**BEAMRAD Tool for data users**

**Title**

1. **Is there a title/acronym available for the dataset?**
2. **Is there information on how to cite the resource?**

**General description**

1. **What are the target entities?** i.e., the anatomical region of subjects from whom the data was acquired (e.g., brain shown in MRI).
2. **What data type is included in this dataset?** Image data (e.g. MRI) , signal data (e.g., ECGs) or tabular data (e.g., EHR)?
3. **Is the principal motivation and objective for the dataset described?** What are the principal motivations behind the development of the dataset?
4. **Are there different versions of this dataset?** If so, is the principal motivation and objective for updating this dataset described? What is the current version?
5. **What are the keywords for the dataset** (e.g., datatype, modality, body part, relevant medical terms).
6. **Is there information about coordinators or organizers involved in the dataset creation? Is there a list of contributors available?**
7. **When was the dataset published (first version)?**
8. **Is the dataset associated with a challenge?** If so, what is the challenge name?

**Dataset usage**

1. **Is there an official website or other place where the dataset is hosted?** If so, what is the location (e.g., a URL).
2. **What is the accessibility of the dataset?** (choose from provided categories). Are there any specific barriers or formal agreements?
   1. Open access – no requirements for access or only sign-in
   2. Open access with barriers – datasets fulfilling the theoretical criteria for open access, but are inaccessible because of unpredictable reasons (e.g., no response to requests or broken hyperlinks)
   3. Regulated access – required the fulfilment of formal agreements or approvals.
3. **Is there information about the ethics approval?** Is there a reference to the ethics approval or waiver.
4. **Is there a data usage agreement? Is there an explicit listing of the license applied?**
5. **Is there information regarding sponsoring/funding provided?**
6. **Are there any publication restrictions?**
7. **Is there a description of when the data included in the dataset was acquired (e.g., between 2005-2020)?**
8. **Is there information about the origin of data** **(e.g., country and hospital)?**
9. **Is it a single-center or multi-center study?**
10. **Are the criteria for data inclusion described? (e.g., consecutive participants, inclusion and exclusion criteria)**.
11. **Are there relevant characteristics (e.g., level of expertise) of the subjects (e.g., surgeon)/objects (e.g., robot) involved in the data acquisition process?**
12. I**s there any information on the data anonymization protocol?**

**Metadata**

1. **Is there information about the participant (e.g., age, medical history)? If relevant, what are the features, variable types, and the number of categories provided for the variable?**
2. **Is any information provided along with the data?** If yes, what kind of information?
3. **Is there information about the hardware used for the data acquisition?** (i.e. the vendor and the specific model)
4. **Is there general information about the data acquisition protocol, such as view or modality?** If so, what kind of information is provided?

**Sample size**

1. **What is the number of participants and the number of data samples?**
2. **If there is more than one inclusion for a participant, what is the number of data samples per participant?**

**Missing data**

1. **Is the information about missing values provided?** What information are missing?

**Training and test set distribution**

1. **Is data divided into training/test sets?** If the dataset consists of training/test sets, how is the data distributed?
2. **Are both the training and test datasets available?**
3. **Is there an explanation for the potential differences between distributions of the data subjects in the training and test sets and the real-world distribution?**

**Data Annotation**

1. **What types of annotations are included in this dataset?**
2. **Are the data annotation methods explained?**
3. **Are there human annotators involved?**
4. **If human annotators were involved, is the annotation instruction protocol available?**
5. **If human annotators were involved, is the number of annotators indicated? If yes, how many individual annotators are involved in the annotations process?**
6. **If human annotators were involved, is there information on how they were selected?**
7. **If human annotators were involved, is the disciplinary background of the annotators indicated, including their years of expertise?**
8. **If there are multiple annotations, is the approach to reaching a consensus on the final labels described by the dataset’s creators?**

**Data preprocessing**

1. **Is the data preprocessed? If yes, how was the data preprocessed?** What were the methods (e.g., cropping, resampling, and contrast enhancement).

**Sources of errors**

1. **Are the most relevant possible sources of error (e.g., error through annotation variability) described in the documentation of the dataset?**
2. **Are the most relevant error sources quantified?** (For instance, by giving specific information for training, validation and test cases?). If so, please provide a list of the quatifications.
3. **Is there information regarding the limitations of the dataset?** What are the limitations?