Annotation guide for selected MICCAI 2023 papers

Annotation Guide: Demographic Information

Purse 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.

Categories:

- Age
- Sex/gender
- Ethnicity
- Geographical Location

Follow these steps:

1. Open the file called "...."
2. Locate the section that describes the datasets.
3. Determine if the datasets include demographic details like age, gender, race, and geographical location.
4. If demographic data is present, annotate this as '1'. If not, annotate as '0'.

categories	age	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 by age, '0' if not. Annotate '1' if there's any information on the ages of the study subjects, '0' if not. Annotate '1' if there's a mention of age in relation to the 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. Labeling by Age: "Participants, aged 18-25, were selected for the study." 2. Age Information: "The dataset comprised individuals predominantly in their 30s and 40s." 3. Mention of Age: "Age distribution played a significant role in the analysis of the data."	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. ocation 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	age', 'young', 'old', age interval, age groups, age kan be in years and in months (for infants), age can be defined as a mean or median	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'

Annotation Guide: Geographical Location details

Purse 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.

Categories:

- Location
- 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 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. Open the file called "...."
2. Locate the section that describes the datasets.
3. Determine if the datasets include details of geographical location.

keywords	location	location as a healthcare facility	name of healthcare facility
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 mentions where the datasets are collected from. Use '0' if this information is not mentioned.	This category encompasses various types of establishments dedicated to providing medical, surgical, and other forms of health care treatment and services. Name of particular healthcare facility is not relevant for this category. Annotate with '1' if a healthcare facility is specified as the location of data collection (regardless of the facility's name). Otherwise, annotate with '0'.	Insert name of <keyword> if mentioned, else None
Example Phrase	"Data was collected from [insert geographical term here]."	"Data is collected from a [hospital/clinic]."	"Data is collected from Mayo clinic"
Keywords to look for	geolocation', 'geographical', 'geographic', 'country', 'countries', 'city', 'cities', 'hospital', 'hospitals', 'clinic', 'clinics', 'continent','province', 'state', 'region', 'town', 'village', 'area', 'district'	'hospital', 'hospitals', 'clinic', 'clinics'	'hospital', 'hospitals', 'clinic', 'clinics'

location as large-scale geographical entity	name of location as large-scale geographical entity	location as subnational geographical entity	name of location as subnational geographical entity
A location as a large-scale geographical entity includes the largest territorial divisions, typically recognized on an international scale. 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'.	Insert name of <keyword> if mentioned, else None	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. Annotate with '1' if the location is defined as a subnational geographical entity like a city, state, or district. Use '0' if no such	Insert name of <keyword> if mentioned, else None
"Data is collected from [country/continent]."	"Data is collected from... in China"	"Data is collected from [city/state/region]."	"Data is collected... New York"
country', 'countries', 'continent', 'continents'	country', 'countries', 'continent', 'continents'	'city', 'cities', 'province', 'state', 'region', 'town', 'village', 'area', 'district'	'city', 'cities', 'province', 'state', 'region', 'town', 'village', 'area', 'district'

Annotation Guide: Dataset details

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 datasets' 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.

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.
- 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.
- 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".
- Sex-specific cancer
    - Annotate '1' if the cancer is sex specific. A sex-specific cancer is determined by male or female reproductive organs.
    - Cancer can be sex-specific in two ways:
        - There is a higher occurrence of the illness in one sex compared to the other, or
        - The illness is determined by/related to underlying biological factors or one sex only.
        - In this case we only annotate '1' if cancer is determined/related to one sex only: use the column "organ" as guideline.

Follow these steps:

1. Open the file "....."
2. Locate the section in the paper where the datasets are described.
3. Determine if the datasets include information on the categories specified: quantity, public or private status, image types, and organs/body parts.
4. Use provided keywords and example phrases to assist in identifying relevant information.

categories	dataset quantity	quantity of public datasets	quantity of private datasets	image type in datasets	organ/body part in datasets	sex-specific cancer
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 public 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.	Annotate the organs/body parts described in the datasets by string name. Type "Unknown" if this information is not mentioned.	Annotate '1' if the cancer is sex-specific: cancer is determined/related to one sex only: use the column "organ" as guideline.
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	"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) described as public/publicly available	"We use the public gbc us dataset [3] (...) we use the publicly available kvasir-seg [17] dataset" From above example there are 0 dataset(s) described as private.	"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"	"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 from sex-specific cancer
Keywords to look for	"We use the dataset...", "the dataset includes", "we trained our model on dataset..."	public', 'publicly', 'open access', 'open', 'published'	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"	breast', 'no organ mentioned', 'gastro', 'colorectal', 'chest', 'skin', 'colon', 'tymph nodes', 'skin', 'prostate'. For more keywords look under sheet "organs"	"Male Reproductive System" : ['penis', 'prostate', 'testis'] Female Reproductive System": ['cervix', 'uterus', 'vagina']