2020 11th International Conference on Electrical and Computer Engineering (ICECE) | 978-1-6654-2254-3/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICECE51571.2020.9393087

Bengali Sign Language Characters Recognition by Utilizing Transfer Learned Deep Convolutional Neural Network

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

***Abstract*—Being the fifth most-spoken native language of the world, Bengali is spoken by approximately 265 million people worldwide. In Bangladesh alone, among nearly 130 million native speakers, 13 million individuals are undergoing the hearing impairment problems. Hence, developing a recognition system of the Bengali sign alphabets has been an area of interest for decades. Previously, numerous researchers contributed by either developing benchmark datasets or providing ideas regarding classifiers for accurate recognition. In most of the cases, either the dataset in consideration was not well-constructed or the recognition accuracy was not satisfactory. In this research, we began with a benchmark dataset of Bengali sign alphabets with 38 signs. We applied augmentation and a modified InceptionV3 architecture. The experimental analysis showed that our trained model achieved an overall accuracy of 94.41% which outperformed the previous best-known outcomes by a noteworthy margin.**

***Index Terms*—Bengali Sign Alphabets Recognition, Transfer Learning, Modified InceptionV3 Architecture, Augmentation, Deep Convolutional Neural Network**

I. Introduction

Hearing injury or loss is an inadequate or total inability to listen to [1]. Hearing impairment may happen in both or individual ears [2]. Till 2013, hearing impairment affected greater than 1.1 billion individuals to remarkable proportions [3]. This created inefficiency in approximately 538 million people and occurred intermediate to severe inability in almost 124 million people [2], [4], [5]. To overcome the communication channel gap with deaf and dumb communities, sign language is utilized which is the most practiced literature in the mute-deaf community to interact with people and distribute views. Sign languages are various sorts of natural expressions that have their custom construction and dictionary [6]. Obviously, these literature are not universal for all populations and these are not commonly recognizable amongst different communities despite the existence of extraordinary relationships among many sign signals [6]. That is why a high-performance computerized classification of sign language is highly demanded by the mute-deaf population.

With a community of more than 130 million [7], Bangladesh is undergoing more than 13 million hearing impairment problems solely amongst which nearly 3 million people are undergoing severe to heavy hearing impairment that motivates them to disability [8]. Because of not being a universal literature Bengali sign language has its own pictorial symbols with the assistance of hand gestures, positions, and movements. In Bengali sign language literature, there are a total of 38 one hand symbols including 9 vowels and 27 consonants, and about 4000 double hand symbols for commonly used words in Bengali [9].

In this research, we started with the most recent and noteworthy work of the Bengali sign alphabets classification [9] where we considered a dataset of 38 sign alphabets. After applying some preprocessing steps and splitting the dataset into the train set and test set, we applied augmentation on the train set. Then, with the assistance of the validation set, we trained our model by utilizing InceptionV3 architecture and measured the performance by applying the model on the test set. Analysis of the outcomes concluded that our model achieved a higher performance accuracy and outperformed the most recent and highest-achieved outcome by a remarkable margin.

II. Literature Review

Having numerous languages in the world, sign language recognition has received a lot of research contributions in the last decades. Several popular sign languages have received most of the contributions including American sign language [10], Indian sign language [11], Turkish sign language [12], and Japanese sign language [13]. Still, hundreds of languages are seeking contributions for near accurate recognition. Despite being one of the most spoken languages in the world, Bengali has not received a remarkable amount of researchers in the field of sign language recognition. A few noteworthy contributions have been made previously. Recognition of the Bengali sign alphabets by utilizing neural networks was proposed in 2012 [14]. In the same year, Rahman et al. proposed an artificial neural network-based approach to convert signs into printed letters [15]. However, in 2016, Support Vector Machine (SVM) was suggested as the best classifier for the recognition of the sign alphabets [16]. In all this researches, the main problem was the datasets. The datasets were not well constructed and there was a lack of alphabets in all of these datasets. However, in 2019, a research work contributed a well-constructed dataset of Bengali sign alphabets containing

978-1-6654-2254-3/20/$31.00 ©2020 IEEE

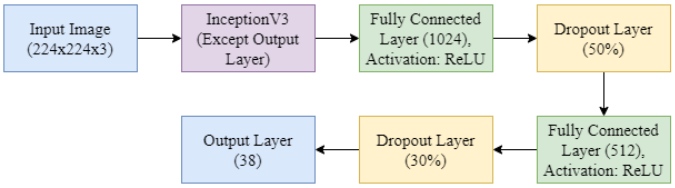

Fig. 1: Proposed modified InceptionV3 architecture

a total of 38 sign alphabets including 9 vowels and 27 consonants. They achieved an overall accuracy of 89.6% by applying VGG19 based convolutional neural network [9]. In this research, we worked with this benchmark dataset and achieved an overall accuracy of 94.41% which outperformed the best-achieved performance. Moreover, in [9], the proposed model of authors overfitted, whereas our model did not overfit while training.

III. Materials and Methods

In this section, first, the dataset description has been provided. After that, a description of transfer learning and the modified InceptionV3 architecture has been introduced. Finally, augmentation has been discussed.

A. Dataset Description

In this research, we considered a publicly available Bengali sign language dataset [9]. The dataset contains 320 samples for each of the 38 classes, a total of 12160 samples. These samples were collected from a total of 340 people among whom 298 people were normal and 42 people were from the deaf community. Also, the ratio of male and female participants was 1.35 and the age-range of the individuals was 12-30. 42 deaf students were from a school in Bangladesh run by the National Federation of the Deaf.

B. Transfer Learning

Transfer learning contracts on gathering data acquired while settling one barrier and performing it to another but similar complication [17]. Suppose, the knowledge collected while learning to distinguish cars could be utilized during the identification of other vehicles like trucks. This department of research illustrates an extraordinary connection to the enduring chronicle of cerebral investigation on the transfer of knowledge, although confirmed relationships among the two spaces are insufficient. From a practical perspective, carrying or dispatching information from formerly succeeded assignments for the training of current jobs has the ability to dramatically magnify the individual efficiency of an agent [18].

C. Modified Inception V3 Architecture

Inception v3 is a broadly accepted image classification architecture which has attained more than 78.1% overall accuracy on the ImageNet dataset [19]. The model is the conclusion of numerous concepts disclosed by many researchers over the ages. It is developed based on the primary paper: ”Rethinking the Inception Architecture for Computer Vision” by Szegedy et al. [20]. The architecture itself is manufactured of symmetric and asymmetric building blocks, including convolutions, max pooling, average pooling, dropouts, concats, and fully connected layers. Batch normalization is utilized broadly everywhere in the architecture and implemented to activation inputs. Loss is measured via Softmax. However, in our research, we applied the InceptionV3 architecture except for the output layer. After applying the inceptionV3 block, we applied a fully connected layer of size 1024 followed by a dropout of 50% and then a fully connected layer of size 512 followed by a dropout of 30%. In both cases of the fully connected layer, ReLU activation was utilized. Finally, there is an output layer of size 38. While applying the architecture no layer was kept frozen. Figure 1 illustrates our proposed methodology.

TABLE I: Class-wise precision, recall, f1-score and accuracy (acc.) illustration

Class	Precision	Recall	F1-Score	Acc.
1	0.95	0.90	0.92	0.90
2	0.83	1.00	0.91	1.00
3	1.00	0.93	0.96	0.93
4	0.97	0.97	0.97	0.98
5	0.85	0.97	0.91	0.98
6	0.97	0.95	0.96	0.95
7	0.93	0.95	0.94	0.95
8	0.95	0.95	0.95	0.95
9	0.95	0.97	0.96	0.98
10	1.00	0.97	0.99	0.98
11	0.93	0.95	0.94	0.95
12	1.00	0.85	0.92	0.85
13	1.00	0.95	0.97	0.95
14	0.95	0.93	0.94	0.93
15	0.95	1.00	0.98	1.00
16	0.67	0.90	0.77	0.90
17	0.97	0.97	0.97	0.98
18	0.95	0.88	0.91	0.88
19	1.00	1.00	1.00	1.00
20	0.95	0.97	0.96	0.98
21	0.97	0.93	0.95	0.93
22	1.00	0.95	0.97	0.95
23	0.98	1.00	0.99	1.00
24	0.77	0.60	0.68	0.60
25	1.00	0.97	0.99	0.98
26	0.98	1.00	0.99	1.00
27	1.00	0.95	0.97	0.95
28	0.95	1.00	0.98	1.00
29	1.00	1.00	1.00	1.00
30	0.86	0.95	0.90	0.95
31	0.97	0.88	0.92	0.88
32	1.00	1.00	1.00	1.00
33	0.94	0.85	0.89	0.85
34	0.97	0.93	0.95	0.93
35	0.95	0.97	0.96	0.98
36	0.93	1.00	0.96	1.00
37	0.97	0.97	0.97	0.98
38	0.97	0.95	0.96	0.95

TABLE II: Comparison of performances among our proposed work and previous work

Classifiers	Overall Accuracy
Deep CNN [9]	89.60%
Modified InceptionV3 Architecture	**94.41%**

D. Augmentation

Insufficient data has constantly been a remarkable obstruction while executing deep learning structures like convolutional neural networks. Moreover, imbalanced data in terms of classes can be a supplementary difficulty. While there may be adequate data for some classes, consistently significant, but the under-sampled classes will undergo ineffective class-specific enforcement or performance. This aspect is constant. If the model determines from a few examples or events of a given class, it is less feasible to foretell the group label or test class. Image augmentation artificially produces training photographs through various methods of processing or association of multiple processing, such as random rotation, flips, shear, and transfers, etc. [21].

IV. Experimental Analysis

In this section, first, preprocessing steps have been introduced. After that, the design of experiments has been discussed. Finally, result analysis is presented with proper evidence.

A. Preprocessing

While feeding pictures to a convolutional neural network, a heavy preprocessing of the pictures was overlooked as Convolutional Neural Network is a compelling tool that can distinguish precious characteristics from raw images. Nevertheless, some preprocessing steps were needed. The input pictures were in the diverse aspects, hence, they were reshaped to 224x224x3. For more precise classification, augmentation was implemented with the compensation of the Augmentor Library [21]. While applying the method of augmentation, the max left rotation, the max right rotation, and the probability of rotation of the rotation function was fixed to 3, 3, and 0.4 respectively. The values of grid width, grid height, probability, and magnitude of the random distortion function were set to 4, 4, 0.4, and 4 respectively. Moreover, the percentage areas and the probability of the zoom random function were adjusted to 0.9 and 0.2 respectively. After the augmentation method, we had 1000 images per class, a total of 38,000 images for 38 classes of Bengali sign alphabets.

B. Experimental Settings

The model was trained for 25 epochs with a batch size of 24 as after that the validation loss shifted to almost constant for the rest of the epochs. Adam optimizer [22] with the learning rate of 0.0001 was employed to maximize the error function. Categorical cross-entropy function was utilized for the loss or error function. For bypassing overfitting, the dropout technique was followed.

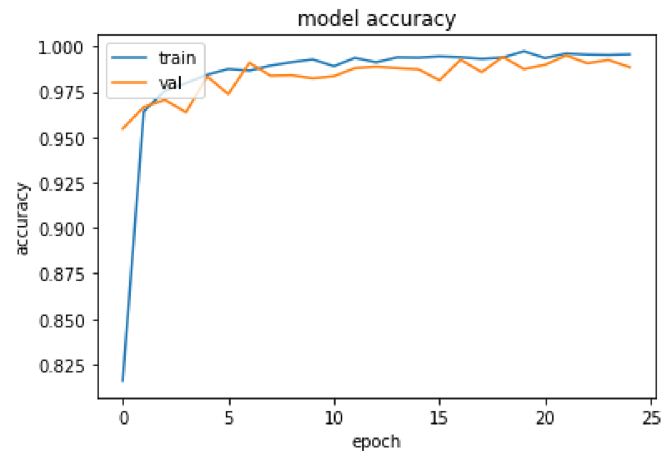

Fig. 2: Training accuracy and validation accuracy of our proposed model

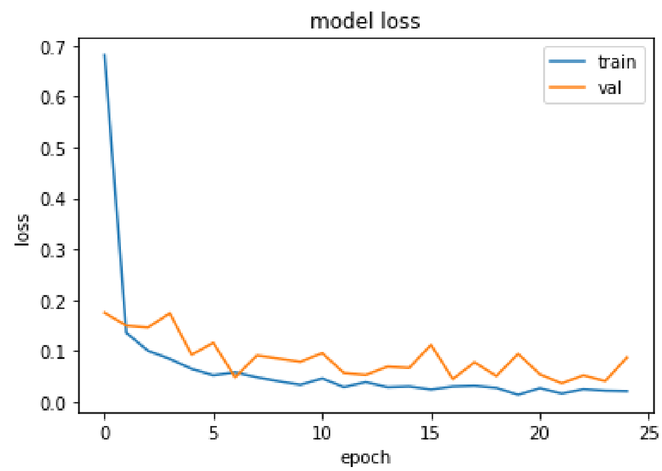

Fig. 3: Training loss and validation loss of our proposed model

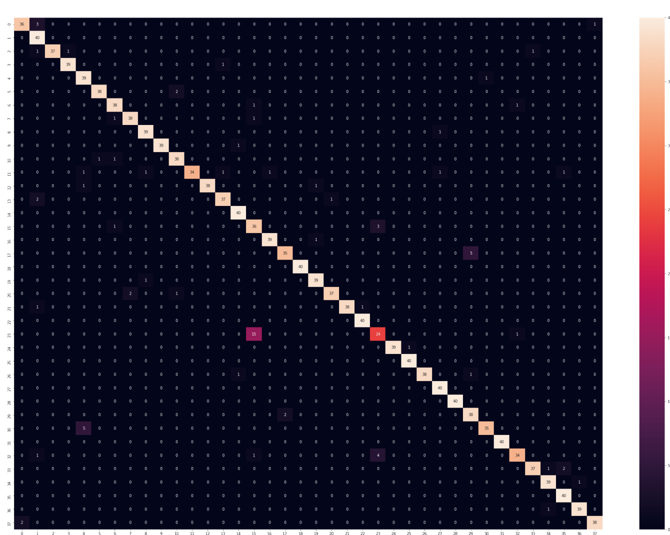

Fig. 4: Obtained confusion matrix after applied on the test set

C. Result Analysis

After applying the initial step of resizing, the dataset was split into the train set and the test set. The test set contains 40 samples for each of the class, a total of 1,520 images. On the other hand, the initial train set had 280 samples for each of the 38 classes. We applied augmentation on the train set and generated 1,000 samples for each of the 38 classes. Among 1,000 samples, 800 samples of each class were kept in the train set and the rest of the 200 samples were kept in the validation set. Hence, the train set had a total of 30,400 samples and the validation set had a total of 7600 samples. We trained our model with the assistance of the modified InceptionV3 architecture. Figure 2 illustrates the training accuracy and validation accuracy while training. Moreover, Figure 3 illustrates the training loss and validation loss while training. After that, we applied our trained model on the test set and obtained 94.41% overall accuracy. Figure 4 illustrates the obtained confusion matrix. To investigate more clearly, in TABLE-I, we showed the class-wise precision, recall, f1-score, and accuracy. In TABLE-II, we compared our results with the previous best work and concluded that our obtained performance outperformed the previous findings by a remarkable margin.

V. Conclusion

In this research, we started with a benchmark dataset that contains 320 samples for each of the 38 Bengali sign alphabets in consideration. After resizing the images, the dataset was split into the primary train set and the independent test set. Then, augmentation was applied on the train set and we split the train set into train set (80% of the primary train set) and validation set (20% of the primary train set). After that, we applied the modified InceptionV3 architecture and measured the performance after applying the trained model on the independent test set. Result analysis revealed that our approach produced an overall accuracy of 94.41% which outperformed the previous best result of 89.60% by a noteworthy margin. In the future, we have plans to apply more sophisticated models on the dataset. Furthermore, more tuning of hyperparameters may reveal some more knowledge regarding the recognition of the Bengali sign alphabets.

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] 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," in *2019 IEEE Global Humanitarian Technology Conference (GHTC)*. IEEE, 2019, pp. 1–7.

[10] K. Mandikhanlou and H. Ebrahimnezhad, "Multimodal 3d american sign language recognition for static alphabet and numbers using hand joints and shape coding," *MULTIMEDIA TOOLS AND APPLICATIONS*, 2020.

[11] A. Tyagi and S. Bansal, "Feature extraction technique for vision-based indian sign language recognition system: A review," *Computational Methods and Data Engineering*, pp. 39–53, 2020.

[12] G. Eryiğit, C. Eryiğit, S. Karabüklü, M. Kelepir, A. Özkul, T. Pamay, D. Torunoğlu-Selamet, and H. Köse, "Building the first comprehensive machine-readable turkish sign language resource: methods, challenges and solutions," *Language Resources and Evaluation*, vol. 54, no. 1, pp. 97–121, 2020.

[13] H. Brock, I. Farag, and K. Nakadai, "Recognition of non-manual content in continuous japanese sign language," *Sensors*, vol. 20, no. 19, p. 5621, 2020.

[14] B. C. Karmokar, K. M. R. Alam, M. K. Siddiquee *et al.*, "Bangladeshi sign language recognition employing neural network ensemble," *International journal of computer applications*, vol. 58, no. 16, pp. 43–46, 2012.

[15] M. A. Rahman, M. Abdullah, S. Mondal *et al.*, "Recognition of static hand gestures of alphabet in bangla sign language," *IOSR Journal of Computer Engineering (IOSRJCE)*, vol. 8, no. 1, pp. 07–13, 2012.

[16] M. A. Uddin and S. A. Chowdhury, "Hand sign language recognition for bangla alphabet using support vector machine," in *2016 International Conference on Innovations in Science, Engineering and Technology (ICISET)*. IEEE, 2016, pp. 1–4.

[17] J. West, D. Ventura, and S. Warnick, "Spring research presentation: A theoretical foundation for inductive transfer," *Brigham Young University, College of Physical and Mathematical Sciences*, vol. 1, no. 08, 2007.

[18] T. George Karimpanal and R. Bouffanais, "Self-organizing maps for storage and transfer of knowledge in reinforcement learning," *Adaptive Behavior*, vol. 27, no. 2, pp. 111–126, 2019.

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

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

[22] Z. Zhang, "Improved adam optimizer for deep neural networks," in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*. IEEE, 2018, pp. 1–2.