# Datasheet for ASL Citizen


Aashaka Desai[β]    Lauren Berger[γ]    Fyodor O. Minakov[α]    Vanessa Milan[α]
Chinmay Singh[α]    Kriston Pumphrey[γ]    Richard E. Ladner[β]
Hal Daumé III[α,δ]    Alex X. Lu[α]    Naomi Caselli[γ]    Danielle Bragg[α]

[α]Microsoft Research    [β]University of Washington    [γ]Boston University    [δ]University of Maryland
{hal3,lualex,dabragg}@microsoft.com
{aashakad,ladner}@cs.washington.edu, nkc@bu.edu


Updated June 12th, 2023

---

## Motivation

**For what purpose was the dataset created?** Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

The dataset was created to help enable research on isolated sign language recognition (ISLR) - i.e. recognizing individual signs from video clips - and sign language modeling more generally.

Specifically, we frame ISLR as a dictionary retrieval task: given a self-recorded video of a user performing a single sign, we aim to retrieve the correct sign from a sign language dictionary. This dataset was created with this framing in mind, with the intent of grounding ISLR research in a practical application useful to the Deaf community. While we believe this dataset is suited for methods development for ISLR in general, this dataset specifically contains signs in American Sign Language (ASL).

In designing our collection mechanism, we sought to address limitations of prior ISLR datasets. Previous datasets have been limited in terms of number of videos, vocabulary size (i.e. number of signs contained), real-world recording settings, presence and reliability of labels, Deaf representation, and/or number of contributors. Some past datasets (in particular scraped datasets) have also included videos without explicit consent from the video creators or signers in the videos.

**Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?**

This dataset was created by Microsoft Research in collaboration with Boston University. Each organization's involvement in collection is detailed below.

Microsoft: platform design, platform engineering, primary ethics board (IRB) review of collection procedures (review of record), additional compliance review of and guidance for the platform (e.g. privacy, security, etc.), platform maintenance and debugging, technical support for participants, hosting of collection infrastructure (website, database, videos, backups, etc.), funding for participant compensation, data processing and cleaning, ethics and compliance board review of the dataset release (e.g. privacy, data cleaning, metadata, etc.), hosting of released assets (dataset, code, other supplementary materials)

Boston University: platform feedback, seed sign recordings, secondary IRB review of collection procedures, participant recruitment, answering or redirecting participant questions, procurement and distribution of participant compensation

**Who funded the creation of the dataset?** If there is an associated grant, please provide the name of the grantor and the grant name and number.


Microsoft primarily funded the creation of this dataset. Microsoft funded building and maintaining the collection platform, data storage and processing, participant compensation, and all time spent on the Microsoft activities listed above.

Boston University funded all time spent on the Boston University activities listed above. Support was provided in part by National Science Foundation Grants: BCS-1625954 and BCS-1918556 to Karen Emmorey and Zed Sehyr, BCS-1918252 and BCS-1625793 to Naomi Caselli, and BCS-1625761 and BCS-1918261 to Ariel Cohen-Goldberg. Additional funding was from the National Institutes of Health National Institute on Deafness and Other Communication Disorders of and Office of Behavioral and Social Science Research under Award Number 1R01DC018279.


**Any other comments?**

None.


1


## Composition

**What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?** Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

The instances are self-recorded videos of participants performing individual signs in ASL. Examples of still frames from this dataset are shown in our paper publication. Distribution of video lengths is shown in Figure 1.

**How many instances are there in total (of each type, if appropriate)?**

There are 83,399 instances of videos. In total, this data represents videos from 52 participants over a vocabulary size of 2,731 signs in ASL. On average, there are 30.5 videos for each sign and 1,604 videos per participant. The distribution of videos per participant is bimodal because we compensated participants for up to 3,000 videos. For 22 participants, the dataset includes 2992 ± 16 videos. For the remaining 30 participants, the dataset includes an average of 586 videos. (These video counts are for dataset Version 1.0, our first publicly released version of the dataset, after processing and cleaning.)

**Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?** If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).

The dataset contains a sample of single-sign videos, covering many fundamental ASL signs, and demonstrated by a sample of Deaf and hard-of-hearing community members in everyday environments.

The ASL vocabulary was taken from ASL-LEX [2], which is a linguistically analyzed corpus of ASL vocabulary, covering many fundamental ASL signs. Specifically, our dataset contains 2,731 distinct signs (or glosses). We chose to adopt this vocabulary set because of the provision of detailed linguistic analysis of each sign, which complements the video set we provide, and allows for a richer set of uses for the videos. Due to ASL-LEX corpus updates across the time taken for data collection, six glosses in our dataset do not have corresponding linguistic information.

The videos themselves contain a sample of the ASL community executing these signs to webcams in home environments. This type of crowdsourced collection has the benefit of not restricting geographic proximity (thus potentially expanding diversity), and capturing signers in their natural environments. Still, we recruited largely from our own Deaf community networks using snowball sampling. This type of convenience sampling can result in biases; for example, our sample of videos contains a high proportion of people who self-identified as female, compared to the population of ASL users at large.

**What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features?** In either case, please provide a description.

Each instance consists of a video file in .mp4 format. Each instance also has an associated gloss (or English transliteration), which was the target value for the signer. These glosses are consistent with a previous lexical database, ASL-LEX [2], and thus can be mapped onto standardized identifiers and phonological properties for the signs provided in this lexical database. Finally, each instance is also associated with an anonymous user identifier, identifying which of the 52 participants performed the sign.

**Is there a label or target associated with each instance?** If so, please provide a description.

The label is the English gloss associated with each sign, as described above.

**Is any information missing from individual instances?** If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

All instances have complete information. However, we not that some users have blank metadata in their demographic information. This is intentional, as provision of this information was entirely voluntary, and some users did not provide some fields.

**Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)?** If so, please describe how these relationships are made explicit.

Yes. Each instance is tagged with a user ID identifying which user performed the sign. This user ID can be further associated with a separate metadata file containing demographic information on each of the users, such as the self-identified gender of the signer. We do not analyze this demographic information in our manuscript, but provide it because it could be useful for studying fairness (and other research).

**Are there recommended data splits (e.g., training, development/validation, testing)?** If so, please provide a description of these splits, explaining the rationale behind them.

Yes. Instances are labeled as either "train" (training set), "val" (validation set), or "test" (test set), containing 40,154, 10,304, and 32,941 videos respectively. The data splits are stratified by user such that each user is

2

unseen in the other data splits. These splits align with our dictionary retrieval task, because we expect users querying the dictionary to be unseen during training and model selection. Some participants contributed data over the entire vocabulary, while others only contributed data for a subset. To balance our test set metrics across the vocabulary, we assigned 11 participants who contributed 3,000 ± 5 videos (i.e. the maximum number of videos participants would be compensated for) to the test dataset. The other 11 participants who contributed 3,000 ± 5 videos, in addition to 30 participants who contributed a smaller number of videos, were otherwise split between the train and val set. We tried to balance gender identities across splits, as seen in Figure 2. Other than these factors, participants were randomly assigned to splits.

**Are there any errors, sources of noise, or redundancies in the dataset?** If so, please provide a description.

We implemented some filters for blank videos, videos not containing people, and videos without signing as described above. However, these filters are basic, and may not have captured all videos with technical problems. Additionally, since these videos are self-recorded, not all users may perform the same sign for a gloss, since the same English gloss can sometimes refer to multiple signs (e.g. when there are regional variations of a sign, or when the English gloss is a homonym). We limited this issue by collecting data in an ASL-first process, where users watched a video of a sign prompt rather than reading an English gloss prompt, but in some instances, users may still not follow the seed signer (e.g. when their regional variation of a sign is not documented in the dictionary, or when the seed signer is using an outdated sign for signs that rapidly evolve). It is also possible that contributors may have made mistakes in signing, or submitted erroneous videos that our filters did not catch.

**Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?** If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a future user? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.

The dataset is self-contained, but the gloss labels can optionally be mapped to ASL-LEX, which provides detailed linguistic analysis of each sign in the vocabulary we used. The linguistic analysis download can be found at `https://asl-lex.org/download.html`, and information about funding and licenses for ASL-LEX can be found at `https://asl-lex.org/about.html`.

**Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals non-public communications)?** If so, please provide a description.

While the videos contain recordings of people signing, all contributors consented to participate in this dataset and agreed to terms of use for our web platform. The consent process provided detailed information about the project's purpose, and explained that the dataset would be released to the public for research purposes.

**Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?** If so, please describe why.

Generally no. The videos reflect signs a viewer would be exposed to in everyday conversational ASL, taken from an established corpus of vocabulary. However, some of the vocabulary may refer to content that some find offensive (e.g. a person's private parts). In addition, because this database is a "snapshot" of the language at time of curation, some signs may be outdated and refer to stereotypes (e.g. around identity) phased out as the language has evolved and continues to evolve.

We also believe the chance of erroneous offensive content is extremely low. We recruited from trusted groups, manually vetted the first and last video submitted by each user on each date of submission to verify good faith effort, passed all videos through libraries to detect and blur appearance of third parties, and finally did a manual review of all videos. We conducted our review and cleaning iteratively, under close guidance from Microsoft's Ethics and Compliance team. We did not identify any offensive content in any of our reviews. All video blurring and omissions was done out of an abundance of care for our dataset participants (e.g. to remove a third party or personal content). However, it is impossible to guarantee that users did not submit videos that some may find offensive.

**Does the dataset relate to people?** If not, you may skip the remaining questions in this section.

Yes.

**Does the dataset identify any subpopulations (e.g., by age, gender)?** If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.

No. We release general aggregated demographics as part of our paper publication, but do not release individual demographics, to help protect participant privacy. These aggregated demographics span gender, age, region, and years of ASL experience. Providing demographic data on the collection platform was fully voluntary (i.e. not required and not tied to compensation) and self-reported.

3

**Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset?** If so, please describe how.

Yes. The videos contain uncensored faces and are generally filmed in the users' home environments. We chose not to censor user faces because facial expressions are critical linguistic components of ASL. Users provided consent for dataset release, and were able to delete videos or opt out of the dataset any time prior to release.

**Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)?** If so, please provide a description.

Not directly, but some sensitive attributes about participants might be guessable from the videos (e.g. race, or relation to the Deaf community).

**Any other comments?**
None.

---

### Collection Process

**How was the data associated with each instance acquired?** Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

Videos were self-recorded and contributed by participants through a crowdsourcing web platform. We built on a platform described in [1], with optimizations to support scale. Demographics could optionally be entered into the platform as part of a user profile. Please see the Supplementary Materials in our paper publication for a detailed description of the optimized design components and rationale.

**What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)?** How were these mechanisms or procedures validated?

Users contributed videos through a web platform that accessed the user's webcam to facilitate recording within the website itself. Contributors used their own hardware for recording (e.g., webcams). This setup is consistent with the type of setup that future dictionaries might have to demonstrate and look up a sign.

**If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?**
N/A

**Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?**
A team of researchers, engineers, and a designer were involved in the data collection. The team designed, built, and maintained the platform, and also managed recruitment, participant engagement, and compensation. The team was compensated through salary, stipend, or contract payment.

Data contributors were also compensated monetarily. The seed signer was paid to record the seed sign videos. The rest of the data was crowdsourced. For every 300 signs recorded, these participants received a $30 Amazon gift card, for up to 3,000 signs.

**Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)?** If not, please describe the timeframe in which the data associated with the instances was created.

Collection of the seed sign videos ran from April-May 2021. The collection of the community replications ran from July 2021 to April 2022.

**Were any ethical review processes conducted (e.g., by an institutional review board)?** If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

Yes. The data collection was reviewed by the two collaborating institutions' Institutional Review Boards (IRBs) – Microsoft (primary, IRB of record #418) and Boston University. The platform itself and the dataset release also underwent additional Ethics and Compliance reviews by Microsoft.

**Does the dataset relate to people?** If not, you may skip the remaining questions in this section.
Yes.

**Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?**
Videos were self-contributed by individuals directly.

**Were the individuals in question notified about the data collection?** If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.
Yes. When participants first visited our web platform, they engaged in a consent process, which

4

provided detailed information about the procedures, benefits and risks, use of personal information, and other details about the project. In addition, the web platform provided an information page that explained the purpose of the project, a list of team members, and contact information. For the exact consent text, please visit https://www.microsoft.com/en-us/research/project/asl-citizen/consent-form/.

In addition to the procedures described in the consent form, participants were prompted with instructions as they viewed prompt signs and recorded their own versions. Screenshots that include the task instructions are provided in Fig. 1 of [1].

**Did the individuals in question consent to the collection and use of their data?** If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.

Yes. (See answer and links above.)

**If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?** If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).

Yes. Users could re-record videos, delete their videos from the collection, as well as withdraw from the dataset any time before public release.

**Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted?** If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

Yes, a Data Protection Impact Analysis (DPIA) has been conducted, including taking a detailed inventory of the data types collected and stored and retention policy, and was successfully reviewed by Microsoft.

**Any other comments?**
None.

<div style="text-align:center">

**Preprocessing/cleaning/labeling**

</div>

**Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)?** If so, please provide a description. If not, you may skip the remainder of the questions in this section.

Yes. First, we removed empty videos automatically, by removing those under 150 KB in size or where YOLOv3 [4] did not detect a person ( 50 videos). We manually reviewed the first and last videos recorded by each participant on each day and random samples

throughout, checking for a list of sensitive content provided by our ethics and compliance board. Three types of personal content were identified for redaction: another person, certificates, and religious symbols. To protect third parties, we used YOLOv3 to detect if multiple people were present. For these videos and others with identified personal content, we blurred the background using MediaPipe holistic user segmentation. We blurred a subset of pixels for one user, since the personal content was reliably limited to a small area. We also removed one user's videos, who recorded many videos without a sign in them, and videos of a written error message. Finally, we manually re-reviewed all videos.

In our reviews, we did not identify any inappropriate content or bad-faith efforts. In total, we blurred the background of 268 videos where a second person was detected automatically, 293 additional videos with sensitive content, and 32 additional videos with a person missed by the automatic detection. We blurred a small fixed range of pixels for 2,933 videos, and omitted 513 videos where the blurring was insufficient or an error message (resulting from the data collection platform) showed.

**Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?** If so, please provide a link or other access point to the "raw" data.

No, not publicly.

**Is the software used to preprocess/clean/label the instances available?** If so, please provide a link or other access point.

No, but these procedures are easily reproducible using public software.

**Any other comments?**
None.

<div style="text-align:center">

**Uses**

</div>

**Has the dataset been used for any tasks already?** If so, please provide a description.

Yes. We provide supervised classification baselines in the manuscript, and show how these classifiers can be used to solve the dictionary retrieval problem.

**Is there a repository that links to any or all papers or systems that use the dataset?** If so, please provide a link or other access point.

Yes. The link is available on our project page at https://www.microsoft.com/en-us/research/project/asl-citizen/.

**What (other) tasks could the dataset be used for?**

Many methods beyond supervised classification can be used to address the dictionary retrieval framing, including unsupervised learning, identification of linguistic features, domain adaptation, etc. To enable these approaches, we provide a larger test dataset. Besides our

5

dictionary retrieval framing, this dataset could be used for a number of purposes both within sign language computing and outside, including pretraining for continuous sign language recognition, or motion tracking.

**Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?** For example, is there anything that a future user might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other undesirable harms (e.g., financial harms, legal risks) If so, please provide a description. Is there anything a future user could do to mitigate these undesirable harms?

Our dataset collection centers a sociolinguistic minority and disability community (the Deaf community) that is already subject to misconceptions, stereotypes, and marginalization. Sign language is a critical cultural component of this community and must be handled respectfully. Some machine learning efforts on sign language proceed without recognition of these existing inequities and cultural practices, and promote harmful misconceptions (e.g. that sign languages are simple, or just signed versions of English), use offensive language or stereotypes (e.g. outdated terminology like "hearing impaired" or "deaf and dumb"), or simply exploit the language as a commodity or "toy problem" without engaging with the community. Some practices that we outline in our paper can avoid these harms: these include including Deaf collaborators and community input in the work, and ensuring that they are compensated; using a critical problem framing that centers useful and culturally respectful applications (e.g. our dictionary retrieval framing); and ensuring that Deaf scholars are cited and their perspectives, concerns, and priorities are integrated into the design of machine learning algorithms.

**Are there tasks for which the dataset should not be used?** If so, please provide a description.

We recommend using this data with meaningful involvement from Deaf community members in leadership roles with decision-making authority at every step from conception to execution. As we describe in our linked paper, research and development of sign language technologies that involves Deaf community members increases the quality of the work, and can help to ensure technologies are relevant and wanted. Historically, projects developed without meaningful Deaf involvement have not been well received[3] and have damaged relationships between technologists and deaf communities.

We ask that this dataset is used with an aim of making the world more equitable and just for deaf people, and with a commitment to "do no harm". In that spirit, this dataset should not be used to develop technology that purports to replace sign language interpreters, fluent signing educators, and/or other hard-won accommodations for deaf people.

This dataset was designed primarily for work on isolated sign recognition; signing in continuous sentences—like what is needed for translating between ASL and English—is very different. In particular, continuous sign recognition cannot be accomplished by identifying a sequence of signs from a standard dictionary (e.g. by matching to the signs in our dataset), due to grammatical and structural difference in continuous signing from sign modulation, co-articulation effects and contextual changes in the meaning of signs. At a minimum, this dataset would need to be used in conjunction with other datasets and/or domain knowledge about sign language in order to tackle continuous recognition or translation.

**Any other comments?**
None.

---

### Distribution

**Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?** If so, please provide a description.

Yes. This dataset is released publicly, to help advance research on isolated sign language recognition.

**How will the dataset will be distributed (e.g., tarball on website, API, GitHub)** Does the dataset have a digital object identifier (DOI)?

The dataset will be made publicly available for download through the Microsoft Download Center. [Details redacted for review]

**When will the dataset be distributed?**

The dataset was released on 06/12/2023.

**Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?** If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

Yes, the dataset will be published under a license that permits use for research purposes. The license is provided at https://www.microsoft.com/en-us/research/project/asl-citizen/dataset-license/.

**Have any third parties imposed IP-based or other restrictions on the data associated with the instances?** If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.

No, there are no third-party restrictions on the data we release. However, the complimentary phonological evaluations of sign vocabulary in our dataset previously pub-

6

lished by ASL-LEX are published under a CC BY-NC 4.0 license (see https://asl-lex.org/download.html).

**Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?** If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.

No.

**Any other comments?**

None.

---

### Maintenance

**Who will be supporting/hosting/maintaining the dataset?**

The dataset will be hosted on Microsoft Download Center.

**How can the owner/curator/manager of the dataset be contacted (e.g., email address)?**

Please contact ASL_Citizen@microsoft.com with any questions.

**Is there an erratum?** If so, please provide a link or other access point.

A public-facing website is associated with the dataset (see https://www.microsoft.com/en-us/research/project/asl-citizen/). We will link to erratum on this website if necessary.

**Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?** If so, please describe how often, by whom, and how updates will be communicated to users (e.g., mailing list, GitHub)?

If updates are neccessary, we will update the dataset. We will release our dataset with a version number, to distinguish it with any future updated versions.

**If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)?** If so, please describe these limits and explain how they will be enforced.

The dataset will be left up indefinitely, to maximize utility to research. Participants were informed that their contributions might be released in a public dataset.

**Will older versions of the dataset continue to be supported/hosted/maintained?** If so, please describe how. If not, please describe how its obsolescence will be communicated to users.

All versions of the dataset will be released with a version number on Microsoft Download Center to enable differentiation.

**If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?** If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to other users? If so, please provide a description.

We do not have a mechanism for others to contribute to our dataset directly. However, others could create comparable datasets by recording versions of the same signs (from ASL-LEX). Such a dataset could easily be combined with ours by indexing on the signs' unique identifiers.

**Any other comments?**

None.

# References


[1] Danielle Bragg, Abraham Glasser, Fyodor Minakov, Naomi Caselli, and William Thies. Exploring collection of sign language videos through crowdsourcing. *CSCW 2022*, 2022.

[2] Naomi K Caselli, Zed Sevcikova Sehyr, Ariel M Cohen-Goldberg, and Karen Emmorey. Asl-lex: A lexical database of american sign language. *Behavior research methods*, 49(2):784–801, 2017.

[3] Michael Erard. Why sign-language gloves don't help deaf people, 2017.

[4] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.



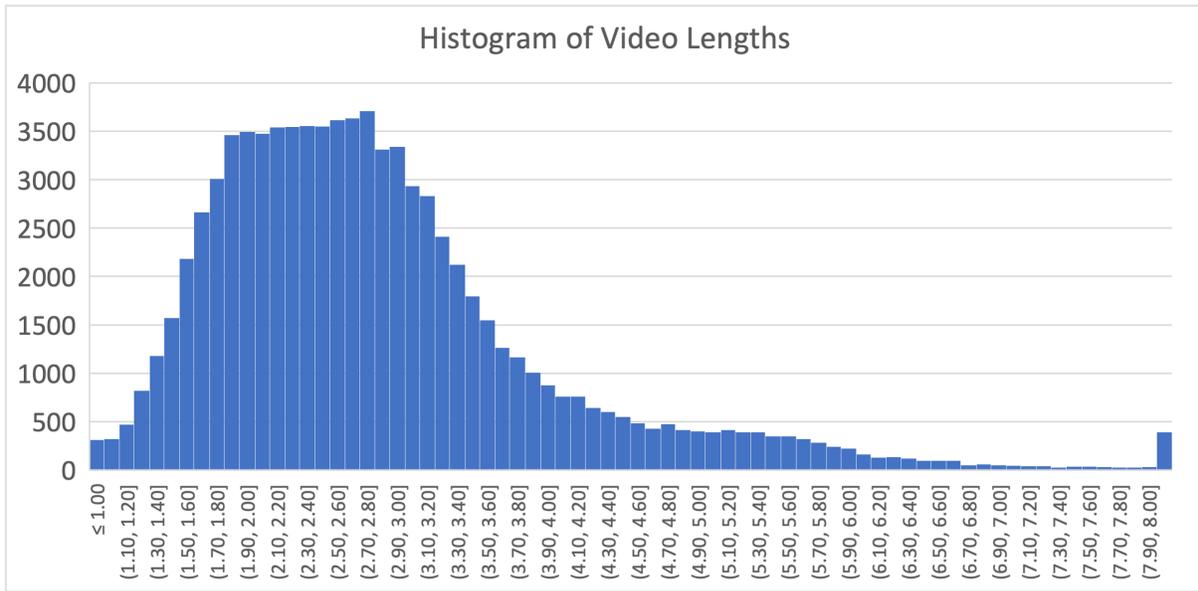
7


Figure 1: Histogram of video lengths in ASL Citizen dataset

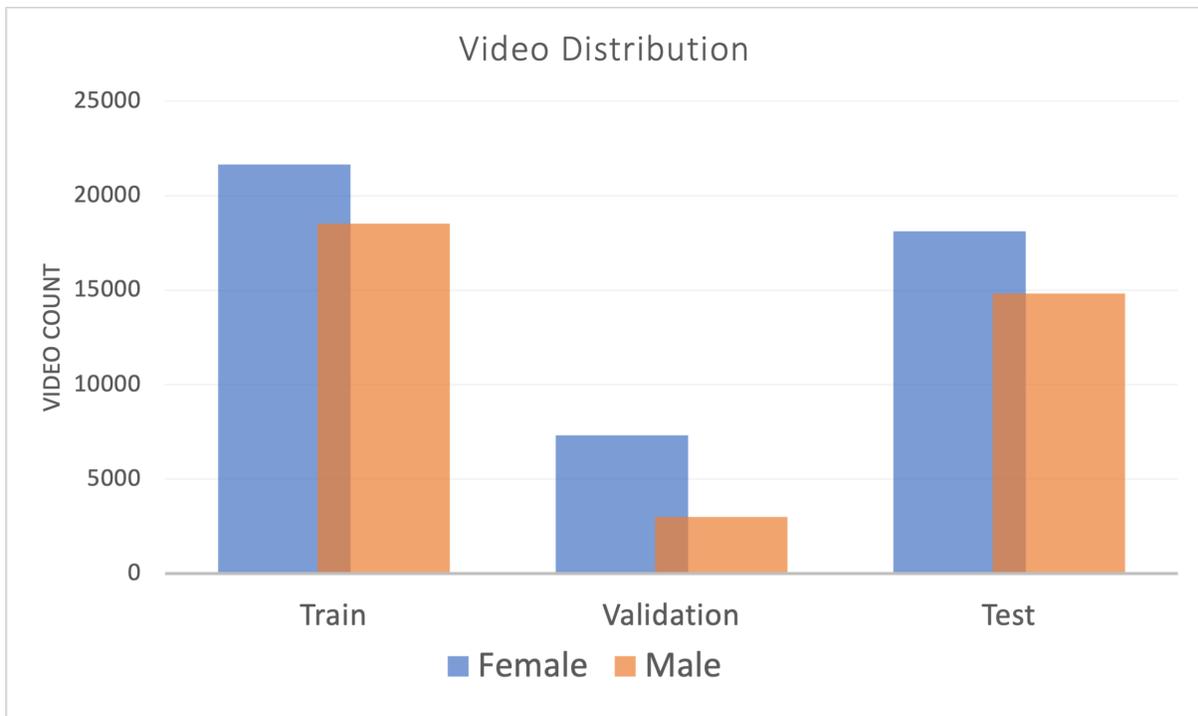



Figure 2: Distribution of videos across different data splits.

8