ISTA
Seed Quality Assurance

ISTA Secretariat
Richtiarkade 18, 8304 Wallisellen, Switzerland
Phone: +41 44 838 60 00 | Fax: +41 44 838 60 01
Email: ista.office@ista.ch
www.seedtest.org

Guideline
Advanced Technology Applications for Seed Testing
Computer vision

Any hardcopy of this document is not controlled copy

CREATED	**REVIEWED**	**APPROVED**
19.12.2024	FEBRUARY 2025	02.05.2025
ATC, PUR, GER, VIG, STA, PTC	ISTA ACCREDITATION AND TECHNICAL DEPARTMENT, ECOM	ECOM

SCOPE

The scope of this document is to assist the ISTA accredited laboratories in applying advanced technologies in seed testing and to establish common expectations, benchmark quality and provide best practices to meet the principles of the ISTA Accreditation Standard.

OBJECTIVES

Produce an operational guide or advice for the development, verification, and application of computer vision (equivalent to analysts) tools in seed testing, such as computer classification of seeds in purity analysis, other seed determination (OSD), seedlings in germination test, or seed in varietal purity analysis.

Provide clear verification expectations of the user laboratories and support the audit team in evaluating the process.

Specify the minimum quality control expectation for the development, verification and application of computer vision (equivalent to analyst) tools in seed testing.

Facilitate the speedy or prioritised applications of computer vision (equivalent to analyst) in seed testing by sharing common or best practices in computer vision application in seed testing and collaborating among users in the seed testing community.

RELATED DOCUMENTS

ISTA Accreditation Standard

International Rules for Seed Testing (ISTA Rules)

PT-P-01-ISTA Standard Proficiency Test

TCOM-P-01-ISTA Method Validation for Seed Testing

RESPONSIBILITIES

ISTA accredited laboratories, Technical Committees (TCOMs), the lead of tool development.

Based on the test application of computer vision or advanced technology, the relevant TCOMs could be consulted for guideline adherence or technical advice in evaluating quality evidence for the reported performance parameters.

The ISTA Accreditation and Technical Department: supports the implementation of the ISTA Accreditation Standard requirements related to new technologies.

ISTA auditors: perform the evaluation of the accredited laboratories, including the use of new technologies, where applicable.

ABBREVIATION

ATC: Advanced Technologies Committee

AI: Artificial intelligence

GER: ISTA Germination Technical Committee

OSD: Other Seeds Determination

PUR: ISTA Purity Technical Committee

RUL: ISTA Rules Technical Committee

STA: ISTA Statistics Technical Committee

VIG: ISTA Vigour Technical Committee

SOP: Standard Operational Procedure

PTC: Proficiency Test Committee

DEFINITION:

Computer Vision: Computer vision is a field of artificial intelligence (AI) that uses machine (deep) learning and neural networks to teach computers and systems to derive meaningful information from digital images, videos and other visual inputs. The AI powered computer system enables a system to think, to see, to analyse, and to make recommendations.

Tool development: In this guide, tool development refers to the process of creating AI software tools or applications that assist in various aspects of software development, testing, and maintenance. Tool development aims to address specific needs and challenges faced by developers through processes such as data collecting, coding, debugging, testing, documentation, and deployment, i.e. the research and development stage of an AI application.

Tool Verification: Tool Verification is a crucial process used to ensure that a product, service, or system meets specific requirements and specifications, as well as fulfils its intended purpose. It plays a vital role in quality management systems, such as the ISTA Accreditation Standard.

Performance Indicator: a type of performance measurement that evaluate the success of a particular activity (such as testing) conducted by the method or system developed.

Analyst: Trained personnel of an accredited seed testing laboratory, who are approved to conduct independent testing work. The analysts must have and maintain the necessary education, training, technical knowledge, demonstrated skills to perform authorised analysis to allow the issue of a seed testing report or certificate, e.g., ISTA certificates in case of ISTA accredited laboratories.

PROCESS DESCRIPTION

Lab-developed or customised software requires verification to meeting required accuracy. According to ISTA accreditation standard, *(5.2.7. Each equipment and its software used for testing that is significant to the result must, when practicable, be uniquely identified. Records must be maintained of each major item of equipment and its software).* Here are recommended records of, but not limited to,

a) Overview of tool application scope and performance indicators (Annex 1)

b) Original observations of tool development on performance indicators (Annex 2)

c) Tool performance equivalency verification with single or multiple laboratories (Annex 3)

However, a developed tool, can be shared, or transferred directly to other laboratories with no changes to the application scope. Care must be taken to make sure that the obtained results are not changed when the shared information is implemented in other laboratories. See more details in tool verification and deployment process. The copy of the original document specified in a) and c) should be on file in the technology transferred laboratories. Any method modification or scope expansion will require the support document specified above.

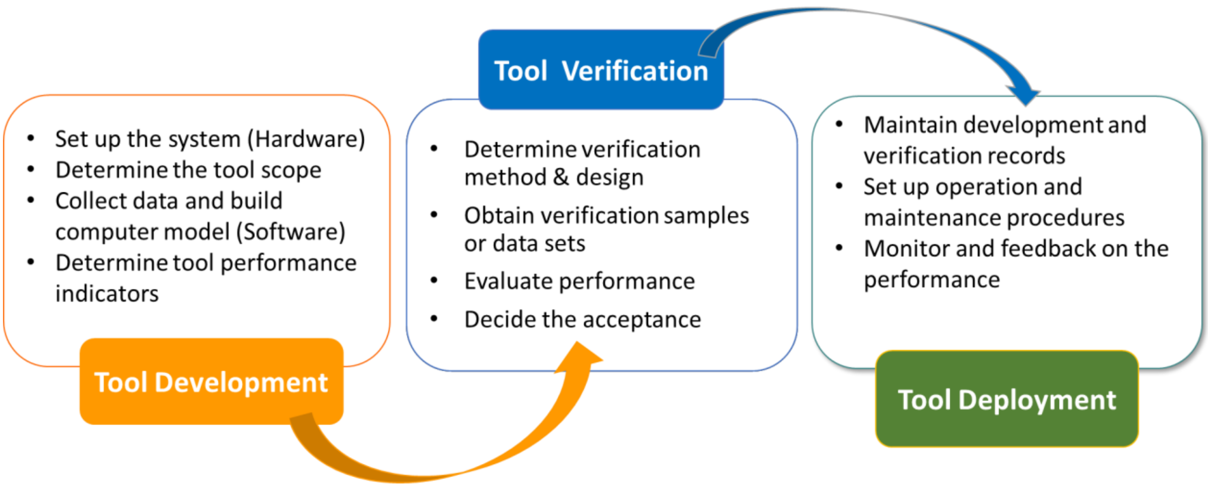


Figure 1: Overview of the tool application process*

*Please note, when a tool requires modification or scope expansion, the same process will need to be repeated but possibly not to the same extent as required in development stage.

General principles

Tool objective and scope intended:

When a tool is in development, the objectives and intended scope, such as kind of analysis and target species must be identified. This clarification will guide the development process for material selection, trial or training plans, and tool deployment steps. For example, if a tool is designed for a full automation system for OSD, sample handling, weighing, and target species must be defined. In contrast, if it is a semi-automation system, a human and machine integration process must be designed. A SOP describing the process is required in QA process.

Because the diversity of plant or crop species involved in seed testing, it is beneficial for an efficient tool development and application with a defined and achievable scope for prioritised business needs. An assessment of the advantage of computer vision vs manual analysis should be conducted to optimise the investment of technology.

In the meantime, tool development should consider the potential needs and plans for possible tool scope expansion, such as increased capacity for analysis and new sample types. The consideration can be reflected in data architecture, storage, labelling, and organisation, as well as computer modelling development.

Equipment and AI Methods:

A computer vision system must include an imaging system, generally referred to as hardware, which will generate raw data, such as taking images of seeds or seedlings using different light sources from whole light spectrum, e.g., visual light (Red, Green, and Blue (RGB) light), multi-, hyper- spectral light, X-ray, infra-red, UV light.

Images of seeds or seedlings from different light sources will be then processed by image segmentation, i.e. datasets, and computer modelling using machine learning or deep learning technologies. These two processes are usually referred to as software development, that is the machine has been trained to achieve the designed function and accuracy of the tool.

Currently, the accuracy of machine (deep-) learning will require a system coordination on similarity in image type (e.g. X-ray or RGB) and image quality (e.g. image resolutions and completeness of image data, such as both sides of asymmetrical seeds or dimorphic seeds) between models' development and end-use applications.

Materials:

To ensure the tool's accuracy, machine (deep-) learning usually requires a large amount of well-labelled and constructed datasets. Human expertise must be integrated into the material selection and preparation, which is the foundation to be equivalent to human analysts testing results. Meantime, natural variation in working samples is to be considered for operational application.

For example:

- Non-cultivated (wild) species usually have large natural variations, the inclusion of a wide range of specimen variations will ensure tool application accuracy in real operations.
- Cultivated species may have distinct variety types, essential for varietal purity testing and the materials for tool development can be defined with this scope or incorporate the varietal types in the tool datasets, for example analytical %purity or OSD testing.
- Classification of testing results, such as seedlings in a germination test or stain of colourations in tetrazolium (TZ) test, as high-quality datasets for tool development, requires the analysts' expertise to interpret the ISTA Rules and to provide a consensus from several experienced analysts. The material preparation should also include the variations of normal and abnormal seedlings or staining patterns. The most efficient way is to develop a shared dataset that passes evaluation by several accredited laboratories or qualified analysts.
- The background interference with the imaging datasets should be considered by selecting an appropriate testing method (e.g. germination paper colour or using transparent agar media if it is permitted by the testing rules used) or filtering out by imaging process (e.g. X-ray to detect roots in the germination media) or image segmentation process (e.g. removing the image background with software).

Outcomes:

The tool using computer vision for seed testing is expected to be equivalent, at minimum, to manual analysis following the ISTA Rules. The development stages must avoid biases of the system. An equivalence evaluation must be conducted at the tool final verification before deployment.

For the accuracy expectation, please see the section “Tool verification” Annexes 2 & 3.

Maintenance:

The system hardware should be calibrated, verified or maintained at the optimum operation status following manufacturer’s recommendation or in-house best practices, e.g. conduct calibration or verification after each software update, developed in-house verification samples based on typical usage, user feedback.

Tool Verification

The tool development stage of a computer vision tool involves extensive training and cross examination with a large and robust dataset. This dataset is comprised of numerous seed samples or specimens with diverse representation across factors that may include but are not limited to different varieties from different production regions, and species (for other seed determination). This extensive work in the tool development stage provides confidence for a robust computer vision tool that can be verified with a smaller well-designed set of samples.

For ISTA accreditation purposes, a minimum of a set of 30 seed samples is recommended with the minimum required working weight or specimens of in species from the defined scope of testing need to be identified by the laboratory for computer vision tool verification. Different processes, steps and assessment criteria are needed for cases where the seed quality characteristic is "known" (e.g. OSD) versus "unknown" (e.g. % germination).

Known Seed Quality Characteristic

In cases where the seed quality characteristic is known, the impure seeds are spiked into the 30 samples at levels that represent a reasonable range in terms of impurity levels and types (i.e. species). The species and impurity levels in samples are unknown (blind) to the lab staff that oversee the computer vision tool measurements. The 30 blind samples are run with computer vision tool and should achieve 90% or better accuracy (correct classification for the tool scope) in the 30 samples scored.

Unknown Seed Quality Characteristic

In cases where the seed quality characteristic is unknown, the 30 samples coming from different seed lots are prepared by the laboratory and divided into a working sample that is evaluated by the computer vision tool measurement and human analyst(s). Samples coming from real production seed lots may be considered for tool verification. To reduce the effects of bias that can be present for human analysts, for example, the thirty samples should be divided among at least three human analysts in the lab (e.g. each of three human analysts evaluates 10 of the samples). The laboratories could use analysts from different laboratories if they choose to further reduce possible bias due to analysts. The 30 paired measurements (by computer vision and by analysts) are analysed using an appropriate statistical test such as a paired t-test to evaluate statistically significant differences between the two measurements at an 0.05 significance level to assess bias. This is followed by an assessment of whether the deviations between the analyst and computer vision for each sample are within reasonable limits. If the two measurements pass these tests, then the computer vision tool passes the verification step. The details of these statistical tests are in Annex 4.

One important point is to remember that the tool verification is not the only evidence of the fitness of the computer vision tool. There is the rigorous tool development work, other audit requirements and participation in ISTA proficiency tests (if possible, when the tool's scope and PT match), that add strength to computer vision tool fitness. This proposed tool verification approach is like the Performance Data Evaluation (PDE) used for accreditation of ISTA GMO testing laboratories.

Tool maintenance and verification, monitoring post verification or deployment

After deployment the continued performance of the tool must be systematically monitored for its on-going testing performance. When testing accuracy is found to be below the expected accuracy (e.g. less than 90% for OSD or out of tolerance results or a trend is detected), C or Below Minimum Performance (BMP) in PTs rating, human intervention must be conducted to investigate the system's performance. Here are some situations where the tool must be re-verified.

a. If testing Rules are updated that has impact on outcome or results of the tool.
b. If the software of the operating system of the hardware has significant updates from the equipment suppliers.
c. If the sample variations are greater than the modelling datasets, e.g. the tested sample is a new variety with significant feature changes.

Tool monitoring is performed based on the ISTA Accreditation Standard requirement 10.3 and must be planned and reviewed according to the laboratory's SOP.

Tool scope modification and expansion

When there is a need for new development, the tool's scope may require expansion.

For example,

a. If testing ISTA Rules are updated to ensure the tool uses the revised rules accurately, e.g. pure seed definition or seedling classification are changed.
b. Testing scope expansion for inclusion of more variations of testing samples, such as new varieties being tested.
c. Testing scope expansion for analysing more crops or species.

Accreditation standard requirements

The laboratory's quality documentation must incorporate the elements when using AI-assisted machine to replace the analyst for example the internal audit, management review, monitoring and training. QA records for the development and on-going verification of the tool must be retained for as long as the tool is in use and a minimum of six years after that.

- The relevant paragraphs of the ISTA Accreditation Standard are: 3.18, 4.3, 4.5, 4.6, 5.2. and 5.3 all paragraphs, 7.1, 7.2, 7.3, 7.5, 8.1, 9.1, 9.2, 9.3, 10.3.1, 10.5.2; there may be paragraphs where it is not specifically written "machinery", but most of the requirements will be applicable to an AI-tool as for an analyst.
- The performance criteria for the AI machine must be the equivalent to analysts (see 4.4).

- The monitoring plan for the machine must be defined and be the same as for analyst (see requirement 10.3.1)
- Corrective actions must be clearly defined and described in case the machine does not match with the performance criteria.
- Keep records of all initial approvals. These records of initial approval must be stored as long as the equipment is in use and a minimum of six years after that to meet the requirement of the ISTA Accreditation Standard.
- Have a document (e.g. Annex 1) for the computer vision on what species of other seed species can be identified, and which cannot (the "unknown" category and must be identified by a human evaluator). This list of other seeds must be in conformity with the *ISTA List of Stabilised Plant Names* (see 4.3).
- Unique identification of the machine must be ensured and also listed in the list of the laboratory equipment with expiry, verification/ monitoring due dates indicated in a list or on equipment labels.
- Make sure the seed used to "train" and subsequently verify the machine has been created by and is controlled by an experienced analyst.

Operator training and training records

a. When a computer vision system is developed for testing, a method description and operation procedure must be available for the operator. Please check Annexes for data/records requirements.
b. An operator must have training records of:
    - Operational procedures for the system, including sample handling and contamination prevention and result interpretation.
    - Hardware operation menu and procedures, i.e. training on how to use the imaging equipment.
    - Monitor the system performance during operational tests or real application and how to provide human intervention when there is a need due to the machine not achieving the expected performance.

Tool calibration, verification and maintenance procedure

a. The laboratory must clearly identify the calibration and verification requirements, procedures, and intervals to ensure the stable and best performance of the computer vision system.
b. Regular calibration or verification must follow the equipment manufacturer's recommendation and/or in-house or quality control practices to ensure stable and the best performance of the computer vision system.
c. Regular maintenance of the tool must be specified in the operational procedures.

Tool software maintenance procedures

a. If the software is updated revalidation is required as specified in the section “Tool Verification”.

b. When the operating system is updated, calibration and adjustment of hardware and/or software performance must be verified.

   If there are issues with the Reference Database data acquisition and storage or system security issues investigation is needed, and revalidation or verification may be necessary.

DISTRIBUTION LIST

ISTA Secretariat
ISTA Auditors
ISTA Technical Committees
ISTA Member and Accredited Laboratories

ANNEXES

Annex 1 Tool overview of application scope and performance indicators
Annex 2 Original observations* of tool development on performance indicators
Annex 3 Tool performance verification* with single or multiple laboratories
Annex 4 Computer vision: verification for unknown seed quality characteristics

REVISION HISTORY

Version #	Changes

Annex 1: Tool overview of application scope and performance indicators
***Notes:** the grey text are hypothetical examples

Category	**Description**	Notes
Name of the tool:	Seed AI Analyzer	
Laboratory Name:	ABC Seed Testing, Inc.	
Scope of the tool:	e.g., Digital RGB computer vision system for OSD in barley seed sample	OSD species: *Brassica* spp. *Cuscuta* spp. *Datura stramonium*
Sample taxon names	e.g., *Hordeum vulgare* L. subsp. *vulgare*	
Name of the tests	e.g., OSD	
Additional specifications, please specify:	e.g., sample percentage purity range: ≥90%	
Performance indicators:		
Accuracy:	e.g., accuracy % in seed id ≥90%	Accuracy verified by: • ISTA PT sample • Blind samples • Peer accepted data set
Others, please specify:	e.g., precision % ≥90%	
Releasing/deployment date:		
Tool Modification, if any		
Modification Date:		
Scope changed:		
Performance indicator if any changes	e.g., accuracy ≥95%	
Operational Procedures		
SOP ID #		
Responsible Person for Tool Operation		
Maintenance requirement	e.g., specified in SOP xxx	
Other notes:		

Conclusion/date:

Responsible:

Annex 2: Original observations* of tool development on performance indicators

***Notes**: 1) When the tool scope is expanded or modified, multiple Annex 2, copies must be kept on record according to ISTA Accreditation Standards

2) Examples, see Annex 1.

Performance Evaluation	Description or original data summary	Notes
Name of the tool		
Laboratory Name		
Scope of the tool:		
Sample taxon names		
Tests		
Sample application range		
Additional specifications, please specify:		
Performance Indicators		
Accuracy		
Others, please specify:		
Tool Development date:		
The lead of the tool development		

Conclusion/date:

Responsible:

Annex 3: Tool performance verification* with single or multiple laboratories

***Notes**: When the tool scope is expanded or modified, multiple copies of Annex 3 must be kept on record according to ISTA accreditation standards.

Performance Evaluation	Description or original data summary	Notes
Name of the tool		
Laboratory Name		
Scope of the tool:		
Sample taxon names		
Tests		
Additional specifications, please specify:		
Tool Verification Method		
Single lab		
Multiple labs		
Sample type or description		
Verification Performance:	e.g., Mean of verification Tool Accuracy Tool Z-score	e.g., Meet >90%
		e.g., Within ±2
Other performance indicators, please specify:		
Verification date		
Tool Verification Organizer:		

Conclusion/date:

Responsible:

Annex 4: Computer vision: verification for unknown seed quality characteristics

At a minimum 30 blind samples (i.e., the lab staff who oversee the computer vision tool measurements and manually analyse the results do not know the origin or specific characteristics of the seed lot) from different lots are evaluated across multiple human analysts (i.e. each sample is evaluated by one of at least three different analysts) and by the computer vision tool. in this example the true value of the seed characteristic is unknown, as in a germination test. To determine the agreement between the two measurements, the approach developed by Bland and Altman, 1986 is used. Let x_i be the measurement on sample *i* by a human analyst, and y_i the measurement on this sample by the computer vision tool.

First, focus on the overall bias, and to do this, perform a paired t-test using x_i and y_i. If the test result is not significant at the 5% significance level, then proceed to the next step; otherwise, the computer vision tool cannot be validated, either there was an issue with the human analysts, or there is a problem with the computer vision tool.

Secondly, assess whether the results agree by determining the Limits of Agreement (LoA):

$$\text{Lower LoA} = \text{Bias} - 1.96 \times s_d$$

$$\text{Upper LoA} = \text{Bias} + 1.96 \times s_d$$

where $\text{Bias} = \frac{1}{30}\sum_{i=1}^{30}(y_i - x_i)$ and $s_d = \sqrt{\frac{1}{30-1}\sum_{i=1}^{30}((y_i - x_i) - \text{Bias})^2}$. Since we expect approximately 95% of the differences to lie within the limits of agreement when the differences are normally distributed, the computer vision tool is validated if no more than two points fall outside these limits.

Consider the following example where the data are in the table below:

Sample #	x_i	y_i	Sample #	x_i	y_i	Sample #	x_i	y_i
1	75.2	75.3	11	75.8	75.8	21	88.8	86.3
2	91.3	92.5	12	93.5	97.1	22	98.1	97.3
3	77.1	78.1	13	82.2	79.7	23	71.3	71.1
4	95.6	97.2	14	86.3	84.3	24	82.2	82.7
5	93.9	94.7	15	80.8	80.6	25	92.7	91.2
6	75.2	78.0	16	90.6	91.3	26	76.9	73.9
7	85.7	84.1	17	78.6	81.4	27	70.2	73.8
8	76.4	78.4	18	94.8	94.2	28	86.5	85.7
9	91.7	97.0	19	75.1	71.1	29	77.9	74.6
10	81.8	80.4	20	96.6	98.6	30	88.3	87.9

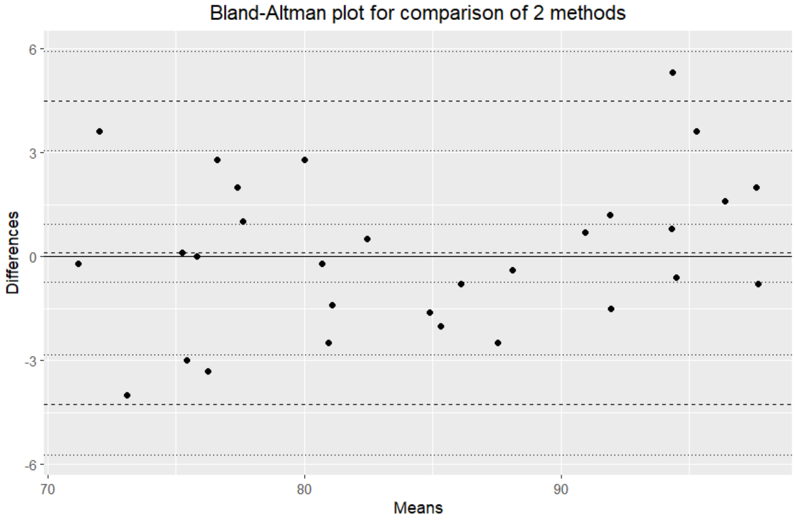
The results are:

- $\text{Bias} = 0.11$, $s_d = 2.23$
- p-value of the paired t-test: 0.7956. The bias is not significantly different from zero.
- $\text{Lower LoA} = -4.27$, $\text{Upper LoA} = 4.49$. One sample falls outside the limits (sample # 9: $y_9 - x_9 = 5.3$).

Therefore, based on these results, the computer vision tool is validated.

The results of the procedure can be illustrated with a Bland-Altman plot: each of the 30 samples is represented on the graph by assigning the mean of the two measurements as the *x*-value, and the difference between the two values as the *y*-value. The bias and the LoAs are shown with dashed lines. This graph was created using the R package *blandr* (Datta, 2017) and the following command:

```
blandr.draw(y, x, ciShading=F)
```



References

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476), 307-310.

Datta, D. (2017). blandr: a Bland-Altman Method Comparison package for R. Version 0.6.0, 2024-06-09, https://github.com/deepankardatta/blandr .

Conclusion: the validation of the computer tool was completed on (date X) and can be used to test X from (date X or Y)

Authorised by responsible person Z: Signed on date Z to allow use of the tool from date X, Y or Z