**Suppl 2: Structured description of a challenge design**

**SUMMARY**

**Item 1: Title**

a) Use the title to convey the essential information on the **challenge mission**.

Click here to enter text …

b) Preferable, provide a short **acronym** of the challenge (if any).

Click here to enter text …

**Item 2: Abstract**

Provide a **summary** of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Click here to enter text …

**Item 3: Keywords**

List the primary **keywords** that characterize the challenge.

Click here to enter text …

**CHALLENGE ORGANIZATION**

**Item 4: Organizers**

a) Provide information on the **organizing team** (names and affiliations).

Click here to enter text …

b) Provide information on the **primary contact person**.

Click here to enter text …

**Item 5: Lifecycle type**

Define the intended **submission cycle** of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.

*Examples:*

* One-time event with fixed submission deadline
* Open call
* Repeated event with annual fixed submission deadline

Click here to enter text …

**Item 6: Challenge venue and platform**

a) Report the **event** (e.g. conference) that is **associated** with the challenge (if any).

Click here to enter text …

b) Report the **platform** (e.g. grand-challenge.org) used to run the challenge.

Click here to enter text …

c) Provide the **URL** for the challenge website (if any).

Click here to enter text …

**Item 7: Participation policies**

a) Define the **allowed user interaction** of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Click here to enter text …

b) Define the policy on the **usage of training data**. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

Click here to enter text …

c) Define the **participation policy for members of the organizers' institutes**. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

Click here to enter text …

d) Define the **award policy**. In particular, provide details with respect to challenge prizes.

Click here to enter text …

e) Define the policy for **result announcement**.

*Examples:*

* Top three performing methods will be announced publicly.
* Participating teams can choose whether the performance results will be made public.

Click here to enter text …

f) Define the **publication policy**. In particular, provide details on ...

* … who of the participating teams/the participating teams’ members qualifies as author
* … whether the participating teams may publish their own results separately, and (if so)
* … whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

Click here to enter text …

**Item 8: Submission method**

a) Describe the method used for result submission. Preferably, provide a link to the **submission instructions**.

*Examples:*

* Docker container on the Synapse platform. Link to submission instructions: <URL>
* Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

Click here to enter text …

b) Provide information on the possibility for participating teams to **evaluate** their **algorithms before submitting** final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

Click here to enter text …

**Item 9: Challenge schedule**

Provide a **timetable** for the challenge. Preferably, this should include

* the release date(s) of the training cases (if any)
* the registration date/period
* the release date(s) of the test cases and validation cases (if any)
* the submission date(s)
* associated workshop days (if any)
* the release date(s) of the results

Click here to enter text …

**Item 10: Ethics approval**

Indicate whether **ethics approval** is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a **reference to the document** of the ethics approval (if available).

Click here to enter text …

**Item 11: Data usage agreement**

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit **listing of the license** applied.

*Examples:*

* CC BY (Attribution)
* CC BY-SA (Attribution-ShareAlike)
* CC BY-ND (Attribution-NoDerivs)
* CC BY-NC (Attribution-NonCommercial)
* CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
* CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

Click here to enter text …

**Item 12: Code availability**

a) Provide information on the **accessibility of the organizers' evaluation software** (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

Click here to enter text …

b) In an analogous manner, provide information on the **accessibility of the participating teams' code.**

Click here to enter text …

**Item 13: Conflicts of interest**

Provide information related to conflicts of interest. In particular provide information related to **sponsoring/ funding** of the challenge. Also, state explicitly who had/will have **access to the test case labels** and when.

Click here to enter text …

**MISSION OF THE CHALLENGE**

**Item 14: Field(s) of application**

State the **main field(s) of application** that the participating algorithms target.

*Examples:*

* Diagnosis
* Education
* Intervention assistance
* Intervention follow-up
* Intervention planning
* Prognosis
* Research
* Screening
* Training
* Cross-phase

Click here to enter text …

**Item 15: Task category(ies)**

State the **task category(ies)**.

*Examples:*

* Classification
* Detection
* Localization
* Modeling
* Prediction
* Reconstruction
* Registration
* Retrieval
* Segmentation
* Tracking

Click here to enter text …

**Item 16: Cohorts**

We distinguish between the *target cohort* and the *challenge cohort*. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on *ex vivo* data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding gender or age (target cohort).

a) Describe the **target cohort**, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

Click here to enter text …

b) Describe the **challenge cohort**, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Click here to enter text …

**Item 17: Imaging modality(ies)**

Specify the **imaging technique(s)** applied in the challenge*.*

Click here to enter text …

**Item 18: Context information**

Provide additional **information given along with the images**. The information may correspond ...

a) … directly to the **image data** (e.g. tumor volume).

Click here to enter text …

b) … to the **patient** in general (e.g. gender, medical history).

Click here to enter text …

**Item 19: Target entity(ies)**

a) Describe the **data origin**, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquiredin the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Click here to enter text …

b) Describe the **algorithm target**, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Click here to enter text …

**Item 20: Assessment aim(s)**

Identify the **property(ies) of the algorithms** **to be optimized** to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (parameter 26), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

* *Example 1:* Find liver segmentation algorithm for CT images that processes CT images of a certain size in less than a minute on a certain hardware with an error that reflects inter-rater variability of experts.
* *Example 2:* Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter 26).

Click here to enter text …

**CHALLENGE DATA SETS**

**Item 21: Data source(s)**

a) Specify the **device(s)** used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

Click here to enter text …

b) Describe relevant details on the imaging process/**data acquisition** for each acquisition device (e.g. image acquisition protocol(s)).

Click here to enter text …

c) Specify the **center(s)/institute(s)** in which the data was acquired and/or the **data providing platform/source** (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

Click here to enter text …

d) Describe relevant **characteristics** (e.g. level of expertise) **of the subjects** (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

Click here to enter text …

**Item 22: Training and test case characteristics**

a) State what is meant by one **case** in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

*Examples:*

* Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
* A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in *data source(s)* (parameter 21) and may include context information (parameter 18). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

Click here to enter text …

b) State the **total number** of training, validation and test cases.

Click here to enter text …

c) Explain **why a total number** of cases and **the specific proportion** of training, validation and test cases was chosen.

Click here to enter text …

d) Mention **further important characteristics** of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

Click here to enter text …

**Item 23: Annotation characteristics**

a) Describe the **method for determining the reference annotation**, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include *manual image annotation*, *in silico ground truth generation* and *annotation by automatic methods*.

If human annotation was involved, state the **number of annotators**.

Click here to enter text …

b) Provide the **instructions given to the annotators** (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the **annotation protocol**.

Click here to enter text …

c) Provide **details on the subject(s)/algorithm(s) that annotated** the cases (e.g. information on **level of expertise** such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Click here to enter text …

d) Describe the **method(s) used to merge multiple annotations** for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

Click here to enter text …

**Item 24: Data pre-processing method(s)**

Describe the **method(s) used for pre-processing** the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

Click here to enter text …

**Item 25: Sources of error**

a) Describe the most relevant **possible error sources related to the image annotation**. If possible, **estimate the magnitude** (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Click here to enter text …

b) In an analogous manner, describe and quantify **other relevant sources of error**.

Click here to enter text …

**ASSESSMENT METHODS**

**Item 26: Metric(s)**

a) Define the **metric(s) to assess a property of an algorithm**. These metrics should reflect the desired algorithm properties described in *assessment aim(s)* (parameter 20). State which metric(s) were used to compute the ranking(s) (if any).

* *Example 1:* Dice Similarity Coefficient (DSC) and run-time
* *Example 2:* Area under curve (AUC)

Click here to enter text …

b) **Justify why** the metric(s) was/were chosen, preferably with reference to the biomedical application.

Click here to enter text …

**Item 27: Ranking method(s)**

a) Describe the **method used to compute a performance rank** for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

Click here to enter text …

b) Describe the method(s) used to manage **submissions with missing results** on test cases.

Click here to enter text …

c) **Justify why** the described ranking scheme(s) was/were used.

Click here to enter text …

**Item 28: Statistical analyses**

a) Provide **details for the statistical methods** used in the scope of the challenge analysis. This may include

* description of the **missing data handling**,
* details about the assessment of **variability of rankings**,
* description of any method used to assess **whether the data met the assumptions**, required for the particular statistical approach, or
* indication of any **software product** that was used for all data analysis methods.

Click here to enter text …

b) **Justify why** the described statistical method(s) was/were used.

Click here to enter text …

**Item 29: Further analyses**

Present further analyses to be performed (if applicable), e.g. related to

* **combining algorithms** via ensembling,
* **inter-algorithm variability**,
* **common problems/biases** of the submitted methods, or
* **ranking variability**.

Click here to enter text …