Static-gesture word recognition in Bangla sign language using convolutional neural network

Kulsum Ara Lipi[1], Sumaita Faria Karim Adrita[1], Zannatul Ferdous Tunny[1], Abir Hasan Munna[1], Ahmedul Kabir[2]
[1]Department of Information and Communication Technology, Faculty of Science and Technology, Bangladesh University of Professionals, Dhaka, Bangladesh
[2]Institute of Information Technology, University of Dhaka, Dhaka, Bangladesh

Article Info

Article history:

Received Sep 17, 2021
Revised Jun 23, 2022
Accepted Jul 01, 2022

Keywords:

BSL
BSL word dataset
Convolutional neural network
Static-gesture signs

ABSTRACT

Sign language is the communication process of people with hearing impairments. For hearing-impaired communication in Bangladesh and parts of India, Bangla sign language (BSL) is the standard. While Bangla is one of the most widely spoken languages in the world, there is a scarcity of research in the field of BSL recognition. The few research works done so far focused on detecting BSL alphabets. To the best of our knowledge, no work on detecting BSL words has been conducted till now for the unavailability of BSL word dataset. In this research, a small static-gesture word dataset has been developed, and a deep learning-based method has been introduced that can detect BSL static-gesture words from images. The dataset, "BSLword" contains 30 static-gesture BSL words with 1200 images for training. The training is done using a multi-layered convolutional neural network with the Adam optimizer. OpenCV is used for image processing and TensorFlow is used to build the deep learning models. This system can recognize BSL static-gesture words with 92.50% accuracy on the word dataset.

Corresponding Author:

Ahmedul Kabir
Institute of Information Technology, University of Dhaka
Dhaka, Bangladesh
Email: kabir@iit.du.ac.bd

1. INTRODUCTION

Bangla is the fifth-most widely spoken language on the planet, spoken by almost 230 million people in Bangladesh and the eastern parts of India. Among them, more than three million are mute or hard of hearing [1]. There is an enormous correspondence gap between those who can speak and listen to the language, and those who cannot. The only way deaf and mute people can communicate is using sign language which uses manual correspondence and body language to pass on significant information. This mode of communication is quite hard to understand for regular people. This is where the field of computer vision is arriving at a potential area to help this communication. Nowadays, computer vision is used for assisting deaf and mute people by automated sign language detection technique. However, these technologies are not so readily available to the people of underdeveloped countries like Bangladesh.

There are not many books where Bangla gesture-based communication can be studied by deaf and mute people. National Centre for Special Education Ministry of Social published a book named "*Bangla Ishara Bhasha Obhidhan*" (Bangla sign language dictionary) edited by Bangladesh sign language (BSL) committee in January 1994, and reprinted in March 1997. This book follows British sign pattern. The centre for disability in development (CDD) published another book named "*Ishara Bhashay Jogajog*" (communication in sign language) in 2005 and reprinted in 2015. Apart from these, there are not many options for people to understand sign language. And this is a huge undertaking that very few people are able to do. If there would be a Bangla sign language recognizer model, general individuals could easily interact with disabled individuals. This would reduce the disparity between people with disabilities and the general population, and ensure a more just society with equal opportunity for all.

This however is a far cry from the current reality for a number of reasons. There is no proper dataset for Bangla sign words for scientific work and progression. There is also not enough successful research on Bangla gesture-based communication. In an attempt to alleviate this situation to some extent, we built up a dataset called BSLword consisting of images of different words in Bangla sign language. This dataset will help in research-based work and improvement of Bangla sign language. Moreover, we utilized the deep learning method called convolutional neural network (CNN) to build a model that can recognize words from the dataset. In this paper, we describe our whole process of dataset construction and model development.

In 2019, Hasan *et al.* [2] proposed an easily understandable model that recognizes Bangla finger numerical digits. Using numerous support vector machines for classifying images, they used the histogram of directed gradient image features to build a classifier. They selected 900 images for training and 100 for testing, respectively, from ten-digit groups. Their system acquired approximately 95% accuracy. Earlier in 2018, Hoque *et al*. [3] proposed a procedure to recognize BSL from pictures that acts continuously. They utilized the convolutional neural organization-based article recognition strategy. Their approach was faster region-based and they obtained an average accuracy rate of 98.2 percent. Their constraint was perceiving the letters, which have numerous likenesses among their patterns. Before that, Uddin *et al.* [4] in 2017 suggested a model of image handling focused on Bangla sign language translation. At first, YCbCr shading segments recognize the client’s skin shade and afterward separates the set of features for each input picture. At last, the separated features are fed to the support vector machine (SVM) to prepare and test. The suggested model showed an average of 86% accuracy for their trial dataset.

Hossen *et al*. [5] proposed another strategy of Bengali sign language recognition that uses deep CNN (DCNN). Static hand signals for 37 letters of the Bengali letter set are interpreted by the technique. Directing tests on three 37 sign arrangements with full 1147 images with shifting the accuracy of feature concentrations taken from each test, they have achieved a robust general recognition rate of 96.33 percent in the training dataset and 84.68 percent in the validation dataset using a deep CNN. In the same year, Islam *et al*. [6] developed a deep learning model to cope with perception of the digits of BSL. In this methodology, they utilized the CNN model to prepare specific signs with a separate preparing dataset. The model was designed and tried with separately 860 training pictures and 215 test pictures. Their training model picked up about 95% precision. Prior to that, in 2016, Uddin and Chowdhury [7] introduced a structure in 2016 to perceive BSL by the use of support vector machine. By analysing their structure and looking at their features, which distinguish each symbol, Bangla sign letters are perceived. They changed hand signs to hue, saturation, value (HSV) shading space from the red, green, blue (RGB) picture in the proposed system. At that point, Gabor channels were utilized to obtain wanted hand sign features. The accuracy of their proposed structure is 97.7%.

Islam *et al* [1], Ishara-Lipi published in 2018, was the primary complete segregated BSL dataset of characters. The dataset includes 50 arrangements of 36 characters of Bangla basic signs, gathered from people with different hearing disabilities, including typical volunteers. 1800 characters pictures of Bangla communication via gestures were considered for the last state. They got 92.65% precision on the training set and 94.74% precision on the validation set. Ahmed and Akhand (2016) [8] presented a BSL recognition system centred on the position of fingers. To train the artificial neural network (ANN) for recognition, the method considered relative tip places of five fingers in two-measurement space, and used location vectors. The proposed strategy was evaluated on a data set with 518 images with 37 symbols, and 99% recognition rates were achieved.

In 2012, Rahman *et al*. [9] proposed a framework for perceiving static hand gestures of the letter set in Bangla gesture-based communication. They prepared ANN with the sign letters’ features to utilize feedforward back propagation learning calculation. They worked with 36 letters of BSL letter sets. Their framework obtains an average precision of 80.902%. Later, in 2015, Yasir *et al*. [10] introduced a computational way to actively recognize BSL. For picture preparation and normalization of the sign image, Gaussian distribution and grayscaling methods are applied. K-means clustering is performed on all the descriptors, and a SVM classifier is applied.

Islam *et al*. [11] proposed hand gesture recognition using American sign language (ASL) and DCNN. In order to find more informative features from hand images, they used DCNN before performing the final character recognition using a multi-class SVM. Cui *et al.* [12] proposed a recurrent convolutional neural network (RCNN) for continuous sign language recognition. They designed a staged optimization process for their CNN model and tuned it using vast amounts of data and compared their model with other sign language recognition models. Earlier, in 2016, Hasan and Ahmed [13] proposed a sign language recognition system for bilingual users. They used a combination of principal component analysis (PCA) and linear discrimination analysis (LDA) in order to maximize data discrimination between classes. Their system can translate a set of 27 signs to Bengali text with a recognition rate of 96.463% on average. In 2017, Islam *et al*. [14] have applied different algorithms for feature extraction of the hand gesture recognition system. They designed a process for real time ASL recognition using ANN, which achieves an accuracy of 94.32% when recognizing alphanumeric character signs.

Huang *et al.* [15] proposed a 3D CNN model for sign language recognition. They used a multilayer perceptron in order to extract features. They also evaluated their model against 3D CNN and Gaussian mixture model with hidden markov model (GMM-HMM) using the same dataset. Their approach has higher accuracy than the GMM-HMM model. In 2019, Khan *et al*. [16] proposed an approach which will shorten the workload of training huge models and use a customizable segmented region of interest (ROI). In their approach, there is a bounding box that the user can move to the hand area on screen, thus relieveing the system of the burden of finding the hand area. Naglot and Kulkarni [17] used a leap motion controller in order to recognize real time sign language. Leap motion controller is a 3D non-contact motion sensor which can detect discrete position and motion of the fingers. Multi-layer perceptron (MLP) neural network with back propagation (BP) algorithm used to recognize 26 letters of ASL with a recognition rate of 96.15%. Rafi *et al*. [18] proposed a VGG19 based CNN for recognizing 38 classes which achieved an accuracy of 89.6%. The proposed framework includes two processing steps: hand form segmentation and feature extraction from the hand sign.

Rahaman *et al*. [19] presented a real-time computer vision-based Bengali sign language (BdSL) recognition system. The system first detects the location of the hand in the using Haar-like feature-based classifiers. The system attained a vowel recognition accuracy of 98.17 percent and a consonant recognition accuracy of 94.75 percent. Masood *et al.* [20] classified based on geographical and temporal variables using two alternative techniques. The spatial features were classified using CNN, whereas the temporal features were classified using RNN. The proposed model was able to achieve a high accuracy of 95.2% over a large set of images. In 2019, Rony *et al.* [21] suggested a system in which all members of a family, if one or more members are deaf or mute members are able to converse quickly and easily. They used convolutional neural networks in our proposed system for hand gesture recognition and classification as well as the other way around. Also in 2019, Urmee *et al.* [22] suggested a solution that works in real-time using Xception and our BdSLInfinite dataset. They employed a big dataset for training in order to produce extremely accurate findings that were as close to real-life scenarios as possible. With an average detection time of 48.53 milliseconds, they achieved a test accuracy of 98.93 percent. Yasir and Khan [23] Proposed a framework for BSL detection and recognition (SLDR) in this paper. They have created a system that can recognize the numerous alphabets of BSL for human-computer interaction, resulting in more accurate outcomes in the shortest time possible. In 2020, Ongona *et al.* [24] proposed a system of recognizing BSL letters using MobileNet.

In this paper, we have built a dataset of BSL words that use a static gesture sign. To the best of our knowledge, this is the first dataset that deals with BSL words. The dataset can be used for training any machine learning model. We used a CNN on the training portion of the dataset and built a model that gained 92.50% accuracy on the test set. The rest of the paper discusses our methodology and results obtained.

2. METHODOLOGY

2.1. Data collection and pre-processing

There are more than a hundred thousand words in the Bangla language, but all of them do not have a corresponding word in sign language. Most sign language words are represented by waving of one hand or both the hands, while some words are represented with static images just like BSL characters. Since this is rudimentary study in this field, we collected only those words which can be understandable by one hand gesture and can be taken with static images. We found 30 such words from the BSL dictionary. The words are shown here in Bangla script with the English transliteration and translation in brackets: দেশ (‘*desh’*, country), স্যার (‘*sir’*, sir), এখানে (‘*ekhane’*, here), কিছুটা (‘*kichuta’*, a little bit), গুণ (‘*gun’*, multiply), বিয়োগ (‘*biyog’*, subtract), দাঁড়াও (‘*darao’*, stand), বাসা (‘*basha’*, house), সুন্দর (‘*shundor’*, beautiful), বন্ধু (‘*bondhu’*, friend), তুমি (‘*tumi’*, you), কোথায় (‘*kothay’*, where), সাহায্য (‘*shahajjo’*, help), তারা (‘*tara’*, star), আজ (‘*aaj’*, today), সময় (‘*shomoi’*, time), সে (‘*she’*, he), সমাজকল্যাণ (‘*shomajkollan’*, social welfare), অনুরোধ (‘*onurodh’*, request), দাড়ানো (‘*darano’*, to stand), বাঘ (‘*bagh’*, tiger), চামড়া (‘*chamra’*, skin), গির্জা (‘*girja’*, church), হকি (‘*hockey’*, hockey), জেল (‘*jail’*, jail), কেরাম (‘*keram’*, carrom), পিয়ানো (‘*piano’*, piano), পুরু (‘*puru’*, thick), সত্য (‘*shotto’*, truth), বৌদ্ধ (‘*bouddho’*, Buddha). The whole data collection method is divided into five separate steps: Capture images, label all data, crop images, resize images, and convert to RGB format.

2.1.1. Capture images

Our dataset contains a total of 1200 static images, 40 images for each of the 30 words. We collected data from several undergraduate students who volunteered for the work. We captured images of different hand gestures with bare hands in front of a white background. A high-quality resolution mobile camera was used to take all the pictures. Figure 1 shows some sample pictures.

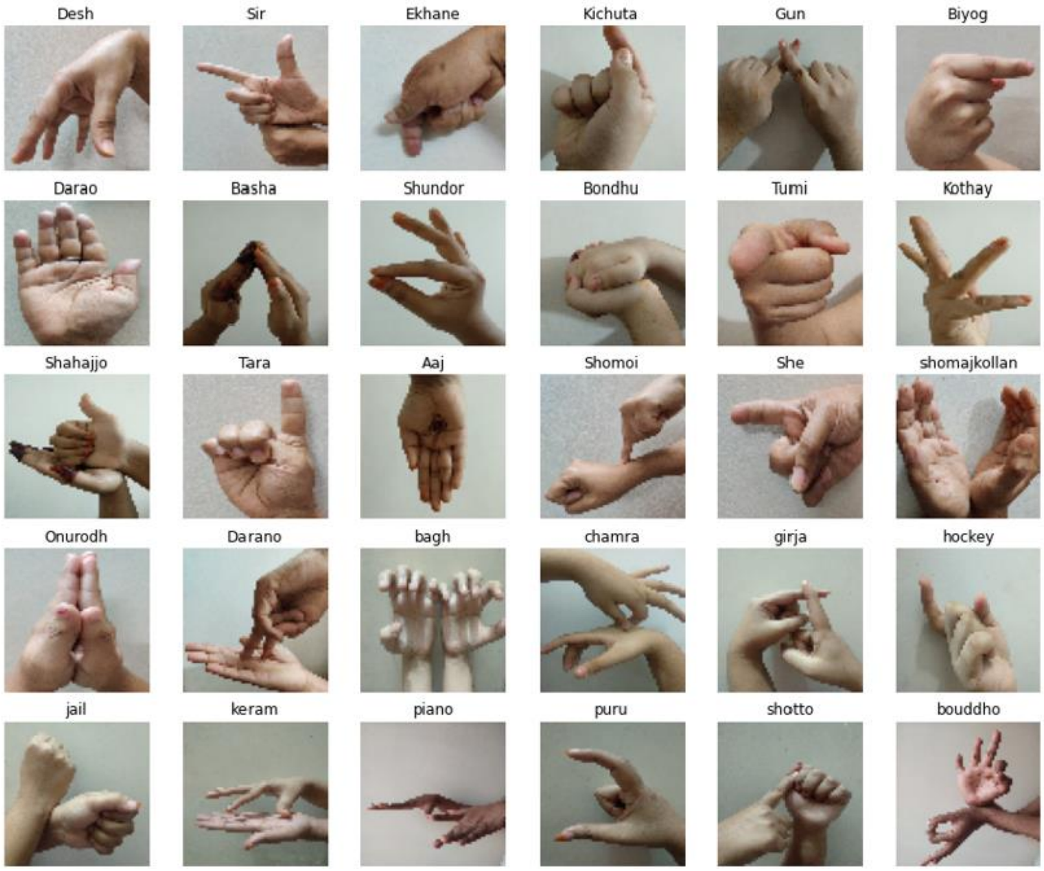

Figure 1. Some captured images (one sample image per word shown)

2.1.2. Label all data

In this step, we categorized all the images and labelled them according to the words. This labelling is important since we are using supervised classification. Our labelling followed a numerical convention from 0 to 29 (0, 1, 2, 3, …, 29).

2.1.3. Crop all images

Due to differences in capturing the images, the hand position within the images is different. Hence cropping is an essential step to use data for continuing the experiment. Uncropped images are all cropped to observe the proportion of width and height for later usage. Figure 2 shows an example of image cropping.

2.1.4. Resize images and converting to RGB

All cropped images are resized to 64×64 images. This step is necessary to make the dataset consistent and to make it suitable to be fed to our deep learning model. Our original pictures are captured in blue, green, red (BGR) color space. So next we convert them to RGB color space.

2.2. Model development

We divided our dataset into two parts using stratified random sampling 80% for training and 20% for testing. We then train our model using the CNN architecture described in the next section. Once the CNN model is created, we can input a random person's Hnd image and the model will detect the sign word.

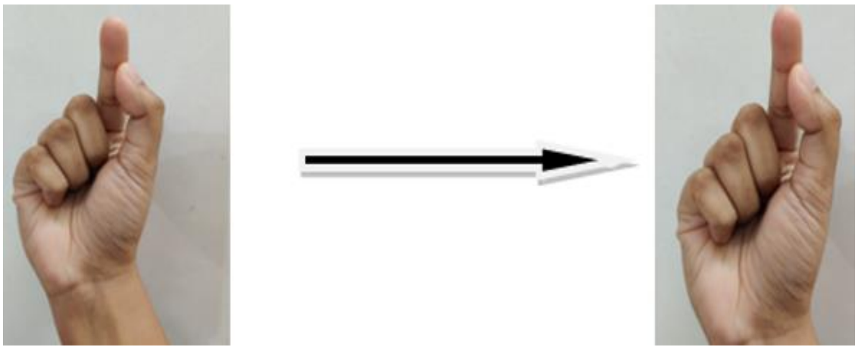

Figure 2. An example of image cropping

2.2.1. CNN architecture

CNN are artificial neural networks that try to mimic the visual cortex of the human brain. The artificial neurons in a CNN are connected to a visual field of the local area, called the receptive field. Discrete convolutions are conducted on the image. The input images are taken in the form of color planes in the RGB spectrum, and the images are then transformed in order to facilitate predictive analysis. High-level features, such as the image edges, are obtained by using a kernel which traverses the whole image starting from top-left and moving towards bottom-right. The CNN model used to recognize these sign words and here, multi-layer convolutional neural networks are used that are connected to each other [25].

In this paper, the proposed model utilizes the Adam optimizer, an expansion of stochastic gradient descent, which is freshly adopted by almost all the computer-vision and natural language processing purposes. For various parameters, the approach calculates a special adaptive learning rate through measurements of first and second gradient moments [26]. The model is trained for 200 epochs for each batch. We used a CNN approach of 12 layers similar to the one used in [1], as shown in Figure 3. For convolution layers 1, 2, 3, 4, 5, and 6, filter sizes are 16, 32, 32, 64, 128, and 256 respectively. The kernel size of each of these layers is 3×3, and the activation function is ReLU. The max pooling layers are each 3×3 as well. Then we use a dropout layer with 50% dropout. After that we have a dense layer with 512 units and ReLU activation. Finally, in the output layer uses ten units with softmax activation.

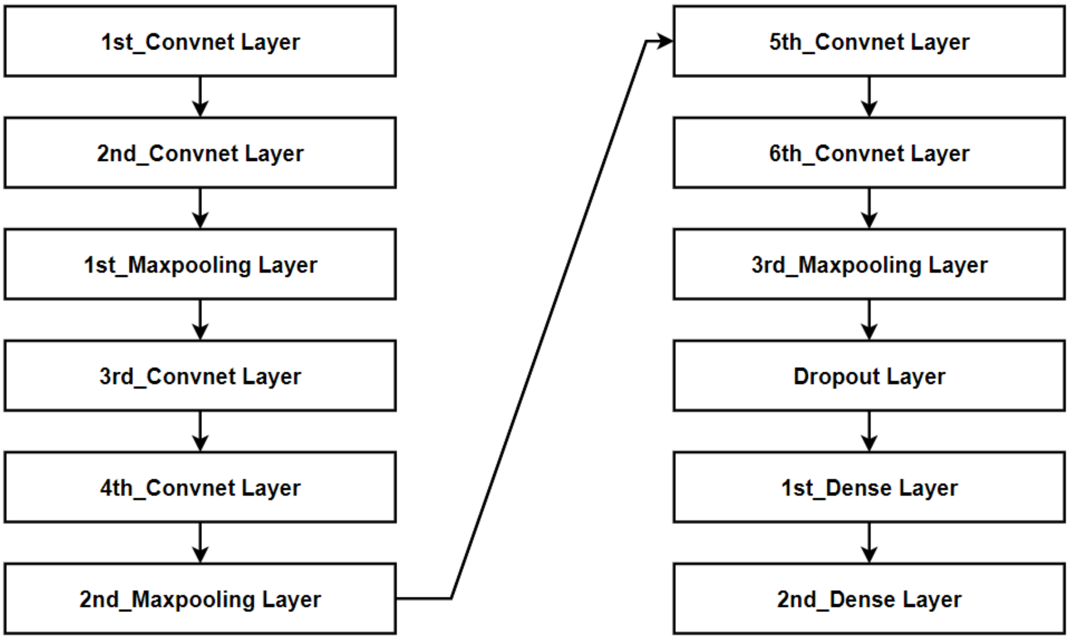

Figure 3. CNN model architecture

3. EVALUATION AND RESULT

As stated earlier, we used 80% - 20% split, resulting in a total of 960 images for training and 240 images for testing. After training the model for 200 epochs using the multi-layered CNN architecture detailed in the previous section, we obtained a test set accuracy of 92.50%. We also calculated the metrics precision, recall and F1-score for each class. The metrics obtained for each class (each of the 30 words signs) are shown in Table 1. It is seen from the table that the performance of the model is quite good for most of the signs. For only a few words, the model fails in some cases to recognize the correct word. Some of these words include কিছুটা (‘*kichuta’*), তারা (‘*tara’*), সময় (‘*shomoi’*), সে (‘*she’*), চামড়া (‘*chamra’*), and অনুরোধ (‘*onurodh’*). Looking at the pictures of these signs (as in Figure 1), we can see that some of them are visually similar and hence prone to confusion by the model. For example, কিছুটা (*‘kichuta’*- row 1 column 4 in Figure 1) and তারা (‘*tara*’- row 3 column 2 in Figure 1) are strikingly similar. The average precision, recall and F1-score are all more than 0.9, so we can say that the overall performance of the model is quite satisfactory.

Table 1. Metrics of each class (sign) in the BSLword dataset. English transliteration of the word is shown

Word	Precision	Recall	F1-score	Word	Precision	Recall	F1-score	Word	Precision	Recall	F1-score
Sir	1.00	1.00	1.00	Darao	1.00	0.80	0.89	Shomajkollan	0.90	1.00	0.95
Shundor	1.00	1.00	1.00	Desh	1.00	1.00	1.00	Hockey	1.00	1.00	1.00
She	0.82	0.75	0.78	Ekhane	1.00	0.90	0.95	Piano	1.00	0.70	0.82
Tara	0.75	0.90	0.82	Gun	1.00	1.00	1.00	Puru	0.88	1.00	0.93
Shotto	1.00	1.00	1.00	Kichuta	0.80	0.89	0.84	Chamra	0.75	0.86	0.80
Shomoi	1.00	0.67	0.80	Kothay	1.00	0.71	0.83	Jail	0.83	0.83	0.83
Aaj	0.80	1.00	0.89	Onurodh	0.80	0.89	0.84	Girja	1.00	1.00	1.00
Basha	1.00	1.00	1.00	Shahajjo	0.80	1.00	0.89	Bouddho	0.89	1.00	0.94
Biyog	1.00	0.80	0.89	Tumi	1.00	1.00	1.00	Bagh	1.00	1.00	1.00
Bondhu	1.00	1.00	1.00	Darano	0.86	1.00	0.92	Keram	1.00	1.00	1.00
Avg. precision = 0.93, Avg. recall = 0.93, Avg. F-1 score = 0.92											

4. CONCLUSION

This paper has introduced a dataset named BSLword, containing 1200 images of 30 static-gesture words in BSL. To the best of our knowledge, this dataset is the very first word-level dataset of BSL. We used a CNN model to correctly identify the words represented by the images in the dataset. The system can recognize BSL static-gesture words with 92.50% accuracy on the word dataset. The average precision, recall and F1-scores are 0.93, 0.93, and 0.92 respectively. We believe that our dataset would be an exceptional asset for BSL recognition specialists. Simultaneously, the dataset can also be beneficial for machine learning and related methods intended for the study of movements for recognizing gestures and signs. We have plans to extend our work in the future in the following ways: currently BSLword only contains a small subset of words of BSL. Our next goal would be to include words with dynamic gestures and make it a comprehensive dataset. This would require not only a huge undertaking in data collection, but also a thorough research to find the most suitable model. Ultimately, our vision is to complete a system that can recognize any word with a reasonable degree of accuracy. If that happens, the mute and deaf people of Bangladesh will no longer suffer from the communication gap that they must endure at present.

REFERENCES

[1] M. S. Islam, S. S. S. Mousumi, N. A. Jessan, A. S. A. Rabby, and S. A. Hossain, “Ishara-Lipi : The First Complete Multipurpose Open Access Dataset of Isolated Characters for Bangla Sign Language,” *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2018, pp. 1-4, doi: 10.1109/icbslp.2018.8554466.

[2] M. M. Hasan and S. M. M. Akhsan, “Bangla Sign Digits Recognition Using HOG Feature Based Multi-Class Support Vector Machine,” *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, 2019, pp. 1-5, doi: 10.1109/EICT48899.2019.9068832.

[3] O. B. Hoque, M. I. Jubair, M. S. Islam, A. -F. Akash, and A. S. Paulson, “Real Time Bangladeshi Sign Language Detection using Faster R-CNN,” *2018 International Conference on Innovation in Engineering and Technology (ICIET)*, 2018, pp. 1-6, doi: 10.1109/ciet.2018.8660780.

[4] J. Uddin, F. N. Arko, N. Tabassum, T. R. Trisha, and F. Ahmed, “Bangla Sign Language Interpretation using Bag of Features and Support Vector Machine,” *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, 2017, pp. 1-4, doi: 10.1109/eict.2017.8275173.

[5] M. A. Hossen, A. Govindaiah, S. Sultana, and A. Bhuiyan, “Bengali Sign Language Recognition Using Deep Convolutional Neural Network,” *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 2018, pp. 369-373, doi: 10.1109/iciev.2018.8640962.

[6] M. S. Islam, S. S. S. Mousumi, AKM. S. A. Rabby, S. A. Hossain, and S. Abujar, “A Potent Model to Recognize Bangla Sign

Language Digits Using Convolutional Neural Network," *Procedia Computer Scince*, vol. 143, pp. 611-618, 2018, doi: 10.1016/j.procs.2018.10.438.
[7] M. A. Uddin, and S. A. Chowdhury, "Hand sign language recognition for bangla alphabet using support vector machine," *2016 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, 2016, pp. 1-4, doi: 10.1109/iciset.2016.7856479.
[8] S. T. Ahmed and M. A. H. Akhand, "Bangladeshi Sign Language Recognition using Fingertip Position," *2016 International conference on medical engineering, health informatics and technology (MediTec),* 2016, pp. 1-5, doi: 10.1109/meditec.2016.7835364.
[9] M. A. Rahman, A. U. Ambia, I. Abdullah, and S. K. Mondal, "Recognition of Static Hand Gestures of Alphabet in Bangla Sign Language," *IOSR Journal of Computer Engineering (IOSRJCE)*, vol. 8, no. 1, pp. 7–13, 2012, doi: 10.9790/0661/0810713.
[10] F. Yasir, P. W. C. Prasad, A. Alsadoon, and A. Elchouemi, "SIFT based approach on Bangla Sign Language Recognition," *2015 IEEE 8th International Workshop on Computational Intelligence and Applications (IWCIA),* 2015, pp. 35–39, doi: 10.1109/iwcia.2015.7449458.
[11] M. R. Islam, U. K. Mitu, R. A. Bhuiyan, and J. Shin, "Hand gesture feature extraction using deep convolutional neural network for recognizing American sign language," *2018 4th International Conference on Frontiers of Signal Processing (ICFSP)*, 2018, pp. 115-119, doi: 10.1109/ICFSP.2018.8552044.
[12] R. Cui, H. Liu, and C. Zhang, "Recurrent convolutional neural networks for continuous sign language recognition by staged optimization," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1610-1618, doi: 10.1109/CVPR.2017.175.
[13] S. M. K. Hasan and M. Ahmad, "A new approach of sign language recognition system for bilingual users," *2015 International Conference on Electrical & Electronic Engineering (ICEEE)*, 2015, pp. 33-36, doi: 10.1109/CEEE.2015.7428284.
[14] M. M. Islam, S. Siddiqua, and J. Afnan, "Real time Hand Gesture Recognition using different algorithms based on American Sign Language," *2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 2017, pp. 1-6, doi: 10.1109/ICIVPR.2017.7890854.
[15] J. Huang, W. Zhou, H. Li, and W. Li, "Sign language recognition using 3D convolutional neural networks," *2015 IEEE international conference on multimedia and expo (ICME)*, 2015, pp. 1-6, doi: 10.1109/icme.2015.7177428.
[16] S. A. Khan, A. D. Joy, S. M. Asaduzzaman, and M. Hossain, "An Efficient Sign Language Translator Device Using Convolutional Neural Network and Customized ROI Segmentation," *2019 2nd International Conference on Communication Engineering and Technology (ICCET)*, 2019, pp. 152-156, doi: 10.1109/ICCET.2019.8726895.
[17] D. Naglot and M. Kulkarni, "Real time sign language recognition using the Leap Motion Controller," *2016 International Conference on Inventive Computation Technologies (ICICT)*, 2016, pp. 1-5, doi: 10.1109/INVENTIVE.2016.7830097.
[18] A. M. Rafi, N. Nawal, N. S. N. Bayev, L. Nima, C. Shahnaz, and S. A. Fattah, "Image-based Bengali Sign Language Alphabet Recognition for Deaf and Dumb Community," *2019 IEEE Global Humanitarian Technology Conference (GHTC)*, 2019, pp. 5-11, doi: 10.1109/GHTC46095.2019.9033031.
[19] M. A. Rahaman, M. Jasim, M. H. Ali, and M. Hasanuzzaman, "Real-time computer vision-based Bengali sign language recognition," *2014 17th International Conference on Computer and Information Technology (ICCIT)*, 2014, pp. 192-197, doi: 10.1109/ICCITechn.2014.7073150.
[20] S. Masood, A. Srivastava, H. C. Thuwal, and M. Ahmad, "Real-time sign language gesture (word) recognition from video sequences using CNN and RNN," *Intelligent Engineering Informatics*, 2018, pp. 623-632, doi: 10.1007/978-981-10-7566-7_63.
[21] A. J. Rony, K. H. Saikat, M. Tanzeem, and F. M. R. H. Robi, "An effective approach to communicate with the deaf and mute people by recognizing characters of one-hand bangla sign language using convolutional neural-network," *2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT)*, 2018, pp. 74-79, doi: 10.1109/CEEICT.2018.8628158.
[22] P. P. Urmee, M. A. A. Mashud, J. Akter, A. S. M. M. Jameel, and S. Islam, "Real-time bangla sign language detection using xception model with augmented dataset," *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, 2019, pp. 1-5, doi: 10.1109/WIECON-ECE48653.2019.9019934.
[23] R. Yasir and R. A. Khan, "Two-handed hand gesture recognition for Bangla sign language using LDA and ANN," *The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014)*, 2014, pp. 1-5, doi: 10.1109/SKIMA.2014.7083527.
[24] T. M. Angona *et al*., "Automated Bangla sign language translation system for alphabets by means of MobileNet," *TELKOMNIKA,* vol. 18, no. 3, pp. 1292-1301, 2020, doi: 10.12928/telkomnika.v18i3.15311.
[25] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," *2017 International Conference on Engineering and Technology (ICET),* 2017, pp. 1-6, doi: 10.1109/icengtechnol.2017.8308186.
[26] Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS),* 2018, pp. 1-2, doi: 10.1109/IWQoS.2018.8624183.

BIOGRAPHIES OF AUTHORS

Kulsum Ara Lipi is pursuing her B.Sc. in Information and Communication Engineering from Bangladesh University of Professionals, Dhaka, Bangladesh. She has her interest in Machine Learning and Data Science. Her current research interest includes Deep Learning and Natural Language Processing. She can be contacted at email: kulsumlipi@gmail.com.

Sumaita Faria Karim Adrita is pursuing her B.Sc. in Information and Communication Engineering from Bangladesh University of Professionals, Dhaka, Bangladesh. She has her interest in Machine Translation specifically in Bangla Sign Language. Her current research interest includes Natural Language Processing and Deep Learning. She can be contacted at email: sumaitafaria@gmail.com.

Zannatul Ferdous Tunny is pursuing her B.Sc. in Information and Communication Engineering from Bangladesh University of Professionals, Dhaka, Bangladesh. She has her interest in the areas of Artificial Intelligence (AI), Robotics, IOT, Data Science. She is also interested in Blockchain, NLP and Computer Vision. She can be contacted at email: zannatulferdous489@gmail.com.

Abir Hasan Munna is pursuing his B.Sc. in Information and Communication Engineering from Bangladesh University of Professionals, Dhaka, Bangladesh. His main areas of interest are Artificial Intelligence (AI), Robotics, IOT, Data Science. He is also interested in Computer Vision, NLP and Blockchain. He can be contacted at email: abirmunna091@gmail.com.

Ahmedul Kabir is an Assistant Professor at Institute of Information technology, University of Dhaka, Dhaka, Bangladesh. His principal areas of interest are Machine Learning and Data Mining. He would like to conduct research in these fields both theoretically and in practical applications for different domains. He is also interested in Software Analytics and Natural Language Processing. He can be contacted at email: kabir@iit.du.ac.bd.