Recognition of Bengali Sign Language using Novel Deep Convolutional Neural Network

Md. Jahangir Hossein, Md. Sabbir Ejaz
Dept. of Computer Science and Engineering, Bangladesh Army University of Engineering and Technology
Email: jahangir7200@live.com, cse.sabbir7@gmail.com

2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI) |978-1-6654-0489-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/STI50764.2020.9350418

Abstract— **On our planet, speech and hearing-impaired people are a part of our society. When an interaction is needed between the impaired and the general people, communication becomes difficult. In several races, impaired people practice various sign languages for communication. For speech and hearing-impaired people, sign language is the fundamental communication method in their lifestyle. However, it is incredibly challenging to desegregate them into the mainstream because the majority part of our community is not aware of their practiced sign language. Nowadays, computer vision-based solutions remain fully appreciated to get their sign language comprehensible to general people. Many analysts are taking a shot at Recognition of Hand Gesture, one of the computer vision-based solutions to recognize sign language. It's been a popular area for research for an extended period now. Some recent studies have reached immense performance using models of deep learning in the region regarding Hand Gesture Recognition in Computer vision. Through this research work, our aim to reduce the communication difficulties among the speech and hearing-impaired people and the rest of Bangladesh by building an appropriate deep learning model that can recognize Bangla Sign Language alphabets precisely. In this work, a different CNN (Convolutional Neural Network) architecture is introduced to identify the alphabets of Bengali sign with the respective Ishara-Lipi database. This architecture accomplished a general precision of 99.86%, which surpassed all prior works regarding Bengali sign alphabet recognition.**

Keywords— Computer Vision, Deep Convolutional Neural Network, Hand Gesture Recognition, Bangla Sign Language.

I. INTRODUCTION

Hearing-impairment or hearing loss is the deficiency or complete inability to listen [1]. The failure may affect only one ear or maybe both ears. The hearing failure affected more further than 1.1 billion identities to a remarkable extent since 2013 [2]. It provoked a loss in almost 538 million personalities and modified to significant inability in nearly 124 million characters [2], [3]. Therefore sign language is practiced to overcome the communication gap among mute-deaf people. It is the commonly used language in the society of mute-deaf to interact with other people and yield their views. Sign language has a moderate structure and rich vocabulary [4]. Although there are excellent correspondences amongst many sign languages, these are not regular languages to every community and not generally understandable with each other. Besides a population above 130 million and Bangladesh is encountering higher than 13 million cases, amongst which roughly 3 million people continue facing severe hearing disasters, which directs them into disability [5]. That's why for sign language, a perception approach of automated including high performance, is too needed for the mute-deaf inhabitants.

Already numerous studies are carried on sign language recognition for diverse communities. For instance, acknowledging American Sign Language accomplished an exactness of 94.32% as of late [6]. Research carried on the Indian sign dialect and achieved considerable correctness of 93% [7]. Another analysis on Spanish sign dialect apprehension recommended an in general exactness of 96% [8]. In any case, not many studies were carried on Bengali sign language identification. To identify the digits of Bangla Sign Language, Sanzidul et al. [9] developed a Convolution Neural Net architecture consisting of 22 layers. They used the multilayer model and achieved 94.88% of accuracy. Based on Fingertip Discoverer Calculation [10] introduced a BD-SL recognition system. The system can classify 46 sorts of sign languages, and this language constitutes consonants, vowels, and numerals. The system achieved 88.69% of the effectiveness to identify the BD-SL sign digits. Shahjalal et al. [11] proposed another Vision-based method for classifying hand sign digits featuring deep learning methods. They adopted data augmentation and gained an accuracy of 92%. Along with Linear Discriminant (FLD) analysis [12] suggested a show step-up upon Principal Component Analysis (PCA). It hybridized by PCA, SVM (Support Vector Machine), and LDA obtained a 96.43% accuracy.

By conventional machine learning (ML) classifiers like Support Vector Machine and K-Nearest Neighbors, barely 86.53% accuracy was attained [13]. Further, a SIFT-based method accomplished an accuracy of about 98% for the Bengali language vowels as of late by the compensation with binary SVM classifier [14]. Nevertheless, very current research obtained an average exactness of 98.84% on 36 Bengali letters that exceeded all the earlier analyses [15].

Here we started amidst the dataset around 36 Bengali letters. To more accurately distinguish the alphabets, we formed and introduced a different deep convolutional neural network representation. To provide accurate identification of the Bengali sign language's alphabets, we obtained the most meaningful features with the support of our preferred model. The model is developed upon the Keras framework. Our proposed demonstrate accomplished a by and considerable test exactness of about 99.86%, which exceeded all the previous studies.

II. WORKING METHODOLOGY

Figure 1 demonstrates our methodology towards the solution to the problem of the Bengali sign language alphabets recognition. For this at first, collect the dataset and preprocess the images using filtering and resizing steps. Then use the augmentation process to enlarge the dataset. Then the dataset divides into train, validation, and test set. Finally, explore the features using a deep convolutional neutral's

978-1-6654-0489-1/20/$31.00©20XX IEEE

to train the model for Bengali sign language alphabets recognition.

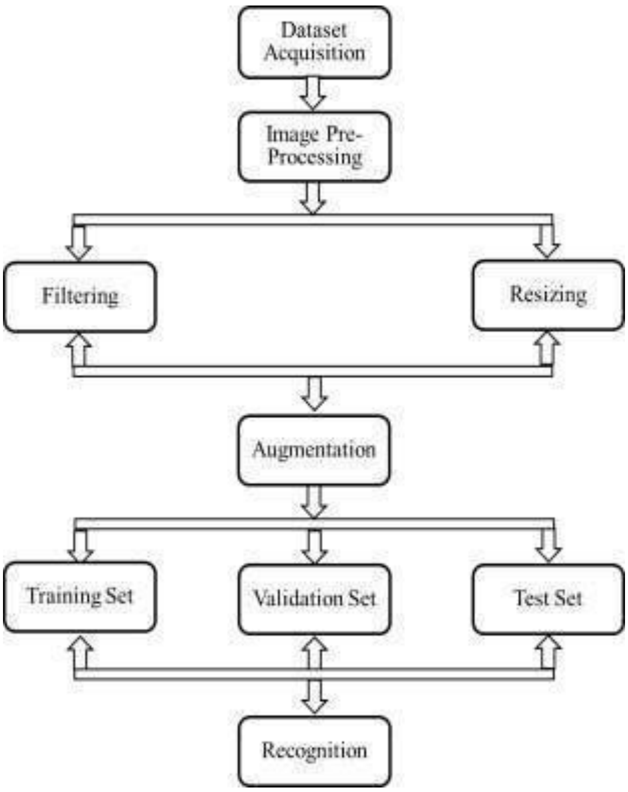

Fig. 1. Flowchart of Bengali sign language alphabets recognition procedure.

A. Dataset Acquisition

During this analysis, we employ the Ishara-Lipi Bengali characters dataset [16] for our research. In Bangla Sign Language (BdSL) it is the earliest complete isolated character dataset. The dataset covers 50 sets concerning Bengali necessary sign characters of 36, attained with several generals' unique cooperation plus deaf volunteers of multiple organizations. Hereunder Bengali sign language (BDSL) letters exist 30 consonants with six vowels to fingerspell each Bengali speech. Dropping mistakes and preprocessing into this Ishara-Lipi Bengali sign language dataset, a total of 1800 letters insights of Bengali sign language implied within the latest dataset. At a glance of Bengali alphabets, figure 2 represents the signs.

B. Preprocessing & Augmentation

In the beginning, the RGB (color) images are transformed toward grayscale images. Then resize all the images into 64X64 size for the model input and use a median filter to remove noise. After that, with the assistance of the augmenter library, Data augmentation is made. During the data augmentation process, the rotation probability, max right rotation, and max left rotation concerning the rotation function are adjusted to 0.4, 3, and 3, respectively. Accordingly, the random distortion function's grid height and width are set to 4, the probability is valued to 0.4, and the magnitude is fixed to 4. Furthermore, the probability of random zoom function is valued at 0.9, and percentage areas are 0.2. Consequently, later the augmentation procedure, 36000 images are for the 36 classes are created with 1000 image examples in every class.

Data Augmentation is used because insufficient data holds a notable limitation in executing the deep learning models similar to convolutional neural networks. Sometimes imbalanced levels cause additional restrictions. Whereas many groups may have enough equally essential data, there will be a possibility of undergoing under-sampled groups for ineffectual class-specific efficiency. It is a natural event. There is a small potential to prophesize the test utilization and group invalidation if this model learns from fewer examples regarding a given group. Different approaches to handling or combining multiple preparings, like the irregular revolution, transformations, flips, shear, etc. help produce training images artificially throughout the image augmentation process [17].

C. Proposed Deep Convolutional Neural Net's Model

In the image identification area, Convolution Neural Net's (CNN) is the most traditional deep neural systems. After CNN's superior execution within the ImageNet challenge, it received its popularity [18]. The Convolutional neural network holds multiple advantages over different neural networks because it has an inherent combination of layers. Fig. 2. Diverse hand signal signs for Bengali alphabets.

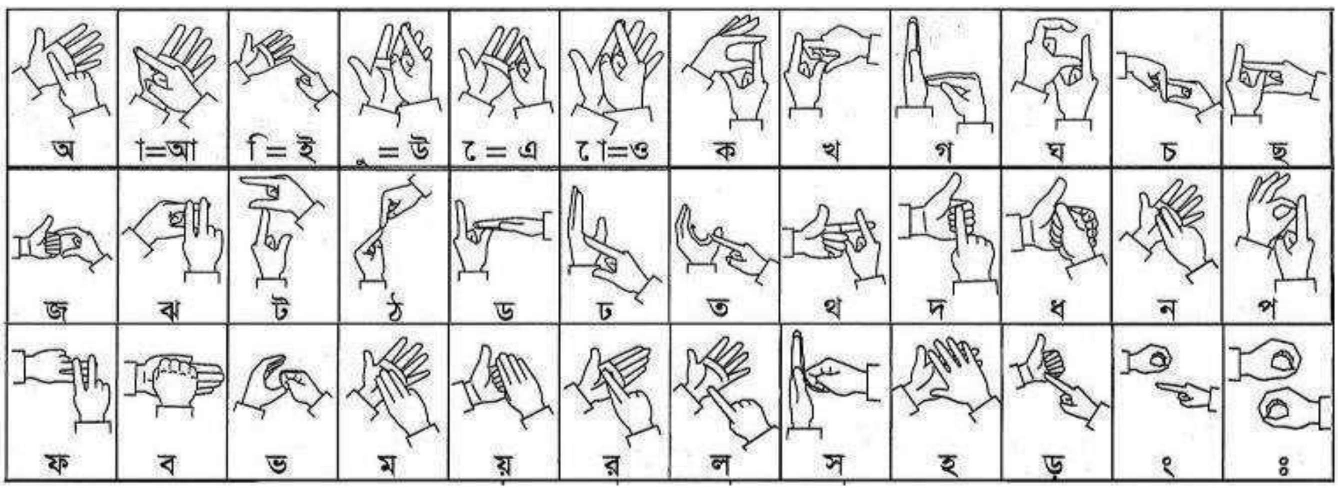

Fig. 2. Different hand gesture signs for Bengali alphabets.

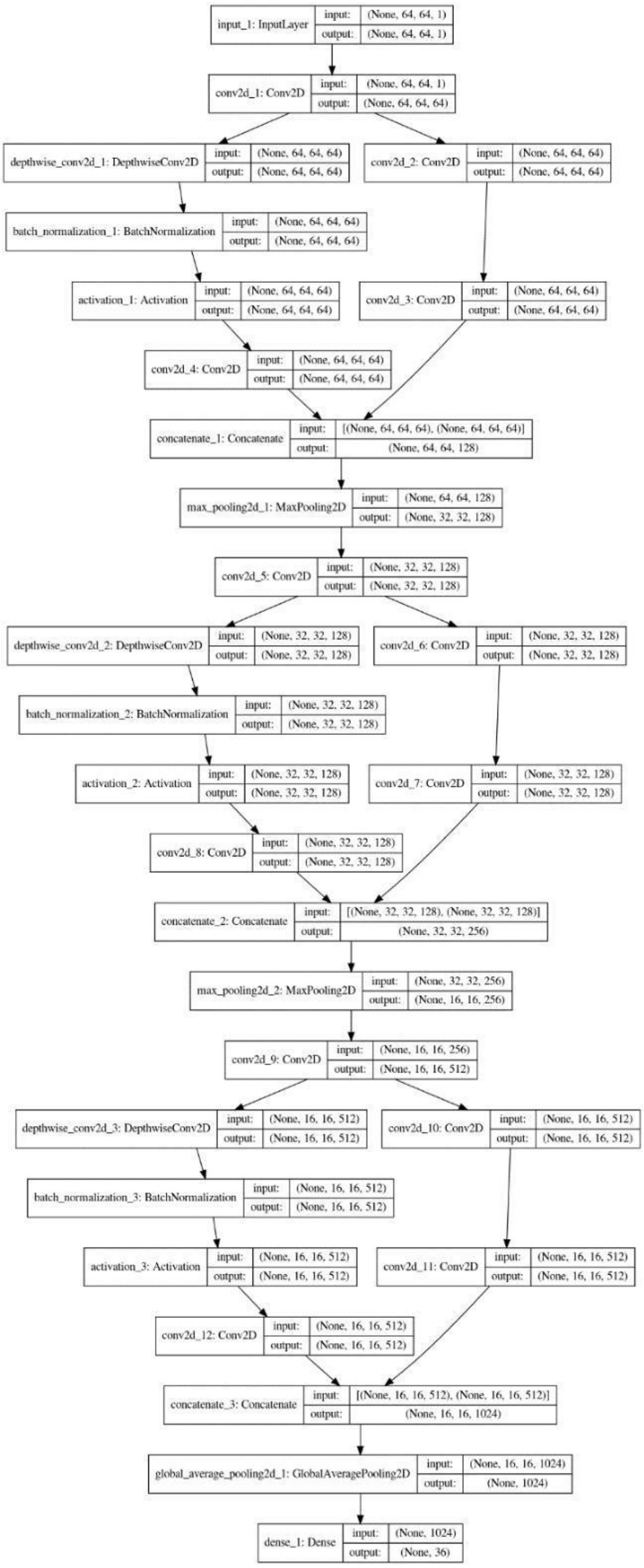

Fig. 3. Proposed deep CNN model architecture for Bengali character recognition.

CNN holds three essential layers, and they are the convolutional layers, the pooling layers, and the fully connected layers. The input matrix is multiplied through different convolutional masks or kernels to create an outline that defines the feature that continues in the image during convolution. Into the pooling layer, scaling and translation variance are implemented to decrease the mapping feature's extent. Lastly, there is a complete associated layer next to the convolutional and pooling layers that clarified the picture's existing characteristics and clarified the picture's non-existence characteristics. These are the three essential layers regarding the Convolutional Neural Network. Correlated to different deep learning network architectures, CNN also has a less amount of parameters. Twelve convolutional layers, another three depth-wise convolutions with batch normalization, and even three more pooling layers have been utilized in this work. The activation function ReLU is used within the convolution layer. Sixty-four convolutional filters are used in the initial two convolutional layers. They are developed within the subsequent two convolutional layers toward 128 to gain more resonant deep characteristics.

Finally, to obtain even more in-depth features from the image, two convolutional layers and 256 convolutional kernels are applied in the latter. A size of 3x3 kernel is involved in every convolution layer, and for specific pooling layer, Max pooling is employed. A 3x3 kernel size of depth-wise convolution and GlobalAveragePool is used for the second final layer by replacing Flatten for reducing the heavy amount of parameter cost. Lastly, the obtained features from the image are conveyed toward the fully connected layer. The ReLU activation function is also implemented in the hidden layer. The hidden layer has 256 nodes to recode each mapping between a fully connected layer input and the output of the output layers. We remain working with 36 characters from Bengali sign language. In figure 3, we represent our suggested deep CNN model architecture.

III. Design of Experiment

The network is prepared with 50 epochs, including the batch size of about 64. To maximize the specific error function, we applied the 'Adam' optimizer with a learning rate of about 0.0001. Concerning the particular loss or error function, a categorical cross-entropy function was engaged. The dropout procedure was followed here to avoid overfitting.

IV. Result Analysis

The database holds 36,000 image examples from 36 alphabetical groups or classes, where 1000 images are available per group or class. Therefore, a consistently distributed and balanced database was created. Next, the training dataset is implemented to our designated CNN network for training. Figure 4 exhibits the training loss and validation loss. Also, figure 5 reveals the corresponding training accuracy and validation accuracy curvature during 50 epochs. The trained network is then utilized to identify the test accuracy. Our model reached 99.86% test accuracy, which outperformed all previous studies. In a different well-defined graphical manner to assess the performance, a more macro-average curve called ROC curve interpretation is done in the work where a well-balanced dataset is examined. For the test set, figure 6 explains the macro-average ROC curve.

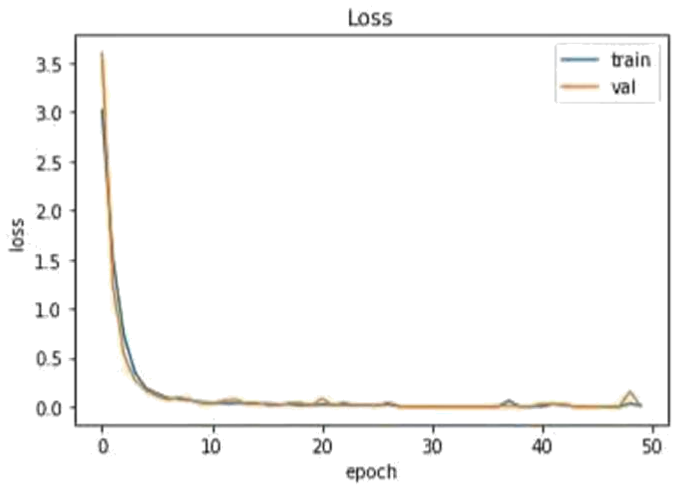

Fig. 4. Training and Validation loss for suggested CNN model during the phase of training.

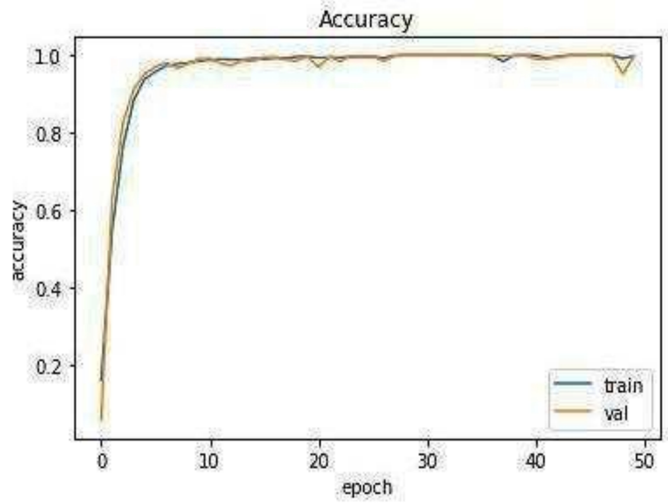

Fig. 5. Training and Val(validation) accuracy for proposed CNN model during the phase of training .

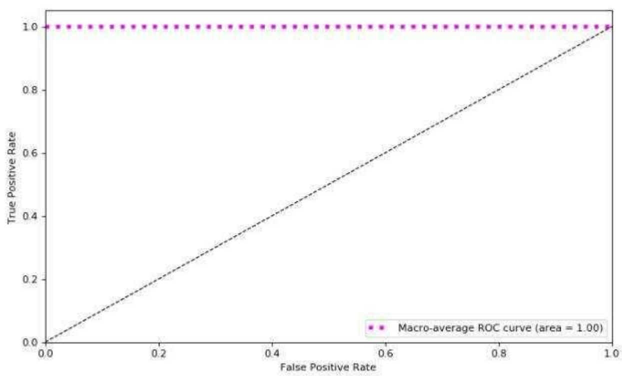

Fig. 6. Macro-average ROC curve for the outcomes of the testing phase.

Conclusively, a comparative study within our work and the early work has been carried here. Table I & II presents the comparison between the early works and our work for each Bengali Sign Language alphabet recognition. And from the table, this signifies that our architecture exceeded the earlier studies' accuracy for all alphabets.

TABLE I. THE CLASS-WISE ACCURACY COMPARISON OF OUR METHOD WITH PREVIOUS METHODS

Class Label	Our Proposed Method	Previous Method [15]
1	100	96.67
2	100	98
3	100	96.33
4	99.95	96.33
5	100	96.5
6	100	96.5
7	99.95	96.67
8	100	94.83
9	100	94.83
10	100	95.17
11	100	94.5
12	100	94.17
13	100	94.33
14	100	94
15	100	97.67
16	100	97.83
17	100	95.5
18	100	93.83
19	100	93.83
20	100	93.83
21	99.85	94.5
22	100	94.33
23	100	93.67
24	100	94.17
25	100	94
26	100	94
27	100	95.67
28	100	95
29	100	94.5
30	100	90.33
31	98.5	91.17
32	100	94.33
33	100	95.83
34	100	95.67
35	100	96
36	99.95	95.5

TABLE II. ISHARA-LIPI DATASET COMPARISON WITH EXISTING WORK

Methodology	Layer	Accuracy (%)
Sanzidul et al. [9]	22	95.35
Tonmoy et al. [19]	7	98.84
Our Proposed Model	12	99.86

V. CONCLUSION

In this work, we began with an alphabet dataset of Bengali signs for accurate identification of the Bengali sign alphabet. We applied our suggested CNN architecture and attained a total test accuracy of about 99.86%, exceeding all previous works' accuracy. We believe that our work will help mute-deaf populations and more advancement in sign language identifications. We accept that this dataset will be an mind blowing resource for investigators in Bangla Sign Dialect affirmation. At the same time, this dataset can be supportive for PC vision and AI strategies outlined for learning signs and approaches for analyzing motion. We acknowledge that this dataset will be viable asset for all clients, understudies, pros of Bangla Sign Dialect and we prospect that the likeness of the BdSL dataset will energize and offer assistance diverse researchers to mull over the trouble of communication by means of signals affirmation, movement affirmation.

REFERENCES

[1] W. H. Organization, 2015, "Deafness and hearing loss fact sheet n 300 updated march 2015," .

[2] B. Bell, A. Bertozzi-Villa, S. Biryukov, T. Vos, R. M. Barber, I. Bolliger, F. Charlson, A. Davis, L. Degenhardt, D. Dicker et al., "Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the global burden of disease study 2013," The Lancet, vol. 386, no. 9995, pp. 743–800, 2015.

[3] K. J. Neumann,J. E. Saunders & B. O. Olusanya, "The global burden of disabling hearing impairment: a call to action," Bulletin of the World Health Organization, vol. 92, pp. 367–373, 2014.

[4] W. Sandler and D. Lillo-Martin, Sign language and linguistic universals. 2006, Cambridge University Press.

[5] M. Alauddin and A. H. Joarder, "Deafness in Bangladesh," in Hearing Impairment. pp. 64–69, Springer, 2004.

[6] M. M. Islam, S. Siddiqua, and J. Afnan, "Real time hand gesture recognition using different algorithms based on American sign language," in 2017 IEEE international conference on imaging, vision & pattern recognition (ICIVPR), IEEE, pp. 1–6. , 2017.

[7] K. Yadav, L. P. Saxena, B. Ahmed, and Y. K. Krishnan, "Hand gesture recognition using improved skin and wrist detection algorithms for Indian sign," Journal of Network Communications and Emerging Technologies (JNCET), vol. 9, no. 2, 2019.

[8] G. Saldaña González, J. Cerezo Sánchez, M. M. Bustillo Díaz, and A. Ata Pérez, "Recognition and classification of sign language for Spanish," Computación y Sistemas, vol. 22, no. 1, pp. 271–277, 2018.

[9] S. Islam, S. S. S. Mousumi, AKM S. A. Rabby, S. A. Hossain, S. Abujar, "A Potent Model to Recognize Bangla Sign Language Digits Using Convolutional Neural Network," Procedia Computer Science, vol. 143, 2018, pp. 611-618.

[10] "An Automated Bengali Sign Language Recognition System Based on Fingertip Finder Algorithm". International Journal of Electronics and Informatics Jarman, Angur & Arshad, Samiul & Alam, Nashid & Islam, Mohammed J. vol. 4.

[11] "Hand Sign to Bangla Speech: A Deep Learning in Vision based system for Recognizing Hand Sign Digits and Generating Bangla Speech," arXiv:1901.05613v1, 17 Jan 2019.Ahmed, Shahjalal & Islam, Md & Hassan, Jahid & Uddin Ahmed, Minhaz & Ferdosi, Bilkis & Saha, Sanjay & Shopon, Md. (2019).

[12] M. Ahmada and S. M. K. Hasan, "A new approach of sign language recognition system for bilingual users," 2015 International Conference on Electrical & Electronic Engineering (ICEEE), Rajshahi, 2015, pp. 33-36.

[13] M. Hasan, T. H. Sajib, and M. Dey, "A machine learning based approach for the detection and recognition of Bangla sign language," in 2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec), IEEE, 2016, pp. 1–5.

[14] Yasir, P. C. Prasad, A. Alsadoon, and A. Elchouemi, "Sift based approach on Bangla sign language recognition," in 2015 IEEE 8th International Workshop on Computational Intelligence and Applications (IWCIA), IEEE, 2015, pp. 35–39.

[15] M. A. Rahaman, M. Jasim, M. H. Ali, and M. Hasanuzzaman, "Bangla language modeling algorithm for automatic recognition of hand-sign-spelled Bangla sign language," Frontiers of Computer Science, vol. 14, no. 3, pp. 143302, 2020.

[16] 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," in 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), IEEE, 2018, pp. 1–4.

[17] M. D. Bloice, C. Stocker, and A. Holzinger, "Augmentor: an image augmentation library for machine learning," arXiv:1708.04680, 2017.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.

[19] Hossain, Tonmoy, F. M. Shah and F. S. Shishi. "A Novel Approach to Classify Bangla Sign Digits using Capsule Network," in 2019 22nd International Conference on Computer and Information Technology (ICCIT), pp. 1-6, IEEE, 2019.

[20] R. A. Dunne and N. A. Campbell, "On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function," in Proc. 8th Aust. Conf. on the Neural Networks, Melbourne, vol. 181. Citeseer, 1997, pp. 185.