ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

Recommendation

ITU-T F.781.1 (06/2024)

SERIES F: Non-telephone telecommunication services

Multimedia services

General framework of quality control of medical images for machine learning applications

ITU-T F-SERIES RECOMMENDATIONS

Non-telephone telecommunication services

TELEGRAPH SERVICE	F.1-F.109
Operating methods for the international public telegram service	F.1-F.19
The gentex network	F.20-F.29
Message switching	F.30-F.39
The international telemessage service	F.40-F.58
The international telex service	F.59-F.89
Statistics and publications on international telegraph services	F.90-F.99
Scheduled and leased communication services	F.100-F.104
Phototelegraph service	F.105-F.109
MOBILE SERVICE	F.110-F.159
Mobile services and multidestination satellite services	F.110-F.159
TELEMATIC SERVICES	F.160-F.399
Public facsimile service	F.160-F.199
Teletex service	F.200-F.299
Videotex service	F.300-F.349
General provisions for telematic services	F.350-F.399
MESSAGE HANDLING SERVICES	F.400-F.499
DIRECTORY SERVICES	F.500-F.549
DOCUMENT COMMUNICATION	F.550-F.599
Document communication	F.550-F.579
Programming communication interfaces	F.580-F.599
DATA TRANSMISSION SERVICES	F.600-F.699
MULTIMEDIA SERVICES	**F.700-F.799**
ISDN SERVICES	F.800-F.849
UNIVERSAL PERSONAL TELECOMMUNICATION	F.850-F.899
ACCESSIBILITY AND HUMAN FACTORS	F.900-F.999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T F.781.1

General framework of quality control of medical images for machine learning applications

Summary

Recommendation ITU-T F.781.1 sets up the initial framework for quality control on medical imaging for machine learning applications, including specifying the workflow of data quality control for machine learning applications, the requirements for medical input images, medical image integration, medical image annotation and criteria on data quality for machine learning applications.

History*

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T F.781.1	2024-06-13	16	11.1002/1000/15928

Keywords

Machine learning, medical images, quality control, workflows.

---

* To access the Recommendation, type the URL https://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, and information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T F.781.1

General framework of quality control of medical images for machine learning applications

1 Scope

This Recommendation gives a framework for the quality control of medical images for machine learning applications. Applications on chest volume computed tomography (CT) and eye disease image are also described.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ISO 12052] ISO 12052:2017, *Health informatics – Digital imaging and communication in medicine (DICOM) including workflow and data management*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 artificial intelligence [b-ISO/IEC 2382:2015]: Branch of computer science devoted to developing data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement.

3.1.2 machine learning [b-ISO/IEC 2382:2015]: Automatic learning, process by which a functional unit improves its performance by acquiring new knowledge or skills, or by reorganizing existing knowledge or skills.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 data annotation: The process of labelling data by humans usable for supervised machine learning.

3.2.2 supervised learning: The machine learning task of learning a function that maps an input to an output based on example input-output pairs.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AI Artificial Intelligence

CT Computed Tomography

ED Euclidean Distance

IoU	Intersection over Union
ML	Machine Learning
MR	Magnetic Resonance
SOP	Standard Operating Procedure

5 Conventions

The following conventions are used in this Recommendation:

– The keywords "should (is required to)" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed;

– The keywords "can (is recommended to)" and "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus this requirement needs not be present to claim conformance;

– The keywords "may (can optionally)" indicate an optional requirement which is permissible, without implying any sense of being recommended. These terms are not intended to imply that the vendor's implementation must provide the option and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with the specification.

6 Background

Digital technologies are considered huge potential promoters in the fields of medicine and public health. While machine learning (ML) and artificial intelligence based technologies hold great promise to reduce human workload, this rapidly developing field also remains uncertain. If the model is poorly designed or the underlying training data are biased or incorrect, errors or problematic results can occur in ML applications. Therefore these technologies can only be used with complete confidence if it has been quality controlled through a rigorous evaluation in a standardized way, especially in the scenarios of the large-scale dataset from multisources.

As one of the most common data modalities in medical diagnosis and treatment, the medical image is targeted in this document to address the following questions of how to:

– Assist the quality control from standard operating procedure;

– Reduce model performance problems caused by inconsistent training data;

– Enable large-scale dataset projects on high diversity of data formats and multi-annotators;

– Improve understanding and train skills for non-professional annotators.

7 Workflow of data quality control

In the process of supervised machine models, data quality is one of the most dependable factors in the performance of computer auxiliary diagnosis. Medical images need to be collected, integrated, processed and annotated in a standardized way, then the raw dataset with manual annotations can be used in training and validation for optimization, as well as testing datasets for the evaluation process. A standard operating procedure for data quality control is shown in Figure 1. During the testing and evaluation of the supervised machine models, unqualified annotations may be identified, which should be relabelled or deleted from the dataset.

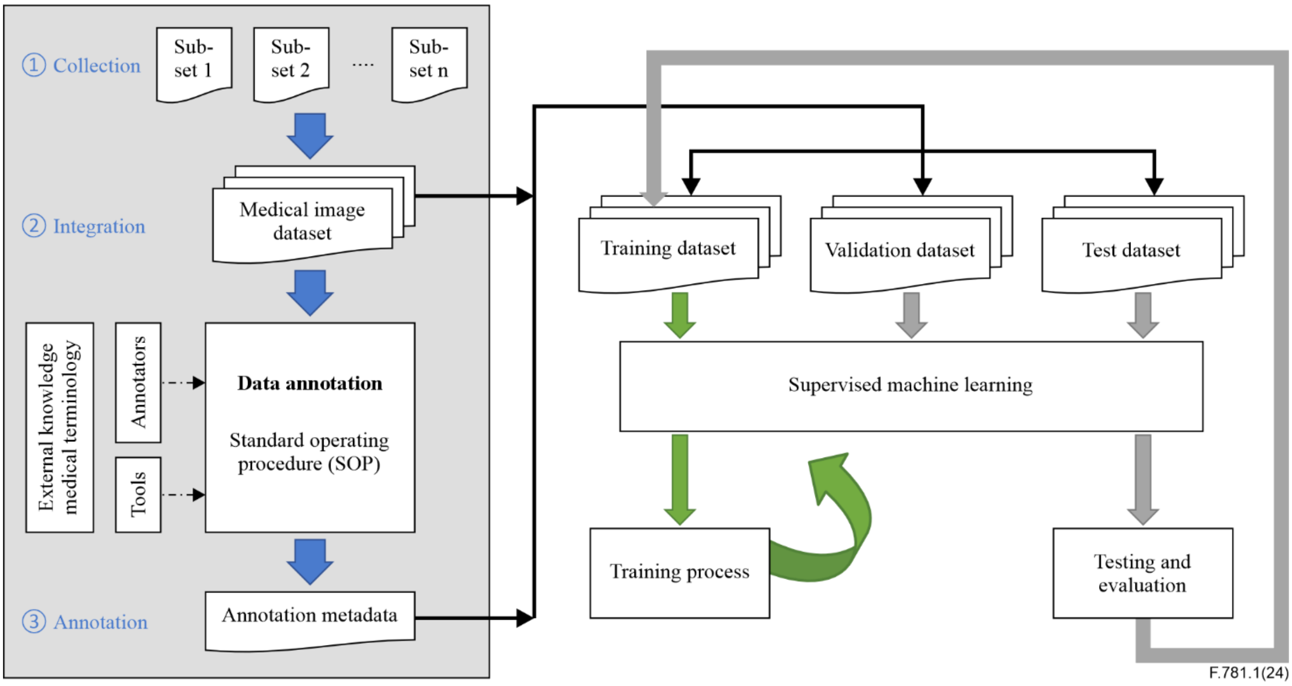


Figure 1 – Workflow of data quality control

8 Requirements of medical input image

Biomedical information evolved with the practice of medicine and engineering technologies at an unprecedented speed through the medical images obtained by human body imaging, high-resolution viewing of cells and pathological specimens. Modalities covered in common measurement include X-ray, ultrasound, magnetic resonance (MR), X-ray, CT, nuclear medicine and high-resolution microscopy, etc. Table 1 refers to the common medical measurement modalities in medical field and Table 2 shows a summary of input data modalities for machine learning applications.

Table 1 – Summary of common medical measurement modalities

	Dimensionality	Description	Anatomies
X-ray	2D, 2D+t	Produces images by measuring the attenuation of X-ray through the body, via a detector array	Most organs
CT	2D, 3D, 3D+t	Creates 2D cross-sectional images of the body by using a rotating X-ray source and detector	Most organs
Ultrasound	2D, 2D+t, 3D, 3D+t	A transducer array emits acoustic pulses and measure he echoes from tissue scatters	Most organs
MR	3D, 3D+t	Use a magnetic field to align protons; radiofrequency and gradient pulses are used to selectively excite protons in tissues and blood in order to measure their spatially encoded unclear magnetic resonance signals	Most organs
Nuclear medicine technology	2D, 3D, 3D+t	Measures the emission of gamma rays from the decay of radioisotopes introduced into the body using external detectors/gamma cameras	All organs with radioactive tracer uptake
Microscopy	2D, 3D, 3D+t	Typically uses an illumination source and lenses to magnify specimens before capturing an image	Primarily biopsies and surgical specimens

Table 2 – Summary of input data modalities for AI4H tasks

Data	Dimensionality	Description	Examples
Image	2D	Two-dimensional medical imaging	– Fundus photos
3D images	3D	Three-dimensional spatial imaging	– Sets of CT slices
4D	4D (3D+t)	3D space imaging changes over time	– Heart film imaging
Video	2D +t	Camera or monitor recording	– Falls among the elderly
Audio/ signal	1D +t	Sound or transmitted in signal form.	– Heart sound/ECG
Text	1D, 2D	Structured/unstructured description in words	– Case history, diagnosis extraction
Single number	1D	Single measurement data	– Blood pressure or respiratory rate

9 Requirements of medical image integration

To avoid bias in medical image collection, it is suggested to consider a balanced data distribution in the process of data preparation and medical image integration.

The image data collected is stored in format compatible with the DICOM system [ISO 12052], for example the one in [b-DICOM PS3.5].

To avoid bias in data distribution, it is suggested that the process be carried out by grouping and crossing, and ensure the effective resolution of inconsistencies. Several annotators (represented as variable *n* in Figure 2) are invited to label the raw dataset. Certain qualifications are required of the annotators, for example, that they be doctors and trained annotators in the specific case domain.

However, for cost considerations, some projects will also set up one annotator (Set *n* to 1) in this parallel independent annotation part, and then go to arbitration if difficulties are encountered.

10 Requirements of medical image annotation

In addition to the quality of the raw image itself, data annotation is one of the most dependable factors in the performance of supervised machine models. The annotation process of medical images is also suggested to be carried out by grouping and crossing. Figure 2 illustrates a recommended standard operating procedure formulated process of data annotation, with much feasibility through variables and configurable threshold, to establish a unified understanding and quality control mechanism in the data annotation process.

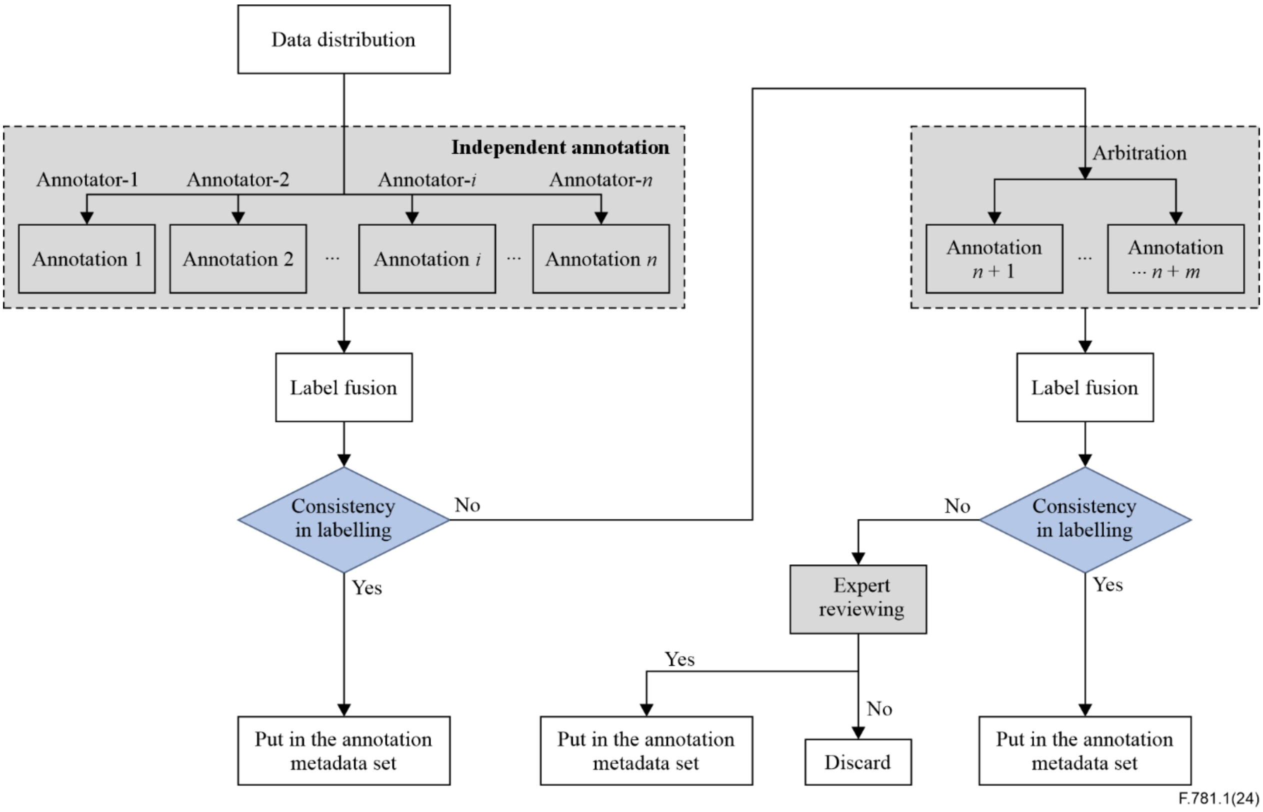


Figure 2 – Data annotation procedure

11 Criteria on data quality for machine learning applications

11.1 Criteria on the type of segmentation for images

For the type of segmentation for images, criteria such as the Jaccard index and Dice's coefficient are recommended to use for detection and segmentation for images. Detailed calculation methods are shown in Table 3.

– The Jaccard index: Jaccard index is also known as intersection over union (IoU) and the Jaccard similarity coefficient, which is a statistic used for gauging the similarity and diversity of sample sets.

– Dice's coefficient: Dice's coefficient is the quotient of similarity and ranges between 0 and 1. This coefficient is not very different in form from the Jaccard index, and they have a connection as $J = D/(2 - D)$, $D = 2J/(1 + J)$.

Table 3 – Criteria on data quality for machine learning

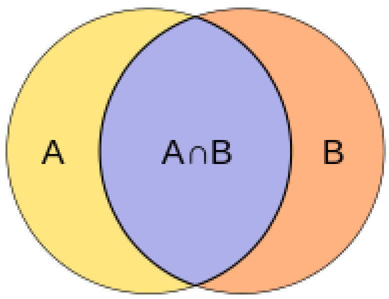
Criteria	Calculation method	Graphical representation
Jaccard index	Numerator represents the area of overlap between two annotations; Denominator represents the area encompassed by two annotations. Dividing the area of overlap by the area of union yields our final score. $J(A,B) = \frac{\lvert A \cap B\rvert}{\lvert A \cup B\rvert} = \frac{\lvert A \cap B\rvert}{\lvert A\rvert + \lvert B\rvert - \lvert A \cap B\rvert}$	

Table 3 – Criteria on data quality for machine learning

Criteria	Calculation method	Graphical representation
		J = $D/(2 - D)$, $D = 2J/(1 + J)$
Dice's coefficient	Numerator represents the double area of overlap between two annotations; Denominator represents the sum of two annotation area. Dividing the area of overlap by the sum area yields our final score. $D = \frac{2\lvert A \cap B \rvert}{\lvert A \rvert + \lvert B \rvert}$	

For the type of localization, criteria like Euclidean distance (ED) are recommended to use for localization. ED is a commonly used definition of distance, which refers to the true distance between two points in the m-dimensional space.

In 2D space, $ED\big((x_1, y_1), (x_2, y_2)\big) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$.

In 3D space, $ED\big((x_1, y_1, z_1), (x_2, y_2, z_2)\big) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$.

11.2 Criteria on data distributions, data consistency and consistency in data annotation process

The data quality standards for machine learning applications should embody the characteristics of authority, scientific rigour, traceability, diversity, closeness and dynamicity.

Authority: Considering that the quality of data depends mainly on the quality of data labelling, the data-collection process should include authorities in the corresponding clinical specialties. Data annotators and annotation disagreement arbitrators should have appropriate and sufficient experience in clinical practice, respectively.

Scientific rigour: To guarantee the training dataset reflect the actual clinical situation, the sample size of the database should be determined by statistical calculations to control sampling error, and the distribution of the samples should be in line with the characteristics of the target diseases, without data augmentation.

Traceability: Quality control procedures documents should be established for data annotation-related activities, which should meet the requirements for traceability. The documents should cover three core phases: requirements clarification and confirmation; staff training and pilot annotation; final quality control and acceptance. Continuous monitoring and checking should be undertaken throughout the process. The project manager should periodically select a certain percentage of the annotated data for detailed quality checks.

Diversity: Data should be sourced from multiple clinical institutions to ensure that the database can be used to evaluate the generalization ability; an appropriate proportion of adversarial data samples may be included to evaluate the robustness of the algorithm, provided that ethical requirements are met.

Closeness: To ensure that the quality of the algorithms can be adequately and objectively evaluated, the database should be kept closed.

Dynamicity: The database should be replaced with a certain percentage of data on a regular basis to ensure that it has continued diversity and closure.

11.3 Criteria on data governance and privacy considerations

A complete framework of privacy policy standards should be established, clarifying the basic content and key terms to be included, providing a reference model for privacy policy formulation and effectively adopting privacy labelling.

It is recommended that the content of privacy terms be displayed in the form of labelling and/or masonry layouts, to enable users to understand the rights and obligations of personal information protection contained in the privacy policy.

Bibliography

[b-ITU FG-AI4H] ITU-T (2023), FG-AI4H Deliverables, *DEL05.3: Data annotation specification*.
https://handle.itu.int/11.1002/plink/1450298673

[b-ISO/IEC 2382] ISO/IEC 2382:2015, *Information technology – Vocabulary*.

[b-DICOM PS3.5] DICOM PS3.5 2024c (2024), *DICOM Part 5: Data Structures and Encoding 2023.*
https://dicom.nema.org/medical/dicom/current/output/pdf/part05.pdf

[b-FDA] United States Food and Drug Administration (2019), *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML) Based Software as Medical Device*, FDA discussion paper.
https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf

[b-NMPA] National Medical Products Administration (2019), *Key Points for the Review and Evaluation of Deep Learning Based Software as Medical Device*, NMPA guidance 28 Jun [in Chinese].
https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww.cmde.org.cn%2Fdirectory%2Fweb%2Fcmde%2Fimages%2Fye62yNGnz7C4qNb6vvay39K9wcbG99C1yO28snzxsDSqrXjLmRvYw%3D%3D.doc&wdOrigin=BROWSELINK

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	**Non-telephone telecommunication services**
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Published in Switzerland
Geneva, 2024