ARTICLE INFORMATION

Article title

Bangla Sign Language Word Dataset for Image-Based Sign Recognition

Authors

Md. Mijanur Rahman, Md. Nadim-ul-Haque*, Fariha Subrin Riza*, Raisa Tabassum Kabir*

Affiliations

*Department of Computer Science & Engineering, Southeast University, 252, Tejgaon

Industrial Area, Dhaka - 1208, Bangladesh

Corresponding author's email address and Twitter handle

2022100000084@seu.edu.bd (Md. Nadim-ul-Haque), 2022100000073@seu.edu.bd (Fariha Subrin Riza), 2022100000032@seu.edu.bd (Raisa Tabassum Kabir)

Abstract

Communication barriers remain a significant challenge for the Bangladeshi Deaf and hard-of-hearing communities, largely due to the limited availability of comprehensive Bangla Sign Language (BdSL) resources. Compared to other datasets for sign language recognition, existing BdSL datasets lack diversity and an adequate number of samples. Most available datasets focus on a narrow set of word types, such as numbers, letters, and commonly used family- or object-related words, resulting in insufficient variation and representation. To address this gap, the proposed dataset includes 376 JPG-format images with a 500×500 resolution, representing 40 unique Bangla words commonly used in daily communication but not properly labeled in existing BdSL datasets. The selected images include diverse variations and were captured following both universal and local forms of Bangla Sign Language. As these words are not typically found in available datasets, this collection is expected to expand the visual resources for BdSL, support research and development in the field, and assist organizations engaged in sign language education for deaf, mute, and hard-of-hearing individuals.

SPECIFICATIONS TABLE

Subject	Computer Science and Engineering
Specific subject area	Focusing on enriching the existing Bangla Sign Language datasets with a properly diverse selection of words, so they can be used in CNN and deep learning models for recognizing words from signs or gestures, benefiting the Bangladeshi deaf, mute, and hard-of-hearing communities.
Type of data	RGB Images (JPG format)
Data collection	376 JPG images with a constant resolution of 500×500 pixels and an average size of 35KB captured in proper lighting and focusing on gesture.
Data source location	Captured in Southeast University campus, and other locations like; Khilkhet, Dhaka; Mohammadpur, Dhaka, and Tejgaon, Dhaka.
Data accessibility	https://drive.google.com/drive/folders/1eQNT9xcF4vZcxMn5jviETJ_rLeW8378J?usp=drive_link
Related research article	1) Recognition Bangla Sign Language using Convolutional Neural Network [1] 2) Bangla Sign Language recognition using convolutional neural network [2] 3) BdSLW60: A Word-Level Bangla Sign Language Dataset [3]

OBJECTIVE

The primary objective of this work is to develop a comprehensive Bangla word sign language dataset that enables machines to interpret and understand corresponding gestures. This aims to reduce the communication gap between hearing individuals and those who are deaf, mute, or hard of hearing in Bangladesh. The dataset is properly preprocessed and presented in a ready-to-use format suitable for CNN and deep learning-based training models, with a focus on sign detection.

VALUE OF THE DATA

- Represents the commonly used words used in day-to-day conversation but not available on existing datasets.

- Helping educators, researcher, organization working on bangla sign-language based fields.

- Variety of words chosen to create a diverse set of words used to express a meaningful sentence or communication.

- This dataset comes with proper labeling, reducing the extra time required for researchers to convert the dataset into labelled data.

- Maintaining strict standards for all images, including dimension and format, to make them ready-to-use.

BACKGROUND

The core of effective Bangla Sign Language recognition system is heavily dependent on the presence of well maintained, organized, variation and clear datasets. Sad to say the existing datasets available to help Bangla sign language system are limited to both variation and range, making it miss many useful and common Bangla words needed for a person's day to day conversation more expressive and meaningful. To resolve this we at first dig down to search the available datasets and listing the commonly used words in those datasets, after that we note down only those variety of words which is not in those used list. A total of 376 images representing 40 words have been selected and categorized into six groups: Action, Emotion, Object, Context, Status, and Other. This categorization aims to address the structural limitations observed in existing Bangla Sign Language datasets. Making it to provide improved feature extraction and enhance generalization ability of machine learning models trained with this data. Also to make sure the images contain a constant characteristic we make sure the images to hold a constant resolution (500x500), labeling, and an average small size. For labeling each image have given a unique label id. A folder structure is also maintained to making it easier for researcher, educators, developers to access and using the dataset in more ease. This dataset supports the development of communication technologies, improves machine learning models for gesture recognition, and aids in creating educational tools for the Bangladeshi Deaf and hard-of-hearing communities.

DATA DESCRIPTION

This Bangla Sign Language (BdSL) Word Dataset contain a total of 376 high-quality JPG images of 40 Bangla vocabulary items of total 6 types used in daily conversations and expressing emotions. Each image captures a sign in a clear, static posture, providing an accurate visual representation of the gesture. All images are stored in (.jpg) format and a mean file size of approximately 35 KB. Resizing images to a consistent resolution has been shown to improve the performance of convolutional neural networks in gesture recognition tasks [4]. Also before finalizing the images we performed data preprocessing as these steps aim to reduce background interference, improve consistency, and simulate real-world variability [5]. Maintain a uniform resolution of 500 × 500 pixels. Data were collected on the campus of Southeast University, Dhaka, Khilkhet, Mohammadpur, Tejgaon using consumer-grade smartphones (e.g., Redmi Note 7 Pro, Honor 200, Poco X3, Oppo A5 pro). For each sign, several photographs were taken from slightly different angles and distances to capture natural variations in signing. Some images include motion-like frames, where the signer's hands show minor positional changes, supporting the development of models robust to small movements. Lighting conditions were kept simple but not strictly uniform, and both indoor and shaded outdoor environments were used. Participants perform the sign/gesture by following video tutorial of instructors from both universal sign language and Bangla sign language forums.

Table 1:
Number of images captured by which device

Image source	Number of original images	Total
Redmi Note 7 Pro (as webcam)	220	220
Honor 200	70	70
Pocco x3	60	60
Oppo A5	26	26
Total	376	376

Table 2:
Metadata file

	Original data	Augmented
File extension	*.jpg,*.jpeg,	*.jpg
Dimension	Various	500x500 pixel

To ensure consistency, after clicking the images we resized all original photographs were resized to 500 × 500 px, while preserving aspect ratio to prevent distortion.
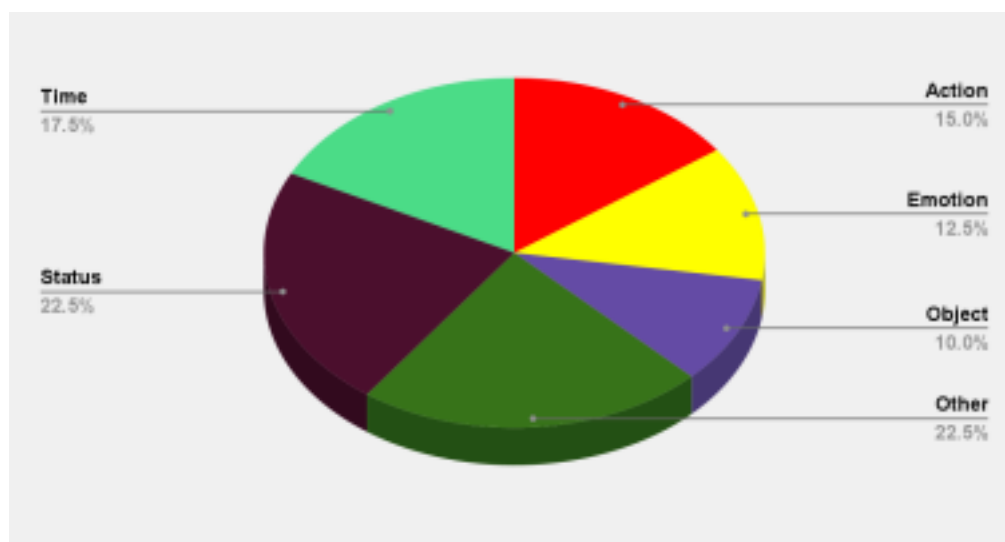

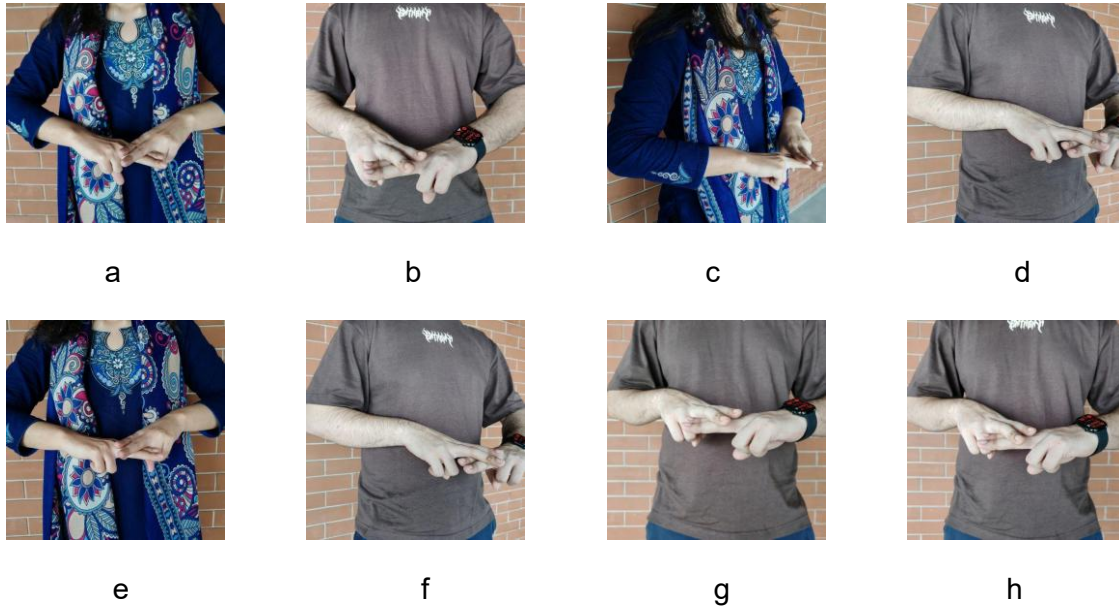Figure 1: Showcasing types of words per category

Figure 2: The example of multiple images of the sign form of the word বসা_sit (a) Female front, (b) Male front, (c) Female right, (d) Male right, (e)Female another version, (f) Male slightly right, (g) Male slightly left, (h) Make another version

EXPERIMENTAL DESIGN, MATERIALS, AND METHODS

For this case the words were selected in reference to make sure these are not available on online datasets. The signs were selected in reference to demonstrations by trusted universal sign language instructors, including video performances by a mute and deaf individual. These demonstrations were re-created and recorded in the form of images.

A total of five participants were involved in reproducing the signs. Image collection was performed using three different mobile devices (Redmi Note 7 Pro, POCO X3, Oppo A5 and Honor 200), ensuring diversity in image quality and resolution. All participants provided informed consent before image capture.

Since high speed, motion blur, different lighting, and viewpoint-dependent appearance have to be tackled [6], in this dataset, to enhance dataset quality, multiple variations of each sign were recorded under different conditions, including variations in hand orientation, lighting, and background. Any unclear or duplicate images were removed through manual screening. All images were resized according to the main gesture region, but no further editing or filtering was applied to preserve the natural appearance of the signs.

For storing data a proper folder structure is maintained for making the searching and using the dataset more comfortable and less tough for users. Universal naming conventions were also followed for resolving confusions of mixing up. For example:

dataset→images→ cateogries(action/context/emotion/object/other/status)→ words under the selected category → image data's

In developing the Bangla Sign Language Word Dataset, we adhered to best practices in data management because properly organized datasets save researchers time and effort, making it easier

to locate, understand, and prepare data for collaboration, reuse, and sharing [7]. Labeling is a cornerstone in supervised machine learning, ensuring that models can learn accurately from data [8]. In this dataset, every image is assigned with a unique label id to maintain consistency, traceability, and reliability during training and testing. For example here with this label id : W10101, we get these information : "W" denotes that it's a word, the first digit refers which type of word (e.g., Action:1, Context:2, Emotion:3, Object:4, Others:5, Status: 6), after this the next two digits indicate serial number of the word within the category, and the final two digits helps to identify the individual image numbering of word. So W10101 indicates: this data is first image of first word under action category

Effective recognition of Bangla Sign Language word-level gestures requires models capable of capturing both the spatial features of hand shapes and the temporal dynamics of gestural movements. A recent review confirms that deep learning, especially convolutional models, continues to set the state of the art for gesture recognition but also emphasizes ongoing challenges for practical, real-world implementation of such systems.[9] In recent time a novel CNNSa-LSTM model, combining CNN, self-attention, and LSTM, achieved 98.7% accuracy by integrating spatial and motion features, showing strong potential for sign recognition [10]

LIMITATIONS

- The dataset is small (376 images) and may not capture the full variety of signs or signer styles in Bangla Sign Language.
- All signs were taken from video demonstrations, so there is limited variation in lighting and background.
- The dataset does not include sequences or motion clips, only single images of each sign.
- The images do not directly convey any emotions of the presenter.

ETHICS STATEMENT

In line with ethical concerns in dataset creation, before capturing images proper consent was obtained from all participants who assists on capturing images. Participants were informed about the purpose of the dataset, how the images would be used, and their right to withdraw at any time. All contributors gave informed consent for the collection and publication of their images for research and educational purposes.

CRedIT AUTHOR STATEMENT

Md. Nadim-ul-Haque: Data collection, dataset preparation, Writing – Original Draft, image annotation, Labelled information, Review, Resizing & Editing. **Fariha Subrin Riza:** Data collection, Investigation,image annotation, Supervision. **Raisa Tabassum Kabir:** Data collection, Writing – Original Draft, image annotation, Review & Editing.

ACKNOWLEDGEMENTS

verification process.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] M. S. Islalm, M. M. Rahman, M. H. Rahman, M. Arifuzzaman, R. Sassi and M. Aktaruzzaman, "Recognition of Bangla Sign Language using Convolutional Neural Network," *2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, Sakhier, Bahrain, 2019, pp. 1-6, doi: 10.1109/3ICT.2019.8910301.

[2] F. Yasir, P. W. C. Prasad, A. Alsadoon, A. Elchouemi, and S. Sreedharan, "Bangla Sign Language recognition using a convolutional neural network," *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Kerala, India, 2017, pp. 49-53, doi: 10.1109/ICICICT1.2017.8342533.

[3] Rubaiyeat, H. A., Mahmud, H., Habib, A., & Hasan, M. K. (2024). BdSLW60: A Word-Level Bangla Sign Language Dataset. *ArXiv*. https://arxiv.org/abs/2402.08635

[4] Wijaya, K., & Widiyanto, E. (2023). Klasifikasi kepemilikan tanda tangan menggunakan convolutional neural network dengan arsitektur AlexNet. MDP Student Conference, 2, 133–143 https://doi.org/10.35957/mdp-sc.v2i1.4328

[5] Aly, M., & Fathi, I. S. (2025). Recognizing American Sign Language gestures efficiently and accurately using a hybrid transformer model. Scientific Reports, 15(1), 1-27. https://doi.org/10.1038/s41598-025-06344-8

[6] Koller, O., Forster, J., & Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. Computer Vision and Image Understanding, 141, 108-125. https://doi.org/10.1016/j.cviu.2015.09.013

[7] Briney KA, Coates H, Goben A (2020). Foundational Practices of Research Data Management. Research Ideas and Outcomes 6: e56508.https://doi.org/10.3897/rio.6.e56508

[8] Fredriksson, T., Mattos, D.I., Bosch, J., Olsson, H.H. (2020). Data Labeling: An Empirical Investigation into Industrial Challenges and Mitigation Strategies. In: Morisio, M., Torchiano, M., Jedlitschka, A. (eds) Product-Focused Software Process Improvement. PROFES 2020. Lecture Notes in Computer Science(), vol 12562. Springer, Cham. https://doi.org/10.1007/978-3-030-64148-1_13

[9] Wu, M. (2024). Gesture Recognition Based on Deep Learning: A Review. EAI Endorsed Transactions on E-Learning, 10. https://doi.org/10.4108/eetel.5191

[10] Baihan, A., Alutaibi, A. I., Alshehri, M., & Sharma, S. K. (2024). Sign language recognition using modified deep learning network and hybrid optimization: A hybrid optimizer (HO) based optimized CNNSa-LSTM approach. *Scientific Reports*, *14*(1), 1-22. https://doi.org/10.1038/s41598-024-76174-7