Classification of Bengali Sign Language Characters by Applying a Novel Deep Convolutional Neural Network

Md. Mehedi Hasan*, Azmain Yakin Srizon† and Md. Al Mehedi Hasan‡
Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh
Email: *mmehedihasann@gmail.com, †azmainsrizon@gmail.com, ‡mehedi_ru@yahoo.com

***Abstract*—With approximately 466 million deaf-mute people in the world, sign language recognition has been an area of interest for researches for a long time now. Previously, many researchers have conducted studies on different sign language datasets, achieving a decent accuracy of recognition. However, not many studies have been conducted on the Bengali sign alphabet recognition dataset despite having approximately 3 million severe deaf cases alone. Previous studies on Bengali sign character detection suggested decent accuracies using traditional machine learning approaches. However, in recent studies, some researches have achieved higher test accuracy using deep learning models. In our study, we've proposed a novel Convolutional Neural Network (CNN) model for the recognition of the Bengali sign alphabets from the Ishara-Lipi dataset. Our model achieved an overall accuracy of 99.22% that has outperformed all previous studies of Bengali sign character recognition for the deaf-mute community.**

***Index Terms*—Bengali Sign Character Recognition, Deep Convolution Neural Network, Augmentation**

I. Introduction

Hearing loss or impairment is an incomplete or total disability to hear [1]. Hearing loss may occur in single or both ears [2]. As of 2013, hearing loss concerned more than 1.1 billion personalities to some extent [3]. This caused an inability in almost 538 million individuals and moderate to critical disability in almost 124 million individuals [2], [4], [5]. To reduce the communication gap with mute-deaf individuals, sign language is practiced which is the most utilized language in the mute-deaf society to communicate with individuals and share views. Sign languages are different types of natural languages that have their custom structure and vocabulary [6]. Of course, these languages are not common for all communities and these are not commonly understandable among each other although there exist remarkable correlations among different sign languages [6]. That is why a high-performance automated recognition of sign language is highly required for the mute-deaf community. With a population of more than 130 million [7], Bangladesh is experiencing more than 13 million hearing cases alone among which approximately 3 million individuals are experiencing from critical to profound hearing loss that leads them to disability [8].

Previously many studies have been conducted on the recognition of sign language for different communities. For example, recognition of American sign language achieved an accuracy of 94.32% recently [9]. A study conducted on Indian sign language detection performed an overall accuracy of 93% [10]. A study based on Spanish sign language detection suggested an overall accuracy of 96% [11]. However, not many studies have been conducted on Bengali sign language classification. With traditional machine learning classifiers like K-Nearest Neighbors (KNN) and Support Vector Machine (SVM), a mean accuracy of only 86.53% was achieved [12]. Moreover, a SIFT-based approach achieved an accuracy of 98% for the vowels of the Bengali language recently with the assistance of the binary SVM classifier [13]. However, a very recent study conducted on 36 Bengali alphabets achieved an average accuracy of 95.83% that outperformed all other previous studies [14].

In this paper, we began with the dataset of 36 Bengali alphabets. To classify the alphabets more precisely, we developed and proposed a novel deep convolutional neural network (CNN) model. With the assistance of our proposed model, the most significant features were extracted to produce a precise recognition of the alphabets of the Bengali sign language dataset. The models have been implemented on Keras framework that ran on top of Tensorflow. Our proposed model achieved an overall test accuracy of 99.22% that outperformed all previous researches.

II. Materials and Methods

In this section,, dataset description, our proposed convolutional neural network model and augmentation have been discussed.

A. Dataset Description

In this research, we've worked on the Ishara-Lipi Bengali characters dataset [15]. This character dataset includes 50 sets of 36 Bengali basic sign letters, obtained with the assistance of several deaf and general volunteers from various institutes. In Bengali sign language characters, there are 6 vowels and 30 consonants by which they can fingerspell all Bengali words. In the Ishara-Lipi dataset, after dropping errors and preprocessing, 1800 character perceptions of Bengali sign language were incorporated in the final dataset. Figure-1 illustrates signs for Bengali alphabets at a glance.

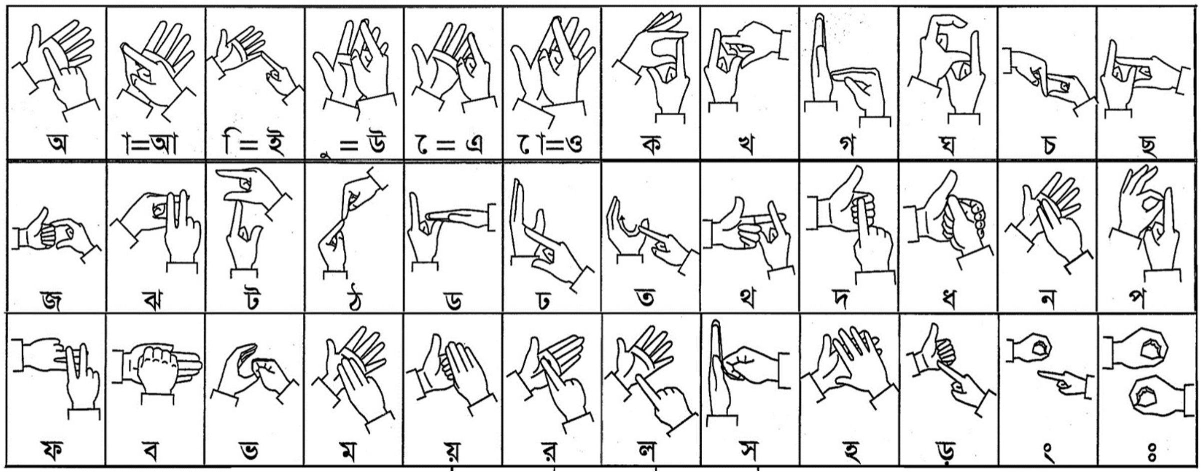

Fig. 1: This figure illustrates signs for Bengali alphabets at a glance.

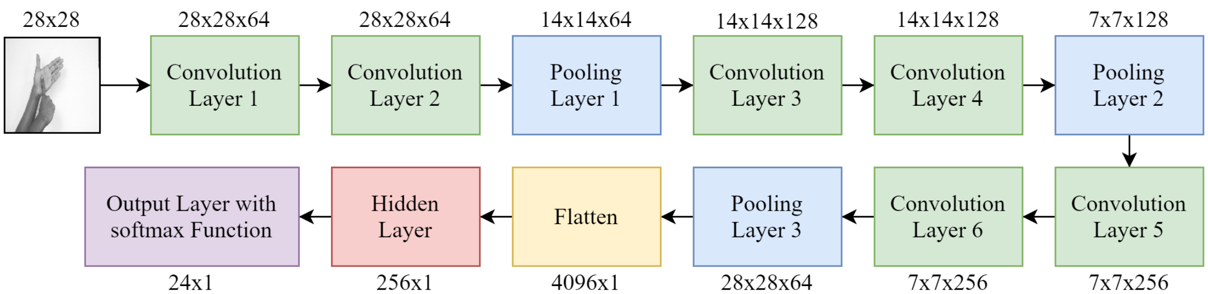

Fig. 2: Illustration of the architecture of our proposed CNN model.

B. *Proposed Convolutional Neural Network Model*

Convolution Neural Network (CNN) is one of the most conventional deep neural networks in the area of image identification. CNN earned its reputation after its notable performance in the ImageNet challenge [16]. A convolutional neural network has several benefits over other sorts of neural networks because of its inherent composition or layers.

CNN has essentially three sorts of layers: the convolutional layer, the pooling layer, and the fully connected layer. In convolutional layers, the input data matrix becomes multiplied with several convolutional kernels or filters to produce a feature map that explains what sort of feature endures in the image. In the pooling layer, translation and scaling variance were provided which also decreased the volume of feature maps. Finally, there is a fully connected layer after all the convolutional and pooling layers which explained what sort of characteristics existed in the picture and what sort of characteristics don't exist in the picture. These three sorts of layers are the basic layers of the Convolutional Neural Network. Moreover, CNN has also a smaller number of parameters compared to other deep learning network structures.

Six convolutional layers and three pooling layers had been used in this study. ReLU activation function was employed in the convolution layer. In the first two convolutional layers, 64 convolutional filters were used which was expanded in the following two convolutional layers to 128 to get more deep characteristics. Finally, in the latter two convolutional layers, 256 convolutional filters were employed in order to extract even more deep features in the image. A 3x3 size filter was applied in all the convolution layer and Max pooling was used for the pooling layer. Finally, the feature plucked from the image was transferred into a fully connected layer. In the hidden layer, the ReLU activation function was applied also. There were 256 nodes in the hidden layer to recode the mapping among the input from a fully connected layer and output from the output layer. The output layer consisted of a total of 24 nodes with softmax [17] activation function as in the classification dilemma, we are dealing with 36 alphabets of Bengali sign language. Figure-2 illustrates the architecture of our proposed CNN model.

C. *Augmentation*

Limited data is a significant barrier in implementing deep learning models like convolutional neural networks. Oftentimes, imbalanced levels can be a supplementary restraint. While there may be enough data for some groups, equally significant, but under sampled groups will undergo inadequate class-specific efficiency. This event is natural. If the model studies from a few instances of a provided group, it is less possible to prophesize the group invalidation and test utilization. Image augmentation artificially produces training pictures

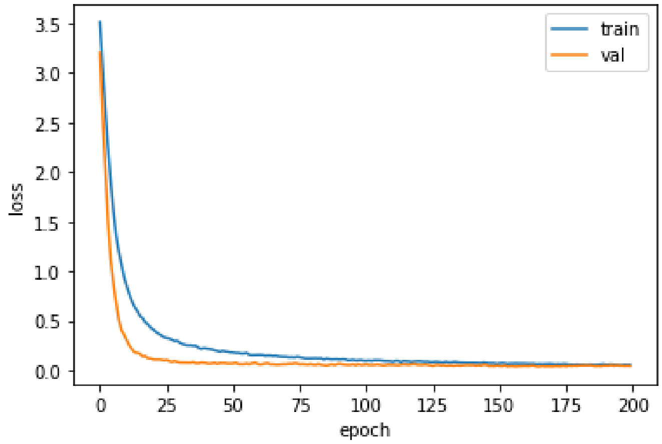

Fig. 3: Validation and training loss for the proposed CNN model while training phase.

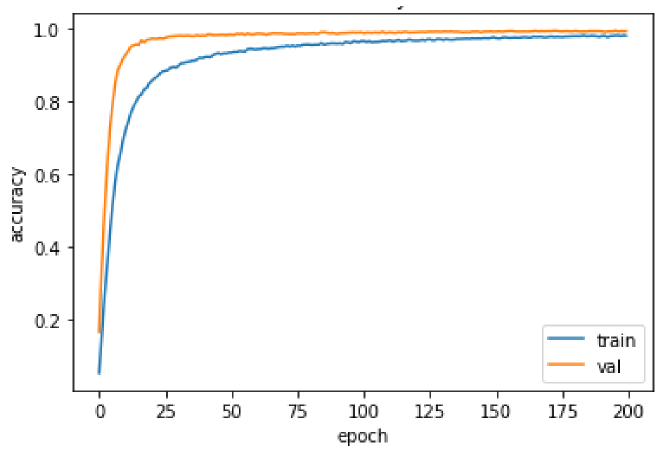

Fig. 4: Validation and training accuracy for the proposed CNN model while training phase.

through various approaches of processing or combination of various processing, such as random rotation, transfers, shear, and flips, etc. [18].

III. Experimental Analysis

A. Preprocessing

Firstly, the RGB images were transferred into grayscale images. After that, augmentation was performed with the cooperation of the Augmentor library. While performing the augmentation procedure, the probability of rotation, max left rotation and max right rotation of the rotation function were fixed to 0.4, 3 and 3 respectively. Probability, grid width, grid height and magnitude of the random distortion function were locked to 0.4, 4, 4 and 4 respectively. Moreover, probability and percentage areas of zoom random function were set to 0.2 and 0.9 respectively. Hence, after the augmentation process, 1000 image samples per class, a total of 36000 images for 36 classes were generated.

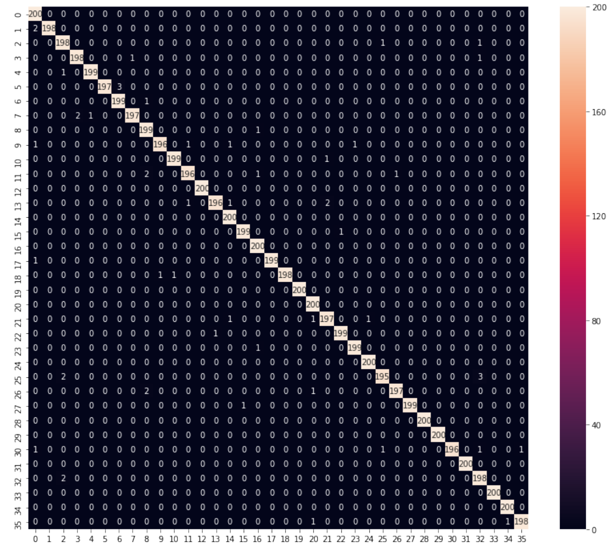

Fig. 5: This figure illustrates confusion matrix for the test set.

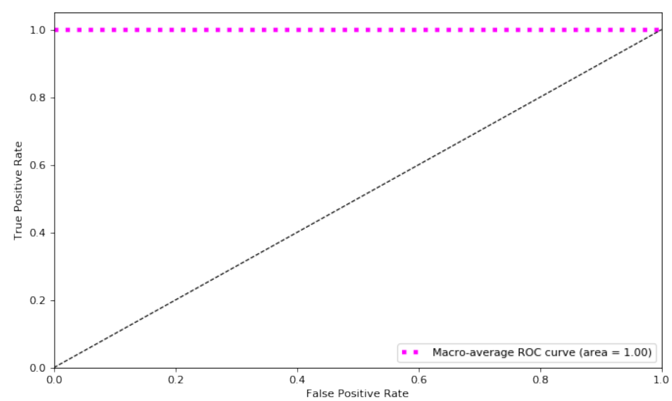

Fig. 6: Macro-average ROC curve analysis for the outcomes of testing.

B. Design of Experiment

The model was trained for 200 epochs with a batch size of 256 as after that the validation loss became nearly constant for the rest of epochs. 'Adam' optimizer with the learning rate of 0.0001 was used to maximize the error function. Categorical cross-entropy function was employed for the loss or error function. For bypassing overfitting, dropout technique was practiced.

C. Result Analysis

The sign language dataset that was used for this research had a total of 36,000 image samples for 36 alphabetical classes while the number of images per class is 1000, hence, a uniformly distributed balanced dataset was generated. After that, our defined CNN model was applied to the train set. Figure-3 shows the training loss and validation loss where Figure-4 shows the training accuracy and validation accuracy curve for 200 epochs. Now the trained model was applied to predict the test accuracy and our model achieved 99.22% test

TABLE I: Comparison between the class-wise accuracy of the previous study and our research

Class Label	Our Model	Previous Study [14]
1	100	96.67
2	99.0	98
3	99.0	96.33
4	99.0	96.33
5	99.5	96.5
6	98.5	95.5
7	99.5	95.67
8	98.5	94.83
9	99.5	97.83
10	98	95.17
11	99.5	94.5
12	98	94.17
13	100	94.33
14	98	94
15	100	97.67
16	99.5	97.83
17	100	95.5
18	99.5	93.83
19	99.0	94.83
20	100	94.83
21	100	94.5
22	98.5	94.33
23	99.5	93.67
24	99.5	94.17
25	100	94
26	97.5	94
27	98.5	95.67
28	99.5	95
29	100	94.5
30	100	90.33
31	98	91.17
32	100	94.33
33	99	95.83
34	100	95.67
35	100	96
36	99	95.5

accuracy which outperformed all previous studies. Figure-5 shows the confusion matrix which was achieved after applying the trained model on the test set. To evaluate the performance in a more precise graphical manner, a macro-average ROC curve analysis was performed as in this research a balanced dataset was considered. Figure-6 illustrates the macro-average ROC curve analysis for the test set. Finally, a comparison between the previous study and our research was conducted. Table-1 shows the comparison between the previous and our research for each Bengali alphabet. And from the table, it is clear that our model outperformed the previous study for each alphabet as well.

IV. Conclusion

Building a sign language recognition system for the mute-deaf community has been an area of interest for the researchers for a long time now. Previously, many studies have been conducted for the recognition of different sign language datasets. However, not many studies were conducted based on the Bengali sign alphabets. A most recent study suggested a mean accuracy of 95.83% [14]. In this study, we started with a Bengali sign alphabet dataset for precise recognition of the sign alphabets. After applying our proposed CNN model, we concluded that our model achieved an overall test accuracy of 99.22% that outperformed all previous studies. We hope our research will be beneficial for the mute-deaf communities and further development on sign language detection.

References

[1] E. Britannica and E. Britannica, "Inc., 2012," *Encyclopædia Britannica Online. Website*, 2012.

[2] W. H. Organization *et al.*, "Deafness and hearing loss. fact sheet n 300. updated march 2015," 2015.

[3] T. Vos, R. M. Barber, B. Bell, A. Bertozzi-Villa, S. Biryukov, 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.

[4] W. H. Organization *et al.*, *The global burden of disease: 2004 update*. World Health Organization, 2008.

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

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

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

[8] M. Amin, "Prevention of deafness and primary ear care (bengali)-society for assistance to hearing impaired children (sahic)," *Mohakhali, Dhaka-1212, Bangladesh*.

[9] 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, 2017, pp. 1–6.

[10] 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) www. jncet. org*, vol. 9, no. 2, 2019.

[11] 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.

[12] 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.

[13] F. 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.

[14] 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, p. 143302, 2020.

[15] M. S. Islam, S. S. S. Mousumi, N. A. Jessan, A. S. A. Rabby, and S. A. Hossain, "Ishara-lipi: The first complete multipurposeopen 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.

[16] 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.

[17] 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, p. 185.

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