ITUPublications

International Telecommunication Union
Standardization Sector

ITU-T Focus Group Deliverable

(03/2023)

Focus Group on Artificial Intelligence for Health (FG-AI4H)

FG-AI4H DEL3

AI4H requirement specifications

ITU-T FG-AI4H Deliverable DEL3

AI4H requirement specifications

Summary

This document defines the lifecycle-based system requirement specifications (SyRS) that explain the informational, functional, behavioural and operational aspects of a generic artificial intelligence (AI) for health (AI4H) system.

SyRS serves as the basis for the system design, system verification and validation plans and procedures for the AI4H system.

System requirements analysis methodology follows a collaborative team-oriented approach, involving all the working groups and topic groups of FG-AI4H, to help the project team identify, control and track various requirements and changes to those requirements during the AI4H system development lifecycle.

Tables are intended to serve as checklists for configuring a basic minimal set of machine learning for health (ML4H) system/product lifecycle requirements specifications, which include the technical, the clinical, the regulatory and the ethical requirements. In the ML4H system/product testing phase, the same tables can be used to generate applicable test cases for verification of requirements specifications to support ML4H product conformity assessment procedures.

Keywords

AI for health (AI4H), health use cases, requirement specifications.

Note

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Change Log

This document contains Version 1 of the Deliverable DEL3 on "AI4H requirement specifications" approved on 16 March 2023 via the online approval process for the ITU-T Focus Group on AI for Health (FG-AI4H).

Editor: Pradeep Balachandran
Technical Consultant (Digital Health)
India
Email: pbn.tvm@gmail.com

Table of Contents

ITU-T FG-AI4H Deliverable DEL3

AI4H requirement specifications

1 Scope

The scope of the system requirement specifications (SyRS) includes a requirements model that defines the informational, functional, behavioural and operational aspects of the AI4H system under consideration.

The SyRS are configured as a quality assessment tool to support the requirements auditing and requirements traceability analysis for the ITU/WHO Focus Group on AI for health topic description documents (FG-AI4H TDDs).

The SyRS are generic in nature and shall be applicable across all domain specialties/topic groups of FG-AI4H. It may be modified, customized or extended appropriately to include the specific requirements and needs of the particular topic group under consideration.

The intended audiences of the SyRS include system analysts, system designers, system developers, system testers, product managers, quality assurance auditors/managers, etc.

The SyRS are subjected to periodic review and revision as per the requirements management process for the verification of its coverage and completeness.

Revisions to the SyRS are to follow a formal change management process defined under the quality management system (QMS) of the system/product manufacturer. Revision shall be performed in an iterative manner based on a rapid incremental delivery (agile process) model to elicit the emergent requirements of the system under consideration as AI systems continue to evolve over time to attain progressive maturity levels.

2 References

[ISO/IEC/IEEE 12207] ISO/IEC/IEEE 12207:2017, *Systems and software engineering – Software life cycle processes.*

[ISO/IEC/IEEE 15288] ISO/IEC/IEEE 15288:2015, *Systems and software engineering – System life cycle processes.*

[ISO/IEC/IEEE 29148] ISO/IEC/IEEE 29148:2018, *Systems and software engineering – Life cycle processes – Requirements engineering*.

[ISO STD 7498-2] ISO 7498-2:1989, *Information processing systems – Open Systems Interconnection — Basic Reference Model. Part 2: Security Architecture*.

[IEEE STD 830-1998] IEEE STD 830-1998, *IEEE Recommended Practice for Software Requirements Specifications.*

[FG-AI4H DEL10] ITU/WHO Focus Group on AI for Health Deliverable 10 (2023), *AI4H use cases – Topic Description Documents*.

3 Definitions

3.1 Terms defined elsewhere

This Technical Report uses the following term defined elsewhere:

3.1.1 system requirements specification [ISO/IEC/IEEE 29148]: The structured collection of the requirements [functions, performance, design constraints and other attributes] for the system and its operational environments and external interfaces.

3.2 Terms defined in this Technical Report

This Technical Report defines the following term:

3.2.1 topic description document (TDD): A document that describes the key aspects for application of AI for health for a specific health use case. They are documented in [FG-AI4H DEL10].

4 Abbreviations and acronyms

This Technical Report uses the following abbreviations and acronyms:

AI	Artificial Intelligence
AI4H	Artificial Intelligence for Health
API	Application Programming Interface
CNN	Convolutional Neural Network
DAISAM	Data and AI Solution Assessment Methods
FDA	Food and Drug Administration
FG-AI4H	Focus Group on AI for Health
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
OS	Operational System
QMS	Quality Management System
REQ-ID	Requirement Identifier
SaMD	Software-as-a-Medical Device
SI	Software Interface
SiMD	Software-in-a-Medical Device
SOP	Standard Operating Procedure
SyRS	System Requirement Specification
TDD	Topic Description Document

5 Conventions

This Technical Report shall conform to the following standard convention of specification language syntax for every requirement specifications statement to indicate its particular significance/compliance level.

Term	Meaning
"*SHALL*"	States a mandatory requirement of this policy
"*SHOULD*"	States are commended requirement of this policy
"*MAY*"	States an optional requirement

The following is the template for the requirement specification IDs (REQ-IDs) used in this Technical Report: <R><hyphen><Acronym for Requirements Type/Sub-Type><Serial Number>.

6 SyRS overview

– System requirements specifications are developed following a generic 'requirements modelling framework' defined under the quality management system (QMS) to guide the process of organizing, promising and tracing the requirements.

– System requirements specifications are broadly organized in terms of (a) Functional requirements, (b) External interface requirements, and (c) Non-functional requirements.

– SyRS is traceable to the respective AI4H lifecycle phase/stage requirement and is verifiable and testable.

– SyRS conforms to applicable regulations, laws, standards, guidelines and best practices of the AI4H domain.

7 Intended use and high-level specification

Table 1 – Intended use and high-level requirements

REQ. ID	Requirement specification	Description
R-PD1	System SHALL have specification for the intended health intervention area/use case for which the AI4H software is used	E.g., health intervention use cases – Public health o Health service o Health systems o Health expenditure o Health inequities o Health surveillance o Health emergencies o Life expectancy and mortality o Cause-specific mortality and morbidity o Communicable diseases o Non-communicable diseases o Civil registration and vital statistics o Other – Clinical health o Prevention o Screening o Diagnosis o Treatment o Research o Other – Non-clinical health o Personal care o Wellness o Education

Table 1 – Intended use and high-level requirements

REQ. ID	Requirement specification	Description
		○ Other
R-PD2	System SHALL have specification for the intended AI-benchmarking class type / AI-Task/ primary product function for which the AI4H software is used	E.g. AI-benchmarking tasks – Classification – Regression/Prediction – Clustering – Association rule learning – Decision Support / Virtual Assistance / Recommendation systems – Matching – Labelling – Detection – Segmentation – Sequential data modelling – Anomaly detection and Fraud Prevention – Compliance Monitoring / Quality Assurance – Process optimization / Automated planning and scheduling – Other
R-PD3	System SHALL have specifications for the intended use of AI4H software within the target health workflow / deployment settings	Describe how the AI4H software fits into the intended health intervention workflow E.g. as autonomous tool, assistive tool, augmentative tool, etc. – as add-on unit to existing system/workflow – as replacement unit for existing system/workflow component – as new standalone system/subsystem/device
R-PD4	System SHALL have specification for product category / type of AI4H software as released in the market	E.g., – Software-as-a-Medical Device (SaMD) – Software-as-a-Medical Service (SaMS) – Software-in-a-Medical Device (SiMD) – Mobile Medical Applications (MMA) – Medical Device Data Systems (MDDS)

Table 1 – Intended use and high-level requirements

REQ. ID	Requirement specification	Description
		– Other
R-PD5	System SHALL have specification for the operation mode of AI4H software	E.g., fully automatic, semi-automatic

8 System functions

Table 2 – Functional requirements

REQ. ID	Requirement specification	Description
R-SF1	System SHALL list the 'functional use cases' for the AI4H software	Functional use cases can be identified in terms of the main functional objectives as stated in the respective AI4H topic description document (TDD). E.g., – TDD-Cardiovascular disease risk prediction – TDD-Outbreak detection – TDD-Symptom assessment – TDD-Dental diagnostics, etc.
R-SF2	System SHALL have description for 'functional use cases 'for the AI4H software	Use case description/diagram shall include information on: – System/Subsystem services/methods – Primary and secondary actors/users – Goals of primary and secondary actors/users – Tasks/Functions performed by primary and secondary actors / users – System data/information acquired, produced or changed by primary and secondary actors/users.
R-SF3	System SHALL have specification for data elements for each functional use case	Data elements include: – data type – data unit – data representation format – data precision/accuracy

Table 2 – Functional requirements

REQ. ID	Requirement specification	Description
		– data range
R-SF4	System SHALL have description for data flow for each functional use case	Data flow description/diagram include information on: – Input/Output data validity checks – Input/Output data sequence of operations – Input/Output data conversion formulas / rules – Data Error/Exception handling and recovery – Data response time
R-SF5	System SHALL have description for process flow for each functional use case	Process flow description/diagram include information on: – Input validity check – Input (stimulus) – Process Algorithm/Formulas – Output (Response) – Process Error/Exception handling and recovery – Process response time

9 User types/classes and characteristics

Table 3 – User type/class requirements

REQ. ID	Requirement specification	Description
R-UC1	System SHALL have specification for the primary and secondary user types/classes / groups for the AI4H software	Primary user types include: – Physician, clinician, lab technician, nurse, pharmacist, domain specialist, data scientist/engineer, business/program/product manager, chief information Officer, other. Secondary user types include: – Software developers, software testers, regulatory affairs and quality managers, risk managers, usability engineers, medical device consultants,

Table 3 – User type/class requirements

REQ. ID	Requirement specification	Description
		service technicians (e.g., update, upgrade, configuration, installation, capturing audit logs, etc.), support staff, other.
R-UC2	System SHALL have specification for educational level of primary and secondary user types/classes/groups for the AI4H software	
R-UC3	System SHALL have specification for target domain experience level of primary and secondary user types/classes /groups for the AI4H software	E.g., experience with target domain, product type, process tools, technology, etc.
R-UC4	System SHALL have specification for technical expertise/ technical skill sets of primary and secondary user types/classes/groups for the AI4H software	
R-UC5	System SHALL have specification for roles of primary and secondary user types/classes/groups for the AI4H software	Primary use roles include: – Physician, clinician, lab technician, nurse, pharmacist, domain specialist, medical expert, data scientist/engineer, computer scientist, business/program/product manager, chief information officer, other. Secondary use roles include: – Software developers, software testers, regulatory affairs and quality managers, risk managers, usability engineers, medical device consultants, service technicians (e.g., update, upgrade, configuration, installation, capturing audit logs, etc.), support staff, other.
R-UC6	System SHALL have specification for system security privilege levels of primary and secondary user types/classes/groups for the AI4H software	
R-UC7	System SHALL have specification for training needed for primary and secondary user types/classes/groups for the AI4H software	

10 Operating conditions/environment

Table 4 – Operating environment requirements

REQ. ID	Requirement specification	Description
R-OC1	System SHALL have a standard operating procedure (SOP) for operations site integration of AI4H software within deployment IT infrastructure	
R-OC2	SOP for operations site integration SHALL specify the run-time environment	E.g., mobile platform, desktop, web/cloud platform, other
R-OC3	SOP for operations site integration SHALL specify the modes of operation	E.g., – programming mode – test mode – troubleshooting mode – monitoring mode – other
R-OC4	SOP for operations site integration SHALL specify the workflow/clinical protocols	
R-OC5	SOP for operations site integration SHALL describe the data processing support functions	E.g., ability to collect and analyse real-time patient data
R-OC6	SOP for operations site integration SHALL describe the backup and recovery operations	
R-OC7	SOP for operations site integration SHALL specify the hardware platform configuration and versions	
R-OC8	SOP for operations site integration SHALL specify the operating system configuration and versions	
R-OC9	SOP for operations site integration SHALL specify the operating site energy efficiency rating	
R-OC10	SOP for operations site integration SHALL specify the installation and acceptance procedure	

11 Design and implementation constraints

Table 5 – Design and implementation constraints

REQ. ID	Requirement specification	Description
R-DIC1	System SHALL list all the 'regulatory policy'/'regulatory standard' related constraints, if any	E.g., Regulatory compliance with country/region /jurisdiction specific data policies, AI geopolitical implications, etc.
R-DIC2	System SHALL list all the 'implementation platform' related constraints, if any	E.g., IT infrastructure implications
R-DIC3	System SHALL list all the 'database' related constraints, if any	
R-DIC4	System SHALL list all the 'network/communication protocol' related constraints, if any	
R-DIC5	System SHALL list all the 'hardware limitations', if any	E.g., timing requirements, memory requirements
R-DIC6	System SHALL list all the 'external interfaces' related constraints, if any	
R-DIC7	System SHALL list all the 'safety and security 'related constraints, if any	
R-DIC8	System SHALL list all the 'cost 'related constraints, if any	
R-DIC9	System SHALL list all the 'accounting and auditing procedures ' related constraints, if any	E.g., Design considerations related to AI transparency, AI trustworthiness, etc.
R-DIC10	System SHALL list all the 'data sharing / replication policy ' related constraints, if any	E.g., patient consent (GDPR), AI gender and race representation, data privacy, trust, ethical and legal considerations, data ownership, data custodianship, data retention policy, etc.
R-DIC11	System SHALL list all the 'technical accuracy to clinical effectiveness mapping' related constraints, if any	E.g., Interpretable AI constraints, explainable AI constraints, algorithmic risk assessment implications
R-DIC12	System SHALL list all the 'business model sustainability ' related constraints, if any	
R-DIC13	System SHALL list, the 'areas of stakeholder conflict', if any	

Table 5 – Design and implementation constraints

REQ. ID	Requirement specification	Description
R-DIC14	System SHALL list all the 'internationalization and/or localization needs' related constraints, if any	
R-DIC15	System SHALL list all the 'specific technologies and/or tools ' to be used in the case of AI4H software	
R-DIC16	System SHALL list all the 'non-clinical data availability 'constraints, if any	E.g., availability of behavioural data, environmental data, patient reported data

12 System interface requirements

Table 6 – System interface requirements

User interface (UI) requirements specification		
REQ. ID	**Requirement specification**	**Description**
R-UI1	System SHALL have description for User Interface (UI)	E.g., GUI features and formats
R-UI2	System SHALL have specification for UI input and output valid range	
R-UI3	System SHALL have specification for UI input and output accuracy	
R-UI4	System SHALL have specification for UI input and output tolerance	
R-UI5	System SHALL have specification for UI input and output - units of measure	
R-UI6	System SHALL have specification for UI input and output timing	
R-UI7	System SHALL describe the UI relationships to other inputs/outputs	E.g., Source of input or destination of output

Table 6 – System interface requirements

User interface (UI) requirements specification		
REQ. ID	**Requirement specification**	**Description**
R-UI8	System SHALL list the UI screen formats/window layout constraints	
R-UI9	System SHALL have specification for UI data formats	
R-UI10	System SHALL have specification for UI command format	
R-UI11	System SHALL define the standard UI widget elements and functions	
R-UI12	System SHALL define the UI standards/style guides	
R-UI13	System SHALL define UI keyboard shortcuts	E.g., programmable function keys
R-UI14	System SHALL define the UI error message display standards	
R-HI1	System SHALL have description for Hardware Interface (HI)	
R-HI2	System SHALL specify HI supported device types	
R-HI3	System SHALL specify the HI source of input	
R-HI4	System SHALL specify the HI destination of output	
R-HI5	System SHALL specify the HI data types	
R-HI6	System SHALL specify the HI control protocols	
R-HI7	System SHALL specify the HI communication protocols	
R-SI1	System SHALL have description for software interface (SI)	
R-SI2	System SHALL specify the SI source of input	
R-SI3	System SHALL specify the SI destination of output	

Table 6 – System interface requirements

User interface (UI) requirements specification		
REQ. ID	**Requirement specification**	**Description**
R-SI4	System SHALL specify the SI input and output data items valid range	
R-SI5	System SHALL specify the SI input and output data items accuracy	
R-SI6	System SHALL specify the SI input and output data items tolerance	
R-SI7	System SHALL specify the SI input and output data items - units of measure	
R-SI8	System SHALL specify the SI input and output data items timing	
R-SI9	System SHALL specify the SI operating systems used	
R-SI10	System SHALL specify the SI tools and libraries used	
R-SI11	System SHALL specify the SI third-party/commercial components	
R-SI12	System SHALL specify the SI services offered	
R-SI13	System SHALL specify the SI communication protocols	
R-SI14	System SHALL specify the SI application programming interface (API) protocols	
R-SI15	System SHALL specify the SI data sharing mechanism	
R-DI1	System SHALL have description for Data Interface (DI)	
R-DI2	System SHALL have specification for the database	E.g., – data schema/ structure – data entities and their relationships

Table 6 – System interface requirements

User interface (UI) requirements specification		
REQ. ID	**Requirement specification**	**Description**
		– accessing capabilities – data integrity constraints – data retention requirements
R-CI1	System SHALL have specification for the web browser used	
R-CI2	System SHALL define network server communications protocols	
R-CI3	System SHALL define the communication standards	
R-CI4	System SHALL have specification for the communication security/encryption mechanisms	
R-CI5	System SHALL have specification for the data transfer rates and synchronization mechanisms	
R-MI5	System SHALL have specification for the primary and secondary memory configurations/limits	

13 Non-functional requirements

Table 7 – Non-functional requirements

Performance requirements specification		
REQ. ID	**Requirement specification**	**Description**
R-PER1	System SHALL have specification for the static performance parameters	
R-PER2	System SHALL have specification for the dynamic performance parameters	E.g., amount of data to be processed within specified time periods for – normal workload conditions – peak workload conditions

Table 7 – Non-functional requirements

Performance requirements specification		
REQ. ID	**Requirement specification**	**Description**
R-PER3	System SHALL have specification for the number of terminals to be supported	
R-PER4	System SHALL have specification for the number of concurrent users to be supported	
R-PER5	System SHALL have specification for the amount of information to be handled	
R-PER6	System SHALL have specification for the type of information to be handled	
R-PER7	System SHALL have specification for the AI algorithmic performance on data types-images, videos, text and natural language	
R-PER8	System SHALL have specification for the AI computational efficiency: accuracy-computational cost trade-offs	
R-PER9	System SHALL have specification for the accuracy standards /acceptable algorithm accuracy rates based on use cases /domain specialization	E.g., – The system shall have a sensitivity of 97% – The system must be able predict the strength of the plaques in the blood to 0.2 mm, etc.
R-PER10	System SHALL have specification for the metrics for continuous improvement	E.g., – workflow impact – patient safety impact – care quality impact – provider/patient satisfaction impact
R-SAF1	System SHALL have specification for the applicable safety standards	
R-SAF2	System SHALL have specification for the applicable safety certifications	

Table 7 – Non-functional requirements

Performance requirements specification		
REQ. ID	**Requirement specification**	**Description**
R-SAF3	System SHALL have specification for the protocols for safety alarms, safety alerts	
R-SEC1	System SHALL define the data vulnerability classification procedure	
R-SEC2	System SHALL define the data validation techniques	
R-SEC3	System SHALL define the data encryption/decryption techniques	E.g., state-of-the-art cryptographic techniques
R-SEC4	System SHALL define the data integrity verification schemes	E.g., Checking data integrity for critical variables
R-SEC5	System SHALL define the user authentication schemes	E.g., Multi-factorial User Authentication
R-SEC6	System SHALL define the user data privacy certifications	E.g., measures adopted to ensure compliance with existing data privacy and management best practices and regulations
R-SEC7	System SHALL define the data access control functions	E.g., authentication, authorization, monitoring logging and auditing of health data registries/repositories
R-SEC8	System SHALL define the audit logs	E.g., for viewing, creation, modification, validation, copying, import, export, transmission, reception, etc
R-SEC9	System SHALL define the data persistence/storage schemes	E.g., safe and secure data storage measures used, data repository compliance with applicable laws
R-SEC10	System SHALL define methods on how to identify and assess cyber vulnerabilities and threats	E.g., compliance with cyber security standards and guidelines such as – AAMI Technical Information Report 57, "*Principles for Medical Device Security – Risk Management*" – OECD Guidelines for the Security of Information Systems and Networks – etc.
R-QTY1	System SHALL define 'reliability' measures and metrics of AI4H software	

Table 7 – Non-functional requirements

Performance requirements specification		
REQ. ID	**Requirement specification**	**Description**
R-QTY2	System SHALL define 'availability' measures and metrics of AI4H software	
R-QTY3	System SHALL define 'adaptability' measures and metrics of AI4H software	E.g., How the AI solution can be generalised to desired range of population with particular consideration of particular class of people (covering diverse backgrounds, cultures and disciplines, etc.)
R-QTY4	System SHALL define 'accountability ' measures and metrics of AI4H software	E.g., – accounting formats and procedures, auditing formats and procedures for different role-based responsibilities – accountable governance practices in compliance with ethical standards
R-QTY5	System SHALL define 'accuracy ' measures and metrics of AI4H software	
R-QTY6	System SHALL define 'flexibility ' measures and metrics of AI4H software	E.g., How the AI tool will integrate into existing system / workflow flexibility of final decision-making capability by the health practitioner taking into account other factors including patient history, options and preferences
R-QTY7	System SHALL define 'interoperability' measures and metrics of AI4H software	
R-QTY8	System SHALL define 'reusability ' measures and metrics of AI4H software	
R-QTY9	System SHALL define 'testability ' measures and metrics of AI4H software	E.g., Requirements Vs Test Plan traceability matrix
R-QTY10	System SHALL define 'usability ' measures and metrics of AI4H software	E.g., Human Factors design Logical Visual Flow charts, Event based Alerts/Alarms/Notifications, etc
R-QTY11	System SHALL define 'robustness ' measures and metrics of AI4H software	
R-QTY12	System SHALL define 'resiliency ' measures and metrics of AI4H software	

Table 7 – Non-functional requirements

Performance requirements specification		
REQ. ID	**Requirement specification**	**Description**
R-QTY13	System SHALL define 'maintainability ' measures and metrics of AI4H software	
R-QTY14	System SHALL define 'portability ' measures and metrics of AI4H software	E.g., – use of portable programming language – use of compiler or language subset – use of operating system
R-QTY15	System SHALL define 'explainability 'measures and metrics of AI4H software	E.g., a documented procedure on how a medical practitioner explains: – how the AI based decision making / result can impact patient care including the limitations of the AI tool – what hardware, software settings, data pre and post processing techniques were used for data sensing modalities (e.g. MRI imaging hardware and software settings) – how the 'ground truth' was established for the training data – how data integrity was verified – with what accuracy was data labelling done – how the AI tool performance was tested and under what all conditions including appropriateness to the target patient group – conditions under which the AI tool was not tested – whether model attained 'saturation' condition during learning phase – whether compliance with regulatory approval requirements achieved or not

14 System design requirements

Table 8 – System design requirements

AI data design		
REQ. ID	**Requirement specification**	**Description**
R-DD1	System SHALL state applicable regulations and policies related to data handling	E.g., Policies on data management, data acceptance, data protection, data sharing, copyright, privacy laws, patient consent and confidentiality, ethics board approved consent procedures for sharing patient data, retention of raw data etc.
R-DD2	System SHALL have specification for 'data provenance'	E.g., – Data acquisition protocol for reproducibility (who, when, where, how, etc.), – Digitization, data migration to other databases, etc – Hardware and software configuration of data acquisition/ data processing device/App: Sensor type, sampling rate, operational system (OS) version, firmware version, etc.
R-DD3	System SHALL define input data source formats	E.g., real and synthetic data sources – Electronic health records (anonymized) – Medical images – Vital signs signals – Lab test results – Photographs – Non-medical data-socioeconomic, environmental, etc.) – Questionnaire responses – Free Text (discharge/summary, medical history/notes, etc.) – PACS, – Web portal – mHealth App – Medical device – Other
R-DD4	System SHALL have details of the data collectors	E.g., – Medical personnel (physician/clinician/nurse/pharmacist/ etc.)

Table 8 – System design requirements

AI data design		
REQ. ID	**Requirement specification**	**Description**
		– Support personnel – Patient (or proxy person) – Machine-generated
R-DD5	System SHALL define input data types	E.g., – Real valued – Integer-valued – Categorical value – Ordinal value – Strings – Dates – Times – Complex data type – Other
R-DD6	System SHALL define input data formats	E.g., – DICOM PS3.0 (latest versions) – for Diagnostic Image (X-Ray, CT, MRI, PET, other pathological slides, etc) – JPEG/PNJ – for Static Image – MP3/OGG – for Audio: – MP4/MOV – for Video – SNOMED – for clinical observations/terminology – LOINC – for laboratory observations – WHO ICD-10 for disease classifications – RxNORM for medication code – Other

Table 8 – System design requirements

AI data design		
REQ. ID	**Requirement specification**	**Description**
R-DD7	System SHALL define output data types	E.g., – Binary/Class output (0 or 1) as in case of classification problems – Probability output (0-1) as in case of classification problems – Continuous valued output as in case of regression problems
R-DD8	System SHALL have specifications for the data encoding-decoding formats	
R-DD9	System SHALL have specifications for the compression and encryption techniques	E.g., Lossy and non-lossy compression techniques, homographic encryption
R-DD10	System SHALL have specifications for the annotation/labelling of ground truth/reference output data	E.g., – Standards for health data vocabulary/labelling for training and test data o Standards for clinical terminology o Laboratory observations o Disease mapping o Procedure mapping o Messaging o Clinical data format o E.g., coding standards (e.g. SNOMED, LOINC, ICD-10, HL7-FHIR, etc) – procedure – to establish the reference or ground truth for the training data (whether based on objective measures, expert group consensus, etc) – labelling accuracy calculation technique – labelling error estimation technique

Table 8 – System design requirements

AI data design		
REQ. ID	**Requirement specification**	**Description**
R-DD11	System SHALL have defined 'data completeness' verification techniques used	E.g., – Data cleaning and correction for ranges, variations, outliers, missing values, etc – Data bias minimization techniques used – Data variance minimization techniques used – Data normalization method – for Data Preparation phase-to account for data variability in terms of data sensing hardware, data sensing software settings, sensor device model and device configuration – Data representativeness: Data to represent different types of population covering diverse backgrounds, cultures and disciplines, vulnerable persons, persons with disabilities, ethnic minorities, women, children, geriatric, refugees and other categories facing risk of exclusion, discrimination, stigmatization, prejudice, abuse, human rights violations, torture, inhumane treatment and marginalization
	System SHALL define 'data de-biasing' techniques used	
R-DD12	System SHALL have specifications for the 'data integrity' mechanisms used	E.g., RAID, Mirroring, Checksum, Digital Signature, etc
R-DD13	System SHALL have specifications for the 'data privacy' mechanisms used	E.g., – patient consent obtained – ethics board approval – anonymization and de-identification methods used – secure data disposal policy/agreement
R-DD14	System SHALL have specifications for the 'data safety and security' mechanisms used	E.g., – Access Control Functions (Authentication, Authorization, Monitoring, Logging and Auditing) – Audit Logs for viewing, creation, modification, validation, copying, import, export, transmission, reception, etc. based

Table 8 – System design requirements

AI data design		
REQ. ID	**Requirement specification**	**Description**
		○ on Blockchain Technology ○ Merkle Trees, etc – Data Repositories compliance with ISO 7498-2 security model and other allied standards for best practice recommendations on information security management – Data sharing through secured channels – Data flow control mechanisms within practice boundaries – Implementing security standards based on digital certificate, SSL, SHA-256, etc.
R-DD15	System SHALL have specifications for the 'data interoperability' mechanisms used	E.g., – Data formats – Messaging coding standards – Application Programming Interfaces (APIs)/web services for data exchange, data loading/importing – Protocols and tools to collect and integrate diverse data
R-DD16	System SHALL define the 'data preparation' methods used	E.g., – Descriptive statistical methods used to summarize the distribution and relationships between variables using visualizations such as charts, plots, and graphs – Statistical methods for data cleaning such as Imputation – for rectifying corrupt or missing values – Data modelling using statistical techniques – encoding, scaling, transforms, etc.
R-DD17	System SHALL define the data splitting criteria for the training, validation, and testing data sets	E.g., Independent data sets to be used for each of the training, validation and testing phases of model development
R-MD1	System SHALL define the type of AI model	E.g., 'Static Model' or 'Continuous/Incremental Learning Model'?

Table 8 – System design requirements

AI data design		
REQ. ID	**Requirement specification**	**Description**
	System SHALL define the algorithm selection method for AI model training	e.g., algorithm selection – for optimization, specialization, generalization
	System SHALL define the choice of particular machine learning method used for AI model training	E.g., – Active learning method – for performance improvement for new data points – Reinforcement learning method – to solve decision-making problems – Genetic algorithms and simulated annealing – for optimization problems – Online/Incremental learning method – for streaming data – Additive Tree method – for AI model interpretability – Federated learning, dynamic thresholding – for ai model performance improvement – Long short-term memory (LSTM) and gated recurrent units (GRU) for resource-constrained, low memory devices, Convolutional neural networks to handle multi-dimensional datasets – Collaborative Filtering method – for Recommended systems – Transductive learning method
R-MD2	System SHALL define the AI model selection criteria	E.g., specific machine learning algorithm and its configuration that is applied on the training dataset in order to learn the model – Supervised learning-based algorithms o Linear regression o Logistic regression o k-nearest neighbours o Decision trees o Random forest o Gradient boosting machines o XGBoost o Support vector machines (SVM)

Table 8 – System design requirements

AI data design		
REQ. ID	**Requirement specification**	**Description**
		o Neural network o other – Unsupervised learning based algos o k means clustering o Hierarchical clustering o Neural network o other – Reinforcement learning-based algorithms – Association rule learning-based algorithms – Apriori algorithm – Eclat algorithm – Deep learning-based algorithms – Convolutional neural network (CNN) – Recurrent neural networks (RNNs) – Long short-term memory networks (LSTMs) – Stacked auto-encoders – Deep Boltzmann machine (DBM) – Deep belief networks (DBN) – Other
R-MD3	System SHALL have specification for test data set design	E.g., – Criteria to ensure proportionate mix of true/false positives and true/false negatives and data disjoint from training set – Algorithmic accountability – Split tests – Multiple split tests – Cross validation

Table 8 – System design requirements

AI data design		
REQ. ID	**Requirement specification**	**Description**
		– Multiple cross validation – Statistical significance validation – Uncertainty estimation – Other
R-MD4	System SHALL define the AI model evaluation metrics used	E.g., – Model accuracy (%) – Model accuracy – Mean and standard deviation – Model accuracy – Box plot summarization – Root mean squared error (RMSE) – Sensitivity (true positive rate) – Specificity (true negative rate) – F1-Score (class wise performance determination) – Confusion matrix – K-fold cross-validation – Gain and lift charts – Kolmogorov Smirnov chart – Gini coefficient – Log loss – Area under the ROC curve (AUC) – Concordant – Discordant ratio – Jaccard coefficient – Pearson Correlation – Other
R-MD5	System SHALL define the AI model optimization techniques used	E.g., – Adding or deleting Features/Attributes of the input data – Aggregating or decomposing features/Attributes of the input data

Table 8 – System design requirements

AI data design		
REQ. ID	**Requirement specification**	**Description**
		– Tuning model hyper-parameters – Normalization and standardization of input data – Changing the learning rate of the algorithm – Examining the statistical significance of results – Recruiting ensemble methods for combining/augmenting the prediction scores of multiple models – Monitoring and tracking API response times and Computational Memory requirements of the serving infrastructure – Procedure to detect whether AI model attained 'saturation' point of learning or not – other
R-MD6	System SHALL have specification for the AI model card/ sheet format	E.g., – assumptions, constraints, dependencies on the algorithm used – current performance figures – expected/optimal performance – major risk conditions – model version – other
R-MD7	System SHALL have specification for the AI model Accuracy, Specificity and Sensitivity, Latency	
R-MD8	System SHALL have specification for the AI model software implementation framework	E.g., AI model training tools, toolkits and software libraries

15 System deployment requirements

Table 9 – System deployment requirements

REQ. ID	Requirement specification	Description
R-DPY1	System SHALL have specification for the AI4H software deployment/run-time environment configuration	E.g., – IT infrastructure requirements for network, storage and computing resources- – Processor (manufacturer, speed, and features), RAM (memory size), hard disk size, communication, display interface, sensors, energy sources, safety features, etc.
R-DPY2	System SHALL have 'assembling and testing 'procedure for the AI4H software deployment/run-time environment	
R-DPY3	System SHALL have specifications for the AI4H software delivery packaging	E.g., – executable software – support data files – support documentations – installation scripts for different computing configurations-hardware, operating system, peripheral devices, networking interfaces – technical support (for functions and features, troubleshooting guidelines, training materials) – product version – optimal operating condition – optimal configuration – efficiency rating (if applicable) – standards compliance/ certification (if any)
R-DPY4	System SHALL have specifications for the distributed computing environment of AI4H software	E.g., – data workflows, pipelines, and Extract, Transform and Load (ETL) processes – production pipelines for training, retraining, data analytics, data visualization, network connectivity, storage, security and scalability
R-DPY5	System SHALL have specifications for the high-performance production environment of the AI4H software	E.g.,

Table 9 – System deployment requirements

REQ. ID	Requirement specification	Description
		– Software Container (Docker) Architectures for Benchmarking Platforms- E.g., CrowdAI, Kaggle, etc. – API response times and Computational Memory configurations – Versioning of code/model/data
R-DPY6	System SHALL define the service levels for the AI4H software deployment environment	
R-DPY7	System SHALL define the AI service utilization metrics for the AI4H software deployment environment	

16 User documentation/training requirements

Table 10 – User documentation and training requirements

REQ. ID	Requirement specification	Description
R-UD1	System SHALL define the user documentation delivery formats/standards	E.g., – User tutorial – Technical guide – User safety guide – Online help

17 Assumptions and dependencies

Table 11 – Assumptions and dependencies

REQ. ID	Requirement specification	Description
R-DPN1	System SHALL state the assumptions, if any, on the selection of the Machine Learning Algorithm based on the available input dataset	E.g., potential vulnerabilities, risks or biases
R-DPN2	System SHALL list the unintended consequences, if any, applicable to the AI4H software	E.g., Unintended consequences due to – technology – patient safety issues – workflow disruptions – inadvertent biases
R-DPN3	System SHALL state third-party/commercial components / licenses used, if any	
R-DPN4	System SHALL list the components reused from other projects, if any	
R-DPN5	System SHALL state the vendor-neutral interoperability standards, if any	

18 Quality process compliance

Table 12 – Quality assurance requirements

REQ. ID	Requirement specification	Description
R-PRO1	System SHALL define the primary quality metrics of the AI4H software	E.g., patient safety, quality of care, workflow efficiency, etc
R-PRO2	System SHALL define a project management process for AI4H software development as per QMS	E.g., for AI model development phase – to oversee implementation and monitoring of system performance and use
R-PRO3	System SHALL define a data management process for AI4H software development as per QMS	E.g., for data management practices during data preparation phase – for representing, accessing, storing and transferring health data

Table 12 – Quality assurance requirements

REQ. ID	Requirement specification	Description
R-PRO4	System SHALL define a regulatory audit process for AI4H software development as per QMS	E.g., for AI model validation phase – to ensure use of AI and practice is in compliance with regulatory, ethical principles and standards
R-PRO5	System SHALL define a software delivery process for AI4H software as per QMS	E.g., building and packaging of AI APIs and Web services
R-PRO6	System SHALL define a quality audit process for AI4H software development as per QMS	E.g., AI model validation phase – for quality assurance related to quality management, risk management, reporting standards, training
R-PRO7	System SHALL define a regulatory, quality and security certification process, if applicable	

19 Risk management requirements

Table 13 – Risk management requirements

REQ. ID	Requirement Specification	Description
R-RIM1	System SHALL define the procedure and metrics for risk assessment	E.g., – Risk identification and characterization – Risk analysis – Risk assessment criteria
R-RIM2	System SHALL define the procedure and metrics for risk control	E.g., – Risk reduction – Risk acceptance
R-RIM3	System SHALL define the procedure and metrics for risk communication	
R-RIM4	System SHALL define the procedure and metrics for risk review	

20 Change management requirements

Table 14 – Change management requirements

REQ. ID	Requirement specification	Description
R-CHM1	System SHALL define a change management plan and procedure	E.g., based on stakeholder change requests, gaps identified, feedback analysis. Supporting data include – Approval/ Authorisation formalities in case of any modification to the original deployed model – Usage traceability record – information on software version, date and time of use, use environment and the patient to whom it was applied – User skill traceability record – User Experience Surveys and User Satisfaction Ratings record – Workload demand and AI operational efficiency record – Cost-benefit-patient outcome analysis report – Software up gradation/change request to meet clinical change management requirements – Service utilization and service compliance report (Periodic/Term-wise) – other
R-CHM2	System SHALL define a change implementation plan and procedure	E.g., – timeline and schedule – stakeholder capacity building and training – stakeholder communication and feedback mechanisms – other

21 System validation requirements

Table 15 – System validation requirements

REQ. ID	Requirement specification	Description
R-VDN1	System SHALL define test plan and procedure for functional testing	
R-VDN2	System SHALL define test plan and procedure for performance testing	E.g., based on – Feature Engineering – Ensample methods – Algorithm Tuning – Minimizing Model Variance-Bias trade-off
R-VDN3	System SHALL define test plan and procedure for hardware and software platform testing	E.g., based on multi-vendor equipment
R-VDN4	System SHALL define test plan and procedure for hardware and software interface testing	
R-VDN5	System SHALL define test plan and procedure for data interface/interoperability testing	E.g., with other health information systems / databases
R-VDN6	System SHALL define test plan and procedure for data quality testing	E.g., for data integrity, data completeness, data bias
R-VDN7	System SHALL define test plan and procedure for data access control testing	E.g., for authentication, authorization, monitoring, logging and auditing
R-VDN8	System SHALL define test plan and procedure for workflow /protocol Integration testing	E.g., to ensure proper AI solution interoperability with clinical workflow setting
R-VDN9	System SHALL define test plan and procedure for safety and security controls testing	E.g., – assessment of the likelihood of a threat, vulnerability in case of device functionality – assessment of the likelihood of a threat, vulnerability in case of user safety – estimation of the type and probability of risks (for device, environment of use and user) – verification of security controls for device (software, hardware, firmware) – verification of security controls for data repositories

Table 15 – System validation requirements

REQ. ID	Requirement specification	Description
		– verification of security controls for data channels – verification of security controls for environment of use – verification of security controls for user
R-VDN10	System SHALL define test plan and procedure for User Group testing	E.g., to ensure the user has adequate and appropriate knowledge, skills and competency level to the use/operate AI system in the given role
R-VDN11	System SHALL define test plan and procedure for Usability testing	E.g., usability assessment report for different user groups
R-VDN12	System SHALL define test plan and procedure for User-Interface testing	
R-VDN13	System SHALL define test plan and procedure for Installation testing	
R-VDN14	System SHALL define test plan and procedure for Stress testing	

22 AI4H topic description to AI4H system requirements traceability matrix

– For the purpose of auditing the coverage of technical requirements for the benchmarking process that are defined in the topic description documents (TDD), AI4H requirements specifications are configured as a requirements verification tool.

– This tool partially serves as a quality assessment tool for the benchmarking process under the AI4H assessment platform.

– The tool is implemented as a 'traceability matrix' that represents a mapping between the TDD (technical sub-topics) and their corresponding AI4H system requirements.

– A template for the traceability matrix is shown in Table 16. The first column in the matrix describes the 'question from TDD. The second column represents the actual ' requirements traceability attribute', that we want to verify for its coverage, completeness and correctness. This traceability analysis will be performed for the different TDDs and results shall be tabulated against their respective TG columns in the matrix.

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
TDD Section 1.1.5 Definition of the AI task		✓	✓	x
What kind of AI task is implemented?	☐ Classification			
	☐ Prediction			
	☐ Matching			
	☐ Libelling			
	☐ Detection			
	☐ Segmentation			
	☐ Anomaly detection and Fraud Prevention			
	☐ Compliance Monitoring / Quality Assurance			
	☐ Process optimization / Automation			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
Which input data are fed into the AI model?	☐ Electronic health records (Anonymized)			
	☐ Medical Images			
	☐ Vital signs signals			
	☐ Lab test results			
	☐ Photographs			
	☐ Non-medical data (Administrative, Insurance, Socioeconomic, Environmental, etc.)			
	☐ Questionnaire responses			
	☐ Free Text (Discharge/Summary, Medical History/Notes, etc.)			
	☐ Genomic, genetic, proteomic data			
	☐ Health data registries			
	☐ Data from medical (clinical) studies			
	☐ Health data from biobanks			
	☐ Data related to behaviour, lifestyle, professional status			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
TDD Section 1.1.6 Current gold standard				
Are there any numbers describing the performance of the current state-of-the-art?	☐ Intra-annotator labelling accuracy			
	☐ Editable automatic pre-annotation tool			
	☐ Annotation tool labelling accuracy			
	☐ Interoperability and Coding standard (e.g., SNOMED, LOINC, ICD-10, HL7-FHIR, etc)			
	☐ Gold standard method-objective measure			
	☐ Gold standard method – expert group consensus			
TDD Section 1.1.7 Relevance and impact of an AI solution				
Which impact of deploying such systems is expected	☐ Risk-benefit-ratio			
	☐ Faster patient classification, diagnosis, treatment, recovery			
	☐ Percentage reduction in professionals' workload (incl. clinician cognitive, routine)			
	☐ Degree of automation / semi-automation introduced			
	☐ Degree of smartness/intelligence augmentation			
	☐ New knowledge discovery			
	☐ Replacement or redefinition of existing gold standard			
	☐ Improved patients´ experience with health system			
	☐ Patient empowerment			
	☐ Improving efficiency			
	☐ Cost optimization			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Reduction of the number of errors and mistakes			
	☐ Improving administrative processes			
TDD Section 1.2 Ethical considerations				
How is the privacy of personal health information protected?	☐ Data privacy /protection policy			
	☐ Data sharing policy (HIPAA)			
	☐ Patient consent (GDPR)			
	☐ Individual vs. administrative consent			
	☐ Electronic informed consent			
	☐ Ethics board approval			
	☐ Anonymization and de-identification method			
	☐ Secure data disposal policy/agreement			
	☐ Data protection officer			
How do we ensure that benchmarking data are representative and that an AI offers the same performance and fairness?	☐ Proportionate sample size of different classes			
	☐ Subpopulation considered relevant and appropriate to your use case: (a) diverse demographics, (b) cultures and disciplines, (c) vulnerable persons, (d) persons with co-morbidities and contra-indications (e) persons with disabilities, (f) ethnic minorities, (g) women, (h) children, (i) geriatric, (j) persons facing risk of any kind of exclusion, discrimination, and marginalization in the social status and educational status			
	☐ Subpopulation data (age-group, gender, ethnicity, etc) distribution descriptive statistical methods to summarize the distribution and relationships between variables (e.g. Histograms) calculation of			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	distributions (histograms), mean / average values, quartiles, joint distribution of features, correlation, etc.			
	☐ Measurement / sampling bias error			
	☐ Data bias minimization / de-biasing technique used			
	☐ Data variance minimization technique			
	☐ Data partitioning / splitting method			
	☐ Ground truth labelling error			
	☐ Data inclusion /exclusion/rejection criteria			
	☐ Equal true positive rates and also false-positive/negative rates on 'protected population subgroup' as compared to that for the 'entire population group'			
TDD Section 1.3.4 Regulatory considerations				
How will the development process of the benchmarking be documented in an effective, transparent, and traceable way?	☐ SOP (configuration and version control) process (for libraries and frameworks)			
	☐ Versioning systems DREAD and CVSS			
	☐ Name / ID of software libraries and frameworks			
	☐ Software libraries and frameworks version			
	☐ User tutorial version			
	☐ Technical guide version			
	☐ Lookup Table for risks (with related probabilities of occurrence with magnitude of severities and tolerance rates)			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
How will the risk management be implemented?	Failure mode and effects analysis for risk communication: lack of availability / robustness: ☐ slow response times ☐ interoperability problems ☐ software using more CPU, GPU, RAM, I/O, bandwidth than specified			
	☐ Procedure and metrics for risk assessment			
	☐ Procedure and metrics for risk control			
	☐ Procedure and metrics for risk communication			
	☐ Procedure and metrics for risk review			
	☐ Risk-based tool validation			
	☐ Risk-based SOUP validation			
	☐ Risks related to client server architecture			
	☐ Risks related to (de)serialization of data			
	☐ Risks related to format and protocol conversions			
	☐ Risks related to multiple API versions and API gateways			
	☐ Risk related specifically for programming language			
	☐ Risks related to compiler and compiler settings			
How is the test data quality ensured (e.g., the process of harmonizing data of	☐ Data pre-processing statistical method or (cleaning) tool			
	☐ Data integrity verification protocol			
	☐ Data access control protocol			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
different sources, standards, and formats into a single dataset may cause bias, missing values, outliers, and errors)?	☐ Data sequence of operations			
	☐ Data conversion formulas/ rules			
	☐ Data error/exception handling and recovery rules			
	☐ Data communication security / encryption mechanisms			
	☐ Data transfer rates			
	☐ Data encryption/decryption technique			
	☐ Data quality assessor			
How is data privacy in the context of data protection regulations ensured, considering regional differences?	☐ Data Ethical Clearance/ Patient Consent obtained			
	☐ Estimates (type and probability) of data safety risks			
	☐ Distributed and federated ML model learning			
	☐ Training over encrypted data (homomorphic encryption)			
	☐ Security protocols /controls for device (software, hardware, firmware)			
	☐ Security protocols / controls for data repositories			
	☐ Security protocols / controls for data channels			
	☐ Security protocols / controls for environment of use			
	☐ Security protocols / controls for user			
TDD Section 3.2 Benchmarking by AI developers				
Which scores and metrics have been used?	☐ Performance metrics			
	☐ Robustness metrics			
	☐ Explainability metrics			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Fairness metrics			
	☐ Safety metrics			
How did they approach the acquisition of test data?	☐ Protocols and tools to collect and integrate diverse data			
	☐ Data acquisition / sensing modality			
	☐ Data acquisition / sensing device type			
	☐ Sensor device-OS version, firmware version			
	☐ Data collection place			
	☐ Data collection time			
	☐ Data collector			
	☐ Data sampling rate			
	☐ Data update version			
	☐ Device identifiers and attributes			
	☐ Device Authentication protocol (Multiple Levels)			
	☐ Device synchronization protocol			
	☐ Data ownership/controller			
TDD Section 4.2.2.1 Benchmarking system architecture				
How does the architecture look?	☐ Software container (Docker) architecture – Technology stack			
	☐ Data workflows, pipelines, and extract, transform and load (ETL) processes			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
What are the most relevant components and what are they doing?	☐ Production pipelines for training, retraining, data analytics, data visualization, network connectivity, storage, security and scalability			
	☐ Modes of operation (test mode, troubleshooting mode, monitoring mode)			
	☐ Application safety control mechanism			
	☐ User safety control mechanism			
	☐ APIs/Web services for data exchange, data loading/importing			
	☐ Data interface ability to collect and analyse real-time patient data			
	☐ Hardware platform configuration and versions			
	☐ Operating system configuration and versions			
	☐ Database configuration			
	☐ Network/communication protocol			
	☐ External interfaces			
	☐ Internationalization and/or localization protocols			
	☐ Primary and secondary memory configurations			
	☐ Real time machine learning based threat malware classification			
What underlying technologies and frameworks have been used?	☐ SQL			
	☐ Hadoop			
	☐ Apache Kafka			
	☐ SaS			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Julia			
	☐ Python			
	☐ Spark			
	☐ Tensorflow			
	☐ Keras			
	☐ PyTorch			
	☐ GPU			
	☐ Google Analytics			
	☐ QlikView			
	☐ Tableau			
TDD Section 4.2.2.2 Benchmarking system dataflow				
How do benchmarking data access the system? (data interfaces)	☐ Data workflows, pipelines, and Extract, Transform and Load (ETL) processes			
	☐ Data formats			
	☐ Messaging coding standards			
	☐ APIs/Web services for data exchange, data loading/importing			
	☐ Data access control functions e.g., authentication, authorization, monitoring logging and auditing			
Where and how (data format) are the data, the	☐ Syntactic level File formats (XML, JSON, PDF, docx, CSV, DICOM, HL7)			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
responses, and reports of the system stored?	☐ Semantic level standards-taxonomies e.g. LOINC (e.g. laboratory values), ATC (drugs), ICD (diagnoses), UCUM (units) and the organizational level (IHE)			
	☐ Semantic standards-nomenclatures e.g. LOINC			
	☐ Database -data schema/ structure			
How are the data sent to the AI systems?	☐ Data Interface protocols -OSI-Protocols such as TCP/IP, HTTPs, SFTP, CAN, RS232, USB			
	☐ Data Interface-Bus-Systems such as CAN, USB			
TDD Section 4.2.2.3 Safe and secure system operation and hosting				
What safety control measures were taken to manage risks to the operating environment?	☐ Security protocols / controls for device (software, hardware, firmware)			
	☐ Security protocols / controls for data repositories			
	☐ Security protocols / controls for data channels			
	☐ Security protocols / controls for environment of use			
	☐ System security privilege levels of primary and secondary user			
	☐ Lookup table for 'hazards and related harms' specification with related probabilities of occurrence with magnitude of severities			
How is the hosting system itself protected against attacks?	☐ MDR, IVDR, other standards compliance for IT security, compliance with national safeguards			
	☐ Encryption protocol (for Data at Rest, Transmission, Use)			
	☐ Restriction of permitted IP and MAC addresses and physical access protection			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Check for incorrect file length, completeness, incorrect character set, unexpected characters, data sent multiple times, outdated/late data, unexpected or incorrect formats, no well-formed XML, invalid JSON files, incorrect data types, XML that does not correspond to the specified schema), other character sets, values that are not included in the intended value range (e.g. in the classification or encoding system), wrong time zone, wrong number format, impossible data			
	☐ Development tools, development environments and libraries (SOUP, OTS components) loaded only from approved sources classified as secure			
	☐ Libraries checked for malware before use			
	☐ Configuration data for the devices, user data (in particular access data), keys, software certificates, program code (including SOUP/OTS)			
	☐ Definition of who (e.g. user, notified body, authority) to be informed and procedure on how to inform			
	☐ Network/interfaces (bandwidth, availability, ports, IP ranges, latencies, encryption, firewalls etc.), virus protection, operating systems, physical access permissions			
	☐ Installation, connection to network, evaluation of audit logs, deletion of unneeded users, exchange of keys or certificates, deletion of temporary files			
	☐ Consent – primary and secondary data usage			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Data Processing Impact Assessment: A Data Protection Impact Assessment (DPIA) to identify and minimize the data protection risks of a project.			
	☐ Principles of privacy by design incorporated			
	☐ Data repositories security model			
	☐ Data flow control mechanism			
	☐ Security standards based on digital certificate, SSL, SHA-256, etc			
How are the data protected against data loss (e.g., what is the backup strategy)?	☐ RAID			
	☐ Mirroring			
	☐ Checksum			
	☐ Digital Signature			
	☐ Rounding error rate			
	☐ Compression, decompression error rate			
	☐ Noise reduction, filtering error rate			
	☐ Normalization, transformation error rate			
	☐ Re-sampling error rate			
How is it ensured that the correct version of the benchmarking software and the AIs are tested?	☐ SOP for configuration and version control process.			
	☐ Name of software			
	☐ Version of software with date and time stamp			
	☐ Version of data is aligned with corresponding software versions (software for processing and product)			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
How are automatic updates conducted (e.g., of the operating system)?	☐ Approval/ Authorization formalities in case of any modification			
	☐ Algorithm update trigger actions (update of risk analysis and re-evaluation of risk-benefit analysis, re-training of algorithm, product recall, implementation of better risk mitigation measures)			
	☐ Threshold values for trigger actions			
	☐ Algorithm update frequency / period			
How and where is the benchmarking hosted and who has access to the system and the data (e.g., virtual machines, storage, and computing resources, configurational settings)?	☐ IT infrastructure requirements for network, storage and computing resources-Processor (manufacturer, speed, and features), RAM (memory size), hard disk size, communication, display interface, sensors, energy sources, safety features, etc			
	☐ User role and access privilege			
	☐ Access control functions (authentication, authorization, monitoring, logging and auditing)			
How is the system's stability monitored during benchmarking and how are attacks or issues detected?	☐ Protocols for safety alarms			
	☐ Risk alert for input data not meeting the requirements			
	☐ Risk alert for inability of the system to meet the non-functional requirements			
	☐ Risk alert for ML algorithms not meeting the quality metrics			
	☐ Risk alert for adversarial attacks			
	☐ Risk alert for software bugs			
	Risk alert for hardware related risks: ☐ CPU, RAM, I/O, hard disk space not as specified ☐ memory, CPU, GPU flaws			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ hard disk full ☐ RAM, CPU, I/O overutilization by other applications.			
	Risk alert for software related risks: ☐ other type or version of operating system, browser, virtualization layer (.NET, JRE, VM), libraries ☐ software patches not installed ☐ software bug.			
	Risk alert for network related risks: ☐ bandwidth, latency not as specified ☐ Endpoints, protocols not supported or blocked			
	Risk alert for interface related risks: ☐ Unspecified data volumes ☐ Wrong data inputs			
	☐ Self-tests verifying/validating the implementation and effectiveness of risk mitigation measures			
How are issues (e.g., with a certain AI) documented or logged?	☐ Audit Logs for (viewing, creation, modification, validation, copying, import, export, transmission, reception, etc.) based on blockchain technology, Merkle trees, etc.			
	Failure mode and effects analysis for risk communication ☐ lack of availability / robustness ☐ slow response times ☐ interoperability problems ☐ software using more CPU, GPU, RAM, I/O, bandwidth than specified			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
TDD Section 4.2.3 AI input data structure for the benchmarking				
What are the general data types that are fed in the AI model?	☐ Data types (images, audio, videos, text and natural language, time series, sensor, etc)			
	☐ Computational data types (real valued, integer-valued, categorical value, ordinal value, strings, dates, times, complex data type, other			
How exactly are they encoded?	☐ Syntactic level File formats (XML, JSON, PDF, docx, CSV, DICOM, HL7)			
	☐ Semantic level standards-taxonomies e.g., LOINC (e.g. laboratory values), ATC (drugs), ICD (diagnoses), UCUM (units) and the organizational level (IHE)			
	Meta data description ☐ Filename, ☐ file format ☐ URL ☐ domain ☐ keywords ☐ type, dataset size, ☐ % of missing cells ☐ license ☐ release date ☐ collection range			
The exact data format with all fields and metadata	☐ MP3 / OGG – for Audio			
	☐ MP4/MOV – for Video			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ DICOM PS3.0 for X-Ray, CT, MRI, PET			
	☐ JPEG / PNJ – for Static Image			
	☐ File formats (XML, JSON, PDF, Docx, CSV)			
	☐ Data version			
	☐ Data processing software version			
	☐ Data source			
	☐ Number of data sources			
	☐ Data collection place			
	☐ Data collection time			
	☐ Data collection author(s)			
	☐ Data directory structure			
	☐ Data backup repository			
	☐ Number of data updates			
	☐ Data registry			
Resolution and data value ranges (e.g., sizes, resolutions, and compressions)	☐ Data resolution / precision			
	☐ Data value range			
Data size and data dimensionality	☐ Data sample size			
	☐ Data dimensionality			
	☐ Data dimensionality reduction technique			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Data compression format			
	☐ Train: Validation (Tuning): Test dataset partitioning ratio			
TDD Section 4.2.4 AI output data structure				
What are the general data output types returned by the AI and what is the nature of the output (e.g., classification, detection, segmentation, or prediction)?	☐ Binary / Class output (0 or 1)			
	☐ Probability output (0-1)			
	☐ Continuous valued output			
How exactly are they encoded? Discuss points like:	☐ Syntactic level file formats (XML, JSON, PDF, docx, CSV, DICOM, HL7)			
	☐ Semantic level standards-taxonomies e.g., LOINC (e.g., laboratory values), ATC (drugs), ICD (diagnoses), UCUM (units) and the organizational level (IHE)			
The exact data format with all fields and metadata (including examples or links to examples)	☐ File formats (XML, JSON, PDF, Docx, CSV)			
What types of errors should the AI generate if something is defective?	UI output display (warning, alert) in case of: ☐ Output data validity checks ☐ Output data sequence of operations			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Output data response time			
	☐ System 'RESET' option specification			
	☐ Data error/exception handling and recovery protocol			
TDD Section 4.2.6 Scores and metrics				
What general criteria have been applied for selecting scores and metrics?	☐ Performance metrics			
	☐ Robustness metrics			
	☐ Transparency including explainability metrics			
	☐ Fairness metrics			
	☐ Safety metrics			
	☐ Accuracy metrics			
	☐ Accessibility and universal design metrics			
What scores and metrics have been chosen/defined for robustness?	☐ Explanation of limitations and residual risks			
	☐ Risk tolerance rate estimation for resource overload			
	☐ Risk tolerance rate estimation for runtime errors			
	☐ Risk tolerance rate estimation for adversarial attack errors			
	☐ User control for 'algorithm change protocol'			
	☐ User option settings to reject, delay or roll-back an algorithm change			
What scores and metrics have been chosen/defined for medical performance?	☐ Sensitivity (true positive rate)			
	☐ Specificity (true negative rate)			
	☐ F1-Score (class wise performance determination)			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Area under the ROC curve (AUC)-operating threshold (classification tasks)			
	☐ Prevalence, Type 1, Type 2 errors, precision and confidence interval specifications			
What scores and metrics have been chosen/defined for non-medical performance?	☐ Risk-benefit ratio			
	☐ Patient benefit estimation			
	☐ Patient care cost estimation			
	☐ Case interpretation efficiency estimation ☐ Decision curve analysis			
	☐ Model – false positive referral rate			
	☐ Specification for 'average time of interpretation' (clinician versus model)			
	☐ Model development cost			
	☐ Model deployment cost			
	☐ Model utilization cost			
	☐ Model maintenance and support cost			
	☐ Cost involved in establishing patient safety, quality of care, workflow efficiency, etc.			
	☐ Faster patient diagnosis / treatment			
	☐ Percentage reduction in professionals' workload (incl. clinician cognitive, routine)			
	☐ Degree of automation/semi-automation introduced			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Degree of smartness/intelligence augmentation new knowledge discovery			
	☐ Enabling replacement or redefinition of existing gold standard, etc.			
What scores and metrics have been chosen/defined for model explainability?	☐ Textual or Visual display formats for model prediction interpretation through LIME (local interpretable model-agnostic explanations), LRP (Layer-wise relevance propagation) explanation graphs			
	☐ Super pixels explanations for the predicted classes (Google's Inception network, highlighting positive pixels)			
	☐ Features that contribute towards the prediction (in one colour code) and non-contributing features (in another colour code)			
	☐ 'General features' learned from pre-trained' network and the new 'Transfer leaned features' from transition and prediction layers of the network			
	☐ Features that are globally relevant/significant and that are locally relevant /significant			
	☐ Learning curves to illustrate the impact of hyperparameter and epochs on quality metrics			
	☐ Data pre-conditions (met/not met) indicator (in case of data leakage scenario, incomplete data sets or data drift, malicious input data, etc.)			
	☐ Data (protected/not protected) verification status			
	☐ Directed acyclic graph (DAG) for indicating feature dependencies			
	☐ Algorithm change (performed / not performed)			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Risk communication indicators in case of non-specified user type, non-specified use environment			
	☐ Model prediction decision thresholds chosen			
	☐ Visualization of the dependency (strength, direction) of the prediction of the feature values: 'Counterfactual' explanation, Sharpley-values, ICE-plots, partial dependency plots (PDP)			
	☐ Class activation mapping-heat maps plot			
	☐ Heat map- spatial resolution specification			
Does it use some kind of approach for correcting dataset bias (e.g., the test dataset usually has a different distribution compared to the distribution of a condition in a real-world scenario? For estimating the real-world performance, metrics need to compensate this difference)	☐ Selection of subject data			
	☐ Selection of study method			
	☐ Selection of efficacy evaluation standard			
	☐ Clinician consensus determination			
	☐ Fatigue, perceptual bias, cognitive bias error estimation			
	☐ QUADAS-2 tool test report			
	☐ Subject/Control group data size calculation formula (statistical hypothesis based)			
	☐ Control study-medical device specifications			
	☐ Success criteria for evaluation results			
	☐ Comparison report of patient outcomes (for patients on whom the AI model is applied) versus (patients on whom the AI model is not applied)			
	☐ Bias and fairness risk estimation			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ PROBAST tool test report			
	☐ FairML tool test report			
	☐ 'Pre-trained' model verification (GoogLeNet, ImageNet, AlexNet, VGGNet-16, ResNet-50 for CNN architectures)			
	☐ Verification of the type of regression method used when (number of variable) is greater than the (number of observation) Lasso regression, Ridge regression			
	☐ Verification status that distinct populations are not inappropriately combined			
	☐ Verification status that no inappropriate performance metric was selected			
TDD Section 4.2.7 Test dataset acquisition				
How have the data been collected/generated (e.g., external sources vs. a process organized by the TG)?	☐ Combined databases			
	☐ Federated learning (e.g., pooling data from primary, secondary and tertiary hospitals)			
Have the design goals for the benchmarking dataset been reached (e.g., please provide a discussion of the necessary size of the test dataset for relevant benchmarking results,	☐ Verification of exclusion of predictors with high correlation			
	☐ Verification of exclusion of predictors with degenerate distribution or variance close to zero			
	☐ Data leakage detection			
	☐ Verification status for exclusion of variables that allow the algorithm to identify the outcome (acting as proxy for the outcome)			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
statistical significance, and representativeness)?	☐ Distributed and federated ML model learning			
	☐ Data leakage detection			
	☐ Verification status for exclusion of variables that allow the algorithm to identify the outcome (acting as proxy for the outcome)			
	☐ Descriptive statistics plots for data dimensionality and variance calculation of distributions (histograms), mean / average values, quartiles joint distribution of features, correlation, etc.			
	☐ Verification for data transformation (flip, rotate, colour map, scale, etc.) done to match/reflect the real-world clinical data consistency profile			
Were they collected in an ethical-conform way? What kind of data anonymization or de-identification has been applied?	☐ Data privacy/protection policy			
	☐ Data sharing policy (HIPAA)			
	☐ Patient consent (GDPR)			
	☐ Individual vs. administrative consent			
	☐ Electronic informed consent			
	☐ Ethics board approval			
How is the bias of the dataset documented (e.g., sampling or measurement bias, representation bias, or practitioner/labelling bias)?	☐ Selection of subject data			
	☐ Selection of study method			
	☐ Selection of efficacy evaluation standard			
	☐ Clinician consensus determination			
	☐ Fatigue, perceptual bias, cognitive bias error estimation			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Subject / Control group data size calculation formula (statistical hypothesis based)			
	☐ Control study-medical device specifications			
	☐ Adherence to regulatory compliance with country/region specific data policies on data protection, data sharing, copyright, privacy laws, patient consent and confidentiality			
	☐ Measurement bias introduced due to selection of data or samples that do not represent the true parameters/ distribution of the population of interest			
	☐ social / representation bias introduced in people-based data sources collected from services / surveys / social media etc. that are subject to inherent bias due to historic decisions, outdated laws, cultures and disciplines, ethnic minorities, human rights violation, etc.			
	☐ Practitioner / labelling bias introduced as a result of subjective bias by practitioners during project design, analyses, or interpreting outputs, etc.			
	☐ subjective practitioner bias arising out of differences in experience in technology application, experience in medical domain, competencies, skill sets, physical and cognitive prerequisites and limitations, cultural and social characteristics, etc.			
	☐ Training data algorithmic bias: biased data results in biased model			
	☐ Algorithmic focus bias refers to the selection or rejection of certain types of input data			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Transfer context bias: a model trained on data generated within a specific context shall not exhibit optimal performance when the same model is applied on a different contextual data			
	☐ Algorithmic tuning: bias introduced as a result of tuning or modifying the parameters of the ML algorithm leading to over-fitting or vice-versa			
	☐ Bias risk estimation			
Have any scores, metrics, or tests been used to assess the quality of the dataset (e.g., quality control mechanisms in terms of data integrity, data completeness, and data bias)?	☐ Data completeness validation protocols ☐ Data cleaning and correction for ranges, variations, outliers, missing values, etc.			
	☐ Data integrity validation protocols Integrity mechanisms – RAID, mirroring, checksum, digital signature, etc.			
How were the data submission, collection, and handling organized from the technical and operational point of view (e.g., folder structures, file formats, technical metadata encoding, compression, encryption, and password exchange)?	☐ Data interface protocols – OSI-protocols such as TCP/IP, HTTPs, SFTP, CAN, RS232, USB			
	☐ Data encryption protocol (at Rest, Transmission, Use)			
	☐ Data Access control functions (authentication, authorization, monitoring, logging and auditing)			
	☐ Data compression format			
	☐ Data privacy and confidentiality protocol			
	☐ Data security protocol			
	☐ Data encoding/ decoding protocol			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Data interface protocol			
	☐ Data digitization tool and protocol			
	☐ Data migration tool and protocol			
	☐ Data porting tool and protocol			
	☐ Security protocols / controls for data repositories			
	☐ Security protocols / controls for data channels			
	☐ Database configuration			
	☐ Database – data schema/ structure			
	☐ APIs/Web services for data exchange, data loading / importing			
Which scores, metrics, and thresholds were used to assess the label quality and the need for an arbitration process?	☐ Intra-annotator labelling accuracy			
	☐ Annotation tool labelling accuracy			
	☐ Eligibility criteria of annotation specialists			
	☐ Perceptual errors and bias estimates			
	☐ Ground truth or reference standard establishment method			
How have inter-annotator disagreements been resolved (i.e., what was the arbitration process)?	☐ Adjudication process leading to consensus grading by specialists			
Were metadata on the annotation process included in the data (e.g.,	☐ Standard health data vocabulary / labelling for training and test data			
	☐ Standards for clinical terminology			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
is it possible to compare the benchmarking performance based on the annotator agreement)?	☐ Clinical history			
	☐ Physical body exam results			
	☐ Laboratory observations			
	☐ Disease mapping			
	☐ Procedure mapping			
	☐ Clinical data messaging formats			
	☐ Data anonymisation/pseudonymisation formats for training and test data			
	☐ Labelling accuracy / error calculation and reporting for training data			
	☐ Intra/inter annotator reliability measurements (with practitioner levels of experience)			
How was access to test data controlled (i.e., to ensure that no one could access, manipulate, and/or leak data and data labels)? Please address authentication, authorization, monitoring, logging, and auditing.	☐ Access control functions (authentication, authorization, monitoring, logging and auditing)			
	☐ Audit logs for viewing, creation, modification, validation, copying, import, export, transmission, reception, etc. based on blockchain technology, Merkle trees, etc.			
	☐ Data sharing through secured channels			
	☐ Implementing Security standards based on digital certificate, SSL, SHA-256, etc			
	☐ Data integrity validation protocols: RAID, mirroring, checksum, digital signature			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ "*Break glass*", i.e., an option to bypass the authorization strategy in order to be able to access important data immediately			
TDD Section 4.2.7.1 Data sharing policies				
Roles and responsibilities	☐ Data provider			
	☐ Data protection officer			
	☐ Data controllers			
	☐ Data processors			
	☐ Data receivers			
Which legal framework was used for sharing the AI?	☐ GDPR, HIPAA			
TDD Section 4.2.10 Result				
What overall performance of the AI systems concerning medical accuracy, robustness, and technical performance (minimum, maximum, average etc.) has been achieved?	☐ Model accuracy (PA) estimate – Confidence interval (e.g., With X % confidence, PA lies in the interval [lower bound, upper bound]			
	☐ Model accuracy – mean and standard deviation			
	☐ Model accuracy – Box plot summarization			
	☐ Root Mean Squared Error (RMSE)			
	☐ Sensitivity (true positive rate)			
	☐ Specificity (true negative rate)			
	☐ F1-Score (class wise performance determination)			
	☐ Confusion matrix			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Gain and lift charts			
	☐ Kolmogorov Smirnov chart			
	☐ Gini coefficient			
	☐ Log loss			
	☐ Area under the ROC curve (AUC)			
	☐ Concordant – Discordant ratio			
TDD Section 4.2.11 Discussion of the benchmarking				
How was the performance of the AI system compared to the baseline?	☐ ML model selection, ML model metric selection and optimal hyper-parameter setting guided on the basis of the cost estimates of ML model errors			
	☐ Model performance fairness estimation on bias relevant population subgroups (females, ethnic subgroups, age)			
	☐ Estimation of statistical measures of significance and uncertainly			
	☐ Specification for 'confidence intervals' and 'metric variation' (error bars)			
	☐ Algorithm augmentation verification (Lagrangian approach)			
	☐ Algorithmic tuning / decision thresholds used for differential diagnosis based on age, gender, ethnicity, etc			
	☐ 'Demographic Parity' verification			
	☐ Equal Opportunity' verification			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ 'Verification Status' - that model makes positive prediction on a 'protected population subgroup' as same as that for the 'entire population group'			
	☐ 'Verification Status' – that model makes equal true positive rates and also False-positive/negative rates on 'protected population subgroup' as those for the 'entire population group'			
	☐ Verification to ensure that higher model coefficients are penalized to lower the model complexity			
	☐ Selection of significant features from 'variable importance chart'			
	☐ Intended user verification (verification of input data distribution for target patient population representativeness			
	☐ Cost estimation for wrong clinical decision making / if outputs do not meet the specified 'quantitative quality criteria'			
	☐ Lookup table for risk acceptance criteria specification (clinical risk and cost of risk)			
	☐ ML model sensitivity analysis for subsets of the population including vulnerable, marginalized and other sensitive categories of the population			
	☐ ML model failure analysis for any subsets of the population			
	☐ Hyper-parameter tuning of ML models to reduce the variance-bias trade-off			
	☐ ML model selection on the basis of finding a balance between model accuracy and model simplicity or interpretability			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ ML model's unacceptable failure rates for any of the subsets of the population			
	☐ ML model's influence by relevant inputs from the target user / domain experts			
	☐ Level of interpretability of the ML model meets the expectation of the target user or not			
	☐ Factors that significantly influenced ML model performance			
	☐ ROC discrimination threshold' estimation (for mapping the output probabilities to binary predictions)			
	☐ Estimation of measure that indicates the response predicted by a model on adding independent variables – Residual deviance value			
	☐ Estimation of measure that indicates percent of variance in a predictor which cannot be accounted by other predictors (multi-co linearity) – 'Tolerance' value (reciprocal of Variance Inflation Factor (VIF))			
	☐ Algorithm update cycle estimation (time and the amount of change is quantified) to enforce, prevent, delay or roll-back changes to algorithms			
	Algorithm update specifications with respective time stamps: In cases of Neural Networks, ☐ fit parameters such as weights of neurons or cut-off of activation function ☐ hyperparameters such as numbers of neurons per layer and number of layers			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Optimizer specification (Stochastic gradient descent (SGD), Adam, Adadelta, Adagrad, L-BFGS, etc.)			
	☐ Estimation of fitness measure which penalizes model for the number of model coefficients – AIC (Akaike Information Criteria) value			
	☐ Conditions under which the ML model was not tested			
	☐ Cost estimation of ML model errors			
	☐ Estimation of the risk probabilities associated with model performance variability due to adversarial attacks (e.g., with manipulation of selected pixels in images)			
	☐ Risk probabilities associated with model performance variability w.r.t change in hardware and software configurations of model development environment			
	☐ Estimation of the risk probabilities associated with model performance variability when tested against the following conditions: (a) patients other than those specified, (b) non-specified use environment, (c) patients of different age, sex, race, co-morbidities, (d) patients with different severity of disease type			
	☐ Model effectiveness re-assessed/re-evaluated/re-calibrated for multiple clinical settings of different health environments (e.g., with variability in workflow, demographics, etc.)			
	☐ Model generalizability consideration – verification of input data representative of variations in data acquisition and reconstruction parameters, target population, operating time scales.			
	☐ Training over encrypted data (homomorphic encryption)			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
Are there any technical lessons?	☐ Prevalence in test database' may not reflect ' real-world prevalence'. addressing by provisioning of combined databases, federated learning (e.g., Pooling data from primary, secondary and tertiary hospitals)			
	☐ Multi-core training offers a measurable benefit on your system			
	☐ K-nearest neighbour model to select those instances of the training set that are most similar to the test set.			
	☐ Different modelling pipelines that result in different sets of predictions, scoring the predictions, then making changes to the pipeline that are expected to result in an improved score			
	☐ Maintaining consistency of model performance parameters w.r.t to change in hardware and software configurations of AI model development environment			
	☐ Traceability matrix to link together design, implementation, testing, and risk management			
	☐ Selection of ML algorithms based on a set of assumptions about the underlying distribution of the input variable			
	☐ Cost and Loss functions of ML algorithms			
	☐ ML algorithm hyper-parameter tuning and configuration through reinforcement learning			
	☐ Inverse relationship between the performance (of AI systems) and the level of explainability and trust			
	☐ Algorithm tuning and ensembles methods for performance improvement			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Re-sampling techniques like k-fold cross-validation to estimate the performance of the model when making predictions on data not used during training			
How was the performance and operational efficiency of the benchmarking itself (e.g., how long did it take to run the benchmarking for all AI models vs. one AI model; was the hardware sufficient)?	☐ Average time taken for the model in assigning class labels for the test data			
	☐ Response time specification as a function of number of users, number of transactions, frequency and amount of input data etc			
	☐ API response times and computational memory configurations			
	☐ Number of cores available to support multi-core training			
	☐ Cost of energy and computation			
	☐ CPU load greater than x%, a data traffic greater than y MB/s, a storage medium that is using than z% of its capacity, more than n login attempts within m minutes, and so on			
	☐ Server operating environment configuration (e.g., Linux server with 'x' Titan X GPUs)			
	☐ Heat map generation time			
	☐ 'System Availability' specification in terms of percentage of time, percentage of usages or as meantime between failure			
TDD Section 5 Discussion				
Did the AI system perform as predicted relative to the baselines?	ML model summary in terms of: ☐ explainability ☐ trust ☐ quality of service			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ generalizability ☐ context applicability ☐ efficiency ☐ safety Implication ☐ risks ☐ limitations ☐ user rating (scale) ☐ recommendations ☐ extensibility to other settings			
	☐ Patient safety impact (e.g., early detection and lowering of disease severity levels, increased coverage under screening programs)			
	☐ Care quality impact (e.g., workflow efficiency, reliability and reproducibility of outcomes, increased accessibility, affordability, increased patient and clinician satisfaction)			
	☐ Interpretability-performance trade-off observed (e.g., lack of model interpretability Vs model performance statistical gains			
	☐ Comparison of model inferential capabilities with that of the user (clinician)			
	☐ Level of tolerance that the clinical setting allows to the misdiagnoses/mis-predictions of the model based on the particular use case			
	☐ Patient 'benefits (demonstrable improvements in patient outcomes like)' outweighs the 'risks' involved			

Table 16 – TDD requirements traceability matrix template

AI4H topic description to AI4H system requirements traceability matrix-template				
AI4H TDD description	**AI4H requirements tractability attribute**	**TG-symptoms (verification status)**	**TG-ophthalmology (verification status)**	**TG-xxx**
	☐ Clinical improvements made by the model outweighs the cost of any work changes that it necessitates			
	☐ Acceptable risk-benefit-ratio			
	☐ Predicted benefit actually correspond to clinically observed benefit			
	☐ Model performance comparable to the performance scores or the level of competence of the clinician/specialist/user in the clinical setting			
	☐ Conditions under which model renders most confusions about output label classes and performs poorly			
	☐ Conditions under which the model outperforms the clinician/specialist/user classification performance			
	☐ Conditions under which the model performs worse than that of the clinician/specialist/user			
	☐ Conditions under which the performances of both the model and clinician/specialist are comparable			

23 AI4H requirements specification for model reporting parameter configuration

The TDD requirements traceability matrix is utilized to define the criteria for configuring the reporting parameters for AI4H model evaluation results and this enables 'custom mode' of reporting service under the AI4H assessment platform. The preliminary set of parameter configuration criteria identified from the TDD requirements traceability matrix is listed below:

– AI evaluation process flow measures

– AI model performance evaluation results

– AI model risks and severity

– AI model bias and fairness measures
– AI model explainability measures
– AI model generalizability measures
– AI model interpretability measures
– AI model robustness measures
– AI model uncertainty measures
– AI assessment platform – computational infrastructure measures
– Intended use
– Device security and privacy
– Patient safety
– Custom/domain specific/clinical effectiveness – scores and measures