Health Policy and Technology 15 (2026) 101149

Contents lists available at ScienceDirect

Health Policy and Technology

journal homepage: www.elsevier.com/locate/hlpt

Original Article/Research

Digital health and artificial intelligence: a research approach to enable sustainable and personalised local healthcare

Monica Moroni [a,1,*], Lisa Novello [a,1], Giulia Malfatti [b], Lorenzo Gios [b], Roberto Bonmassari [c], Maurizio Del Greco [c], Massimiliano Maines [c], Michele Moretti [c], Sandro Inchiostro [c], Federica Romanelli [c], Elisabetta Racano [c], Tania Elena Maggi [c], Valentina Fiabane [c], Adele Compagnone [c], Lorena Filippi [c], Roberta Pasquini [c], Marta Betta [c], Lucia Pavanello [c], Andrea Manica [c], Diego Cagol [c], Enrico Santoprete [c], Simone Cecchetto [c], Giorgia Bincoletto [d], Diego Conforti [e], Andrea Nicolini [b], Giuseppe Jurman [a,f]

[a] Data Science for Health Unit, Fondazione Bruno Kessler, Trento, Italy
[b] TrentinoSalute4.0 – Competence Center for Digital Health of the Province of Trento, Trento, Italy
[c] Azienda Provinciale per i Servizi Sanitari (APSS) di Trento, Trento, Italy
[d] Faculty of Law of the University of Trento, Trento, Italy
[e] Provincia autonoma di Trento, Italy
[f] Department of Biomedical Sciences, Humanitas University, Milan, Italy

ARTICLE INFO

Keywords:
Artificial intelligence
Public healthcare
Personalised medicine
Digital health

ABSTRACT

Background: The integration of Artificial Intelligence (AI) into healthcare services and technologies offers substantial potential for personalised medicine. The Autonomous Province of Trento (Italy) provides a unique setting for AI-driven healthcare research, due to its unified healthcare system, advanced IT infrastructure, and strong public-private collaborations. This paper explores an initiative aimed at improving healthcare accessibility and promoting innovation through AI in three clinical domains: Cardiology, Diabetic Retinopathy, and Paediatric Ophthalmology.

Methods: The project employs a structured approach, involving specialised working groups addressing clinical needs, AI techniques, legal and ethical compliance and data management. The initiative aims to develop predictive models aligned with European and national data protection regulations.

Results: Three primary clinical objectives were defined: estimating individual risk profiles in heart failure patients, personalising screening intervals for diabetic retinopathy, and supporting early diagnosis of anterior segment opacities in infants. Data relevant for the selected outcomes were identified. A dedicated platform for compliant, secure and structured access to data was developed. A data analysis plan was designed, including data processing, models selection, optimization and evaluation. All research protocols were approved by the local Ethics Committee.

Discussion: The initiative investigates the AI potential to improve clinical outcomes and establish a sustainable, personalised healthcare system. Key challenges include data accessibility, regulatory compliance, and adherence to ethical standards. The project's comprehensive framework offers a model for broader applications. Future research will focus on model validation and expanding the initiative to other clinical domains.

Public Interest Summary: This article presents the "Digital Health and Artificial Intelligence" project, an initiative funded by The Autonomous Province of Trento (Italy) to enhance healthcare accessibility and foster innovative healthcare models using technology and Artificial Intelligence (AI). The current work presents the design and preparatory work for the implementation of three AI-based solutions for research purposes, encompassing three areas: i) Cardiology, ii) Diabetic Retinopathy, and iii) Paediatric Ophthalmology. The paper outlines the legal and organizational frameworks, mathematical modelling and data management emphasising the necessity of cross-disciplinary endeavour and collaboration. Overall, this project represents a forward-looking initiative

* Corresponding author at: Fondazione Bruno Kessler, Via Sommarive, 18, 38123 Trento TN.
E-mail address: mmoroni@fbk.eu (M. Moroni).
[1] These authors contributed equally to this work

https://doi.org/10.1016/j.hlpt.2025.101149

Available online 12 December 2025

promoting research conducted on citizen data to address healthcare needs through innovative AI-driven approaches in line with legal and ethical standards.

Introduction

The international landscape of Artificial Intelligence (AI) research in healthcare presents both significant potential alongside complex challenges and limitations [1,2]. AI's impact is expanding so rapidly that it is poised to become a cornerstone of future medical practice. It offers new pathways to enhance sustainability, optimise healthcare delivery, and advances a predictive, preventive and personalised approach to medicine. This objective is widely advocated by international institutions [3].

The healthcare sector is characterised by the vast amounts of potential available data, which serves as a foundation for AI applications. There is an increasing number of research initiatives applying AI across public health, preventive care and specific clinical domains [4–6]. At the same time, the deployment of AI in healthcare should be subject to a robust legal and ethical framework. At both national and European levels, efforts are intensifying to regulate the use of AI in research and healthcare delivery, as demonstrated by the recent European Artificial Intelligence Act [7].

Despite these opportunities, many critical challenges need to be addressed to fully leverage AI in healthcare. Within this framework, the Autonomous Province of Trento in Italy offers a unique environment for piloting and advancing a value chain that combines AI research and innovation in healthcare, thanks to a set of distinctive conditions hereafter described.

One major challenge is the alignment of clinical needs, research opportunities, and regulatory requirements. Technological advancements in the medical field enable the acquisition of large data volumes essential for impact research. However, the requirement to obtain consent for each specific data processing purpose introduces significant procedural challenges. This is particularly evident in real-world clinical settings in prospective studies, where consent for research and data are collected during routine care activities, slowing visits and clinical procedures. Similarly, retrospective studies exhibit their own challenges in collecting informed consents, both organizationally (e.g. contacting patients not attending clinical follow-ups) and legally (e.g. dealing with cases with deceased or unreachable patients). Furthermore, informed consent documents should be formulated according to transparency and accessibility principles, allowing patients to understand research goals even without technical or scientific expertise. Moreover, research outputs must demonstrate reliability and transparency to build trust and be clinically relevant. Balancing innovation with compliance and ethical integrity is crucial for the successful implementation of AI technologies, and shortens the value chain between research and impact in terms of innovative services [7]. Another significant challenge concerns data access and data management. The General Data Protection Regulation (GDPR – EU Regulation 2016/679) imposes specific requirements on the processing of personal data and requires the development of appropriate organisational and technical measures. Data platforms should be built to facilitate research and the implementation of AI pipelines while ensuring secure and efficient data management [7,8]. In this regard, the Province of Trento's unified healthcare system offers a strategic advantage, ensuring high-quality standards and comprehensive control over clinical service delivery pathways and procedures, facilitating the mitigation of the above-mentioned challenges. Additionally, a top-notch and evolving IT infrastructure supports both healthcare professionals and citizens, enhancing the potential of new technologies in delivering healthcare services [9]. Among the different infrastructures, TreC+ is the core one: an integrated digital health platform that empowers patients and functions as a personal health data repository. It allows provincial residents to access, manage, and share health information through telemedicine, and has already been used in studies involving diabetes, heart failure, and pregnancy. Finally, the institution of multi-professional and inter-institutional collaborations is required to ensure an effective and ethical use of data in AI research and the respect of patient privacy, data protection, data security, and ethical principles [2].

These efforts must involve both private and public stakeholders at local, national and international levels. In this context, the Province of Trento benefits from a long-standing collaboration between public and private institutions engaged in research and healthcare, under a public-driven cooperation. A central role is played by TrentinoSalute4.0 (TS4.0), the "Competence Center on Digital Health" endorsed by the Autonomous Province of Trento. TS4.0 connects the local healthcare trust (Azienda Provinciale per i Servizi Sanitari - APSS), the Bruno Kessler Foundation (FBK), the University of Trento, and the wider community. The Center facilitates collaboration and coordination among stakeholders by integrating three elements: healthcare planning directives from the Autonomous Province of Trento, clinical innovation needs from APSS, and applied research, including AI.

Further distinctive aspects make Trento a favourable context for clinical research. Its small geographical size and relatively low population are paired with a streamlined governance structure. This, combined with the Province's special autonomy in health and research under the Italian Constitution, allows for rapid and flexible decision-making.

In this paper, the authors report the design phase and preparatory work of an AI-based healthcare initiative that fosters research aimed at improving healthcare accessibility and promoting innovative healthcare models using digital technologies and AI for predictive modelling, early detection, and personalised interventions within a public-driven research agenda. The "Digital Health and Artificial Intelligence" project is a 24-months initiative funded by Provincial Government Resolution No 2475 on December 22, 2022 [10]. It aligns with the provincial healthcare priorities, focusing particularly on three key clinical areas: i) Cardiology, ii) Diabetic Retinopathy, and iii) Paediatric Ophthalmology, selected for their significance in addressing prevalent health issues and their alignment with international and national research agendas, particularly in the management of chronic diseases and primary prevention.

The project develops on two levels:

- organizational innovation: involving the introduction and validation of organizational processes that leverage on technological advancements to support healthcare professionals and citizens;
- AI model development: focused on the design and development of diagnostic and predictive AI models based on digital data collected in routine clinical practice.

Overall, the "Digital Health and Artificial Intelligence" project represents a forward-looking strategy to address healthcare needs through innovation, multidisciplinary collaboration, and the responsible use of AI and digital technologies.

Methods

The core objective of the project is to develop predictive AI models for research purposes addressing high-level research questions, whilst the secondary objective is to validate a cross-institutional framework for AI research. The participating centres and entities are TS4.0, APSS, FBK and the University of Trento, whose overall roles are reported in Table 1.

Table 1
Participating entities and role in the "Digital Health and Artificial Intelligence" project. TS4.0: TrentinoSalute4.0, FBK: Bruno Kessler Foundation, APSS: Azienda Provinciale per i Servizi Sanitari.

Institution	Department	Role
TS4.0		Project coordination and management
APSS	Cardiology Units (Trento and Rovereto)	Clinical center for Cardiology
	Ophthalmology Unit	Clinical center for Diabetic Retinopathy and Paediatric Ophthalmology
	Diabetes Centers	Clinical center for Diabetic Retinopathy
	Orthoptists APSS	Clinical center Paediatric Ophthalmology
	Primary Care Pediatricians	Clinical center Paediatric Ophthalmology
	IT Department	Data extraction and data management
FBK	Data Science for Health Unit, Digital Health and Wellbeing Center	Data analysis and modelling
University of Trento	Faculty of Law	Regulatory compliance

Study planning and design

In the methodological planning of the research project, several specialised working groups are established to address distinct aspects of the study. Each group oversees the achievement of specific objectives to ensure a comprehensive and coordinated approach. Roles, responsibilities, and composition of each working group are described hereafter and summarized in Fig. 1.

The "study design team" focusing on clinical and methodological issues consists of clinical experts, AI specialists, and project managers and is designed with the following tasks. Clinical experts identify relevant research questions, ensuring alignment with pressing clinical needs and gaps in current knowledge. Together with AI specialists, they identify relevant information for the algorithms. This process includes selecting essential variables to answer the research questions and identifying appropriate digital and medical devices to gather reliable clinical, biometrical, and imaging data. Project managers coordinate efforts to maintain alignment with the overall project goals and the writing of the scientific protocols and related documents to be submitted to the local Ethical Committee.

The "AI and data analysis team" is composed of AI specialists and project managers and focuses on the technical development and integration of AI models within the research framework. Tasks of this team include building AI solutions tailored to the selected research questions and data sources, developing and finalising operational and research parts of the protocols.

The "privacy and data governance team" is responsible for the compliance of all project's aspects with relevant privacy and data protection laws. Its tasks include several key functions: identifying appropriate scenarios to meet privacy requirements; outlining workflows to safeguard participants' rights; and conducting Data Protection Impact Assessments (DPIA) to mitigate data management risks. This includes programming robust data protection and security measures to be implemented throughout the data flows and lifecycle. The group consists of data protection experts, Data Protection Officers (DPO), and IT specialists focusing on data management and IT risk analysis. Project managers ensure these activities integrate smoothly into the overall project plan.

The "data management team" is composed of IT experts in data management, data protection, security, and legal compliance, alongside project managers and focuses on identifying and confirming the sources of data necessary for the research, ensuring availability and reliability from clinical records, devices, and other sources. Furthermore, this group ensures that data handling procedures are secure, efficient, and adhere to regulatory standards and requirements by setting up a dedicated IT platform to manage the data lifecycle, from collection to processing and storage. This platform must ensure compliance with the GDPR and cybersecurity requirements and generate datasets in a form that is useful for research purposes.

This structured approach, delineating specific tasks and team compositions, aims to facilitate effective project execution while addressing the multifaceted requirements of clinical, methodological, data protection, and data management issues. The entire drafting process was coordinated by TS4.0, ensuring a cohesive and comprehensive strategy that leverages the strengths and expertise of all involved entities to coherently plan the project's objectives. In the following, we present the project design according to the CRISP-MED-DM framework [11], which provides a standardized approach to data mining studies, while being

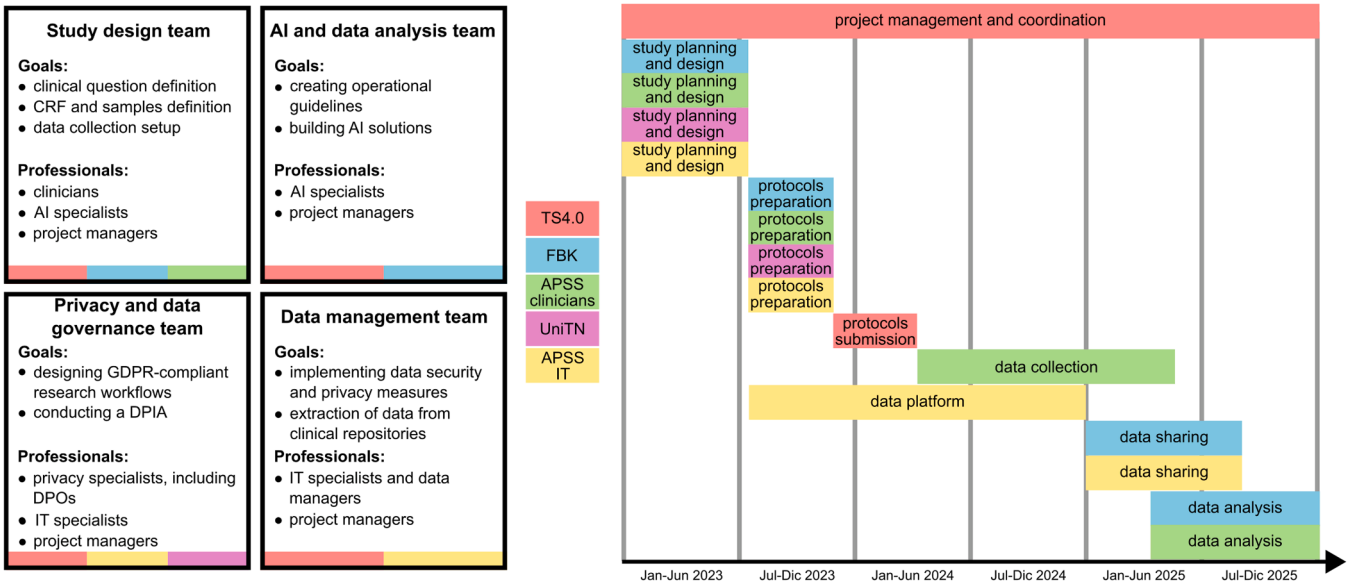

Fig. 1. caption. Working groups description and project Gantt Chart. TS4.0: TrentinoSalute4.0, FBK: Fondazione Bruno Kessler, APSS: Azienda Provinciale per i Servizi Sanitari, UniTN: University of Trento, CRF: Case Report Form, DPIA: Data Protection Impact Assessments, DPO: Data Protection Officer.

tailored to address peculiarities of the medical domain such as privacy concerns.

Project scope definition: outcomes definition

In terms of clinical endpoints, the selected objectives for the cardiology domain include estimating the individual risk profile for heart failure patients to develop complications, such as re-hospitalization, occurrence of arrhythmic events, stroke and atrial fibrillation within 12/24 months. Additionally, a specific aim is identified in terms of estimating the risk profile for sudden cardiac death.

For diabetic retinopathy, the identified goal is to estimate the individual risk profile for disease development or progression, thereby personalising screening intervals while maintaining consistent risk levels. A secondary goal is to develop an algorithm for the staging of diabetic retinopathy.

In paediatric ophthalmology, the selected objective is to detect anterior segment opacities, including congenital cataracts, from infrared video-refractometry images in infants. Additionally, the aim is to estimate the onset risk of these opacities based on medical history data alone and in combination with video-refractometry images.

To achieve these goals data from over 200 heart failure patients, over 100 diabetic patients and over 80 infants, will be collected for each of the applications mentioned above. Since the approval of the ethical committee and signing of partners agreements, informed consent has been collected for all three clinical domains and preliminary tests of data extractions have been performed. Extracted data have been used to perform preliminary data exploration to assess data quality and to start developing processing pipelines.

Project scope definition: data collection and dataset description

Clinical variables potentially relevant for the identified outcomes are selected by the clinical investigators based on two criteria: i) evidence in the published literature and ii) availability of data collected through standard routine practice or ad-hoc devices. Selected variables are acquired from different sources and include multiple modalities, as reported in Table 2.

For cardiology, they include demographics and clinical data from Electronic Health Records (EHRs), self-reported data collected through the TreC+ mobile app within a Patient-Generated Health Data (PGHD)-approach, data from implantable devices, smartwatches and wearable devices certified as medical devices.

For diabetic retinopathy data include fundus images together with demographics and longitudinal clinical variables from EHRs and lifestyle habits self-reported through a questionnaire.

For paediatric ophthalmology collected data include images from infrared (IR)-reflex assessment together with anamnestic information reported by paediatricians.

A detailed description of the collected data, together with references to published evidence of their relevance can be found in Supplementary Tables 1–3.

Table 2
Data description. For each clinical area, the different data modalities are reported together with some examples of collected variables for each data type. For a detailed description of variables and data sources, see Supplementary Tables 1–3. EHRs: electronic health records, MD: medical device, ECG: electrocardiogram, BMI: body mass index, IR-reflex: infrared-reflex.

Clinical area	Data modality/data source	Example data
Cardiology	Demographics	Age, sex
	Clinical variables from EHRs	Medications
	Self-reported data from mobile app	Lifestyle, symptoms, adherence to medications
	Data from implantable devices	ECG signals
	Data from smartwatch	Steps, heart rate
	Data from wearables certified MD	Oxygen saturation
Diabetic retinopathy	Images	Fundus oculi images
	Demographics	Age, sex, ethnicity
	Clinical variables from EHR	BMI, laboratory tests
	Self-reported questionnaire	Physical activity, smoking habits
Pediatric ophthalmology	Images	IR-reflex images
	Demographics	Age, sex
	Anamnestic information reported by paediatricians	Birth conditions, family medical history

Project scope definition: data management and data platform

For each clinical endpoint, data and corresponding sources were identified from various repositories, including the Personal Health Record (PHR), medical imaging archives, the provincial Citizen Clinical Records platform (TreC+), and PGHD, possibly collected via mobile devices (e.g. via TreC+ - mobile version). Variable selection and data management procedures adhere to the core principle of privacy by design [12].

All data is pseudonymized and made accessible to authorized researchers through a dedicated gateway specifically designed for research purposes, known as the "Data Platform" (Fig. 2). The current components of the Data Platform support the handling of data flows through an ETL (Extract-Transform-Load) process (Fig. 2). This setup allows research entities to access high-value data while fully complying with data protection regulations and adhering to the principles of “privacy by default” and “privacy by design”.

Decoupling the data source from the recipient provides the necessary abstraction to make data usable by researchers as if derived from a single large repository based on data managed by APSS. From a technical process perspective, APSS initially accesses the data source through diverse methods, ranging from reading one or more tables on a database (previously processed by the source system) to reading files generated by source applications, to reading information through API sources. After the Extraction phase, data retrieved from the sources is transformed, normalized and stored in the core module of the data platform. This process consists in separating the "transaction/event" data from the personal/demographic information, which undergoes pseudonymisation. Where possible, the system also encodes data to ensure field standardization and the use of data from controlled dictionaries and corporate registries. The subsequent phase consists in the structured packaging of data for each individual research project: this entails the aggregation of the required information into one or more pseudonymized, normalized, and neutralized tables, specifically prepared for research use.

The final phase pertains to the controlled delivery of data to the designated research center: this is achieved through a dedicated component of the Data Gateway, which ensures access to the data “buckets” within a predefined time window. All access and data consultation activities are systematically logged and monitored.

Project scope definition: ethical aspects and data protection

To participate in the scientific projects, participants receive proper information on the study and must provide written informed consent. Information sheets clarify that no direct use of the algorithms is performed on the individuals to support clinical decisions and that AI models are developed solely for research purposes.

In terms of data protection compliance, the project has considered both the rules of the GDPR and of national Legislative Decree No 196/2003 - Italian Personal Data Protection Code. Detailed privacy policies and consent forms have been created according to Articles 12–14 of the GDPR to ensure informed, voluntary, and specific consent for data collection. The APSS and FBK serve as joint data controllers. Under Art.

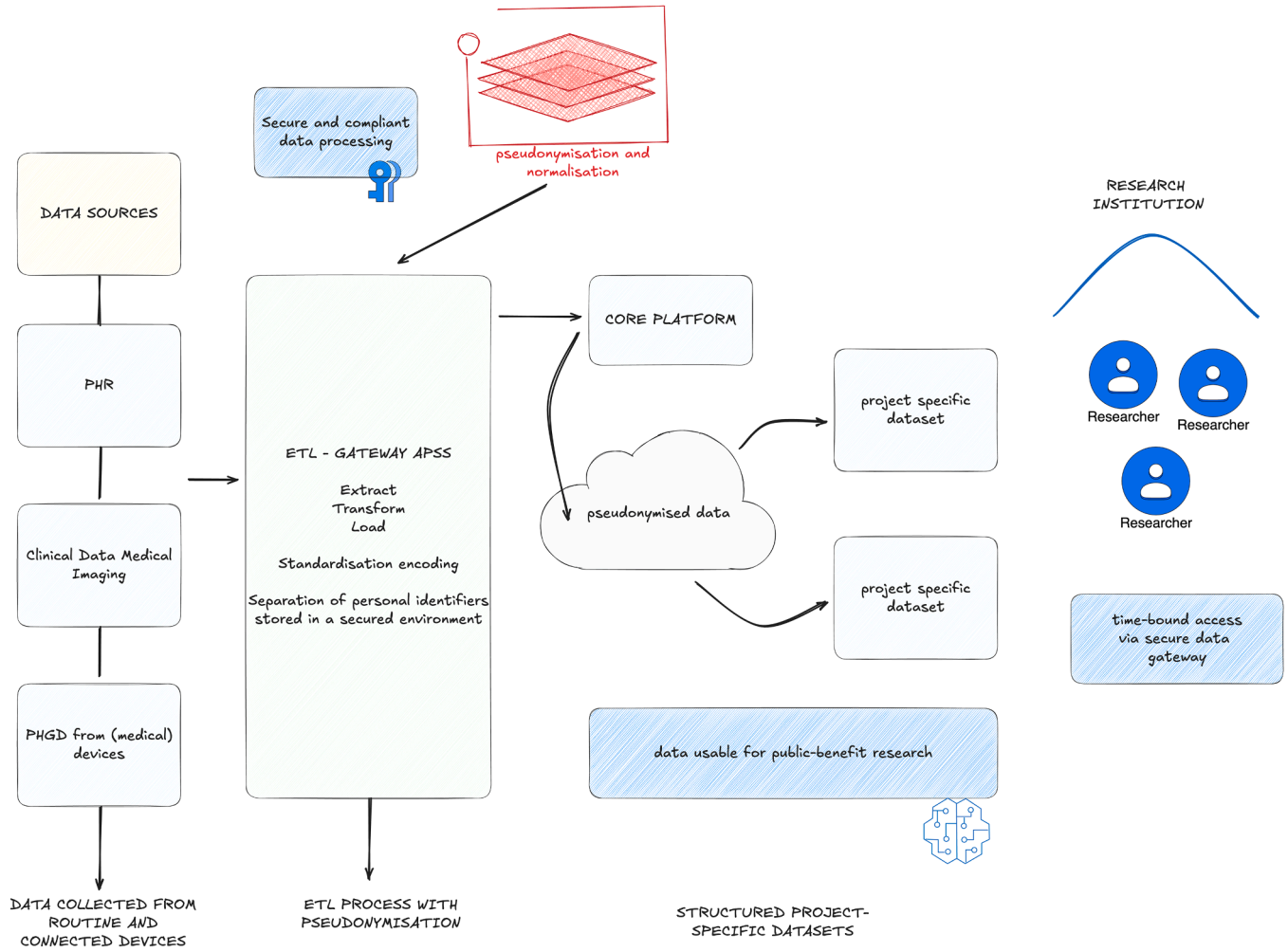

Fig. 2. caption: Data sources and data platform. Schematic representation of the adopted data sharing solution.

26 of the GDPR, joint controllers conclude an agreement to determine their responsibilities for compliance, as regards the exercising of the data subjects' rights and the obligations to provide information. Data processing agreements are established in accordance with Article 28 of the GDPR. A DPIA has been carried out under Article 35 of the GDPR to identify data protection risks and determine the necessary technical and organizational measures to safeguard the collected personal data. Pseudonymization techniques have been applied to separate unique identifiers from medical information before data delivery DPOs of the controllers were involved in drafting all documents and the DPIA. All collected datasets remain confidential and secure, accessible only by authorized personnel, and will not be disclosed or disseminated, except in anonymized or aggregated form for publication purposes.

Data preparation

The wealth of data from healthcare digitalization offers new opportunities. It allows researchers to better characterize patients' disease trajectories by harnessing different data types and capturing different aspects of their health status. While different data types (images, tabular data from EHRs, electrocardiogram (ECG) time series, etc.) are often processed according to individual and specific workflows, generally all data undergoes an initial preparatory step before being further parsed by either unimodal or multimodal models. The data preparation phase encompasses data preprocessing and, whenever possible, data visualisation as fundamental steps to check and improve data quality and to avoid the development of biased AI solutions [13]. Examples of standard preprocessing techniques include cropping, masking and resizing for images [14], missing values imputation, distribution visualization and standardization for tabular data and resampling, normalisation, band-pass filtering, and the generation of spectrograms from wavelet transformations for waveforms [15].

Modeling

Unimodal analyses will independently consider individual data modalities, adopting suitable model architectures for each data type. Tabular data will be handled with shallow learning methods, which are well established for this data type and might be more suitable in case of reduced numerosity [16]. For image-related tasks, more advanced Deep Learning (DL) approaches will be considered, which have proven to be successful on images in a variety of domains but are also challenged by the issue of data availability. To deal with possible data scarcity, data augmentation will be performed and, if convenient and feasible, data will be integrated with i) public datasets [17], for which descriptions and statistics of available demographic data and features of interest will be reported, and ii) synthetic data [18]; to assess the reliability of the generated data, evaluations on whether relationships between targets and inputs observed in real data are maintained in synthetic data will be carried out, and more generally explainability techniques will be adopted to assess whether the algorithms exhibit biases potentially present in the datasets [19,20]. Regarding time series of physiological signals such as ECG data, both shallow learning and DL methods will be investigated. The main difference between these two approaches is that

shallow learning methods require the "a priori" extraction of time series features (e.g. features describing ECG waveform morphology, or the duration of its constituent waves [21]), while DL approaches can identify predictive information from the time series themselves, or their spectrograms [22], without the need of deriving "hand-crafted" features. However, these latter approaches potentially require larger cohorts and might benefit from the integration of collected data with additional synthetic and/or public data.

Motivated by the need to combine information from multiple domains and scales to potentially capture complex interplays between pathological processes and different body districts, multimodal models will consider different data sources (i.e. images and tabular data) to inform decisions [23].

Models' optimization and evaluation

Shallow learning and DL models might present some pitfalls, among which overfitting. To minimise the risk of overfitting, rigorous data handling will be observed, dividing data in a training and a test set and assigning all samples from the same subject either to the training or to the test set. The training set will be used to define the preprocessing pipeline and fit the models with a cross-validation approach. The model reaching the best cross-validated performance on the training set will be then applied to the hold out test set, never used during training and model optimization. Since data might be unbalanced with respect to the target variable, the performance on the test set will be assessed using multiple validation metrics, such as balanced accuracy, area under the Receiver Operating Characteristic curve, Matthews Correlation Coefficient [24] and F1 score, to ensure a comprehensive evaluation.

Particular attention will be paid to the interpretability and explainability of the developed solutions, favouring the adoption of transparent models and explainability techniques (e.g. Shapley Additive exPlanations method [25] for tabular data or saliency heatmaps [26] for images). The model output will not be limited to risk factors but also to the identification of how the most relevant features affect model responses [27,28].

Results

Research protocols were developed and submitted to the Ethics Committee in the Autonomous Province of Trento in November 2023, with approvals received in March 2024.

So far, a major outcome of the ongoing "Digital Health and Artificial Intelligence" project has been the establishment of an interdisciplinary, public-driven research framework that structurally integrates clinical and mathematical expertise. This collaborative model involves healthcare professionals, data scientists, legal experts, and technical personnel. This synergy enables the co-development of clinically relevant research questions and the design of AI-based predictive solutions aligned with actual healthcare needs. This strategic approach has allowed for the design of a unified methodological architecture, supporting a replicable model for designing, developing and implementing future AI-driven health research.

A second critical achievement lies in the development of a dedicated research gateway, embedded within the provincial Data Platform, designed to enable compliant, secure, and structured access to health data for scientific purposes. This gateway ensures alignment with the GDPR (particularly Articles 5, 6, 9, and 35), by integrating core principles such as data minimization, purpose limitation, privacy by design and by default. The pseudonymization of personal data and the abstraction of source repositories (via ETL framework) ensure a clear separation between operational and research domains. This enhances both legal compliance and data usability. The gateway currently facilitates controlled access to various types of data, including EHRs, medical imaging, PGHD, and structured inputs from mobile devices (via the TreC+ platform). The successful implementation and operational status of this data platform serves as a primary measurable indicator of progress, fulfilling the project's secondary objective to validate a cross-institutional framework for AI research.

The dedicated DPIA could also be identified as a relevant result of the research so far, particularly when considering the multi-clinical approach adopted in the study targeting three different clinical domains. The datasets, the related protocols and DPIA have been designed based on two criteria, namely the availability of routinely collected data and supporting evidence from existing clinical literature. The infrastructure and methodological innovations from this phase are intended to serve as a scalable model for future AI research within healthcare systems, particularly in settings with integrated public health governance.

At this stage, data collection is ongoing and initial results of predictive modeling are expected by the end of 2025. However, interim data (end of August 2025) are available for the diabetic retinopathy and paediatric ophthalmology research lines and hereafter reported, whereas data collection for the cardiology domain is ongoing and will be presented in future dedicated publications.

For the diabetic retinopathy domain, informed consent and data were collected from $n = 70$ patients. Self-reported questionnaires were used to explore the distribution of demographics and lifestyle habits, revealing the prevalence of Caucasian ethnicity, non-smokers and physically active patients (Table 3). Data exploration of EHRs variables revealed the need of multiple cleaning steps, including harmonization of laboratory tests names and unit of measurement and harmonization of comorbidities descriptions, which had been manually inserted by clinicians in EHRs. After data cleaning, laboratory test results were averaged on a yearly basis and exploratory data analysis was performed, revealing multiple variables with a high proportion of missing values (Fig. 3A). Most populated variables included information about blood glucose, glycated haemoglobin, cholesterol, blood pressure, and weight which are considered as relevant predictors for diabetic retinopathy [29]. Harmonized variables distributions were further visualized to inspect the presence of outliers (Fig. 3B-D).

For paediatric ophthalmology, data were collected from $n = 185$ patients. The available IR-reflex images ($n = 244$) were used to develop an automated pipeline for the identification of pupils (Fig. 3E), consisting in image equalization, contours detection and identification of pupils as circular shapes in the contours image. By visual inspection, the pipeline has a success rate of 96.92 %, representing a key measurable indicator of technical progress toward the diagnostic support objective.

Table 3
Descriptive statistics for the diabetic retinopathy cohort. Demographics and lifestyle habits for patients from the diabetic retinopathy cohort, self-reported through a questionnaire.

Variable	Class	N (%)
Sex	Female	33 (44 %)
	Male	42 (56 %)
Ethnicity	Arab/Middle Eastern	5 (6.67 %)
	Caucasian	68 (90.67 %)
	Hispanic/Latino	2 (2.66 %)
Smoke	Never	41 (54.67 %)
	No (from <6 months)	1 (1.33 %)
	No (from >6 months)	24 (32 %)
	Yes	9 (12 %)
Physical activity	Never	6 (8 %)
	Daily	26 (34.67 %)
	2–3 times a week	29 (38.67 %)
	2–3 times a month	14 (18.66 %)
	Mean ± *std*	*[min; max]*
Age	67.36 ± 13.70	[32; 89]
Weight	74.15 ± 18.29	[0; 143]
Years of disease	11.82 ± 10.43	[0; 50]

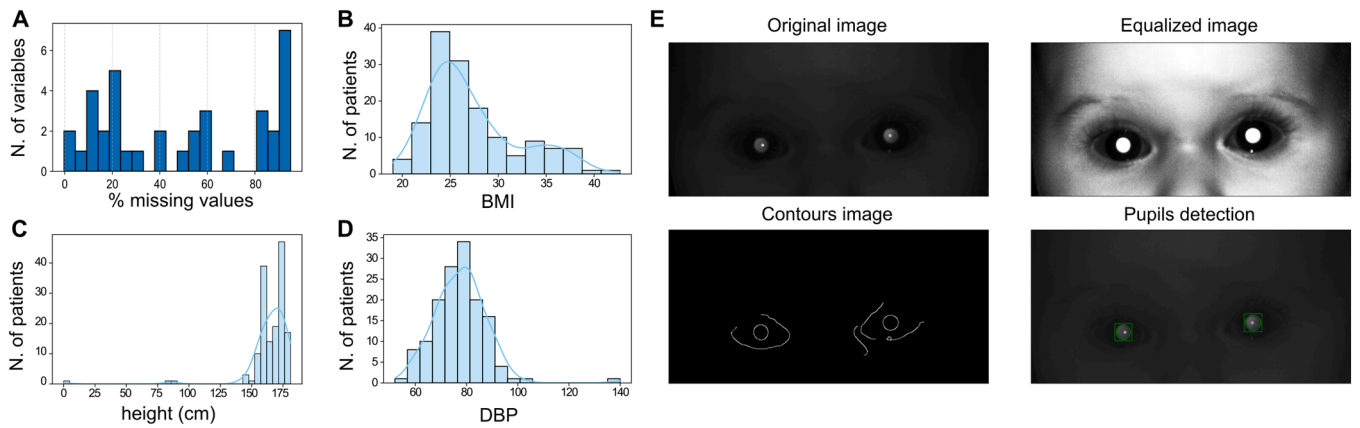

Fig. 3. caption. Preliminary data processing. A. Missing values proportions for laboratory tests data collected for the diabetic retinopathy domain from an initial batch of $n = 70$ patients. B-D. Example distributions of some laboratory tests variables. E. Automated pipeline for pupils detection on IR-reflex images. The original image (top-left) is equalized (top-right) and contours are detected (bottom-left). Circular elements in the contour image are detected as pupils (bottom-right).

Discussion

The "Digital Health and Artificial Intelligence" project aims to develop predictive models for research, focusing on both technical innovation and scientific advancements. It aligns with the growing body of evidence supporting AI potential to enable personalized healthcare across diverse clinical domains including cardiology [30–32], diabetic retinopathy [29,33] and paediatric ophthalmology [14,34,35]. Additionally, the project seeks to validate a comprehensive AI research framework covering technological, procedural, and privacy aspects, as highlighted in the current scientific literature [1,2,19,27]. This endeavour represents an initial step towards formalising research collaborations among provincial entities, with a strong focus on the active involvement of citizens, clinicians, and researchers. The entire initiative promotes research based on citizen data, hand in hand with a robust foundation for advancing healthcare outcomes through innovative GDPR-compliant and ethically grounded AI-driven approaches.

Within the European context, the project is aligned with the EU's strategic vision for digital transformation, as envisioned by the European Health Data Space Regulation and promoted through the establishment of the European Digital Innovation Hubs Network [36]. While our project focuses on specific clinical domains, its underlying philosophy echoes larger-scale national strategies, such as Finland's AuroraAI [37], which also aims to create a sustainable and ethical ecosystem for AI-driven public services.

Already from the preparatory and initial project phases, we might identify some areas of strengths and weaknesses for this initiative. In terms of strengths, the involvement and collaboration of multi professional working groups ensures a comprehensive and coordinated approach. In particular, the involvement of clinical specialists endorses a patient-centered approach. Furthermore, the partnership among the Autonomous Province of Trento, FBK, APSS, the University of Trento and TS4.0 ensures the project is well-supported and aligned with provincial healthcare priorities. An additional strand of advantage is represented by the advanced IT infrastructure and the strong public-driven collaborations: the province's robust IT infrastructure and history of effective public-private partnerships provide a conducive environment for AI research in healthcare.

In terms of weaknesses, the project faces challenges due to complex administrative processes, data protection compliance, and time-consuming activities. Patient enrolment and data management require meticulous planning. Furthermore, the initiative's broad scope across three clinical areas demands significant resource allocation, both in personnel and technological infrastructure. These complexities have been tackled through close collaboration between project managers, administrative staff and privacy experts. This joint effort resulted in the definition of a formal procedure that will serve as a blueprint for future innovative initiatives. Further limitations emerged in the preliminary project's phase: limited interoperability across data source and non-standardized data format; data sparsity in specific clinical subsets due to the rarity of some clinical conditions but also due to difficulties in recruiting patients due to timely rigid organizational models. Such difficulties underscore the relevance of initiatives designed to make health records available for research, such as the one implemented by the Andalusian Public Health System [38]. While that initiative focuses on broader data availability, our project's federated, project-specific data gateway represents a complementary model that allows for rapid, compliant implementation within the current legal framework, addressing immediate research needs. Additionally, the collected datasets are limited to a small geographical area, resulting in imbalanced population statistics and disease distributions. This raises concerns about potential algorithmic bias. Addressing such risks requires careful monitoring and mitigation strategies. In future phases, enhancing the generalizability of findings will be essential, for example by expanding collected datasets or integrating them with external data sources.

Overall, the initial phase of this ambitious project already provides some early insights. The need for strong coordination mechanisms among institutional partners is essential to avoid inefficiencies and to support administrative procedure. Organizational models allowing effective use of innovative systems in clinical practice are as relevant as a good research plan. The early involvement of privacy experts was crucial in shaping the project's documentation allowing smoother alignment with GDPR. Lastly, the constant dialogue between researchers, developers and clinicians was essential to aligning innovation needs with clinical expectation.

In conclusion, the "Digital Health and Artificial Intelligence" project aims to investigate AI's potential to improve health outcomes and establish a sustainable, personalized healthcare system in Trento. By leveraging multi-professional expertise and public-private partnerships, the project underscores ethical standards, regulatory compliance, and collaboration, paving the way for future advancements after clinical investigations. These early reflections may be useful for other Institutions or regions planning to develop AI-based healthcare innovation within similar governance and regulatory frameworks.

Ethical issues and patient consent

Authors confirm that authorization signed by the patient has been obtained in compliance with the current laws regarding patient authorizations relating to the use or disclosure of protected health information of the jurisdiction(s) to which the patient and the physician are subject. Informed consent and research protocols have been approved by the

local ethics committee with the following titles and references: Cardiology / Artificial Intelligence Systems for Detecting Health Trajectories in Patients with Heart Failure / SMART-SC 2023 (Reg A959; approved 20 March 2024, N.3/2024); Diabetic Retinopathy / Personalized Screening Interval for Diabetic Retinopathy Detection - INSPIRED (Reg A958); Early Detection of Anterior Segment Opacities in Infants / InPOSA (Reg A957).

Authors' roles

Mo.Mo., L.N., R.B., M.D.G., M.Ma., Mi.Mo., S.I., F.R., E.R., A.C., L.F., R.P., M.B., L.P., G.M., L.G., D.Co., A.N., G.J. conceived the project design and research questions. R.B., M.D.G., M.Ma., Mi.Mo., S.I., F.R., E.R., T.E. M., V.F., A.C., L.F., R.P., M.B., L.P. recruited patients and collected data. A.M,. D.Ca., E.S., S.C. developed the secure data platform. Mo.Mo. performed data analysis. Mo.Mo., G.M., S.I., F.R., E.R., A.C., L.F., R.P., M.B., L.P. contributed to the interpretation of the results. G.B., G.M., L. G. ensured compliance of the project with privacy and data protection regulations. G.M., L.G., A.N. coordinated the project. Mo.Mo., L.N., G. M., L.G. G.B., wrote the manuscript. All authors revised the final version of the manuscript.

Funding

This research was funded by the Autonomous Province of Trento. The "Digital Health and Artificial Intelligence" project, a 24-months initiative approved and funded by Provincial Government Resolution No. 2475 on December 22, 2022, aims to enhance healthcare accessibility and foster innovative healthcare models through the use of technology and AI.

Acknowledgments

The authors would like to thank all the colleagues involved in the preparatory stages of this very initiative from the different participating institutions: Provincia Autonoma di Trento, Azienda Provinciale per i Servizi Sanitari, Fondazione Bruno Kessler, Universita di Trento.

Data availability

Data collection is currently ongoing and analyses presented in the current study are preliminary.

Declaration of competing interest

Authors declared no conflict of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.hlpt.2025.101149.

References

[1] MdA Rahman, E Victoros, Ernest J, Davis R, Shanjana Y, MdR Islam. Impact of artificial intelligence (AI) technology in healthcare sector: a critical evaluation of both sides of the coin. Clin Med Insights Pathol 2024;17. 2632010X241226887.
[2] Murphy K, Di Ruggiero E, Upshur R, et al. Artificial intelligence for good health: a scoping review of the ethics literature. BMC Med Ethics 2021;22:14.
[3] Golubnitschaja O, Topolcan O, Kucera R, et al. 10th Anniversary of the European association for predictive, preventive and personalised (3P) medicine - EPMA World Congress Supplement 2020. EPMA J 2020;11:1–133.
[4] Kubin A-M, Huhtinen P, Ohtonen P, Keskitalo A, Wirkkala J, Hautala N. Comparison of 21 artificial intelligence algorithms in automated diabetic retinopathy screening using handheld fundus camera. Ann Med 2024;56:2352018.
[5] Kim M, Kang D, Kim MS, et al. Acute myocardial infarction prognosis prediction with reliable and interpretable artificial intelligence system. J Am Med Inform Assoc 2024;31:1540–50.
[6] Malaguti MC, Gios L, Giometto B, et al. Artificial intelligence of imaging and clinical neurological data for predictive, preventive and personalized (P3) medicine for Parkinson Disease: the NeuroArtP3 protocol for a multi-center research study. PLoS ONE 2024;19:e0300127.
[7] Meszaros J, Minari J, Huys I. The future regulation of artificial intelligence systems in healthcare services and medical research in the European Union. Front Genet 2022;13:927721.
[8] Prakash S, Balaji JN, Joshi A, Surapaneni KM. Ethical conundrums in the application of artificial intelligence (AI) in healthcare—a scoping review of reviews. JPM 2022;12:1914.
[9] Eccher C, Gios L, Zanutto A, Bizzarri G, Conforti D, Forti S. TreC platform. An integrated and evolving care model for patients' empowerment and data repository. J Biomed Inf 2020;102:103359.
[10] Autonomous Province of Trento (2022, December 22). Provincial Government Resolution No. 2475, Sanità digitale e intelligenza artificiale - Strumenti per avvicinare il Servizio sanitario ai cittadini e per lo sviluppo del "sistema provinciale". Retrieved from https://delibere.provincia.tn.it.
[11] Niakšu O. CRISP data mining methodology extension for medical domain. Balt J Mod Comput 2015.
[12] Bincoletto G. Data protection by design in the E-Health care sector: theoretical and applied perspectives. Nomos Verlagsgesellschaft mbH & Co. KG; 2021. https://doi.org/10.5771/9783748929895.
[13] Volovici V, Syn NL, Ercole A, Zhao JJ, Liu N. Steps to avoid overuse and misuse of machine learning in clinical research. Nat Med 2022;28:1996–9.
[14] Liu X, Jiang J, Zhang K, et al. Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network. PLoS ONE 2017;12:e0168606.
[15] Safdar MF, Nowak RM, Pałka P. Pre-processing techniques and artificial intelligence algorithms for electrocardiogram (ECG) signals analysis: a comprehensive review. Comput Biol Med 2024;170:107908.
[16] Grinsztajn L., Oyallon E., Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? 2022. URL http://arxiv.org/abs/2207.08815 Accessed 2 November 2022.
[17] Farahat Z, Zrira N, Souissi N, et al. Diabetic retinopathy screening through artificial intelligence algorithms: a systematic review. Surv Ophthalmol 2024. S0039625724000511.
[18] Coyner AS, Chen JS, Chang K, et al. Synthetic medical images for robust, privacy-preserving training of artificial intelligence. Ophthalmol Sci 2022;2:100126.
[19] Giuffrè M, Shung DL. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. Npj Digit Med 2023;6:186.
[20] Lenatti M, Paglialonga A, Orani V, Ferretti M, Mongelli M. Characterization of synthetic health data using rule-based artificial intelligence models. IEEE J Biomed Health Inf 2023;27:3760–9.
[21] Tseng AS, Noseworthy PA. Prediction of atrial fibrillation using machine learning: a review. Front Physiol 2021;12:752317.
[22] Wagner P, Mehari T, Haverkamp W, Strodthoff N. Explaining deep learning for ECG analysis: building blocks for auditing and knowledge discovery. Comput Biol Med 2024;176:108525.
[23] Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med 2022;28:1773–84.
[24] Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom 2020; 21:6.
[25] Lundberg S., Lee S.-I. A unified approach to interpreting model predictions. 2017. doi:10.48550/ARXIV.1705.07874.
[26] Bora A, Balasubramanian S, Babenko B, et al. Predicting the risk of developing diabetic retinopathy using deep learning. Lancet Digit Health 2021;3:e10–9.
[27] Holzinger A., Biemann C., Pattichis C.S., Kell D.B. What do we need to build explainable AI systems for the medical domain? 2017. URL http://arxiv.org/abs/1712.09923 Accessed 31 July 2024.
[28] Goebel R, Chander A, Holzinger K, et al. Explainable AI: the new 42? In: Holzinger A, Kieseberg P, Tjoa AM, Weippl E, editors. editors *machine learning and knowledge extraction*. Cham: Springer International Publishing; 2018. p. 295–303.
[29] Aspelund T, Þórisdóttir Ó, Ólafsdottir E, et al. Individual risk assessment and information technology to optimise screening frequency for diabetic retinopathy. Diabetologia 2011;54:2525–32.
[30] Pieszko K, Shanbhag AD, Singh A, et al. Time and event-specific deep learning for personalized risk assessment after cardiac perfusion imaging. Npj Digit Med 2023; 6:78.
[31] Inan O.T., Baran Pouyan M., Javaid A.Q., et al. Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients. Circ: Heart Failure 2018;11:e004313.
[32] Selder JL, Kolste HJT, Twisk J, Schijven M, Gielen W, Allaart CP. Accuracy of a standalone atrial fibrillation detection algorithm added to a popular wristband and smartwatch: prospective diagnostic accuracy study. J Med Internet Res 2023.
[33] the ISDR Study Group, Broadbent D.M., Wang A., et al. Safety and cost-effectiveness of individualised screening for diabetic retinopathy: the ISDR open-label, equivalence RCT. *Diabetologia* 2021;64:56–69.
[34] Long E, Lin H, Liu Z, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nat Biomed Eng 2017;1:0024.
[35] Lin H, Li R, Liu Z, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. EClinicalMedicine 2019;9:52–9.

[36] Bagić M, Pintarić K, Ivanko P, Keber M. Organizing an AI innovation challenge within the European digital innovation hub in Croatia, 2025. Eur J Public Health 2025;35. ckaf161.829.
[37] Leikas J, Johri A, Latvanen M, Wessberg N, Hahto A. Governing ethical AI transformation: a case study of AuroraAI. Front Artif Intell 2022;5:836557.
[38] Carriazo AM, Alonso-Trujillo F, Vázquez-Granado FJ, Túnez I, Del Moral-Leal ML. Innovation ecosystems in health and care: the Andalusian Reference Site as an example. Front Med 2024;11:1482830.