Recognition of Bangla Sign Language for Letters and Words using Hand Gestures and Predictive Analytics

Mohammad Sabbir Musfique[1], Asir Ahbab Raiyan[2], Munjib Hasan Chowdhury[3], Md. Enamul Hoque Marzun[4], Md. Abdus Sattar[5], and Muhammad Nazrul Islam[6]*

[1] Department of Computer Science and Engineering, Military Institute of Science and Technology, Mirpur Cantonment, Dhaka-1216, Bangladesh

{sabbir.musfique01, ahbab.raiyan, munjibhasan, enamul96209620}@gmail.com, masattar@cse.mist.ac.bd, nazrul@cse.mist.ac.bd*

Abstract. Sign language is the primary means of communication for people who are deaf. Despite the availability of enough research work for English Sign Language, research work on Bangla Sign Language (BdSL), particularly including both the letters and words, is limited. This paper assesses various models for the recognition of BdSL letters, words, and the combination of them through Machine Learning (ML) and Deep Learning (DL) models. Initially, the BdSL47 dataset (47,000 images for 47 different signs) was used to assess all the models. Following that, the 30-word BdSL dataset consisting of 1,200 images was augmented to contain 4,800 images. Finally, the augmented 30-word BdSL dataset was merged together with the BdSL47 dataset. Again, all the various models proposed in the paper were used for model assessment. Models used for the assessment in the proposed paper work include CNN model, MobileNetV2 model, VGG 16 model, KNN model, and Random Forest model. Deep Learning models deliver the best results for the merged vocabulary dataset in the proposed research work by giving 98.44% accuracy compared to the ML models like Random Forest and KNN.

Keywords: Communication, BdSL, Image Augmentation, Convulational Neural Network (CNN), KPI, MobileNet, VGG16

1 Introduction

Sign language serves as a visual gesture-based primary communication medium for hearing-impaired individuals. Deaf people can use sign language to share their feelings and express their emotions. Over five percent of the population are deaf in the world. These people deal with difficulties in interacting with others especially when joining the workforce, education, healthcare, and transportation [3]. Approximately 15% of the world's population have some degree of hearing loss, and many of them are children [4]. Approximately 71 million people worldwide use the spatial movements-based language for their primary interactions [5]. In Bangladesh [1], there are 153,776 vocal disabled people, 73,507 hearing disabled people, and 9625 hearing and visually disabled people.

A digital Bangla Sign Language Interpretation system can surpass this communication barrier between vocally hearing disabled people and a common person. Bangla Sign Language (BdSL) is a unique form of communication which is often overlooked and is different from spoken or written languages. It uses hand shapes, palm orientation, complex body gestures, and facial expressions to show meaning. Many BdSL signs require both hands which eventually increase computational complexity [11]. There also exists the challenge of capturing the hand sign images in a proper background where high illumination of light is present [10].

While Machine Learning (ML) has significantly advanced sign language recognition for languages like ASL, its application to Bangla Sign Language (BdSL) remains limited. Currently, studies on BdSL mostly employ traditional computer vision methods for isolated letter recognition while modern ML approaches for both letter and continuous word recognition are yet to be explored [6]. The study addresses the gap by systematically examining state-of-the-art ML models for comprehensive interpretation of BdSL.

The objective of the research are: (i) To explore the existing ML and DL models of recognizing Bangla sign for letters, (ii) To explore the existing prediction models of recognizing Bangla sign for words and (iii) To find out the best performed machine learning algorithms for predicting both Bangla Sign Language for both letters and words.

To attain these objectives, two open-source datasets were collected; one for word and another for letters. The datasets were merged, including BdSL47 (47 signs, 47,000 images) and a newly curated dataset of 30

© The Author(s) 2026
M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_87

Bengali words (1200 images). The dataset was enhanced using image augmentation techniques (grayscale, Gaussian blur, high/low contrast). As such, the outcome of this study showed that: (a) CNN algorithm performs better compared to other algorithms and is also consistent with respect to the background image. (b) Similarly, the KNN model performs consistently better while recognizing both letters and words of BdSL. The ML algorithms such as CNN, MobileNet v2 and VGG16; and also the DL algorithms which include KNN, RandomForest were deployed to different sets of data. The performance of each model is evaluated independently, with no inter-model dependencies affecting the comparative analysis.

The rest of the article is organized as follows. Section 2 discusses the related researches previously done in this area. Section 3 presents the research methodology and the implementation of proposed system. Section 4 discusses about the results and analysis of the conducted research work. Finally, Section 5 concludes the article by highlighting the research contributions, research limitations and future plans.

2 Literature Review

Existing study highlights two foundational datasets for BdSL to address the communication gap including the lexical Bangla Ishara Bhasha Obhidhan [3] and the conversational Ishara Bhashay Jogajog [4], which address communication gaps. The first dataset named "Ishara-Lipi" was mentioned in [7], for isolated Bangla characters. In the article [5], it was mentioned that the authors presented a real-time computer vision-based BSL recognition system with a vowel recognition accuracy of 98.17%. Most of the existing models have focused on letters or numerical digits. Most of the approaches are not scalable for dynamic gestures or larger vocabularies of BdSL [5]. Previous works mostly concentrated on alphabet and digit recognition, thus leaving the topic of detecting static-gesture words in BdSL largely unexplored.

Research on Bangla Sign Language recognition is relatively less as compared to other sign languages like American Sign Language or Indian Sign Language. 2D and 3D tracking sensors for depth information and segmentation Machine learning models, including but not limited to HMM, CRF, and SVM, have been used for the identification, feature extraction from gestures, and gesture recognition. Deep learning approaches, especially CNNs, have become popular owing to their high-level feature extraction capability and higher accuracy. VGG16, VGG19, and custom CNN architectures, pre-trained models were applied on static and dynamic gestures and showed high accuracy in isolated datasets [15]. Now, there is some clear evidence that different types of studies were conducted on sign language detection around the world, most of them based on American Sign Language, Thai Sign Language, and Arabic Sign Language. Methods involving YOLOv3 for the real-time conversion of ASL and CNNs for the detection and speech generation from Arabic Sign Language have also given hopeful results. There are methods called SIFT and PCA-based Feature Extraction. That used to detect 38 Bangla signs; these methods converted the images from RGB to HSV color space. There was also Incomplete and a lack of diversity in BdSL datasets. One of the popular datasets for hand gesture classification is "Isharalipi Dataset" but not suitable for real-time object detection as it is of low resolution [3].

There was another article which stated that unlike widely studied sign languages like American Sign Language (ASL), BSL possesses complex grammar and limited resources, hence making the detection and translation difficult. A studies was conducted for real-time BSL alphabet recognizer using deep learning and utilized a dataset of 3,000 images categorized into original, binary, and segmented formats [6]. In [8], authors achieved accuracy for Bengali words and alphabets were 92.5% using a dataset. It was also found that most of the earlier studies considered small or single datasets. Thus, generalization of such models was limited [2].

Very few studies were conducted on Bangla Sign Language as well as focused on static gestures for alphabets and digits. While some efforts were directed to translating BSL into text and the identification of static hand gestures, research regarding sentence construction and dynamic gestures remains few and far between. Large and diverse datasets for word-level or sentence-level recognition are also limited [6]. There are no video dataset available for BdSL in real-time except for a few image-based datasets, that creates a gap in research on video-based BdSL [8]. Again, a dataset named BdSL47 is a comprehensive dataset that can be a valuable resource for the researchers working on computer vision-based Bangla sign language recognition [14].

The researchers and developers can explore the use of multimodal deep learning architectures to correctly identify Bangla hand signs because the dataset contains both RGB images and depth key-points of each sign for analysis. A critical limitation is the scarcity of diverse, high-resolution datasets for BdSL. While some resources exist, yet they are very low in resolution and consisting of limited vocabulary which hinders real-time applications. Recent works were based on Bangla sign alphabets only and also they were

conducted on a small-scale dataset which is indeed an issue in recognizing large number of Bangla sign alphabets and words.

3 Exploring Predictive Models

The methodological overview of the research work is depicted in Figure 1 and the main steps are discussed below sequentially.

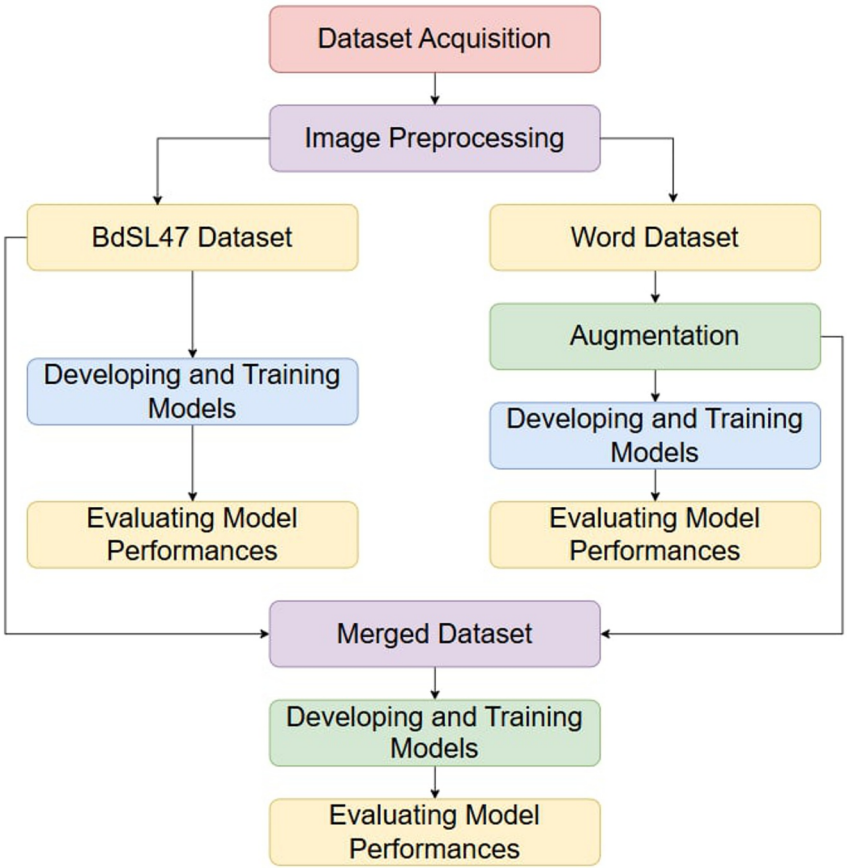

Fig. 1: Methodological overview

3.1 Dataset Collection

Acquiring the data was a crucial part. At first, two datasets consisting of hand images were collected which expressed sign words and letters. The first one is BdSL47 which contains 47000 RGB input images of 47 signs (10 digits, 37 letters) of Bangla Sign Language [14]. And the second dataset consists of 1200 images, categorized into 30 different classes. Each class represents a distinct sign in the Bangla Sign Language(BSL),

with each class containing 40 images. The images in the dataset are in RGB color space. Both are available and collected from open source resources [12].

After that, both datasets were merged to increase the robustness of the model. The merger is supposed to have equal balanced gesture classes and have the real world variations better represented. This dataset was primarily used to evaluate various deep learning models as well as evaluating different machine learning models.

3.2 Image Preprocessing

In the next step, all images obtained found after merging of the data sets were pre-processed. All input images were resized and normalized in the range of 0 and 1 to ensure consistency and to facilitate the machine learning and deep learning models to learn correctly and effectively. We have also made some adjustments in the brightness of the images to enhance the robustness of machine learning models. In general, fine-tuning was done only when available images of the dataset are not drastically different in context from the dataset on which the pre-trained model is already trained. After that as part of image augmentation, they were converted into four distinct types:

1. *Grayscale*: All the collected images were converted to grayscale in order to remove color dependencies and reduce computational complexity. The images were resized to 128×128 pixels.
2. *Gaussian Blur*: The model applied a slight blur using a 5×5 kernel to introduce slight variations in image texture.
3. *High Contrast*: Next, the brightness (alpha=1.2, beta=30) was increased of the corresponding images.
4. *Low Contrast*: After that, gradual decrease of brightness by setting alpha=0.8 and beta=-30 of the hand-gesture images took place.

The first picture of every row(a) in Figure 3 is indicating the original hand-gesture image of the sign words, which is basically the colored images. After collecting such samples, the dataset was enhanced using four key augmentation techniques: (a) Grayscale conversion, (b) Gaussian Blur, (c) High Contrast, and (d) Low Contrast. These augmentations are placed sequentially ((b, c, d & e) of every row) in the picture as mentioned in Figure 3. Such preprocessing certainly improves the generalization of the model by introducing variations in lighting, texture, and noise conditions.

Preprocessed images were saved in a structured directory hierarchy based on their class labels. This preprocessing enriches the dataset by incorporating slight variations while retaining the essential features of the images. Each class of static hand gesture images was stored in a separate subdirectory, with the directory name used for the class label (e.g., “Aaj”, “Baagh”, “Basha”, “Biyog”, “Bondhu”).

3.3 Development of the Model

Next, we focused on the development of a deep learning model for recognizing static hand gestures that can work well in the context of Bengali Sign Language (BSL). The model architecture was specifically devised with CNNs as the most fitting framework for image-based tasks. The augmentation techniques (rotation, flipping, scaling, and contrast adjustment) made the models more robust.

A CNN architecture was formed by a stack of distinct layers that transform the input volume into an output volume (for instance, holding the class scores) through a differentiable function as illustrated in Figure 4. The hyperparameters that were tuned include the learning rate, batch size, and number of epochs-for obtaining optimal performance.

3.4 Hyperparameter Tuning

To enhance the performance of both traditional machine learning (ML) and deep learning (DL) models, hyperparameter optimization was utilized. In the K-Nearest Neighbors and the Random Forest algorithm, since they fall in the class of machine learning models, the GridSearchCV and the 5-fold CV search was utilized in the search for optimal hyperparameters, including the number of neighbors in KNN, the Euclidean and Manhattan distances, and the weights in the KNN algorithm, and the number of trees and depth in the Random Forest algorithm. These search strategies utilize the ’accuracy’ scorings metric.

In the cases of the deep learning models (CNN, MoblieV2, and VGG16), hyperparameters like the batch size, the number of epochs, the dropout rate, and the learning rate were optimized. Furthermore, the training process was monitored using the validity accuracy. The use of early stopping and model checkpointing was adopted. This provided the ability to optimize each model appropriately in terms of their learning ability, hence an increase in the evaluation accuracy, recall, and F1 measures.

Fig. 2: Examples of the merged dataset

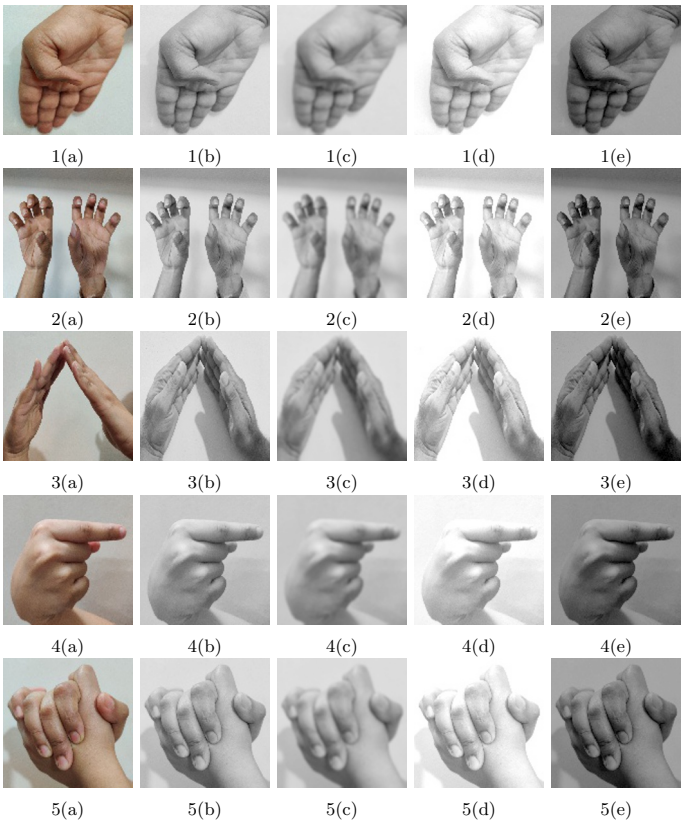

Fig. 3: Preprocessed images of some Bangla sign words: 1. Aaj, 2. Bagh, 3. Basha, 4. Biyog, 5. Bondhu.

3.5 Train-Test Split

Finally, the dataset was divided into two parts. One part for training and the other part for testing. The train-test split in 80:20 ratio was maintained to ensure unbiased evaluation. Preprocessing was critical in maintaining data consistency; all input images were resized to 128x128 pixels so that proper learning could be achieved by the models. Grayscale conversion, thus helping in computation, was done without losing the prime information about the gestures. Training continued for multiple epochs, while the performance of the model across training and validation datasets was monitored. Early stopping was implemented to stop training when validation performance had plateaued to prevent any overfitting. The assigned checkpoint ensured that the model with the best validation accuracy was saved for the purposes of final evaluation. Finally, the testing phase was checked for generalization capability, and performance metrics indicated that deep learning models outperform conventional machine learning models in recognizing BdSL static gestures.

3.6 Prediction of the Model

After having trained the machine learning and deep learning models properly, we had performed extensive testing to measure how well the models could distinguish Bengali Sign Language gestures. The testing was done with some performance metrics that included accuracy, recall, F1-score, and confusion matrix assessment.

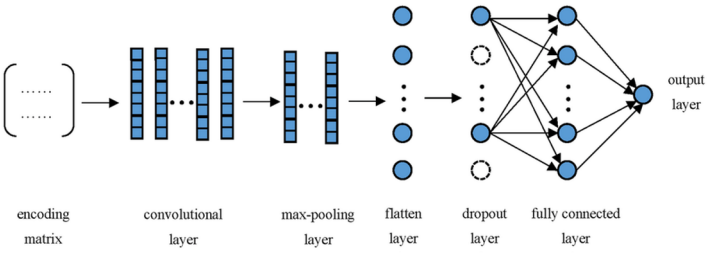

Fig. 4: CNN architecture

The evaluation procedure involved mapping predicted labels for gestures to true labels in order to evaluate classification performance. The evaluation's findings informed subsequent model architecture, guaranteeing peak performance for real-time Bengali Sign Language identification.

4 Results and Analysis

The system has been evaluated with 9640 test images in 77 classes which have not been used to train. We have calculated metrics (Recall, F1-Score, and Accuracy) to assess the performance of the hand gesture detection model. In the experiment, implementation of various deep learning models was taken as mentioned in table 1.

Table 1: Result for the letters using deep learning algorithms

Algorithm Name	**Test Accuracy(%)**	**F1 score(%)**	**Recall(%)**
CNN	98.09	98.09	98.09
MobileNet v2	96.57	96.56	96.57
VGG16	94.64	94.64	94.64

Consequently, it was found that the CNN (Convolutional Neural Network) model provides the highest test accuracy (98.09%) along with the highest F1 score and recall among other deep learning algorithms for the letters and VGG16 had provided the least accuracy (94.64%).

Table 2: Result for the words using deep learning algorithms

Algorithm Name	**Test Accuracy(%)**	**F1 score(%)**	**Recall(%)**
CNN	77.92	77.96	77.92
MobileNet v2	94.58	94.62	94.58
VGG16	97.08	97.07	97.08

In case of word dataset it was found that VGG16 had performed much better compared to other deep learning models, as mentioned in table 2. The highest accuracy obtained here is 97.08%, where CNN had provided the least accuracy which was 77.92%.

After merging both the datasets, the system using the previously mentioned deep learning algorithms. The KPI values obtained in this case are mentioned in table 3. CNN was found to perform better than all other deep learning algorithms that we evaluated in the system. It had given the highest test accuracy (98.44%) along with the highest F1 score and recall. In addition, MobileNet v2 had the lowest test accuracy (94. 2%) and the overall lowest F1 score (89. 67%) and recall (89. 64%) were obtained from VGG16. Better results were achieved compared to some existing studies [2] [15].

Table 3: Result for the merged dataset using deep learning algorithms

Algorithm Name	Test Accuracy(%)	F1 score(%)	Recall(%)
CNN	98.44	98.44	98.44
MobileNet v2	94.2	94.17	94.2
VGG16	97.08	89.67	89.64

Table 4: Result for the merged datasets using machine learning algorithms

Algorithm Name	Test Accuracy(%)	F1 score(%)	Recall(%)
KNN	96.05	96.01	96.05
RandomForest	96.24	96.19	96.24

From table 4 it is evident that RandomForest had comparatively provided better results than KNN algorithm.

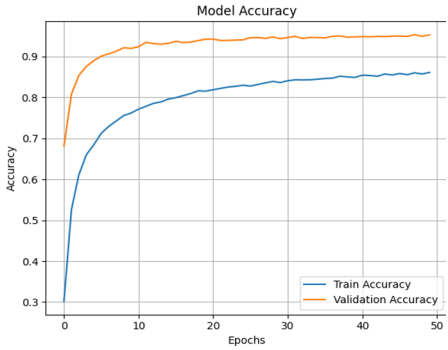

Fig. 5: Model Accuracy of augmented images training set

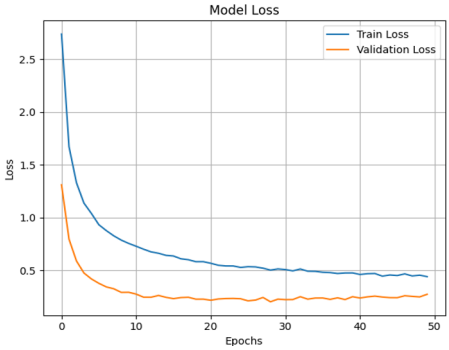

Fig. 6: Model Loss of augmented images training set

It is evident from Figure 5 that the model's performance improves as the number of epochs increases. The training accuracy (blue curve) steadily rises and converges near 90%, while the validation accuracy (orange curve) stabilizes above 90%. Figure 6 illustrates the loss of the model that was discovered in the system for Bengali sign language recognition. Both the blue line (train loss) and orange line (validation loss) are trending downwards, meaning the model is learning well. The initial loss is high but gradually reduces immediate in the first phases and then starts to settle. The validation loss is less than the train loss, meaning it suggests good generalization performance with very low overfitting.

Despite the success achieved by both models in the controlled dataset context, certain aspects may also affect the model's ability to generalize. Background lighting, camera viewpoints, signer identity, and gesture complexity might all impact the model's behavior concerning the precision of its predictive capabilities. At the current level, the dataset lacks signer variability, whereby the gestures used in the dataset are merely rigid representations.

The study results showed that the accuracy of CNN was the highest among all other algorithms that were implemented in the research work. MobileNet V2 has the least accuracy obtained in the work. Similarly it was also obtained that CNN had achieved the highest F1-score of 98.44%, indicating that it maintains a strong balance between precision and recall irrespective of image background. VGG16 and Random Forest follow closely with 97.54% and 96.90%, respectively, showing their ability to perform well in both identifying and correctly classifying signs of Bangla Sign Language.

The merged dataset improved robustness across all models while keeping the overfitting minimized and generalization maximized, especially for the DL models. Consistently, DL models did better than ML models with respect to accuracy, F1 score and Recall whereas ML models were more time-efficient at training and had a lower computational requirement, suitable for resource-constrained environments.

5 Conclusion

From the research work, the potential of both Machine Learning Models and Deep Learning Models in effectively recognizes Bangla Sign Language for individual words and letters. The research work proves that the BdSL recognition capabilities by the CNN algorithm are better compared to other models. Also, The KNN model also shows consistent accuracy for both letters and words. All evaluations were conducted independently, ensuring unbiased performance comparisons. One limitation of the study is that performance may vary when gestures are taken from complex or dynamic environments, or when signer-specific variations occur. In future work, the dataset will be expanded by adding more samples, additional word categories, and a greater diversity of signers to strengthen generalization. These advancements are intended to support real-time BdSL translation applications and make the system more practical for day-to-day use.

Acknowledgement

The authors used standard language editing software to enhance the flow and readability of the text. All suggestions were manually reviewed and verified by the authors to ensure accuracy. Furthermore, no artificial intelligence tools were utilized for data analysis, image generation, or the interpretation of results in this study.

References

1. Podder, K.K., et al.: Bangla Sign Language (BdSL) alphabets and numerals classification using a deep learning model. Sensors 22(2), 574 (2022). doi:10.3390/s22020574
2. Siddique, S., Islam, S., Neon, E.E., Sabbir, T., Naheen, I.T., Khan, R.: Deep learning-based Bangla sign language detection with an edge device. Intelligent Systems with Applications 18, 200224 (2023). doi:10.1016/j.iswa.2023.200224
3. Talukder, D., Jahara, F.: Real-time Bangla sign language detection with sentence and speech generation. (2020)
4. Urmee, P.P., Al Mashud, M.A., Akter, J., Jameel, A.S.M.M., Islam, S.: Real-time Bangla sign language detection using Xception model with augmented dataset. In: Proc. IEEE WIECON-ECE, pp. 1–5. IEEE (2019)
5. Lipi, K.A., Adrita, S.F.K., Tunny, Z.F., Munna, A.H., Kabir, A.: Static-gesture word recognition in Bangla sign language using CNN. TELKOMNIKA 20(5), 1109 (2022)
6. Shurid, S.A., et al.: Bangla sign language recognition and sentence building using deep learning. In: Proc. IEEE CSDE. IEEE (2020)
7. Rahaman, M.A., Jasim, M., Ali, Md.H., Hasanuzzaman, Md.: Bangla language modeling algorithm for automatic recognition of hand-sign-spelled Bangla sign language. Frontiers of Computer Science 14(3) (2019)
8. Sams, A., Akash, A.H., Rahman, M.: SignBD-Word: Video-based Bangla word-level sign language and pose translation. (2023)
9. Bin Munir, M., Alam, F.R., Ishrak, S., Hussain, S., Shalahuddin, Md., Islam, M.N.: A machine learning-based sign language interpretation system for communication with deaf-mute people. In: Proc. XXI Int. Conf. on Human-Computer Interaction (2021)
10. Rafi, A.M., Nawal, N., Nur, L., Nima, C., Shahnaz, C., Fattah, S.A.: Image-based Bengali sign language alphabet recognition for deaf and dumb community. In: Proc. IEEE Conference (2019)
11. Uddin, M.A., Chowdhury, S.A.: Hand sign language recognition for Bangla alphabet using SVM. IEEE Xplore (2016)
12. Jim, A.A.J., Rafi, I., Akon, Md.Z., Biswas, U., Nahid, A.A.: KU-BdSL: An open dataset for Bengali sign language recognition. Data in Brief 51, 109797 (2023). doi:10.1016/j.dib.2023.109797
13. Tiku, K., Maloo, J., Ramesh, A., R., I.: Real-time conversion of sign language to text and speech. In: Proc. ICIRCA. IEEE (2020)
14. Rayeed, S.M.: BdSL47: A complete depth-based Bangla sign alphabet and digit dataset. Mendeley Data 3 (2023)
15. Islam, M.S., Rahman, M.M., Rahman, Md.H., Arifuzzaman, M., Sassi, R., Aktaruzzaman, M.: Recognition of Bangla sign language using CNN. In: Proc. 3ICT. IEEE (2019)
16. Rahaman, M.A., Jasim, M., Ali, Md.H., Hasanuzzaman, Md.: Bangla language modeling algorithm for automatic recognition of hand-sign-spelled Bangla sign language. Frontiers of Computer Science 14(3) (2019)

17. Roy, P., Uddin, S.M.M., et al.: Bangla sign language conversation interpreter using image processing. IEEE Xplore (2019)
18. Jin, C.M., Omar, Z., Jaward, M.H.: A mobile application of American sign language translation via image processing algorithms. IEEE Xplore (2016)
19. Shahid, M., et al.: CNN–Squeeze Excitation Network for Bengali sign language recognition and mobile deployment. Sensors 24(16), 5351 (2024)
20. Miah, M., et al.: BenSignNet: Bengali sign language alphabet recognition using concatenated segmentation and CNN. Applied Sciences 12(8), 3933 (2022)
21. Ahmed, R.T., et al.: Computer vision-based Bangla sign language recognition using transfer learning. (2023)
22. Rubaiyeat, M., et al.: BdSLW60: A word-level Bangla sign language dataset for sentence-level understanding. arXiv:2402.08635 (2024)
23. Shahgir, A., et al.: Spatio-temporal graph neural networks for Bangla sign language word-level recognition. arXiv:2401.12210 (2024)
24. Abedin, T., Islam, M.M., Shuvo, M.R., Rahman, A.: A hybrid CNN and hand pose estimation approach for Bengali sign language recognition. arXiv:2107.11818 (2021)
25. Das, D., Roy, R., Islam, M., Sarker, S.: A hybrid model for Bengali sign language recognition using transfer learning and random forest. Expert Systems with Applications 200, 118914 (2022)
26. Basnin, S., Nahar, S., Hossain, M.A.: Bangla sign language recognition using CNN–LSTM model. In: Proc. ICCCS, pp. 663–671. Springer, Singapore (2021)
27. Haque, M., Nahian, M.T., Molla, M.S., Hasan, M.M.: BdSpell: A YOLOv5-based spelling system for Bangla sign language recognition. arXiv:2309.13676 (2023)
28. Pranto, F., Siddique, M.R.: A real-time Bangla sign language translator using Mediapipe and LSTM. arXiv:2412.16497 (2024)
29. Shawon, S., Sakib, M., Alam, M.N., Hasan, S.N.: Exploring video transformer models for Bangla sign language word recognition. arXiv:2506.04367 (2025)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.