving able to improve accuracy of the classification of healthy versus cardiomyopathy cases** compared to the use of clinical parameters alone.
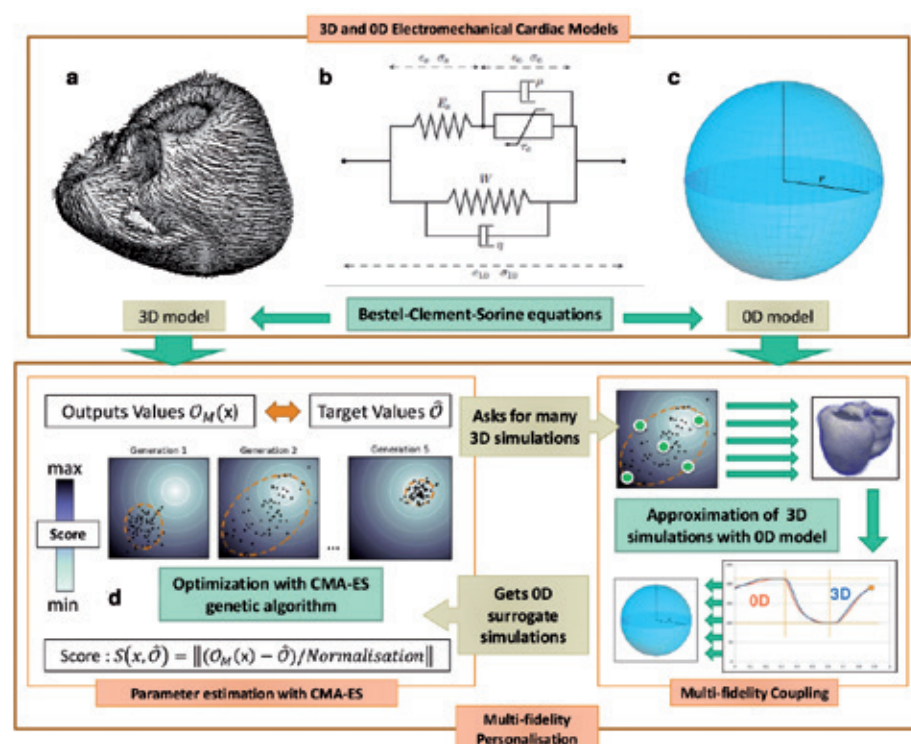
1. Marchesseau, S., Delingette, H., Sermesant, M. & Ayache, N. Fast parameter calibration of a cardiac electromechanical model from medical images based on the unscented transform. Biomech. Model. Mechanobiol. 1–17 (2013).

2. Mollero, R., Pennec, X., Delingette, H., Ayache, N. & Sermesant, M. A Multiscale Cardiac Model for Fast Personalisation and Exploitation. in International Conference on Medical Image Computing and Computer-Assisted Intervention 174–182 (Springer, 2016).

Figure 1. 3D and 0D cardiac models (a, b, c). The multi-fidelity personalization method performs parameter estimation in the 3D model with CMA-ES (d), based on 0D simulations obtained through a multi-fidelity coupling between the two models.

Hemodynamic modelling and simulation

Tobias Heimann
Siemens Healthineers Technology Center

Constant interaction of the blood flow with cardiac walls, valves, and adjacent arteries and veins can trigger and influence various cardiac pathologies. Therefore, **the hemodynamic models are an important component of the personalized cardiac model** developed by Siemens Healthineers. In the course of the project, the team implemented two different approaches for hemodynamic modelling: a whole-body circulation model and a 3D cardiac model.

The whole-body circulation model simulates systemic and pulmonary circulations using a reduced order flow and pressure model. While an electromechanical model with Windkessel boundary conditions describes the circulation downstream of the heart, an elastance model computes the atrial circulation upstream of the heart. Heart valve dynamics are physically based, with reduced order valves opening and closing due to the pressure gradients across them. This setup allows to simulate valvular disease conditions such as stenosis or insufficiency, cardiac conditions impeding the regular cardiac cycle (e.g. arrhythmias), or irregular hemodynamic events. As the whole-body model can be efficiently updated, it enables interactive prediction of the ejection fraction based on heart rate and blood pressures, and has been implemented in the "Cardio Predictor" prototype (Figure 1), aimed at personalized drug therapy optimization.

The 3D cardiac model simulates the 3D blood flow through the left ventricle, and relies on the Lattice Boltzmann Method (LBM), which describes physics of fluid flow at a mesoscopic scale by taking into account molecular interactions between flow particles. While the initial baseline model accounted for a limited one-way interaction of the fluid with the solid parts of the simulation (i.e. geometric changes by the cardiac walls and valves were directly imposed on the fluid without any feed-back), at later stages a comprehensive fluid-structure interaction (FSI) model has been developed to tie fluid and solid components and allow bi-directional feed-back and influences. The final 3D model enables a detailed insight into the local blood flow (e.g. vortices, pressures), currently possible only by means of time-consuming MRI examinations (Figure 2), and provides information for clinical decision making particularly in those cases where cost or time constraints impedes MRI acquisition.
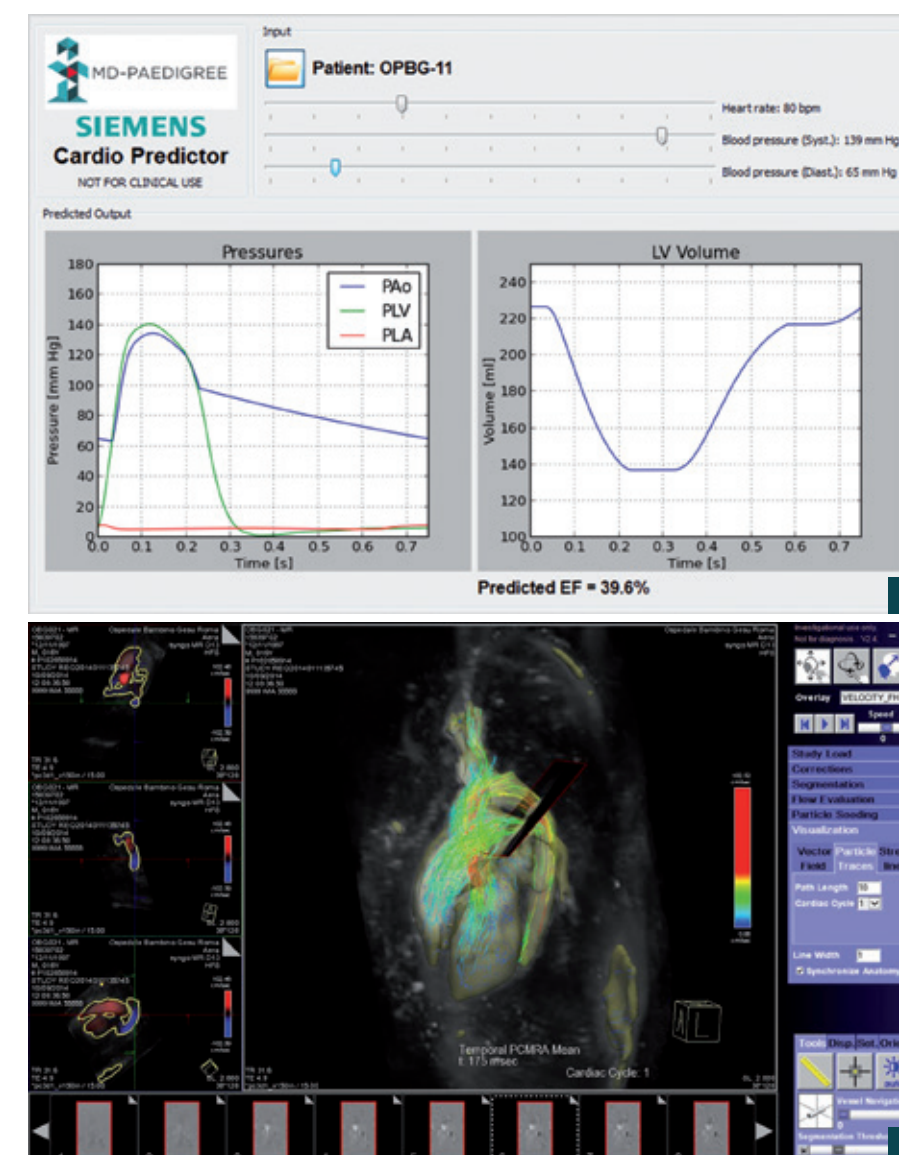
Figure 1. Cardio Predictor prototype.
Figure 2. Blood flow as captured by MRI.

Model-driven paediatric cardiomyopathy pathways: a clinical impact assessment

Rainer Thiel
Empirica Communication
& Technology Research

Karl A Stroetmann
Empirica Communication
& Technology Research

The clinical impact assessment performed in MD-Paedigree illustrated how the translation of bio-computational modelling into a future patient flow will supplement and improve the current management of specific diseases. The ultimate goal behind the socio-economic assessment perspective was indeed to facilitate the testing of clinical application scenarios for bio-computational models, and deliver support tools as well as empirical evidence for health system actors and decision makers, to favour proper exploitation planning and business modelling.

Clinical partners collected comprehensive data on paediatric cardiomyopathy and heart failure, in regard to treatment options, incidence and prevalence, as well as prognoses for different expected outcomes. Based on this knowledge, a detailed clinical pathway model was developed and validated in comparison with the current clinical workflow at Ospedale Pediatrico Bambino Gesù. The model combines three disease stages and various treatment options, together with probability estimates of a child moving from one stage to another. Also, in view of estimating the clinical cost savings to be possibly obtained with the new technologies, clinicians have also been collecting data which allows quantification of the care scenario impact: these include data on (i) costs of the technologies used in each diagnostic care management pathway, current and new; (ii) length of stay and number of consultation concerning each care pathway; (iii) probability of patients going through each pathway, arriving at certain endpoints, or (iv) moving from one pathway to another.

To reflect the complexity of the initial clinical decision making process, a three-stage Markov model was combined with a decision tree approach - a Markov decision process (Figure 1). A Markov Chain simulation tool was applied to compare estimates of transition probabilities and costs of current standard treatment options, for a cohort of children over a ten-year timeframe, together with expected improvements derived from using a clinical decision support tool [1]. The resulting data on expected clinical impact, dealing with changes in transition probabilities between severe and less severe health states and improved outcomes, are based on initial clinical validations and interviews with involved medical experts.

Early results, while indicating a slight increase in capital expenses for licenses and cost of ownership of clinical decision support systems (CDSSs), point to significant ROIs due to improvements in clinical care. In the short term, advanced stratification of patients allows for more accurate placement of patients in appropriate care path. Less aggressive treatment and diagnostic protocols in low risk cases, as indicated by the model, lead to reduced resource consumption, while more intense clinical care reserved for more server cases, is projected to increase children's life expectancy and quality of life.

1. Stroetmann, K. A. & Thiel, R. Model-Driven Paediatric Cardiomyopathy Pathways-A Clinical Impact Assessment. Stud. Health Technol. Inform. 234, 309 (2017).
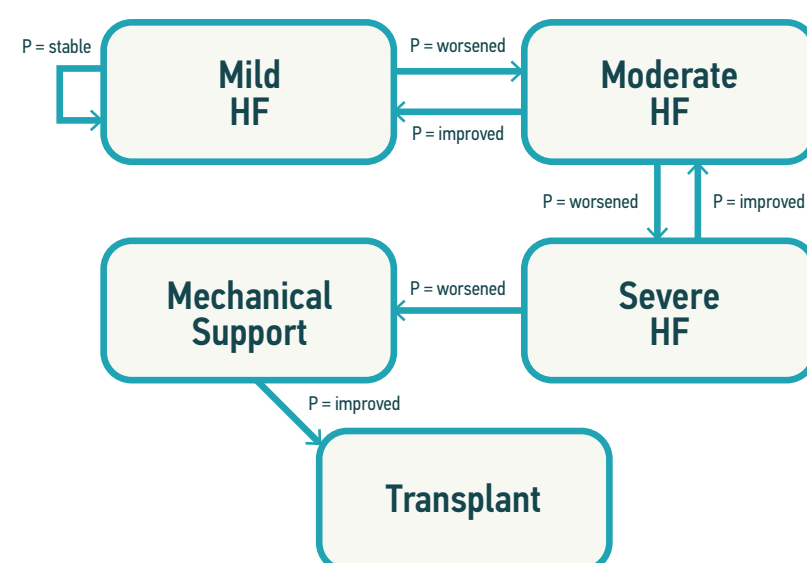
Figure 1. MD-Paedigree Markov Chain Model for cardiomyopathy.

CARDIOVASCULAR DISEASE RISK IN OBESE CHILDREN AND ADOLESCENTS

Cardiovascular disease (CVD) is a class of disorders involving the heart and blood vessels, and includes coronary heart diseases (heart attacks, such as angina and myocardial infarction), cerebrovascular disease (stroke), diseases of the aorta and arteries, such as hypertension and peripheral artery disease, rheumatic heart disease, congenital heart disease, cardiomyopathies, cardiac arrhythmias and heart failure.

With 17.5 million deaths worldwide estimated in 2012 by the World Health Organisation, corresponding to about 31% of global deaths, CVDs are among the leading causes of death and disability in the world. Beside some few non-modifiable factors (age, gender, ethnicity, family history), major risk factors for CVD include tobacco exposure, obesity, physical inactivity, high cholesterol, diabetes, hypertension, an unhealthy diet and harmful use of alcohol. Among these, obesity represent one of the primary risk factors, accounting for approximately one third of all coronary heart diseases and ischaemic strokes and almost 60% of hypertensive diseases in developed countries, according to the World Health Report 2002. As the levels of lipids, blood pressure, and obesity in the young have shown to be directly associated with the extent of early atherosclerosis of the aorta and coronary arteries, the significant increase in childhood and adolescent obesity over the last decade raises increasing concern.

In this context, MD-Paedigree was meant to integrate the variety of known biomarkers for CVD risk assessment into one common framework and analyse their interdependencies. The project aimed at developing predictive computational models to better understand the mechanism of CVD development, as well as allowing the patient-tailored simulation of therapeutic interventions for personalised outcome prediction. Among the many project's accomplishments, is worth mentioning the body fat quantification method developed by the Fraunhofer Institute for Computer Graphics Research IGD team, which enabled direct quantification of subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT) and organ specific fat based on MRI image data (next page). At the same time, the research team of Siemens Healthineers and University of Oxford have been able to implement a machine learning-based approach able to integrate SAT and data from a range of other sources (clinical questionnaires, lab results, microbiome and genetic data, clinical assessments and MRI images) for early identification of young patients at risk of cardiovascular disease (pages 15-16). These tools will likely constitute a strategic support to achieve a timely and accurate CVD risk assessment, supporting a more effective patient management and minimizing the probability of CVD onset.

Automated assessment of body fat distribution from MRI data

Klaus Drechsler
Fraunhofer Institute
for Computer Graphics
Research IGD

To date, the body mass index (BMI) is still the primary measure to assess the degree of obesity for clinical diagnostics. A major drawback of this measure is that it doesn't estimate the fat distribution but only the general adiposity of the subject, while clinical studies have shown that even **individuals with normal BMI might suffer from cardiovascular diseases (CVD): for instance, elevated Visceral Adipose Tissue (VAT) levels are often associated with CVD.** While manual delineation of adipose tissues in 3D data appears of major importance for early diagnosis and risk assessment, it still represents a highly time-consuming process, highlighting the need of automating this task to enable processing of large amounts of data.

The basis for quantification fat distribution is the acquisition of fat and water with separated MR images. Within MD-Paedigree, the Fraunhofer Institute for Computer Graphics Research IGD (FhG) team adopted two different MR acquisition sequences: T2* IDEAL (faster, but with lower resolution) and DIXON (slower, but with higher resolution). While implying a higher amount of technical work, taking into account two different protocols allowed to develop more robust and widely applicable algorithms.

FhG's approach to extract Subcutaneous Adipose Tissue (SAT), VAT and organ specific fat is a mixture of model-based segmentation and classical image processing. Statistical models are employed to capture the appearance and shape variance of the organ of interest from image data and outline an organ-specific segmentation mask [1], which enables to quantify the organ specific fat percentage. SAT consists of the fat deposits under the skin, while VAT is located inside the abdominal cavity. Simple segmentation methods don't allow to extract both tissue types separately, as the two may appear as connected due to partial volume effects causing nearby structures to merge into each other. FhG's approach uses a ray-based method that scans the preprocessed MR images from different directions to detect the inside of the body and separate VAT and SAT. Preprocessing includes correction of magnetic field inhomogeneities, removal of noisy background and arms and thresholding of adipose tissues.

So far, this approach has been used to quantify liver fat in approximately 200 datasets (Figures 1-3), and the quantification of SAT ad VAT are currently in progress. The whole automated detection and quantification process for liver fat, SAT and VAT takes around 5 minutes, and roughly 30 to 90 minutes per dataset are spent for manual refinements. As next step, the team has planned to investigate the trade-off between the efforts for manual refinement and fat quantification results.

1. Drechsler, K., Knaub, A., Laura, C. O. & Wesarg, S. Liver segmentation in contrast enhanced MR datasets using a probabilistic active shape and appearance model. in Computer-Based Medical Systems (CBMS), 2014 IEEE 27th International Symposium on 523–524 (IEEE, 2014).

Figures 1-3. Axial, Sagittal and coronal views of an example segmentation of the liver (red), SAT (green) and VAT (blue).
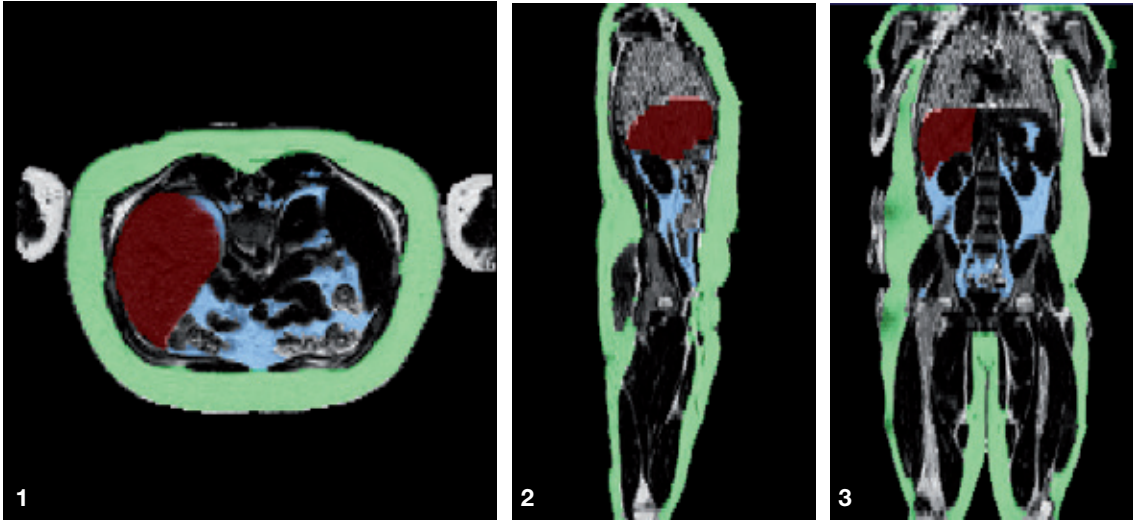
A comprehensive screening tool for the assessment of cardiometabolic health in obese children and adolescents

Olivier Pauly
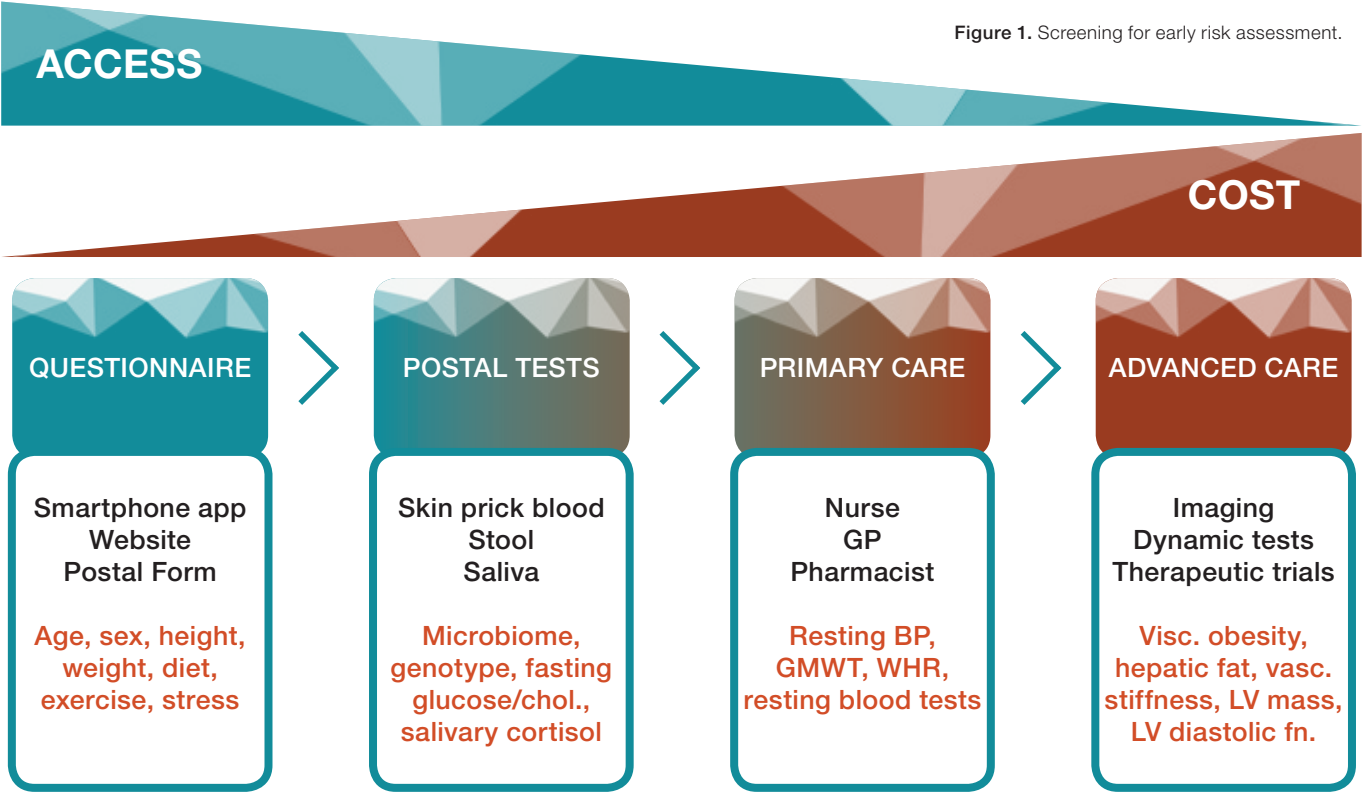Siemens Healthineers
Technology Center

Alexander Jones
University of Oxford

Obesity is a complex disorder and a known major risk factor for the development of cardiovascular disease in both children and adults. It has become clear that many of the detrimental physiological processes associated with adult disorders, such as heart disease and stroke, begin in childhood and are worsened by obesity. However, these processes remain effectively occult because of a reliance upon simplistic measures that do not sufficiently summarise cardiometabolic risk, particularly in the young. **For example, many children, even those with significant obesity, may have normal resting physiological parameters such as blood pressure or fasting glucose, despite evidence from more comprehensive assessment that their cardiometabolic health is compromised.** Such assessment is costly and unsuited for use in large population screening programs, but may have value if reserved for cases at the greatest risk. Identification of those cases is challenging: Siemens Healthineers (SH) together with the clinical partners from University College London, Ospedale Pediatrico Bambino Gesù and Charité Berlin utilised **a complex multi-variable dataset and advanced analytical tools with the overarching goal of identifying young people with early physiological derangements that sig-** nify high cardiovascular risk and that cannot be reliably detected by traditional approaches.

As depicted by Figure 1, **this novel screening approach relies on a sequential strategy where patients at risk are identified through the acquisition of increasingly complex phenotypic measures drawn from a range of different sources, such as questionnaires, stool and blood samples, clinical assessments and advanced medical imaging.** At each stage, predictive models based on deep learning facilitate a risk assessment that determines whether to recall a given patient for the next stage of more advanced examination. The most advanced but also most resource intensive stage depends on complex imaging techniques that quantify body fat distribution and comprehensively measure cardiovascular function. Novel image processing and computational models then enrich this assessment by providing cardiovascular parameters that cannot be measured directly. These ultimately allow to diagnose early cardiometabolic disorders, such as vascular stiffness, ventricular hypertrophy, ectopic fat deposition and insulin resistance. These intermediate outcomes are known to be associated with frank pathology in later life, but may not be exposed by simpler risk assessments.

Figure 1. Screening for early risk assessment.

ACCESS

COST

| QUESTIONNAIRE	POSTAL TESTS	PRIMARY CARE	ADVANCED CARE
Smartphone app Website Postal Form	Skin prick blood Stool Saliva	Nurse GP Pharmacist	Imaging Dynamic tests Therapeutic trials
Age, sex, height, weight, diet, exercise, stress	Microbiome, genotype, fasting glucose/chol., salivary cortisol	Resting BP, GMWT, WHR, resting blood tests	Visc. obesity, hepatic fat, vasc. stiffness, LV mass, LV diastolic fn.

JUVENILE IDIOPATHIC ARTHRITIS

To validate the proposed sequential approach, **a pilot multi-centric study was conducted to collect a cross-sectional dataset of approximately 160 patients, including questionnaire, anthropometrics, genetic, clinical as well as imaging data**. To extract intermediate outcomes characterizing cardiac function and fat distribution within the body, advanced image processing and machine learning tools have been developed yielding each patient's specific parameters. Based on computational modelling approaches developed at SH and INRIA for cardiomyopathy (page 10), **the team has personalized heart models to the MRI data from obese children to extract parameters characterizing different aspects of the cardiac function such as left ventricular mass index and vascular stiffness.** Using the tool for automated fat characterization developed by Fraunhofer Institute for Computer Graphics Research IGD (page 14), the team has extracted parameters characterizing liver, subcutaneous as well

as visceral fat distribution from those obese children.

To address the challenge of combining these complex multimodal data sources with varying scale and distribution, the team adopted **a multi-task deep learning approach: while it permits to perform inference for a considered intermediate outcome associated with elevated risk, it also encodes the original multi-modal patient data into a compact signature that has better properties in terms of scale, distribution and sparseness than the raw dataset.** The properties of the compact signature constitute a better substrate for further analysis with techniques, such as similarity searching, that are known to be compromised without these improvements in data quality. Once plugged into a similarity search engine, it enables the retrieval of similar patients providing important evidences for understanding system's reasoning, assessing prediction confidence, and finally supporting clinical decision.

**Figure 2.** Computational heart model adapted to the obese children.
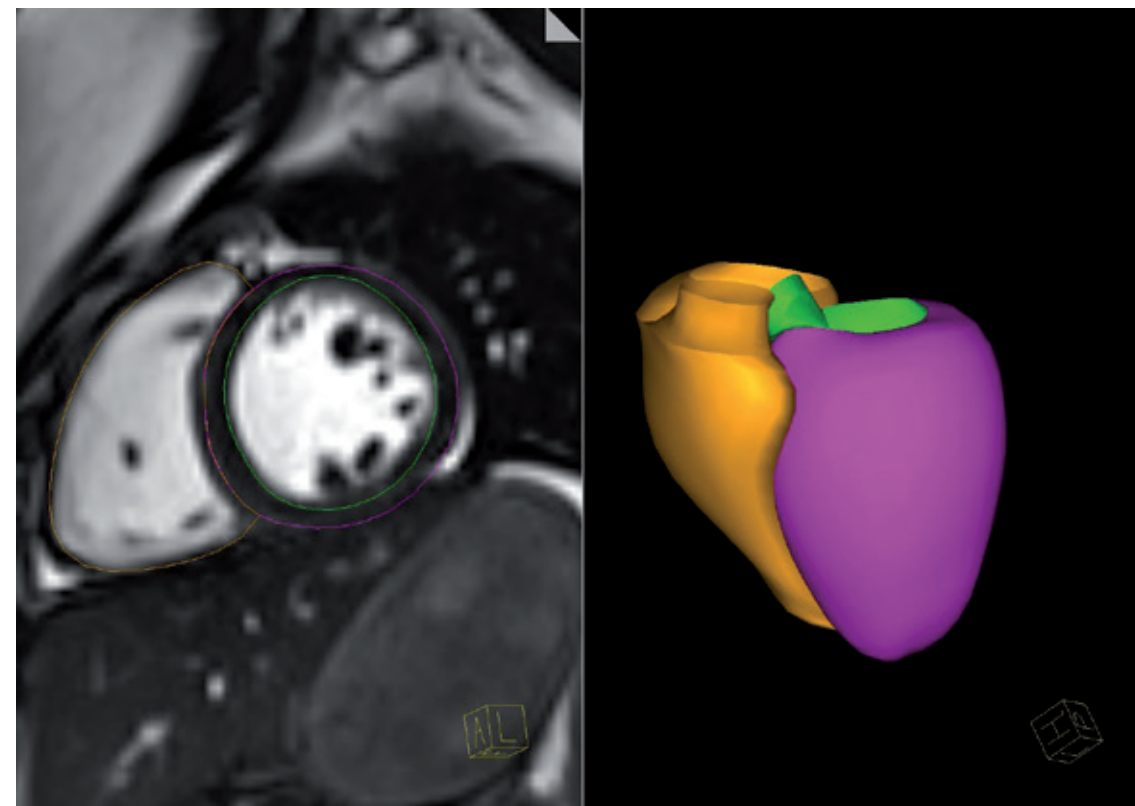
Juvenile idiopathic arthritis (JIA), also known as juvenile rheumatoid arthritis, is an inflammatory auto-immune disorder and represents the most common chronic rheumatologic disease in children**, but the term indicates a broad range of clinically heterogeneous pathologies sharing certain features, including an onset before 16 years of age, a duration longer than six weeks, and an unknown origin (to which the attribute "idiopathic" refers). **The disease affects approximately one in 1,000 children/year, constituting the leading cause of childhood disability from a musculoskeletal disorder.** JIA appears as a joint inflammation inducing swelling, warmth and reduced articulation movement: the affected joints are interested by proliferation of the synovial membrane (the connective tissue lining the joint capsule) and infiltration by inflammatory cells, ultimately leading to destructive lesions of joint structures. Its causes are still poorly understood, and they likely include both environmental and genetic and components: more than a dozen genetic markers have been identified so far, and hundreds are under investigation. Researchers have hypothesized that some type of trigger, such as a pathogen-induced inflammation, may activate the disease process in children with genetic predisposition. **Current classification of JIA disorders is based on clinical criteria (e.g. number and location of the affected joints at disease onset), but it remains largely unsatisfactory, as a wide heterogeneity is still present within the different subtypes.** The present ability of predicting disease course and outcomes is also still limited.

In this context, the goal of MD-Paedigree has been to take into account and extend ICT image-based tools for diagnosis and scoring of JIA developed within the previous FP6 Health-e-Child project, to attain **model-based integration of different data sources covering morphology, gait analysis and bio-genetic data for diagnosis, risk assessment and treatment planning purposes.** Among the most relevant accomplishments, it can be encountered the development of **patient-specific biomechanical models by the Fraunhofer Institute for Computer Graphics Research IGD team (next page)**, which relying on sequential MRI datasets are able to automatically reproduce bone segmentation. Their image outputs constitute **strategic tools for the analysis of joint loading alteration in the JIA disease course**, also allowing automatic detection and quantification of inflamed regions around affected joints. Besides, researchers at University of Sheffield implemented **a mechanistic musculo-skeletal modelling pipeline to combine MRI and gait analysis data to define patient musculo-skeletal geometry and movement patterns, quantify movement alterations and intra-articular forces (pages 19-20).** In this way, it was possible to identify and clinically validate 50 putative biomarkers as possible candidate prognostic predictors. Ultimately, researchers at Istituto Giannina Gaslini, Ospedale Pediatrico Bambino Gesù and University Medical Centre Utrecht investigated the potential role of microbiota in **the pathogenesis of JIA by comparing gut microbiota of JIA patients and healthy subjects (page 21)**. By highlighting an **actual heterogeneity in the microbiota composition between JIA patients in respect to healthy subjects**, this study opens interesting perspectives to better define JIA disease causes, and foresee innovative approaches for therapeutic intervention.
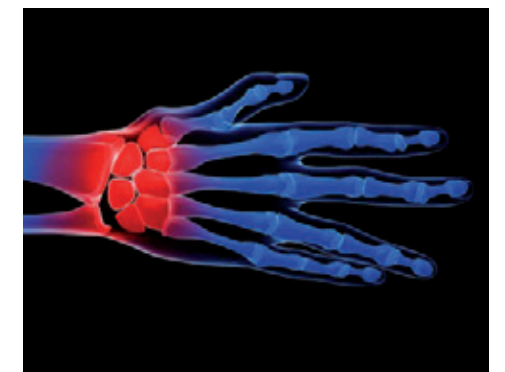
Patient-specific anatomical modelling based on image data and automatic biomarker extraction

Anqi Wang
Fraunhofer Institute for Computer Graphics Research IGD

Stefan Wesarg
Fraunhofer Institute for Computer Graphics Research IGD

Within the MD-Paedigree project, the Fraunhofer Institute for Computer Graphics Research IGD (FhG) developed **a bone segmentation and analysis method for MRI data sets acquired at different time points**. The generated patient-specific geometrical model serves as input for biomechanical models, which in turn are employed **to explore the influence of altered joint loading on the disease course of juvenile idiopathic arthritis (JIA).** Another aspect of FhG's work was to investigate the possibility of automatically detecting and quantifying inflamed regions around affected ankles.

The team implemented three different Coupled Shape Models - statistical model representations of anatomical variation in the examined patient population - to assess bone segmentation: **left and right foot models (Figure 1), containing bone structures and skin, and a third model containing leg, pelvic bones and skin**. For this purpose, FhG's framework for geometrical modelling of multiple coupled structures was employed. The MRI-specific challenges as well as incomplete training data – such as scans with missing anatomical structures – required an extension of the framework, finally leading to the creation of an automated

segmentation tool (Figure 2). Throughout the project, the team has processed more than 100 datasets, and clinical partners have confirmed quality and accuracy of the segmentations. While manual segmentation of a single dataset takes around 50 hours, the automatic segmentation is performed within minutes: thus, time savings are substantial.

In addition, these segmentations are the basis for an automatic detection of joints and a subsequent inflammation region examination (Figure 3). For the latter, pre- and post-contrast medium injection SPIR sequences are used, where regions with high intensity values represent potentially inflamed synovial tissue. Other similarly bright structures (e.g. vessels and tendons) or imaging anomalies are filtered out and, as inflammation usually occurs in and around joints, the remaining candidates are checked for size and proximity to one of the previously detected joints (Figure 4). Ultimately, an output file is generated containing the number of all detected inflammations, the affected joints, the total and single volume of each region. Overall, the detection and quantification process takes around 20 minutes for each dataset. So far, the team has analyzed 33 patient datasets, and the results are still under evaluation.
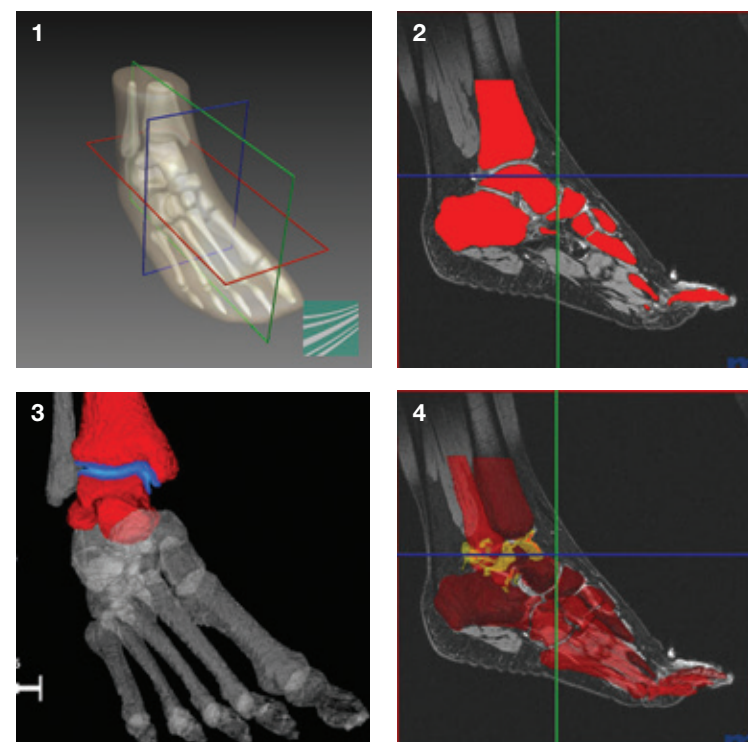
Figure 1. Model of the right foot.
Figure 2. Bone segmentation.
Figure 3. Joint example.
Figure 4. Detected inflammation (yellow).

Biomechanics as a prognostic tool for the investigation of JIA

Marco Viceconti
The University of Sheffield

Claudia Mazzà
The University of Sheffield

Luca Modenese
The University of Sheffield

Erica Montefiori
The University of Sheffield

Imagine the following clinical scenario: a child that presents at the paediatric rheumatologist with pain and swelling in the knee and ankle joint and difficulty in walking. Treatment with non-steroidal anti-inflammatory drugs had no effect. Based on general physical examination, articular examination, blood tests, plain X-ray of the joints, musculoskeletal ultrasound, a diagnosis of oligoarticular juvenile idiopathic arthritis (JIA) has been made by a rheumatologist.

The following questions then arise:
1. What is the prognosis of this child? Is the probability of achieving and maintaining disease remission high or low?
Nowadays, no instruments exist to predict the prognosis of a child with JIA. We used a multi-dimensional approach to provide an answer this question.
2. Are changes in the biomechanical properties of the affected joint, due to the active disease or to the initial cartilage damage, responsible for a structural damage progression?
Altered intra-articular forces might be responsible for the progression of the joint damage and ultimately lead to permanent alterations in joint structures and physical disability. Having an estimate of these forces will answer the question.
3. Does the monolateral joint involvement put the child at risk to develop arthritis also at the contralateral side?
An altered gait pattern and joint loading forces at the contra-lateral side may be risk factors for the development of arthritis in the knee and ankle at that side.

With these questions in mind, the clinical partners at Ospedale Pediatrico Bambino Gesù and Istituto Giannina Gaslini collected **a large set of quantitative anatomical, functional, and biomechanical biomarkers, obtained through direct measurement, data processing, and mechanistic modelling.** The biomechanical biomarkers have been obtained by the University of Sheffield from personalised musculo-skeletal models, which accurately quantify movement alterations and estimate the intra-articular forces. These multidimensional biomarkers are being analysed to evaluate their clinical accuracy, i.e. their ability to answer each of the above clinically relevant question.

The designed workflow [1] combines MRI and gait

analysis datasets (Figure 1), which have been collected synchronously at six-monthly intervals during one year follow-up. Researchers **built highly patient-specific anatomical and biomechanical models accounting for both patient's musculo-skeletal geometry as well as their movement patterns.** These models provide as output an estimate of the joint contact forces, not directly measurable with non-invasive procedures but needed to tackle our clinical questions. They aim to increase the understanding of the disease mechanisms, as well as elucidating cost-effective and clinically simple biomarkers that can be used to predict disease progression.

We collected data from 23 children enrolled by two paediatric hospitals, using innovative and particularly challenging protocols. During the MRI exam, the patient must remain perfectly still while in a noisy and claustrophobic space; the more detailed the exam, the longer its duration. Like in every paediatric study, we had to find a trade-off between the highest possible quality of the images, and the need to keep the children inside the scanner as little as possible. Similarly, the gait analysis protocol was defined to enable a detailed investigation of the foot while still allowing comparison between data from different hospitals and registration with the MRI, still considering the age of the patients.

Despite their complexity and length, our juvenile patients tolerated the protocol well and we had a very good retaining percentage, with data missing at the various time-points for only 13 out of the expected 138 datasets. However, in about 30% of the cases part of the data were either missing for technical problems or deemed of not good enough quality for the post-processing. Nonetheless, this still left us with **the largest modelling database of this kind ever collected for children with JIA, with a total of 124 usable datasets.**

During the data collection, we have developed all the needed image processing and modelling tools, which have been continuously refined and improved. We have also tuned our techniques for the complex task of dealing with partially incomplete data. All the tools have been thoroughly validated in terms of accuracy, reproducibility and sensitivity. Last but not least, we have focused on reusability: our experi-

mental and modelling pipeline can be reapplied to the analysis of any MSK pathology where the estimate of muscle and internal joint forces can be of interest.

We have now built all the personalised models and run 80% of the simulations. **A preliminary statistical analysis of the results has confirmed the excellent discrimination power of this extraordinary data collection and offered 50 anatomo-** functional biomarkers, out of the 235 we tested, as possible candidate prognostic predictors. While it will take months to complete the in-depth analyses, the use of subject-specific modelling may finally shed a light on this complex paediatric syndrome.

1. Prinold, J. A. I. et al. A patient-specific foot model for the estimate of ankle joint forces in patients with juvenile idiopathic arthritis. Ann. Biomed. Eng. 44, 247–257 (2016).
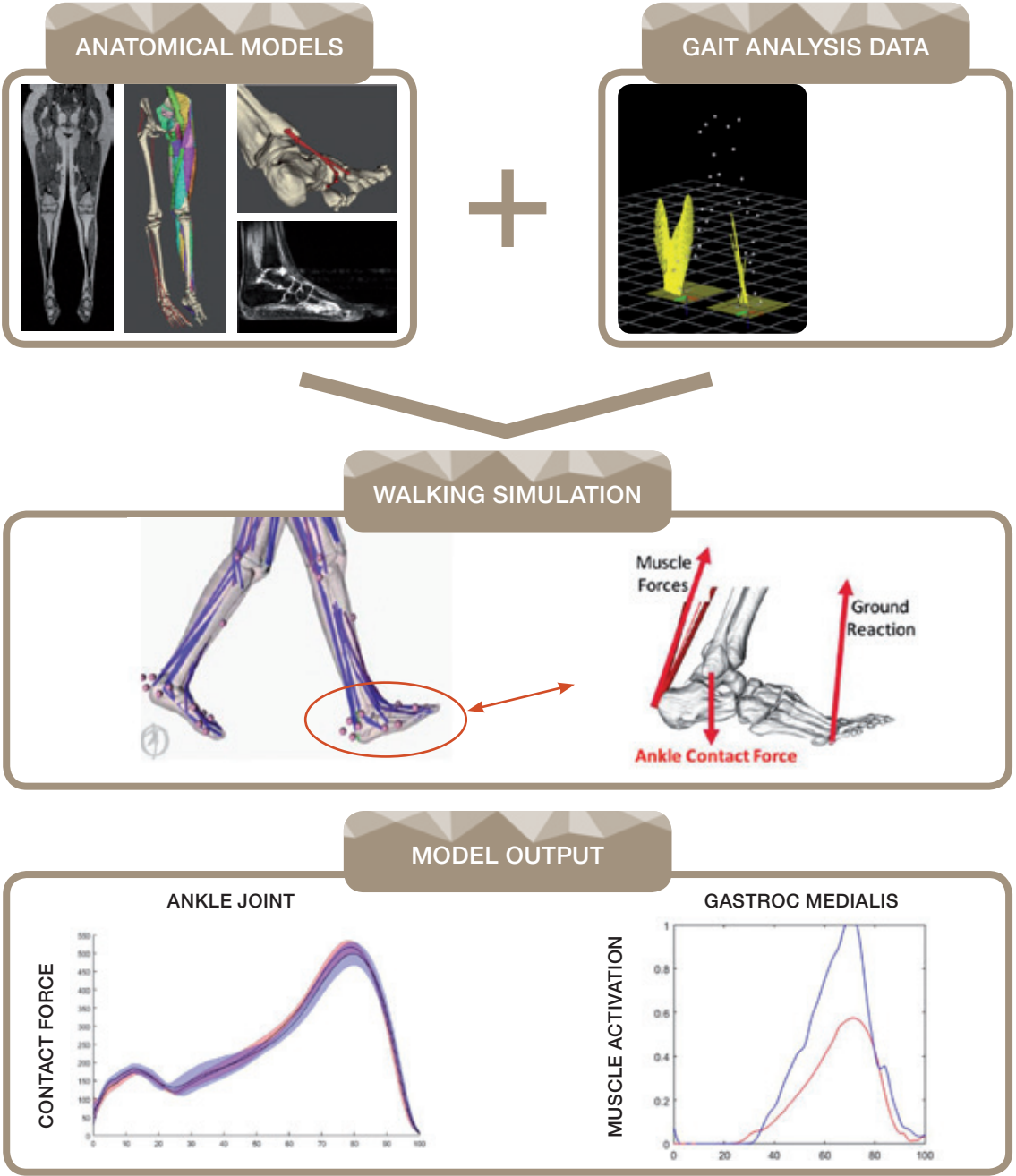
ANATOMICAL MODELS

GAIT ANALYSIS DATA

+

WALKING SIMULATION

Muscle Forces

Ground Reaction

Ankle Contact Force

MODEL OUTPUT

ANKLE JOINT

CONTACT FORCE

GASTROC MEDIALIS

MUSCLE ACTIVATION

Figure 1. The image-based musculoskeletal modelling pipeline.

Gut microbiota and JIA: a debated and intriguing connection

Pieter Van Dijkhuizen
University Medical Centre Utrecht,
Istituto Giannina Gaslini

The human body is colonized by billions of microorganisms, which are mainly present on the skin and in the gut, the so-called microbiota. Most of these are harmless to the host, and some are even beneficial. In recent studies, the composition of the gut microbiota was different in patients affected by autoimmune diseases with respect to healthy subjects: these discoveries have gained increasing attention in the scientific community, leading to the hypothesis **that microbiota could be actively involved in the pathogenesis of autoimmune diseases.**

To shed light on this issue, **researchers at Istituto Giannina Gaslini, Ospedale Pediatrico Bambino Gesù and University Medical Centre Utrecht investigated the gut microbiota composition of 78 Italian and 21 Dutch children with juvenile idiopathic arthritis (JIA)** at the onset of disease (i.e. prior to the start of treatment) as well as during follow up, both in cases where therapy induced inactive disease, and in cases where the disease remained active despite therapeutic interventions. The patients were compared with 107 healthy children of analogous age, gender and geographical origin. The gut microbiota composition was determined by collecting a faecal sample and analysing the relevant sequence (16S) of the bacterial ribosomal RNA.

The results showed clear differences in the gut microbiota composition between patients and healthy controls, confirming earlier findings in other autoimmune diseases (Figure 1). Overall, the microbiota composition appeared less diverse among JIA patients rather than among healthy controls. The analysis also showed differences between Italian and Dutch patients and between different age groups. No differences were found between the various disease activity states, suggesting that the microbiota profile is related to the disease and not to its activity. Indeed, at baseline, no association was found between the microbiota profiles and various disease activity parameters.

As the differences in the composition of gut microbiota were associated with many factors, at the moment it appears still impossible to ascribe a pathogenic role to microbiota. Nonetheless, our results provide **support to the hypothesis that an altered gut microbiota is actually present in JIA patients**: this finding opens up new opportunities of exciting research into their role in human health and disease, potentially leading towards a better understanding of autoimmunity and the development of novel therapeutic strategies.
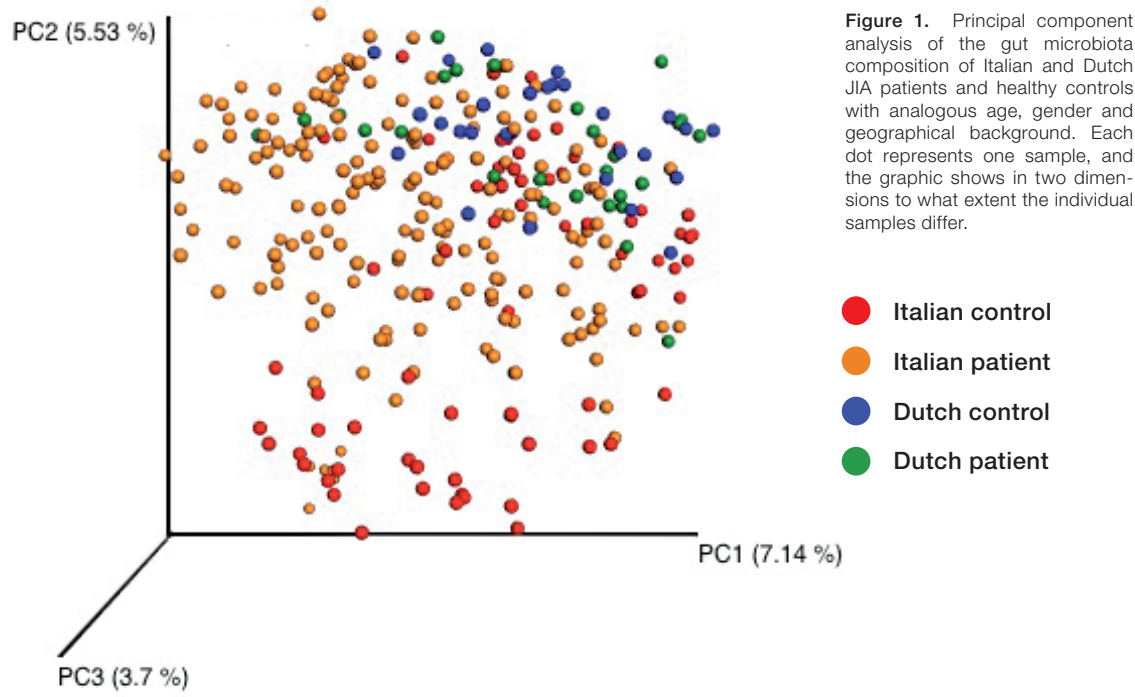
Figure 1. Principal component analysis of the gut microbiota composition of Italian and Dutch JIA patients and healthy controls with analogous age, gender and geographical background. Each dot represents one sample, and the graphic shows in two dimensions to what extent the individual samples differ.

PC2 (5.53 %)

PC1 (7.14 %)

PC3 (3.7 %)

● Italian control
● Italian patient
● Dutch control
● Dutch patient

NEUROLOGICAL AND NEUROMUSCULAR DISEASES

J

Neurological and neuromuscular diseases (NND), within the MD-Paedigree project, indicate a range of disorders affecting voluntary muscular function, particularly body movement and muscle coordination, due to degeneration of nervous fibers or muscles, or secondary to brain injury or malformations in the developing brain. Among these, **Cerebral Palsy (CP) is common, with an incidence ranging between 2 to 3.6 per 1,000 live births.** CP is itself a wide category encompassing several **non-progressive, motor impairment syndromes affecting body movement and muscle coordination**, caused by lesions in the sensory-motor cortex and corticospinal tract disrupting the brain's ability to control movement and maintain posture and balance. The second main target disorder is **Charcot Marie Tooth disease (CMT)**, so called after the three doctors who first identified it. **CMT is among the most common inherited nerve disorders, affecting around 2.6 million people (1 in 2,500) worldwide.** By damaging the peripheral nerves, it impedes proper nervous signalling, inducing **progressive weakening of feet, legs, and hands over time.** The third disorder, **Duchenne Muscular Dystrophy (DMD), is a genetic disease characterized by muscle weakness and progressive degeneration.** Like other dystrophies, DMD is due to a defective mutation in the gene encoding for dystrophin, a major structural component of the muscle cell membrane. **DMD represents the most common and severe form of muscular dystrophy, with an incidence of about 1 in 3,600 juveniles, with about 20,000 new cases each year worldwide.**

For these diseases, efforts have been made to develop strategies to improve walking function, particularly **clinical gait analysis (CGA)**, used to evaluate the joint and muscle function during gait performance. As walking involves a complex interaction of sensorial inputs, computer-aided models quantitatively assessing muscle function and joint loads are crucial to **delve into the primary causes of patients' pathological gait pattern and achieve accurate disease severity assessment, outcome prediction and treatment planning.** In this view, Siemens Healthineers and The University of Sheffield implemented a pipeline to extract **anatomical structures from MRI and generate personalized models for functional gait simulation (next page).** In parallel, Motek and Technical University of Delft adapted the **Motek's musculoskeletal Human Body Model (HBM), personalizing it to the patient's morphology with MRI (page 24),** to support CGA and, in prospect, motion patterns prediction applied to "what-if" scenarios. The Motek's HBM has also been employed **for real-time assessment of individual response to gait perturbations in children with CP (page 25),** evaluating patients' capacity to adapt to specific challenges. All these systems showed high degree of accuracy in the prediction of anatomies and motions, and are poised to be applied in clinical care to aid physicians' decisions and patient care.

Personalized and disease-specific anatomical models from MRI for gait functional simulation in children with neuromuscular disease

Maria Jimena Costa
Siemens Healthineers Technology Center

Jose M Pozo
The University of Sheffield

Accurate modeling of the legs and their function could revolutionise medical treatments, rehabilitation and surgical procedures aimed at regenerating or halting nerve and muscle degeneration. The anatomical modelling research groups of Siemens Healthineers (SH) and The University of Sheffield (TUoS) focused on pediatric patients from three neuromuscular disease groups, namely Cerebral Palsy, Charcot-Marie Tooth and Duchene Muscular Dystrophy.

The pipeline allows generating, from an MRI image of the patient lower limbs, a complete personalized anatomical model for fuctional gait simulation in children comprising 73 lower limb muscles and bones, 185 ligaments and muscle-tendon lines of action, and 9 joint centres and axis. The method was adapted to each disease group, and dramatically reduces the time required for defining these 267 structures, thus becoming a more time-efficient and accurate alternative to rescaling generic models.

As initial step, SH implemented a novel method to extract subject-specific anatomical leg structures from MRI, including muscles, bones and skin. The method underwent special adaptations for each disease group to accommodate disease-specific deviations (Figure 1). Initially, one atlas was built for each disease group including manually annotated individual meshes for every muscle, bone and skin (a). Then, for each patient MRI, the structure is personalized by adapting the corresponding reference by non-rigid deformation with multi-affine initialization for the articulated segments (b-d).

As second step, TUoS developed a new method which complete the anatomical model with the muscle, ligament and tendon attachment points and lines of action, and joint centres and axes, required for the fuctional simulations but non-visible in MRI. For this purpose, a template containing all this information was generated by integrating geometries from the reference MRI segmentation and the publically available TLEM2.0 model. A mesh morphing technique [1] was adapted to propagate the additional anatomical structures from the template to each case (Figure 2). The completed model allows extracting all the patient-specific parameters required for musculoskeletal modeling, e.g. muscle volumes, ligaments and muscle-tendon lines of action, and joint centres and axis [1].

1. Castro-Mateos, I., Pozo, J. M., Lazary, A. & Frangi, A. F. Automatic construction of patient-specific finite-element mesh of the spine from IVDs and vertebra segmentations. in SPIE Medical Imaging 97881U–97881U (International Society for Optics and Photonics, 2016).
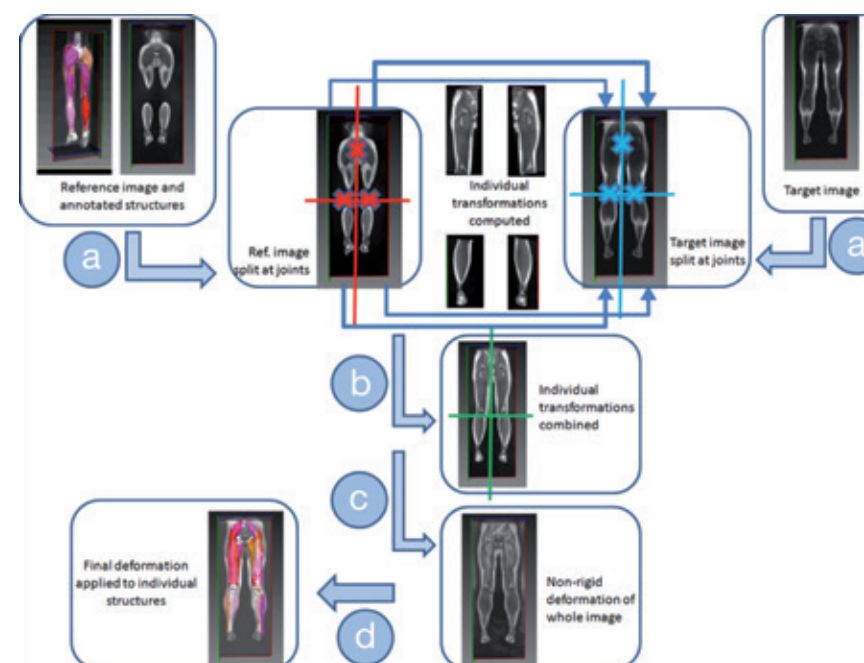
Figure 2. One example for each disease group of the final patient-specific complete anatomical model.

Figure 1. Overview of the bone, muscle and skin extraction method.

A modelling pipeline for personalised musculoskeletal models

Frans Steenbrink
Motek Medical

Johannes Gijsbers
Motek Medical

Thomas Geijtenbeek
Technical University of Delft

In paediatric patients with neurological and neuromuscular diseases (NND), musculoskeletal modelling is a promising tool to support clinical decision-making, and plan or evaluate interventions. However, currently available models are developed by obtaining morphological data from cadavers of adults, which do not match paediatric patients. Within the MD-Paedigree project, Motek and Technical University of Delft (TU Delft) aimed to **adapt Motek's existing real-time musculoskeletal model, the Human Body Model (HBM) [1], personalizing it to the single patient's morphology by the use of MRI data.**

Clinical partners, Siemens Healthineers (SH), The University of Sheffield (TUoS), TU Delft and Motek collaborated to create **a modelling pipeline (Figure 1) from MRI to a personalized HBM model to be used to evaluate a patient's walking pattern by means of clinical gait analysis (Figure 2).** Clinical partners provided MRI, gait analysis and clinical data of patients with NND. SH segmented the MRI images and TUoS extracted modelling parameters to a common file format. On this base, TU Delft created personalized OpenSim models which Motek used to generate personalised real-time HBM models. Gait analysis results using the personalized models were then uploaded to the MD-Paedigree infostructure probabilistic modelling for clinical validation.

The results are encouraging: the pipeline works well for all the available MD-Paedigree datasets with MRI, and clinical validation revealed a good fit between patients' physiology and model predictions. **Comparison of HBM results to other models support the validity of HBM, indicating its potential for use in clinical decision support systems and in real-time feedback applications.**

The inverse simulations described above have provided useful insights into human motion, but rely on recorded data and cannot predict new behaviour, which limits their clinical applications. To overcome this restriction, the team has recently developed **predictive forward simulations able to compute the motion and external reaction forces that optimally perform a given task [2].** Predicted motion patterns can take into account both musculoskeletal and neurological constraints, including neural delay and pathologies such as muscle weakness or spasticity (Figure 3). This would allow research and clinicians to evaluate "what-if" scenarios, which could for example predict the effectiveness of surgical interventions.

1. Van den Bogert, A. J., Geijtenbeek, T., Even-Zohar, O., Steenbrink, F. & Hardin, E. C. A real-time system for biomechanical analysis of human movement and muscle function. Med. Biol. Eng. Comput. 51, 1069–1077 (2013).
2. Geijtenbeek, T., van de Panne, M. & van der Stappen, A. F. Flexible muscle-based locomotion for bipedal creatures. ACM Trans. Graph. 32, 206 (2013).
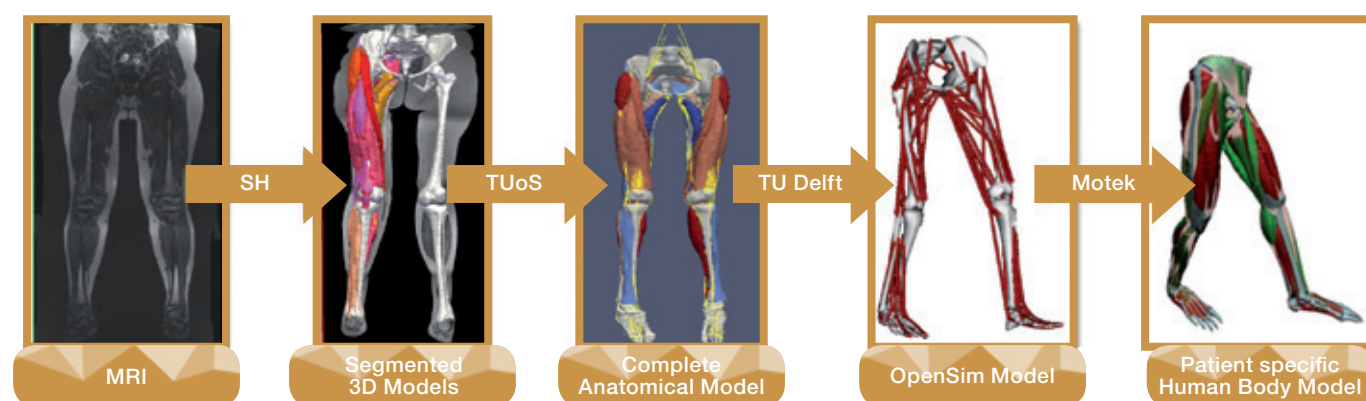

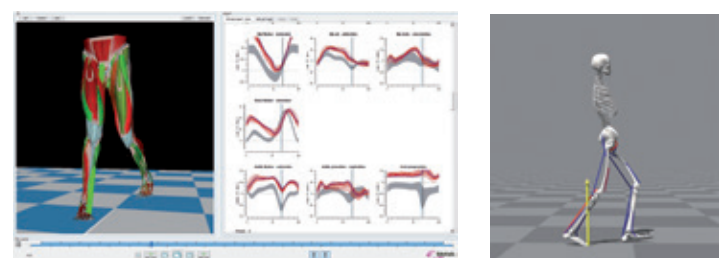**Figure 1.** Modelling pipeline to create the patient-specific musculoskeletal model from MRI.

Figure 2 (left). Gait analysis with personalised musculoskeletal model in the Gait Offline Analysis Tool illustrating the differences in joint angles between a normative data-set and this specific patient case.

Figure 3 (right). Predictive forward simulation model developed by Thomas Geijtenbeek (TU Delft).

Real-time feedback to actively perturb gait in cerebral palsy

Adam Booth
Motek Medical

Frans Steenbrink
Motek Medical

Jaap Harlaar
VU University Medical Centre

Marjolein van der Krogt
VU University Medical Centre

Annemieke Buizer
VU University Medical Centre

In the MD-Paedigree project, Motek and VU University Medical Centre (VUmc) have collaborated on the development of **a model for gathering real-time feedback on gait parameters to support the assessment of clinical gait problems in children with cerebral palsy,** the most common paediatric disorder within the neurological and neuromuscular disease area.

As children with cerebral palsy often display walking impairments, understanding the exact cause of the problem is crucial for determining optimal treatment in each case. For this reason, clinical gait analysis is often used to identify gait limitations and guide treatment planning, contributing to the improvement of patient symptoms. However, gait analysis in children with cerebral palsy typically examines the patient's usual gait, where the child may adopt compensation strategies, possibly masking the underlying gait defects and hindering proper identification of relevant causes. An alternative strategy is to **actively perturb patients challenging them outside of their comfortable self-selected gait, by providing real-time feedback on key gait parameters.** In this way, the individual patient's response may expose underlying impairments, enhancing the information gained from gait analysis.

An application for real-time feedback on clinical gait parameters has been implemented by Motek using the Human Body Model (HBM), a patient-personalised model scaled to the individual's size, gait features and abilities [1]. Through the HBM, feedback can be targeted to specific aspects of an individual's gait, and the difficultly level of feedback is set relative to an individual's gait ability.

As final step, the model has been tested for clinical validation at VUmc with children with cerebral palsy. **The results were promising: children proved able to adapt their gait, reaching clinically significant improvements in regard to a series of gait features, such as hip and knee extension [2] (Figure 1), ankle power generation (Figure 2), step length and pelvis tilt [3].** Also, several children showed compensatory movements in response to the feedback: for example, some children were not able to achieve increased knee extension without increased internal hip rotation. The ability of individual children to adapt specific gait parameters as well as their compensation strategies represent important clinical elements for physicians, which could help them confirm effectiveness of treatment options or adjust current therapies, ultimately contributing to improve their patients' outcomes.

1. Van den Bogert, A. J., Geijtenbeek, T., Even-Zohar, O., Steenbrink, F. & Hardin, E. C. A real-time system for biomechanical analysis of human movement and muscle function. Med. Biol. Eng. Comput. 51, 1069–1077 (2013).
2. van Gelder, L. et al. Real-time feedback to improve gait in children with cerebral palsy. Gait Posture 52, 76–82 (2017).
3. Booth, A., Steenbrink, F., Buizer, A., Harlaar, J. & van der Krogt, M. Real time visual feedback to alter gait in children with CP. Dev. Med Child Neurol. (2017). In press.
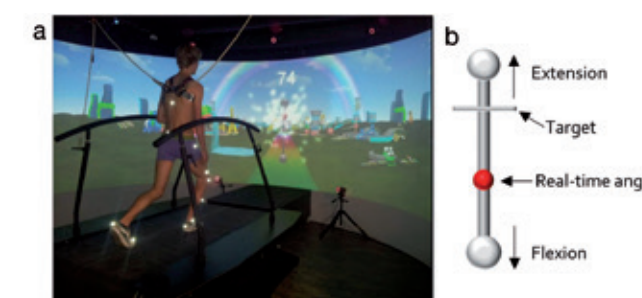

Figure 1. a) GRAIL with feedback. b) Expanded display of feedback given on hip and knee angle.
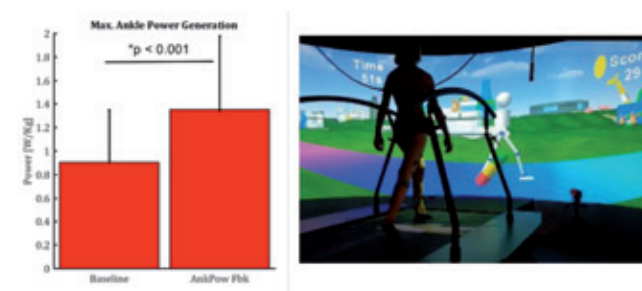

Figure 2. In response to feedback on ankle power, children with CP were able to reach large improvements in ankle power generation of over 45% improvement.

BIG DATA ANALYTICS

B uilding on top of previous EU-funded research efforts (such as Health-e-Child and Sim-e-Child), **the MD-Paedigree Infostructure has been designed as a cloud-based federated architecture**, hosting pseudonymised copies of clinical datasets coming from the incoming affiliated clinical centers, so as to allow **on-site running of advanced decision support algorithms in the clinical setting, thus maximising the value of large biomedical datasets for clinical and research purposes.** To address these requirements, specific attention has been given to the development of applications directed to a double goal: on one side, **an optimal dataset management obtained through pre-processing, profiling and curation/cleaning of incoming data records;** on the other, **to enhance data value for clinical purposes, by the development of analytics designed for specific medical use-cases.** Particular attention has been given to the **development of retrieval tools, as to enable clinicians to consult a vast record of clinical cases in a straightforward and automated manner.** These efforts have led to the implementation of a wide range of applications, which have been all tested by physicians for their clinical validity, computational efficiency and user-friendliness.

Among these, ATHENA RC developed a **unified dataflow processing system, integrating the Data Curation and Validation for data processing and the AITION knowledge discovery tools (next page).** While the former allows a (semi)-automatic data profiling and cleaning process, facilitating the detection of numeric outliers, missing values and various types of inconsistencies, enabling timely detection and removal of dataset errors, the latter employs advanced machine learning and similarity search techniques to achieve an accurate clustering of clinical cases on computationally "learnt" categories. Similarly, **a series of machine learning classifiers, namely Random Forests, Support Vector Machines, a Multilayer Perceptron and Boosting**, have been developed within the HES-SO team to support the classification of clinical cases into their disease category by relying on videos and sensor data regarding clinical gait analysis of paediatric patients **(page 30)**. Other tools, instead, have been implemented for case-retrieval purposes by HES-SO: while the **Shangri-La and Shambala search interfaces (page 30)** allows combined text-based and clinical image-based search and retrieval, **the Case-Based Retrieval service (page 29)** enables clinicians to search for episodes of care based on either unstructured data (i.e. discharge summaries) and structured data (i.e. demographic data).

The route from biomedical analytics to clinical validation

Omiros Metaxas
ATHENA RC
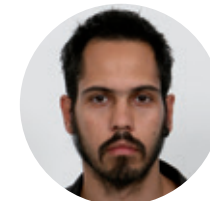
Orfeas Aidonopoulos
ATHENA RC

Harry Dimitropoulos
ATHENA RC

T he ATHENA RC team's goal throughout the project has been to develop **a web-based, end-to-end data profiling, curation/cleaning, pre-processing, analytics and knowledge discovery platform for big data healthcare, including the Data Curation and Validation (DCV) [1] and the AITION knowledge discovery (KDD) [2] tools.** The platform runs on top of ATHENA's EXAREME [3] dataflow processing system, which provides distributed processing and parallelization of resource/time-consuming algorithms related to knowledge discovery, simulation and data mining. The entire platform has only one point of integration with the MD-Paedigree platform.

The DCV module of the platform offers an advanced (semi)-automatic data profiling and cleaning process, providing mechanisms for facilitating the detection of numeric outliers, missing values, inconsistencies, as well as alphanumeric typographical errors of tabular datasets. A user-friendly interface allows defining and running data cleaning rules over a relation such as functional dependencies, conditional functional dependencies and denial constraints. DCV's functionality for computing new derived columns either through discretisation criteria or by computing and executing arithmetic operations (e.g. for computing medical scores), is also very useful for clinicians, saving time and eliminating errors. In addition, DCV keeps a history of all actions that affect the values of data. The user can undo/redo history or save workflows and re-run them in other projects or with other data.

The KDD module of the platform (Figures 1-3) incorporates several well-established machine learning techniques targeting clustering, classification, dimensionality reduction and similarity search. Applying machine learning methods, users can fit their research data on computational models which "learn" to classify patient data on different categories. Clinicians often try to manually develop a model for classifying patients, but implementing models with validated classification rules takes time. The platform provides clinicians with a user-friendly environment where they can build and validate models on their own. Models don't only automate the classification steps but also provide additional information and insights, such as clustering of samples in cohorts, or extraction of features that play a key-role on each use-case. Additionally, clinicians can observe post-analysis metrics to assess their results and have a holistic view of their analysis. In order to integrate the KDD module with the DCV tool, the whole interface layout was redesigned and new pre-processing features were added. Each user can

Figure 1. Main interface of knowledge discovery platform. Parameters have been added to create a classification model. The tool automatically recognises whether this is a binary or multiclass classification task. The number of patients are also shown in case users want to include a grouping variable in their analysis.
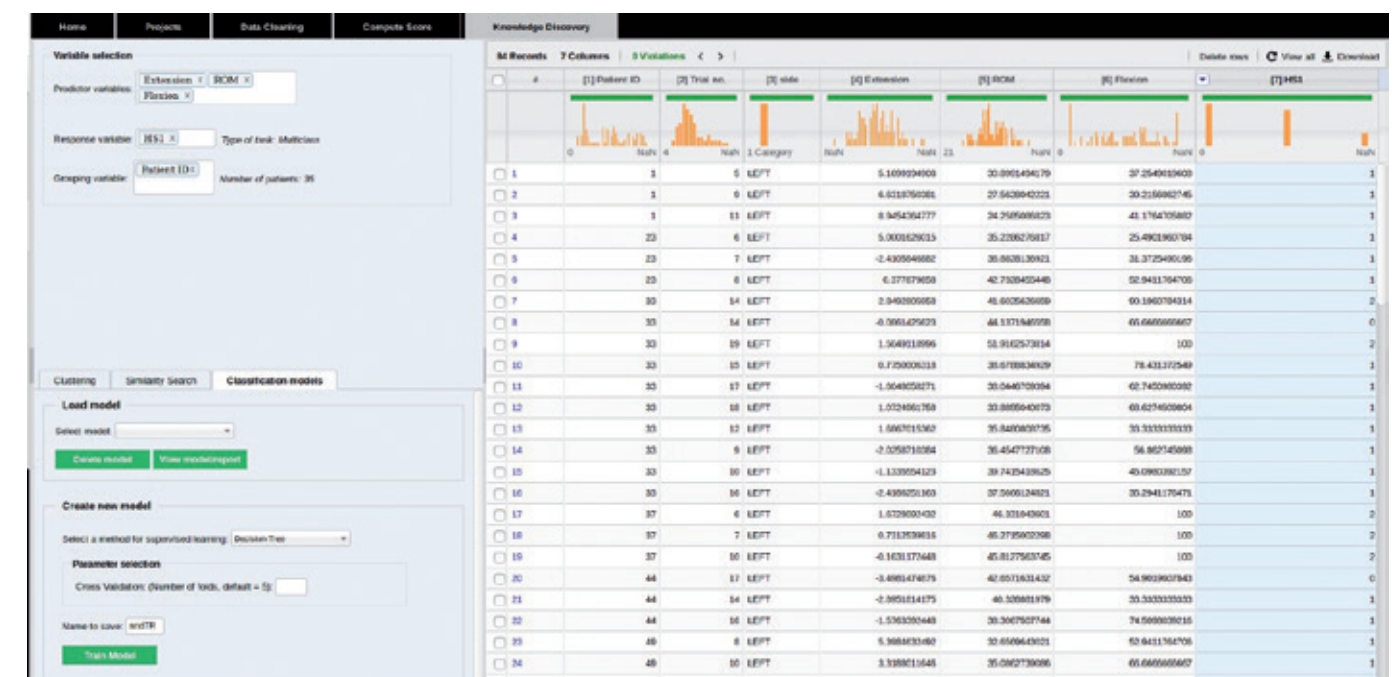
now export data regions of interest through interactive visualisations, or can save a subset of the dataset by deleting selected rows and columns. Missing values can be imputed to handle the missing values in the model's training phase. Similarity search was also enriched with new parameter selection options, and posterior probabilities tables were added after a classification analysis.

Herein, a neurologic and neuromuscular disease use-case is illustrated. Having selected three predictors (extension, range of motion and flexion of the hip joint in the left foot of the patient), a computational model is developed. Once the model fits the data, a comprehensive report is provided. A new classification column is produced to compare with the clinician's manual classification. By using white-box strategies, detailed classification rules are visualized so clinicians can compare their rules with those the model developed.

An end-to-end, user-friendly, web-based analytics platform is hence ready to tackle the most challenging biomedical tasks. From data curation to knowledge discovery, clinicians can explore and identify statistical profiles of their datasets. In cooperation with clinicians, the ATHENA RC researchers are now working on incorporating the platform and the related produced models within the regular clinical validation process.

1. Gogolou, A., Kyriakidi, M. & Ioannidis, Y. Data Exploration: A Roll Call of All User-data Interaction Functionality. in Proc. of the Third Int'l Workshop on Exploratory Search in Databases and the Web (ExploreDB) (2016).
2. Metaxas, O., Dimitropoulos, H. & Ioannidis, Y. AITION: A Scalable KDD Platform for Big Data Healthcare. in IEEE-EMBS International Conference on Biomedical and Health Informatics (2014).
3. Chronis, Y. et al. A Relational Approach to Complex Dataflows. in Proceedings of the Workshops of the EDBT/ICDT 2016 Joint Conference (2016).

Figure 2 (right). the platform gives an insight view to the clinicians. Classification rules, based on which the whole procedure was developed, are provided. Clinicians can thus compare their rules with those the model developed to classify the patient cohorts.

Figure 3 (below). Confusion Matrix (left): how the samples are distributed among classes during the training phase of the algorithm. Predictors' significance (right): the user can easily see that 'ROM' does not contribute to the classification accuracy while 'Extension of the Hip' seems to be a significant player for NND classification tasks.
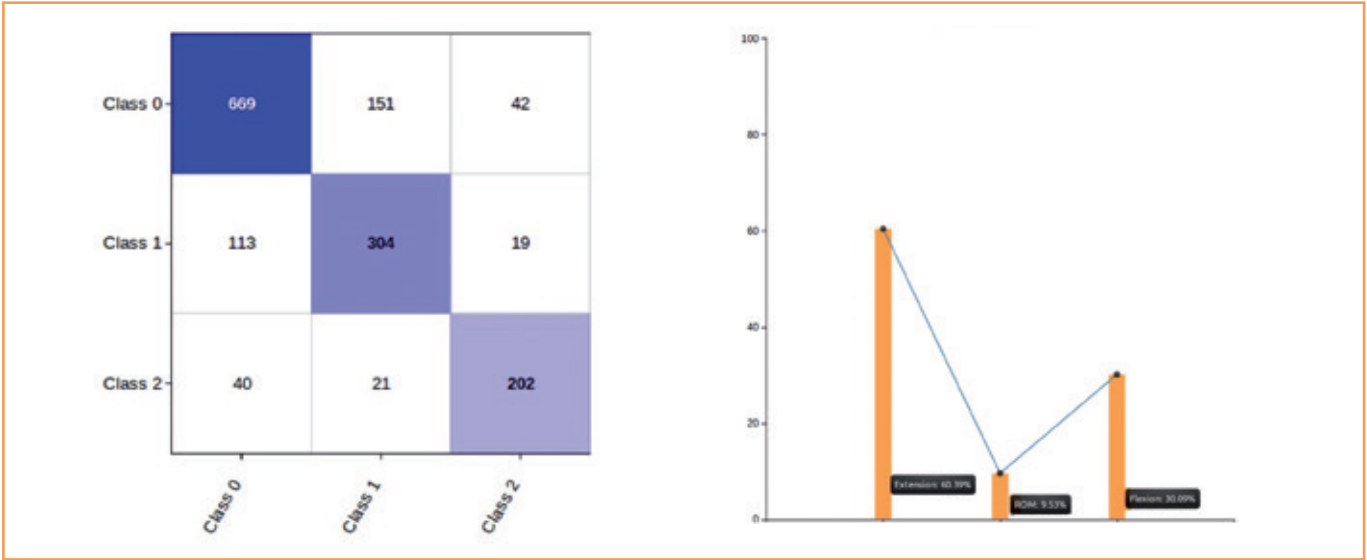
A case-based retrieval tool for physicians

Emilie Pasche
HES-SO

Physicians, particularly when facing complex diseases, show a great interest in finding populations of patients similar to theirs, to observe the response of a particular treatment and learn about the outcomes at different time points in a given clinical pathway. However, a substantial part of the clinical information is stored in unstructured textual contents. Therefore, tools likely to make sense of clinical narratives are essentials to enable the retrieval of similar cases.

To this aim, the University of Applied Sciences of Western Switzerland (HES-SO) developed **a Case-Based Retrieval (CBR) service able to find similar episodes of care based on different modalities: unstructured data** (i.e. discharge summaries) and structured data (i.e. demographic data). It is based on a set of about 43,000 episodes of care relative to patients consulting for cardiac pathologies. The system harvests electronic health records (EHRs) from the MD-Paedigree Infostructure, and EHRs are automatically processed: 1) terminological descriptors are assigned using categorizers [1]; 2) specific fields (e.g. ejection fraction, weight, etc.) are extracted from unstructured textual content (i.e. discharge summaries). Data are then indexed using Apache Solr. At query time, the user inputs a patient description, either in English or in Italian. The query

is processed (Figure 1) and the Solr search engine outputs similar cases in EHR (Figure 2). The user can then review the retrieved episodes of care and validate their respective relevance. These judgements are used to reformulate the query with additional keywords based on a Rocchio algorithm [2]. Finally, the user can obtain refined results.

The system has been qualitatively and quantitatively evaluated during a test session by an end user (i.e. a MD specialized in paediatric cardiology). 40 queries were tested and the top-10 results were assessed. **In more than half of the cases and for up to two thirds of them, the system was able to suggest a similar episode of care at first rank.** Moreover, the feedback relevance service enabled a slight further improvement in the output precision. The feedback obtained from the qualitative evaluation, despite the known rarity of the group of diseases analysed, was sufficient to improve the application usability. From a quantitative point of view, the current results are already regarded as fair to support a case-based retrieval application in a clinical environment.

1. P. Ruch, Automatic assignment of biomedical categories: toward a generic approach, Bioinformatics 22(6) (2006), 658-64.
2. P. Ruch, I Tbahriti, J. Gobeill, A.R. Aronson, Argumentative feedback: A linguistically-motivated term expansion for information retrieval, Proceedings of the COLING/ACL (2006):675-82.
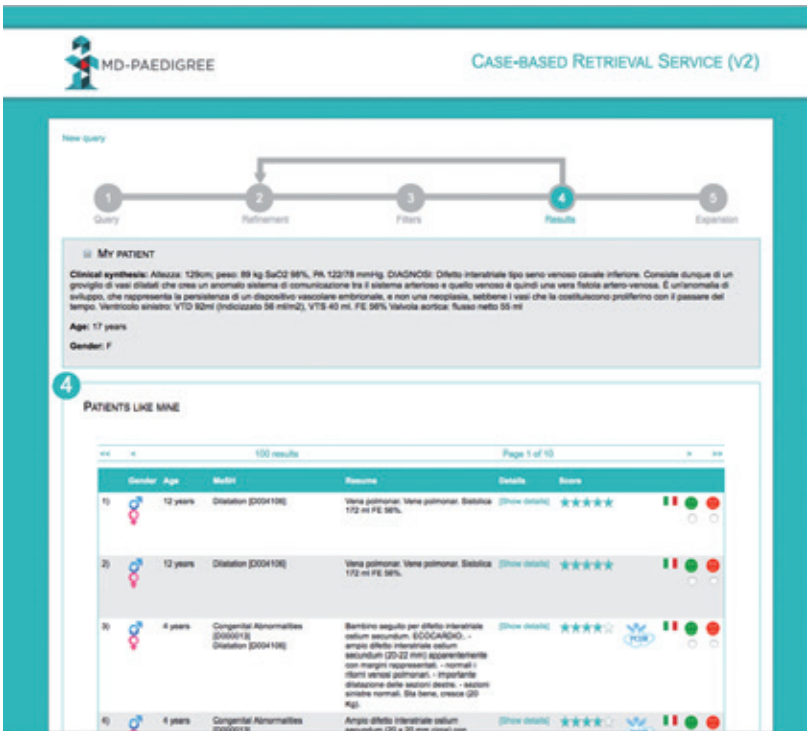
Figure 1. The system reformulates the query displayed on the top and extracts all numeric fields from unstructured text. The physician can, among others, define ranges considered as relevant for each entity. The similar episodes of care are displayed in a table, ranked by relevance.

Content-based retrieval and image analysis tools

Ranveer Joyseeree
HES-SO

Henning Müller
HES-SO

Image-based and text-based retrieval services represent valuable tools to respond to information needs in clinical routine, as they enable clinicians, for example, to find past patients similar to the present ones, to be used as clues to interpret complex clinical cases. To achieve this goal, the HES-SO provided **a combined text-based and image-based retrieval service adapting two innovative search interfaces, entitled Shangri-La [1] and Shambala [2]** that were integrated with the MD-P infostructure. The development of a textual/visual search tool for similarity search has been focused on images from articles appearing in the medical literature: the visual search tool, recently finalized and presented at the CBMI conference, uses visual characteristics (texture and colour features) extracted from the images of articles and compares them to clinical images, and then ranks the results based on visual similarity [3].

In addition to the visual/textual search, in the MD-P disease areas where no multimodal data are available, a focused work has been done to develop classification tools able to rely on other types of data. This focused work started with gait analysis [4], based on videos and sensor data of children walking that are in C3D format. Using JIA and NND C3D data (Figure 1), **a series of machine learning classifiers, namely Random Forests, Support Vector Machines (SVMs), a Multilayer Perceptron (MLP) and Boosting,** were trained on a data subset and then tested on the entire available data. This allowed to obtain **a classification of clinical cases into their disease category with an accuracy between 96.4% and 100%, only based on gait data.**

In a second approach, the team succeeded in developing **an innovative method to estimate left-ventricular (LV) mass using only ultrasound data.** Most clinical protocols include ultrasound data of the heart but not always MRI, currently needed for calculating the LV mass. This fully automatic analysis employs data from the CVD disease area, as shown in Figure 2: several key values are estimated from a combination of frames of the ultrasound video overlaid on each other. First results show that mass estimates are close to the true values, demonstrating feasibility of replacing the costly MRI analyses with data automatically derived from ultrasound videos.

1. García Seco de Herrera, A., Schaer, R. & Müller, H. Shangri-La: a medical case-based retrieval tool. J. Am. Soc. Inf. Sci. (2017). In press.
2. Schaer, R. & Müller, H. A Modern Web Interface for Medical Image Retrieval. in Swiss e-health summit (2014).
3. Joyseeree, R., Schaer, R. & Müller, H. A Demo of Multimodal Medical Retrieval, Workshop on Content-based Multimedia Indexing. in Workshop on Content-based Multimedia Indexing (CBMI) (2016).
4. Joyseeree, R., Abou Sabha, R. & Müller, H. Applying Machine Learning to Gait Analysis Data for Disease Identification. in Medical Informatics Europe (MIE) (2015).
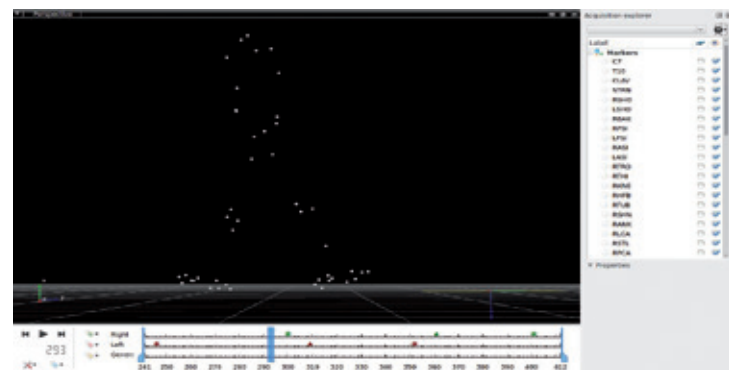
Figure 1. Visualisation of a C3D file using the open-source toolkit: b-tk. The white dots on the black background represent the position of markers on the body of the patient in 3D grid space. Their visibility may be toggled on and off using the menu on the right. Events such as heelstrike may be viewed on the timeline at the bottom.
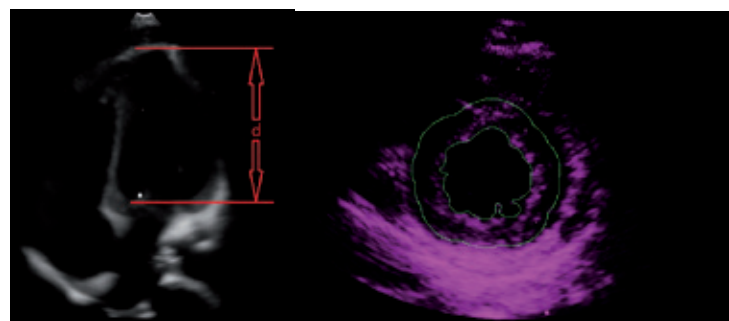
Figure 2. The estimation of the mass of the left ventricle of a patient using only ultrasound data is carried out by manually estimating the vertical height, d, of the left ventricle in an apical 4-chamber view of the heart (left side of figure) and automatically segmenting and measuring its cross-sectional area, A, in a parasternal short axis view (the region in between the green contours on the right side of the figure). The mass estimate is a function of d and A and the function is analytically determined using a leave-one-patient-out (LOPO) cross validation technique.

The way ahead: MD-Paedigree exploitation perspectives

One key challenge of the EU-funded research project is making its results sustainable after the grant period, **extracting value from these results through business initiatives or further research activities**. For this reason, since its inception MD-Paedigree explored the relevant market, focusing on two nascent and dynamic ones: **the advanced Clinical Decision Support System (CDSS) and the more recent clinical data valorisation markets**.

Regarding the former, even though the current adoption of advanced decision support systems is still marginal, demand for automation in clinical decision making is rapidly growing as the medical IT environment becomes increasingly mature, thus providing the necessary ground for the next generation of advanced systems. **CDSSs will be globally the fastest mounting market segment in eHealth, expected to grow at a CAGR of 21.5% to 26.5% from 2016 to 2019 (Figure 1)**. In this context, the CDSSs developed within MD-Paedigree could be the basis for **new products, to be licensed to third party vendors or directly commercialised by their inventors**. Still, at present, a real online marketplace where to download (or directly run) model-based CDSSs is yet to come, despite the appearance of some examples of marketplace mainly for statistical CDSSs, such as the US-based Apervita platform.

Coming to what we called the new **"clinical data valorisation market", clinical data can surely have enormous value for several stakeholders, starting from pharma and medical devices companies**. However, a clear business model indicating how hospitals and citizens/patients will be able to extract value from exchanging and sharing their datasets is still not well defined. Also, key enabling technologies for making this exchange possible, while guaranteeing patients' privacy and data security, are not yet at a sufficient maturity level.

To this regard, one of the key MD-Paedigree exploitation initiatives has been the participation (with its IT Infostructure and wealth of datasets) in **the newly-funded EU project MyHealthMyData (pages 5-6), which explores blockchain and advanced security solutions adoption for healthcare applications and clinical data management**, while also evaluating the most recent regulatory requirements (GDPR). At the same time, further "infonomics" analysis need to be completed regarding the most appropriate data valorisation models for very diverse data assets. Real data commoditisation is not here yet, but MD-Paedigree is well placed for extracting the maximum value from its datasets and technologies, as soon as this hazy landscape will be eventually cleared.

Figure 1. Global CDDS and EHR Markets - Forecast 2016-2021 - CAGR CDSS: 27.1% [1].

GLOBAL CDDS AND EHR MARKETS - FORECAST 2016-2021   [VALORE] B$

— EHR   — CDSS
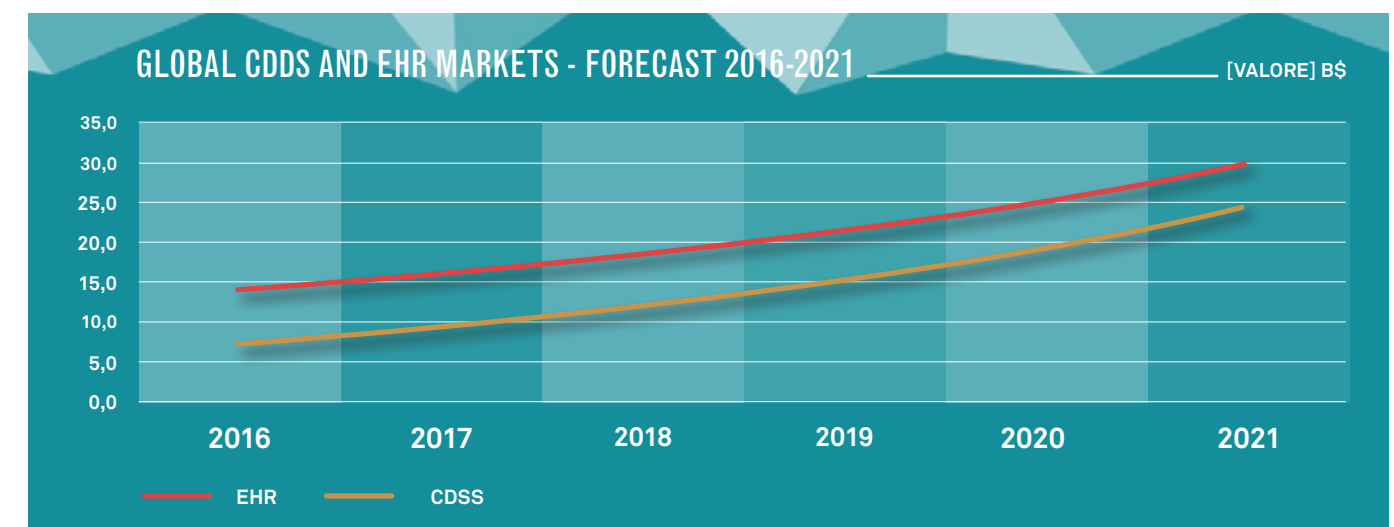
1. MarketsandMarkets. Healthcare Analytics/ Medical Analytics Market by Application (Clinical, RCM, Claim, Fraud, Supply Chain, HR, PHM), Type (Prescriptive), Component (Service, Software), Delivery (On-premise, Cloud), End User (Hospital, Payer, ACO, TPA) - Forecasts to 2021. (2016).

MD-PAEDIGREE PARTNERS

MD-Paedigree is an international collaboration between 22 partners across Europe from industry, academia and healthcare.
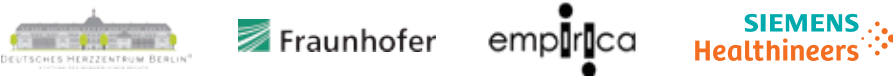
Project Coordinator:

Bambino Gesù
OSPEDALE PEDIATRICO

BELGIUM

KU LEUVEN

FRANCE

Inria

gnúbila

GERMANY

Deutsches Herzzentrum Berlin

Fraunhofer

empirica

SIEMENS Healthineers

GREECE

ATHENA

ITALY

GASLINI

BMR Genomics

SAPIENZA Università di Roma

LYNKEUS
STRATEGY CONSULTING | BLOCKCHAIN & SMART CONTRACTS | DATA ANALYTICS

NETHERLANDS

UMC Utrecht

VU UNIVERSITY AMSTERDAM

Motek Medical
IMPROVING HUMAN PERFORMANCE

TU Delft
Delft University of Technology

ROMANIA

TRANSILVANIA University of Brasov

USA

SIEMENS Healthineers

SWITZERLAND

Hes·so GENÈVE
Haute École Spécialisée de Suisse occidentale

UK

UCL

The University Of Sheffield.

NEWSLETTER INFO

Editorial Board
+ Anna Rizzo (Lynkeus)
+ Mirko De Maldè (Lynkeus)
+ Davide Zaccagnini (Lynkeus)

Contributors
Orfeas Aidonopoulos, Harry Dimitropoulos, Omiros Metaxas (Athena RC)
Karl A Stroetmann, Rainer Thiel (Empirica Communication & Technology Research)
Klaus Drechsler, Anqi Wang, Stefan Wesarg (Fraunhofer Institute for Computer Graphics Research IGD)
Ranveer Joyseeree, Henning Müller, Emilie Pasche (HES-SO)
Roch Mollero (Inria)
Mirko De Maldè, Ludovica Durst, Edwin Morley-Fletcher, Anna Rizzo, Davide Zaccagnini (Lynkeus)
Adam Booth, Johannes Gijsbers, Frans Steenbrink (Motek)
Maria Jimena Costa, Tobias Heimann, Olivier Pauly (Siemens Healthineers)
Thomas Geijtenbeek (Technical University Delft)
Pieter van Dijkhuizen (University Medical Centre Utrecht)
Alexander Jones (University of Oxford)
Claudia Mazzà, Luca Modenese, Erica Montefiori, Jose M Pozo, Marco Viceconti (The University of Sheffield)
Annemieke Buizer, Jaap Harlaar, Marjolein van der Krogt (VU University Medical Centre)

Subscription
The MD-Paedigree newsletter is published twice a year by the MD-Paedigree consortium and is distributed free of charge.
All issues of the newsletter will be available on our website. For subscription to the newsletter please go to md-paedigree.eu

Disclaimer
The MD-Paedigree newsletter is funded by the European Commission under the Seventh Framework Programme. The content of this newsletter cannot be considered as the European Commission's official position and neither the European Commission nor any person acting on behalf of the European Commission is responsible for the use which might be made of it; its content os the sole responsibility of the MD-Paedigree project partners.
Although the MD-Paedigree consortium endeavours to deliver high-quality, no guarantee can be given regarding the correctness and completeness of the content of this newsletter due to its general information character.

Contacts:
Email: info@md-paedgree.eu
Twitter: @mdpaedigree
Web: www.md-padigree.eu

WWW.MD-PADIGREE.EU