



DPC: 24 / 30504604 DC

BSI Group Headquarters

389 Chiswick High Road London W4 4AL

Tel: +44 (0)20 8996 9000

Fax: +44 (0)20 8996 7400

www.bsigroup.com

Date: 13 November 2024

Origin: International

Latest date for receipt of comments: 1 January 2025

Project No. 2024/03260

Responsible committee: CH/62/1 Common aspects of medical equipment, software, and systems

Interested committees:

Title: BS EN IEC 63524 Artificial Intelligence enabled Medical Devices - Computer assisted analysis software for pulmonary images - Algorithm performance test methods

Please notify the secretary if you are aware of any keywords that might assist in classifying or identifying the standard or if the content of this standard

- i) has any issues related to 3rd party IPR, patent or copyright
- ii) affects other national standard(s)
- iii) requires additional national guidance or information

**WARNING: THIS IS A DRAFT AND MUST NOT BE REGARDED OR USED AS A BRITISH STANDARD.
THIS DRAFT IS NOT CURRENT BEYOND 1 January 2025**

This draft is issued to allow comments from interested parties; all comments will be given consideration prior to publication. No acknowledgement will normally be sent. **See overleaf for information on the submission of comments.**

No copying is allowed, in any form, without prior written permission from BSI except as permitted under the Copyright, Designs and Patent Act 1988 or for circulation within a nominating organization for briefing purposes. Electronic circulation is limited to dissemination by e-mail within such an organization by committee members.

Further copies of this draft may be purchased from BSI Shop <http://shop.bsigroup.com> or from BSI Customer Services, Tel: +44(0) 20 8996 9001 or email cservices@bsigroup.com. British, International and foreign standards are also available from BSI Customer Services.

Information on the co-operating organizations represented on the committees referenced above may be obtained from <http://standardsdevelopment.bsigroup.com>

Responsible Committee Secretary: Ms Lindsey Ferrari (BSI)

Direct tel:

E-mail: lindsey.ferrari@bsigroup.com

Introduction

This draft standard is based on national and international discussions. Your comments on this draft are invited and will assist in the preparation of the consequent standard.

For international standards, comments will be reviewed by the relevant UK national committee before sending the consensus UK vote and comments to the international committee, which will then decide appropriate action. If the international standard is approved, it is usual for the text to be published as a British Standard.

For national standards, comments will be reviewed by the relevant UK national committee and the resulting standards published as a British Standard.

UK Vote

Please indicate whether you consider the UK should submit a negative (with supporting technical reasons) or positive vote on this draft. Please indicate if you are aware of any reason why this draft standard should not be published as a British Standard.

Submission of Comments

- **Annotated drafts are not acceptable and will be rejected.**
- All comments should be submitted online at <https://standardsdevelopment.bsigroup.com/>. You will need to register in order to comment.

Template for comments and secretariat observations

Date: xx/xx/20xx	Document: ISO/DIS xxxx
------------------	------------------------

1	2	(3)	4	5	(6)	7
MB	Clause No./ Subclause No./Annex (e.g. 3.1)	Paragraph/Figure/ Table/Note	Type of comment	Commend (justification for change) by the MB	Proposed change by the MB	Secretariat observations on each comment submitted
	3.1	Definition 1	ed	Definition is ambiguous and needs clarifying.	Amend to read '...so that the mains connector to which no connection...'	
	6.4	Paragraph 2	te	The use of the UV photometer as an alternative cannot be supported as serious problems have been encountered in its use in the UK.	Delete reference to UV photometer.	



62B/1370/CD

COMMITTEE DRAFT (CD)

PROJECT NUMBER:

IEC 63524 ED1

DATE OF CIRCULATION:

2024-11-08

CLOSING DATE FOR COMMENTS:

2025-01-31

SUPERSEDES DOCUMENTS:

62B/1321/NP, 62B/1335/RVN

IEC SC 62B : MEDICAL IMAGING EQUIPMENT, SOFTWARE, AND SYSTEMS

SECRETARIAT:

Germany

SECRETARY:

Ms Regina Geierhofer

OF INTEREST TO THE FOLLOWING COMMITTEES:

TC 62, SC 62A, SC 62C, SC 62D

HORIZONTAL FUNCTION(S):

ASPECTS CONCERNED:

Safety

This document is still under study and subject to change. It should not be used for reference purposes.

Recipients of this document are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

Recipients of this document are invited to submit, with their comments, notification of any local regulations or technical reasons that may exist and should be considered should this proposal proceed, recognizing that failure to address such requirements could result in the need for new or existing "In Some Countries" clauses. (SEE [AC/22/2007](#) OR [NEW GUIDANCE DOC](#)).

TITLE:

Artificial Intelligence enabled Medical Devices - Computer assisted analysis software for pulmonary images - Algorithm performance test methods

NOTE FROM TC/SC OFFICERS:

Copyright © 2024 International Electrotechnical Commission, IEC. All rights reserved. It is permitted to download this electronic file, to make a copy and to print out the content for the sole purpose of preparing National Committee positions. You may not copy or "mirror" the file or printed version of the document, or any part of it, for any other purpose without permission in writing from IEC.

CONTENTS

FOREWORD	3
1 Scope	5
2 Normative references	5
3 Terms and definitions	5
4 Abbreviated terms	7
5 Overview	7
6 Algorithm performance test methods	8
6.1 Test methods for different application scenario	8
6.1.1 Computer-aided detection	8
6.1.2 Segmentation and measurement	10
6.1.3 Classification	12
6.1.4 Multifunctional Combination Scenario	16
6.1.5 Follow-up Assessment Scenario	16
6.1.6 Patient Triage Scenario	17
6.2 Test methods for quality characteristics	17
6.2.1 General requirement for generalizability test	17
6.2.2 Robustness test	17
6.2.3 Repeatability	18
6.2.4 Consistency	18
6.2.5 Efficiency	18
6.2.6 Analysis of algorithm errors	19
Annex A (informative) General consideration on preparation of algorithm test	20
A.1 Test environment	20
A.2 Test resource	20
A.2.1 General requirement on test set	20
A.2.2 Sample size	20
A.2.3 Test set configuration	20
A.2.4 Data synthesis	21
A.2.5 Phantom and devices	21
A.3 Test tools	21
A.4 Performance metrics and passing criteria	22
A.5 Test procedure	22
A.6 Reporting of test results	23
Annex B (normative) Example of testing set description of chest CT pulmonary nodule	24
B.1 Overview	24
B.2 Application Scope of Datasets	24
B.3 Data collection	24
B.4 Distribution and Composition of Dataset	25
B.5 Annotation Rules of Datasets	25
B.6 Estimation of Sample Size	27
B.7 Testing Set Bias Analysis	27
Annex C (normative) General consideration of Performance Indicators and Statistical Analysis	28
C.1 Overview	28

C.2	Scenario I: Test results are variables of binary classification	28
C.3	Scenario II: Test results are ordered or continuous variables	29
C.3.1	Ordered variables	29
C.3.2	Continuous Variables	31
C.3.3	Scenario III: Test results involve image positions	32
C.3.4	Hypothesis Test of Main Indicators	33
C.4	Sample size determination for test set	33
Bibliography		35
Figure B.1 – Flow Chart for Annotation of Pulmonary Nodule		26
Table 1 – n-classification Confusion Matrix		12
Table 2 – Binary-classification Confusion Matrix		13
Table 3 – Binary confusion matrix deducted from multi-classification situation		14
Table B.1 – Diversity Statistics of Data Sources		24
Table B.2 – Statistics of Distribution of Pulmonary Nodules		25
Table C.1 – Confusion matrix of binary classification		28
Table C.2 – Ordered data structure of diagnostic test		29

INTERNATIONAL ELECTROTECHNICAL COMMISSION

**ARTIFICIAL INTELLIGENCE ENABLED MEDICAL DEVICES - COMPUTER
ASSISTED ANALYSIS SOFTWARE FOR PULMONARY IMAGES -
ALGORITHM PERFORMANCE TEST METHODS –**

FOREWORD

- 1) The International Electrotechnical Commission (IEC) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, IEC publishes International Standards, Technical Specifications, Technical Reports, Publicly Available Specifications (PAS) and Guides (hereafter referred to as "IEC Publication(s)"). Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
- 2) The formal decisions or agreements of IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested IEC National Committees.
- 3) IEC Publications have the form of recommendations for international use and are accepted by IEC National Committees in that sense. While all reasonable efforts are made to ensure that the technical content of IEC Publications is accurate, IEC cannot be held responsible for the way in which they are used or for any misinterpretation by any end user.
- 4) In order to promote international uniformity, IEC National Committees undertake to apply IEC Publications transparently to the maximum extent possible in their national and regional publications. Any divergence between any IEC Publication and the corresponding national or regional publication shall be clearly indicated in the latter.
- 5) IEC itself does not provide any attestation of conformity. Independent certification bodies provide conformity assessment services and, in some areas, access to IEC marks of conformity. IEC is not responsible for any services carried out by independent certification bodies.
- 6) All users should ensure that they have the latest edition of this publication.
- 7) No liability shall attach to IEC or its directors, employees, servants or agents including individual experts and members of its technical committees and IEC National Committees for any personal injury, property damage or other damage of any nature whatsoever, whether direct or indirect, or for costs (including legal fees) and expenses arising out of the publication, use of, or reliance upon, this IEC Publication or any other IEC Publications.
- 8) Attention is drawn to the Normative references cited in this publication. Use of the referenced publications is indispensable for the correct application of this publication.
- 9) IEC draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). IEC takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, IEC [had/had not] received notice of (a) patent(s), which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at <https://patents.iec.ch>. IEC shall not be held responsible for identifying any or all such patent rights.

IEC 6XXXX has been prepared by subcommittee 62B: Medical imaging equipment, software, and systems, of IEC technical committee **TC62**: Medical equipment, software, and systems. It is an International Standard.

a) ...;

The text of this International Standard is based on the following documents:

Draft	Report on voting
XX/XX/FDIS	XX/XX/RVD

Full information on the voting for its approval can be found in the report on voting indicated in the above table.

The language used for the development of this International Standard is **English** [change language if necessary].

This document was drafted in accordance with ISO/IEC Directives, Part 2, and developed in accordance with ISO/IEC Directives, Part 1 and ISO/IEC Directives, IEC Supplement, available at www.iec.ch/members_experts/refdocs. The main document types developed by IEC are described in greater detail at www.iec.ch/publications.

The committee has decided that the contents of this document will remain unchanged until the stability date indicated on the IEC website under webstore.iec.ch in the data related to the specific document. At this date, the document will be

- reconfirmed,
- withdrawn, or
- revised.

ARTIFICIAL INTELLIGENCE ENABLED MEDICAL DEVICES - COMPUTER ASSISTED ANALYSIS SOFTWARE FOR PULMONARY IMAGES - ALGORITHM PERFORMANCE TEST METHODS –

1 Scope

This document describes algorithm performance test methods of computer assisted analysis software for pulmonary radiological images based on artificial intelligence technology.

This document describes the method of standalone performance test.

This document is applicable to medical device that uses artificial intelligence to support the post-processing of pulmonary radiological images. The algorithm input include but is not limited to X-ray, CT (computed tomography) and MRI (magnetic resonance imaging)

This document is not applicable to medical device that uses artificial intelligence to support pre-processing or procedure optimization of clinical workflow.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

IEC 62304:2006/AMD1:2015, *Amendment 1 - Medical device software - Software life cycle processes*

ISO/IEC TS 4213,

ISO 14971,

IEC 63521 Machine Learning-enabled Medical Device – Performance Evaluation Process

IEC CD 63450 Testing of Artificial Intelligence / Machine Learning-enabled Medical Devices

ISO CD 24971-2 Medical devices — Guidance on the application of ISO 14971 — Part 2: Machine learning in artificial intelligence

3 Terms and definitions

3.1

baseline image

image used as the reference among a series of examinations of the same patient

3.2

follow-up image

image acquired at scan during follow-up visit

3.3

repeat screening

screening repeated at regular intervals

3.4 sign

clues and indications that can be usually obtained through objective measurement and provided for doctors to learn about medical progress and disease status during physical or pathological examination

3.5 signs in radiology

signs obtained through radiology means

3.6 stress test

test in which long-term extreme input or environment is applied

Note 1 to entry: Examples of such input or environment include but are not limited to heavy computational load, boundary values, large capacity data, rare data.

Note 2 to entry: Inputs in stress test are often called stress samples, which may exist in the real world.

3.7 reference standard ground truth

benchmarks to compare with the output of AI algorithm

Note 1 to entry: Often used interchangeably with ground truth.

Note 2 to entry: There are different approaches to acquire reference standard, such as diagnosis, treatment and annotation.

[SOURCE: Modified from [IEEE 2802:2022 \[1\]](#)]

3.8 repeatability

the degree of consistency between multiple independent test/measurement results obtained by the same operator in the same way, using the same test or measurement facilities, from the same subject

[SOURCE: Modified from IEEE 2802-2022 [IEEE 2802:2022 \[1\]](#)]

3.9 reproducibility

the degree of consistency between the independent test / measurement results obtained by different operators using a standardized methodology, using the same test or measurement facilities or equipment, from the same subject

[SOURCE: Modified from [IEEE 2802:2022 \[1\]](#)]

3.10 synthetic data

data that is artificially generated rather than produced by real-world events

[SOURCE: IEC CD 63450 Testing of Artificial Intelligence / Machine Learning-enabled Medical Devices]

3.11 augmented data

data created from an original data set by introducing minor transformations to the original data set to increase the size and diversity of the data set

Note 1 to entry: Augmented data is a subset of synthetic data.

3.12

data augmentation

a method applied to a limited set of data to increase the data set by creating modified copies using digital transformations that introduce minor changes to the original data set

Note 1 to entry: Data augmentation is usually conducted in an understandable manner, such as rotation, segmentation, superposition noise/artifact, superposition filter and reconstruction of radiological images.

3.13

mark-labelling

the scoring criterion used to determine whether a region of interest predicted by AI model correctly correlates with the region of interest annotated by human

4 Abbreviated terms

AI	artificial intelligence
AIMD	artificial intelligence medical device
AP	average precision
AUROC	area under the receiver operating characteristics curve
AUT	algorithm under test
FN	false negative
FP	false positive
FROC	free-response receiver operating characteristics curve
IMDRF	International medical device regulation forum
MAP	mean average precision
ML	machine learning
ROC	receiver operating characteristics curve
ROI	region of interest
SAMD	software as medical device
SIMD	software in medical device
SUT	software under test
TN	true negative
TP	true positive

5 Overview

General testing principles for medical devices software as medical device(IEC 62304:2006/AMD1:2015), software in medical device(SiMD) and artificial intelligence medical device(IEC CD 63450 Testing of Artificial Intelligence / Machine Learning-enabled Medical Devices and IEC 63521 Machine Learning-enabled Medical Device – Performance Evaluation Process) are applicable to artificial intelligence medical device. This standard covers particular requirements for the testing of artificial intelligence algorithm intended to analyze pulmonary radiologic images.

As an important step for verifying and validating the computer assisted analysis software for pulmonary images, the algorithm performance test is generally based on test sets and quantitatively compares the output result of the algorithm with the ground truth so as to obtain specific indicators such as false positive and false negative, repeatability and reproducibility, robustness, efficiency, etc.

This document describes the method of standalone performance test. To fulfill IMDRF essential performance principles, quality characteristics specific to [AIMD](#) is also considered in this

document, including generalization, robustness, repeatability, consistency, and efficiency. Analysis of algorithm errors during the test is also recommended in this document, which may provide further information to evaluate other quality attributes.

To avoid intentional optimization or overfitting on a test set, reuse of test sets are limited to keep the software under testing(SUT) from learning through the testing process.

General consideration to prepare algorithm test is described in Annex A.

6 Algorithm performance test methods

6.1 Test methods for different application scenario

6.1.1 Computer-aided detection

6.1.1.1 Mark-labelling

The product with target detection function should output the algorithm marking target and the reference standard of test sets should include the corresponding ROIs. Testing personnel should record the matching method and threshold between the algorithm marking target and the reference standard target in the test plan, wherein the matching method and threshold should be claimed by relevant manufacturer.

NOTE 1 The matching threshold is different from the probability of detection (POD) of the algorithm.

Examples of common mark-labelling criteria include:

- a) Region overlapping: Determine the matching result by calculating the overlapping degree (such as Dice coefficient and Jaccard coefficient) between the algorithm marking target and the reference standard region and setting the matching threshold.
- b) Distance from the central point: Determine the matching result by calculating the distance between the algorithm marking target and the center of the reference standard region and setting the matching threshold.
- c) Central point hitting: Determine the matching result by judging whether the center of the region of interest (ROI) labeled by the algorithm falls into the range of the reference standard region.

NOTE 2 The selection of the central point is associated with the ROI. The shape of the ROI is usually determined by signs, especially the signs in radiology. For example, the central point of a pulmonary nodule (usually convex) is the intersection between the long diameter and short diameter within the range of the detection region. The long diameter is defined as the distance between the two farthest points of the maximum cross-sectional space within the range of the detection region. The short diameter is defined as the longest distance perpendicular to the long diameter in the nodule.

The matching results are divided into three cases:

- a) True positive, number of positive instances that have been properly classified as positive, recorded as **TP**
- b) False positive, number of negative instances that have been wrongly classified as positive, recorded as **FP**;
- c) False negative, number of positive instances that have been wrongly classified as negative, recorded as **FN**.

Treatment under special situation: When many-to-one matching occurs, the matching relationship should be subject to the following priority considerations:

a) If the region overlapping method is adopted, the higher region overlapping degree should be selected;

b) If the central point distance method is adopted, the smaller central point distance should be selected;

c) If the central point falling method is adopted, the smaller central point distance should be selected.

6.1.1.2 Test steps

a) Import the test set to the SUT and export results predicted by AI. The format of AI output should be compatible with reference standard.

b) Obtain TP, FP and FN according to the rule in 6.1.1.1;

c) Calculate performance metrics as described in 6.1.1.3 to 6.1.1.8. Statistical calculation can be based on different units such as lesions and cases. If a case contains multiple lesions, the performance metric within a case can be calculated.

6.1.1.3 Recall

See ISO/IEC 4213 6.2.4

6.1.1.4 Precision

See ISO/IEC 4213, 6.2.4

6.1.1.5 F-score

See ISO/IEC 4213, 6.2.5

6.1.1.6 Average precision

Change the setting of algorithm thresholds and calculate the precision and recall rate corresponding to each threshold. Take the recall rate as x-coordinate and the precision as y-coordinate to generate a precision-recall rate curve, and calculate the integral area under the curve, namely the average precision. Whether the curve is smoothed and the method adopted should be described in the test records.

6.1.1.7 Mean average precision

If multi-target detection is applicable, the average precision of various targets should be calculated, the mean value of which should be recorded as the mean average precision.

6.1.1.8 fROC curve

Change the setting of algorithm thresholds and calculate the recall and non-lesion location rate corresponding to each threshold. The fROC curve is plotted by setting the recall as y-coordinate and the non-lesion location rate as x-coordinate. The value of the x-coordinate is generally set as a geometric progression, i.e., 0.5, 1, 2, ..., , where the upper limit should be greater than the average number of lesions of the individual case. Taking the auxiliary detection of pulmonary nodules as an example, assuming that a single case has an average of 7.5 pulmonary nodules, the value of N should not be less than 8.

The non-lesion location rate should be expressed as NLR. See [Formula \(1\)](#) for its expression: