

MICCAI 2023 Article Annotation Guide

Essential Tools and Software

- PDF Reader: Adobe Reader Pro or any capable PDF reader.
- Data Entry Software: Microsoft Excel.

Resources Provided (2 files and 1 folder)

File #1: annotation_guide_details.doc

This file contains detailed instructions on the annotation process.

File #2: annotation_guide_and_scheme.xlsx with 4 sheets

Sheet 1: Annotation guide details

- a. Overview of all categories with annotating examples and keywords for searching after relevant sentences

Sheet 2: All categories

- b. Overview of all categories in the annotation scheme

Sheet 3: Annotation scheme

- c. A structured template with predefined categories for you to use when annotating.
- d. The follow categories provide article details such as:
 - i. *Title*: Name of the article for reference.
 - ii. *Paper Number*: Unique identifier for the article.
 - iii. *Volume Number*: Volume in which the article is published.

Sheet 4: Keywords organs & image types

- e. A list of keywords to assist in locating relevant sections for annotation.

Article Folder: A directory containing 100 selected PDF articles from MICCAI 2023.

- Naming Convention: Files are titled in the format 'paper_<number>_vol_<number>'.
 - Utilize the **Title** and/or **Paper Number** and **Volume Number** to find articles within the folder.
-

Annotation Steps

1. *Article Location*: Access the designated folder containing the MICCAI 2023 PDF articles.
2. *Article Identification*: Use the list of titles (under "**title**") and filenames (from "**paper name**" and "**vol**") provided in the Annotation Scheme to locate your assigned article(s) within the folder.
3. *Annotation Process*:
 - a. Open the corresponding article.
 - b. Begin the annotation process by adhering to the categories outlined in the Annotation Scheme (**File #2: annotation_guide_and_scheme.xlsx, Sheet 3: Annotation scheme**).
 - c. Employ the provided Keywords List to efficiently identify pertinent sentences for each annotation category (**File #2: annotation_guide_and_scheme.xlsx, Sheet 4: Keywords organs & image types**).

Annotation Guide - Article information

General Information				
Purpose of annotation	To ensure annotations are made on the right selection of articles. Selection of articles were made by excluding non-cancer research/AI models			
sub categories	title	vol	Does the article contain any of these keywords: cancer/tumor/tumour?	notes
General Annotation: 1.Find the title of paper in the folder 2.Determine if you are annotating the right paper 3.Add the name of the paper name and vol Authors is the authors of the paper	<title of the paper>	<insert the number of vol>	Annotate '1' if the article contain one of the keywords, '0' if not.	1. Add sentences from the article that support your findings/annotations. 2. Add any remarks you find fitting. 3. Add a comment if there is categories you find hard to annotate. Please provide why it was difficult.

Categories

Does the article contain any of these keywords: cancer/tumor/tumour?

Excluded sections:

- I. Authors, departments, affiliations
- II. Acknowledgements, References
- III. Only text starting from Abstract and ending with the last line of Conclusion

Notes

Here you can add comments and sentences that “justifies” your annotations. For example:

- If you add ‘1’ for a category, please provide the sentence.
- If you add the name of a healthcare facility, please provide the sentence.

If you find any categories hard to annotate, please add this as a comment. These findings are also important as they show whether the articles provide clear and consistent information or come across as being ambiguous.

Examples:

- It was difficult to determine how many different datasets articles are using.
- Add these sentences for further analysis, or others to back up your annotations.

Extra categories (not mandatory)

- Information found outside article: information provided as Supplementary Material and/or by references.

PET-Diffusion: Unsupervised PET Enhancement Based on the Latent Diffusion Model

Caiwen Jiang¹, Yongsheng Pan¹, Mianxin Liu³, Lei Ma¹, Xiao Zhang¹, Jiameng Liu¹, Xiaosong Xiong¹, and Dinggang Shen^{1,2,4(✉)}

¹ School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
{jiangcw,panysh,dgshen}@shanghaitech.edu.cn

² Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200230, China

³ Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China

⁴ Shanghai Clinical Research and Trial Center, Shanghai 201210, China

Acknowledgment. This work was supported in part by National Natural Science Foundation of China (No. 62131015), Science and Technology Commission of Shanghai Municipality (STCSM) (No. 21010502600), The Key R&D Program of Guangdong Province, China (No. 2021B0101420006), and the China Postdoctoral Science Foundation (Nos. BX2021333, 2021M703340).

References

1. Buades, A., Coll, B., Morel, J.: A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 60–65 (2005)
2. Chen, S., Ma, K., Zheng, Y.: Med3D: transfer learning for 3D medical image analysis. arXiv preprint [arXiv:1904.00625](https://arxiv.org/abs/1904.00625) (2019)
3. Cui, J., et al.: PET image denoising using unsupervised deep learning. Eur. J. Nucl. Med. Mol. Imaging **46**(13), 2780–2789 (2019)
4. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising with block-matching and 3D filtering. Image Process. Algorithms Syst. Neural Netw. Mach. Learn. **6064**, 354–365 (2006)
5. Dosovitskiy, A., Brox, T.: Generating images with perceptual similarity metrics based on deep networks. In: Advances in Neural Information Processing Systems, vol. 29 (2016)
6. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851 (2020)
7. Hofheinz, F., et al.: Suitability of bilateral filtering for edge-preserving noise reduction in PET. EJNMMI Res. **1**(1), 1–9 (2011)

Annotation Guide - Demographic Information

Purpose of annotation

To identify and record specific characteristics of the population included in study datasets. These characteristics typically include age, gender, race, ethnicity, and geographical location.

When examining papers, your task is to identify and annotate demographic information provided in the 'Datasets' or 'Experiments' sections, or sections with a similar title. This information helps us understand the diversity and representativeness of the datasets used in the research.

Demographics				
Purpose of annotation	<i>To identify and record specific characteristics of the population included in study datasets. These characteristics typically include age, gender, race, ethnicity, and geographical location.</i>			
sub categories	Study subject labelled as 'patient/patients' in dataset(s)	sex/gender	ethnicity	geographical location
General Annotation: 1. Locate the section that describes the datasets. 2. Determine if the datasets include demographic details like age, gender, race, and geographical location. 3. If demographic data is present, annotate this as '1'. If not, annotate as '0'. Authors is the authors of the paper	Annotate '1' if the authors label study subjects as 'patient/patients', '0' if not. Annotate '1' if authors speak of patients in their results/findings in relation to their dataset, '0' if not.	Annotate '1' if study subjects are labeled by gender, '0' if not. Annotate '1' if information about the gender of study subjects is provided, '0' if not. Annotate '1' if there's any mention of gender in relation to the dataset, '0' if not.	Annotate '1' if study subjects are categorized by ethnicity, '0' if not. Annotate '1' if information on the ethnicities of study subjects is included, '0' if not. Annotate '1' if there's a discussion of ethnicity in relation to the dataset, '0' if not.	Annotate '1' if study subjects are labeled by their geographical location, '0' if not. Annotate '1' if there is information about the geographical locations of the study subjects, '0' if not. Annotate '1' if there's any mention of geographical location in relation to the dataset, '0' if not.
Example Phrase	1. "The cohort consists of 141 patients with pancreatic ductal adenocarcinoma, of an equal ratio of male to female patients." 2. "Following IRB approval for this study, we search for patients with metastatic breast cancer who had a breast cancer MRI performed between 2010 and 2020"	1. Labeling by Gender: "The survey responses were divided by gender, with 60% female and 40% male participants." 2. Gender Information: "Our study's gender breakdown revealed a higher number of male respondents." 3. Mention of Gender: "Gender differences were evident in the dataset, affecting the outcome of the study."	1. Categorization by Ethnicity: "Data was categorized into four ethnic groups: Caucasian, African American, Hispanic, and Asian." 2. Ethnicity Information: "Participants identified their ethnicity, allowing for an ethnically diverse dataset." 3. Discussion of Ethnicity: "The research highlights ethnic variations in the data, pointing to significant disparities."	1. Labeling by Location: "Data collection was conducted in various geographical locations, including urban and rural areas." 2. Location Information: "The dataset reflects a geographical spread from the northern to the southern regions." 3. Mention of Geographical Location: "Geographical disparities are considered in the dataset analysis, given the spread of participants' locations."
Keywords to look for	patient', 'patients'	gender', 'sex', 'women', 'woman', 'female', 'male'	ethnicity', 'ethnicities', 'race', 'white patients', 'black patients', 'nationality'	geolocation', 'geographical', 'geographic', 'country', 'countries', 'city', 'cities', 'hospital', 'hospitals', 'clinic', 'clinics', 'continent', 'province', 'state', 'region', 'town', 'village', 'area', 'district'

Categories

- Study subject labelled as 'patient/patients' in dataset(s)?
- Age (of study subjects)
- Sex/gender (of study subjects)
- Ethnicity (of study subjects)
- Geographical Location (of study subjects)

Follow these steps.

1. Locate the section that describes the datasets by searching in the PDF for "dataset", "datasets", "data", "data sets".
2. Determine if the dataset(s) include demographic details like age, gender, race, and geographical location of study subjects.
 - a. Search by using the provided keywords.

b. If demographic data is present, annotate this as '1'. If not, annotate as '0'.

Annotation Guide - Dataset information

Purpose of annotation

To concisely catalogue the data used in studies, including the number of datasets, their public or private status, the types of images, and any mentioned organs. These annotations help understand the dataset's scope, critical for assessing research replicability, potential bias, and results applicability. Essentially, it ensures transparency and promotes further research.

Annotators will review research papers to collect specific information about the datasets used by the authors. This includes details about dataset quantity, access status (public or private), image types, and organs/body parts featured.

Dataset/Data collection Information			
Purpose of annotation	To identify availability of & contents of collected data/datasets		
sub categories	dataset quantity	quantity of private datasets	image type in datasets
General Annotation: 1. Locate the section that describes the datasets. 2. Determine if the datasets include details of described categories Authors is the authors of the paper	Annotate the number of dataset(s) used in the paper. Annotate the number by whole integer. Use '0' if this information is not mentioned.	Annotate the number of dataset(s) described as private in the paper. Annotate the number by whole integer. Use '0' if this information is not mentioned.	Annotate the image type described in the dataset(s) by string name. Type "Unknown" if this information is not mentioned.
Example Phrase	"We use the public gbc us dataset [3] (...) we use the publicly available kvasir-seg [17] dataset" From above example there are 2 dataset(s) describes and used in the paper		"TCGANSCLC dataset includes two subtypes in lung cancer, Lung Squamous Cell Carcinoma and Lung Adenocarcinoma, with a total of 1,054 WSIs" From above example the dataset contains the image type "WSI"
Keywords to look for	"We use the dataset...", "the dataset includes", "we trained our model on dataset..."	private, 'privately', 'not publised', 'not open'	image, 'images', 'ct', 'cts', 'mri', 'wsis', 'wsi', 'ultrasound', 'pet', 'scan', 'scans', 'x-ray', 'x-rays', 'xrays', 'gastroscopy', 'colonoscopy', 'endoscopy' For more keywords look under sheet "image types"
sub categories	organ/body part in datasets		Female and/or male organs?
General Annotation: 1. Locate the section that describes the datasets. 2. Determine if the datasets include details of described categories Authors is the authors of the paper	Annotate the organs/body parts described in the datasets by string name. Type "Unknown" if this information is not mentioned.		Annotate 'male' if the data contains male-specific organs Annotate 'None' if not disclosed Annotate 'Unknown' if you are unsure
Example Phrase	"TCGANSCLC dataset includes two subtypes in lung cancer, Lung Squamous Cell Carcinoma and Lung Adenocarcinoma, with a total of 1,054 WSIs" From above example the dataset contains data from the organ "lung"		"774 consecutive bi-parametric prostate mri examinations are included in this study" From above example the dataset contains data male-specific organs.
Keywords to look for	breast, 'no organ mentioned', 'gastro', 'colorectal', 'chest', 'skin', 'colon', 'lymph nodes', 'skin', 'prostate'.		"Male Reproductive System": [penis, 'prostate', 'testis'] "Female Reproductive System": [cervix, 'uterus', 'vagina']

Categories

- Dataset Quantity
 - Count and annotate the number of datasets mentioned in the paper. Record as a whole integer. Annotate '0' if no datasets are mentioned.
- Public Datasets
 - Identify and annotate the number of datasets described as public or publicly available. Use whole numbers and '0' if none are public.
 - A dataset is public if the word "public", "publicly", "open" or "open access" is used.
- Private Datasets

- Count and annotate the number of datasets described as private. Record this information as a whole integer, using '0' if there are no private datasets.
- A dataset is private if the word "private", "privately" or "in-house" is used.
- Image Types
 - Annotate the specific types of images described in the datasets (e.g., MRI, CT, X-ray). If not specified, annotate as "Unknown".
- Organs/Body Parts
 - Annotate the names of organs or body parts mentioned in relation to the datasets. If none are mentioned, annotate as "Unknown".
- Does the dataset contain sex-specific organ(s)?
 - Annotate '1' if data contains sex specific organ(s). A sex-specific organ is determined by male or female reproductive organs.

Follow these steps.

1. Locate the section that describes the datasets by searching in the PDF for "dataset", "datasets", "data", "data sets".
2. Determine if the datasets include information on the categories specified: quantity, public or private status, image types, and organs/body parts.
 - a. Search by using the provided keywords.
 - b. If you find keywords outside the scope of the article, such as in the section for "References", which shows that a referred dataset is open/public or private, add this as a comment in "notes".
3. Use provided keywords and example phrases to assist in identifying relevant information.

Annotation Guide - Geographical Location of Datasets

Purpose of annotation

To identify dataset sources, examining definitions of 'location', understanding that 'location' can extend beyond geographical areas to specific establishments, and recognizing implicit references to location.

When annotating for geographical location information, our goal is to capture details about where each dataset originates.

1. Source Annotation:

Check if the dataset source is specified. Annotate any mention of where the dataset was gathered or sourced. This could be a country, city, institution, or any other place of origin provided in the documentation.

2. Location Definition Annotation:

Look at how each paper defines 'location.' Remember, 'location' isn't just about maps and coordinates; it can also refer to types of places. For example, if a paper talks about data from a 'hospital,' 'clinic,' or 'health center,' annotate this as the location, even though it doesn't give an exact address or city.

Always pay attention to the context — sometimes, the type of place (like a healthcare facility) tells us about the location, even when specific geographical details are not given.

Geographical Location of Datasets/Data collections				
Purpose of annotation	<i>To identify dataset sources by exploring how 'location' is defined beyond geography, including specific establishments and subtle references</i>			
sub categories	location	name of healthcare facility	location as other	name of location as other
General Annotation: 1. Locate the section that describes the datasets. 2. Determine if the datasets include details of geographical location. Authors is the authors of the paper	Annotate with '1' if the paper discloses the source location of the data and/or where the dataset is provided/collected from. Otherwise, annotate with '0'.	Insert name of <keyword> if mentioned, else None	Annotate with '1' if a location is specified as other of data collection (regardless of the name). Otherwise, annotate with '0'.	Insert name of <keyword> if mentioned, else None
Example Phrase	"Data was collected from [insert location term here]."	"Data is collected from Mayo clinic"	"Data is collected from a [department/university/institution/online/laboratory]"	"Data is collected from South Sun Pathology Laboratory"
Keywords to look for	geolocation', 'geographical', 'geographic', 'country', 'countries', 'city', 'cities', 'hospital', 'hospitals', 'clinic', 'clinics', 'continent', 'province', 'state', 'region', 'town', 'village', 'area', 'district', 'department', 'university', 'institution', 'online', 'laboratory'	hospital', 'hospitals', 'clinic', 'clinics'	department', 'university', 'institution', 'online', 'laboratory'	department', 'university', 'institution', 'online', 'laboratory'
sub categories	location as large-scale geographical entity	location as subnational geographical entity	name of location as subnational geographical entity	
General Annotation: 1. Locate the section that describes the datasets. 2. Determine if the datasets include details of geographical location. Authors is the authors of the paper	Annotate with '1' if the data collection location is identified as a large-scale geographical entity such as a country or continent. If not, annotate with '0'.	Annotate with '1' if the location is defined as a subnational geographical entity like a city, state, or district. Use '0' if no such detail is provided.	Insert name of <keyword> if mentioned, else None	
Example Phrase	"Data is collected from [country/continent]."	"Data is collected from [city/state/region]."	"Data is collected... New York"	
Keywords to look for	country', 'countries', 'continent', 'continents'	city', 'cities', 'province', 'state', 'region', 'town', 'village', 'area', 'district'	city', 'cities', 'province', 'state', 'region', 'town', 'village', 'area', 'district'	

Categories

- Location (of dataset/data collection)
- Location as healthcare facility
 - This category encompasses various types of establishments dedicated to providing medical, surgical, and other forms of health care treatment and services. Name of healthcare facility is not relevant for this category.
- Name of healthcare facility
 - Insert name of healthcare facility only
- Location as other (center/department/laboratory/university/institution/online)
 - This category encompasses other types of establishments outside the definition of healthcare facility (types of establishments providing medical, surgical, and other forms of health care treatment and services). Name of other is not relevant for this category.
 - Other types of establishments could be a center, institution, department, laboratory, university and/or online database.
- Name of location as other (center/department/laboratory/university/institution/online)
 - Insert name of other
- Location as large-scale geographical entity
 - A location as a large-scale geographical entity includes the largest territorial divisions, typically recognized on an international scale.
- Name of location as large-scale geographical entity
 - Insert name of country, countries and/or continents
- Location as subnational geographical entity
 - A location as a subnational geographical entity includes entities that are subdivisions of a country, ranging from larger areas like states and provinces to smaller localities like towns and villages.
- Name of location as subnational geographical entity
 - Insert name of city, cities, province, state, region, town, village, area and/or district

Follow these steps.

1. Locate the section that describes the datasets by searching in the PDF for “dataset”, “datasets”, “data”, “data sets”.
2. Determine if the datasets include details of geographical location.
 - a. Search by using the provided keywords.