2022 3rd Int. Conf. on Innovations in Science, Engineering and Technology (ICISET)
26-27 February 2022, Chittagong, Bangladesh

2022 International Conference on Innovations in Science, Engineering and Technology (ICISET) | 978-1-6654-8397-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICISET54810.2022.9775908

Bangla Sign Language Recognition from Hand Gestures using Convolutional Neural Network

Sadia Sultana
Dept. Of CSE
International Islamic University
Chittagong, Bangladesh
hpctgbd@gmail.com

Umme Subrina Jannat
Dept. Of CSE
International Islamic University
Chittagong, Bangladesh
subrinajerin65@gmail.com

Rounok Afza Doha
Dept. Of CSE
International Islamic University
Chittagong, Bangladesh
dohafza@gmail.com

Mohammad Mahadi Hassan
Dept. Of CSE
International Islamic University
Chittagong, Bangladesh
mahadi_cse@yahoo.com

Muhammed J.A. Patwary
Dept. Of CSE
International Islamic University
Chittagong, Bangladesh
jamshed_cse_cu@yahoo.com

***Abstract*—In modern society, sign language has become the primary language of the deaf and dumb community. To communicate with others, they employ a variety of signs. Sign language recognition is a new field of study that aims to improve communication with the deaf and dumb. Many studies on the identification of Bangla sign language have been reported in the literature. However, we found very little research with high accuracy for the identification of Bangla sign language. In Bangladesh, there is also a sizable deaf and dumb population. In this research, we build a modified convolutional neural network for recognizing Bangla sign language. We have 37970 data for 59 classes in our dataset. Finally, we looked at our model's performance separately. We got 100% accuracy for digits, 99.84% for alphabets, and 99.5% for combined use.**

***Index Terms*—Bangla sign language, Convolutional neural network, Dataset, Recognition**

I. Introduction

Machine learning algorithms are used in a wide range of applications [2,5,6], including medicine, email filtering, speech recognition, and Sign Language Recognition,when it would be difficult or impossible to design typical algorithms to execute the essential tasks [20,21,22]. Image categorization becomes incredibly simple when using Computer Vision.

Language is the most important medium for communicating with others. This is the communication style of ordinary people in our society. But there is another community living in our society who are not able to share their feelings with others through language. And then there is a communication gap between these two communities in our society. We are talking about the deaf and dumb community of our society. Now the question arises of how they too can communicate with others.

In this study, a Convolutional Neural Network is employed, which is a supervised machine learning algorithm[7], to detect Bangla Sign Language.

The following are the primary contributions of our work:
1. A datset has been created for Bangla Sign Language Recognition. This can be quite useful for future research on this topic.
2. We implement a CNN model that allows us to recognize BdSL with greater accuracy.

There are five sections to our study. The introduction is the initial section. The scope of the study and a literature review were described in the second section. In the third section, we addressed the proposed model's technique in detail.In the fourth section, we addressed the performance of our proposed model. The final portion contains our overall conclusions as well as our future work plan.

II. Literature Review

There is a scarcity of publically accessible dataset for working on Bangla sign language recognition.And this is the field's primary research restriction. It is also worth noting that no one has worked on all bangla alphabets hand signs in the literature since every character hand sign was not included in the dataset.This study presented a dataset that included all of the characters hand signs used in Bangla Sign Language Recognition.Thankfully, we were able to obtain a dataset which shown in figure 3 from the paper's author [1]. This, however, did not include all of the essential characters of the Bengali alphabet. As a result, we've expanded the dataset to include all of the basic characters of the Bengali alphabet. In this field, a lot of research has been done. However, the communication gap still exists. The lack of a reliable dataset for Bangla sign language could be the cause. Bangla is one of the world's most frequently spoken languages. In Bangladesh,

there are around 2.4 million deaf people [13].

Study found in [1,3,4] and [8] to [11], they all used Convolutional Neural Network for recognizing Bangla sign language. But according to their research, no one has worked on all Bangla alphabets. P. Haque et al. [12] proposed a PCA and KNN framework. On 104 test images, they got 77.8846% accuracy. Md. Uddin et al. [13] established a methodology for detecting BdSL using Support Vector Machine. The dimensionality was then reduced using the principal component analysis technique. The classification method employed was SVM. They were able to detect BdSL with 97.7% accuracy from 4800 images. The approach proposed in [14], to recognize BdSL hand motions, a linear discriminant analysis and an artificial neural network were used. Rahman et al [19] created a method to detect BdSL via videos. The tone and brightness values of human skin color were calculated by the technique, and the tone and brightness parameters of human skin colour were used to depict the hand gesture. Following normalization, it was converted to a binary picture, and a KNN classifier was used to recognize it. From 7200 samples, the system obtained an overall accuracy rate of 96.46% .

III. PROPOSED METHOD

All the steps of our proposed methodology have been included below diagram:

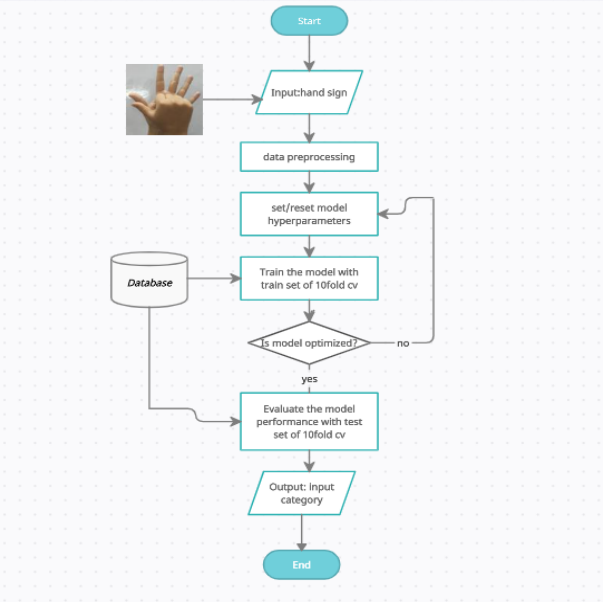

Fig. 1. Flowchart of the proposed methodology

A. Preprocessing image:

At the beginning of preprocessing, we convert RGB images to grayscale. After that, the images were normalized by dividing them by 255. Then we scaled the images to 60x60. There is a lot of information in the RGB image that may not be necessary for processing. So that, we converted in grayscale to reduce storage.

B. Convolutional neural network:

A Convolutional Neural Network is consists of convolution layers withone or more layers that are entirely connected, as in a multilayer neural network. CNN has typically three components. Input, Feature Extraction, Classification & Output. In the input layer, an input image is given to the network. Feature extraction is done in the convolution layer. A kernel operates on an input image to generate feature maps from it. Once a feature map is created, then it passes through a nonlinearity such as ReLU. The pooling layer downsamples the output of the previous layer. In the classification and output layer, the output of the last Convolution layer or pooling layer is flattened to a 1D array and Each input is linked to each output by learnable weight in one or more completely connected layers. The probabilities of each class for classification are mapped by a subset of fully connected layers to the output layer. The last FC layer contains the equal output nodes as the present classes in the dataset.The architecture of the proposed model has been shown in figure 2. In this figure Conv2D stands for Convolution2D, ReLU is the Rectified linear unit, and BN is for Batch Normalization.

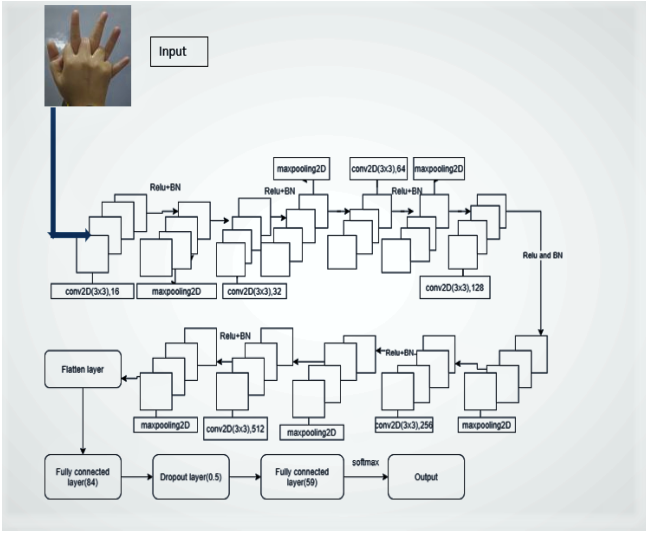

Fig.2. The architecture of the suggested CNN model

- Input Layer: After preprocessing, the images are uploaded to the system. This layer holds 3600 nodes for inputs size of 60×60.

- Convolutional Layer: The input images pass through the convolution layer, which creates a feature map with 3x3 kernels sliding with the input image. We used batch normalization [15] after the ReLU. That is very helpful to reduce internal covariate shifts and speed up the training procedure.

- Activation Function: All CNN models rely heavily on activation functions. In our CNN, we used the activation function ReLU. ReLU is very fast compared to other activation functions in training and can help to solve the problem of vanishing gradients. ReLU can be represented as:

$$ReLU(x) = (0, x) \tag{1}$$

Here, x is the input to a neuron.

- Pooling Layer: In this work, max-pooling2D used as the pooling layer. The max-pooling operation decreases the feature maps dimension. This permits us to reduce the number of parameters, which speeds up our training process and also combats the problem of overfitting. We have used 2x2 pool size in our model. And the stride is also 2. The output dimension (N_{out}) of this operation can be calculated by equation (2)

$$N_{out} = floor(\frac{N_{in} - F}{S}) \tag{2}$$

where:

N_{in} is the input image dimension,
F is the kernel size and S is the stride size.

- Fully Connected Layer: The fully linked layer takes the output of the max-pooling layer as input. In a completely linked layer, each neuron in one layer is coupled to all neurons in another layer. The FC layer uses the retrieved features from the preceding layers to conduct classification. Units 84 and 59 were employed in two FC layers.

- Dropout Layer: Dropout is a linearization approach that is utilized to remove overfitting from our model. We choose the dropout ratio of 0.5 and used it after the first FC layer.

- Output Layer: A CNN's output layer is in charge of calculating the likelihood of every category arising from the input hand sign image. We constructed the last FC layer should have the equal number of neurons in the classes, in order to attain these probabilities. The result of the final FC layer, which generates a vector with components that add up to one, was processed using the softmax activation function. The following formula can be used to compute it:

$$\sigma(X)_j = \frac{e^{xj}}{\sum_{k=1}^{N} e^{x_k}} \tag{3}$$

where:

j = 1, 2, 3,..., N, X_j are the inputs to each Softmax and
N is the number of final FC layer nodes.

C. Training Details

- Data Augmentation: Data augmentation[17] refers to the increasing amount of samples in a dataset for images. The more data we collect, the better our model will perform. We augmented out images such as rotation fixed at 10 degrees, the shift in width is 10% and height is also 10%. We choose the shear range as 10% with a zoom range of 10%.

- Training and Evaluation: We used 10-fold stratified cross validation for our proposed model and used the sparse categorical cross-entropy as a loss function with the Adam optimizer [18]. We set the batch size to 32. We implement our model on Google Colaboratory. Experiments were conducted on a PC with an Intel Core i5 2.4 GHz CPU, 500 GB HDD and 8 GB RAM. We capture the hand sign image for the dataset using the Redmi note9 pro with a 48MP camera.

IV. Results And Discussion

Table 1 shows the results of the suggested model.The outcomes of the model on the combined dataset is slightly lower than the others, perhaps due to the large number of classes that the dataset contains .Maybe the modification in the architecture of the proposed model can increase it.

A. Dataset

We have worked on a dataset consisting of 37,970 data. Of these, 30,916 data come from the authors of the paper [1]. In their dataset, they have not considered all the basic alphabets of Bengali characters. They worked on 45 classes. So, we have created the remaining alphabets hand signs on our own and now we have a total of 59 classes. Our dataset is divided into two parts: Digits with 7052 data and Alphabets with 30,918 data.

Fig.3. Dataset of 45 classes from paper[1]

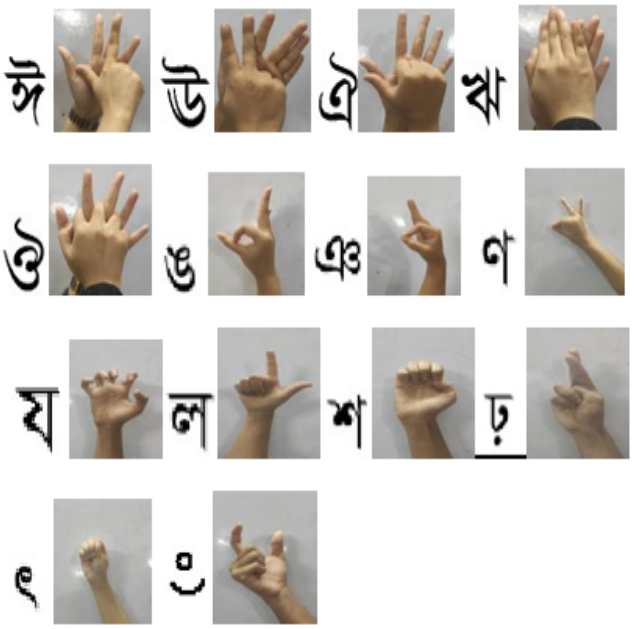

Fig.4. Extended hand sign images for remaining 14 classes

B. *Performance of the model*

First, we observed performance on the digits, achieving a test accuracy of 100%. For alphabets, we gained a test accuracy of 99.84%. Finally, we combined numbers and alphabets and gained a test accuracy of 99.5%. Our model's performance is shown in Table 1. Table 2 compares the results of our proposed CNN model with those of others.

Table 1 Results of our proposed model

Dataset	Train Acc	Test Acc
Digits	99.6%	100%
Alphabets	99.81%	99.84%
Combined	99.7%	99.53%

Table 2 Comparisons with others model

Model	Dataset	Accuracy
PCA&KNN[12]	104	77.88%
SVM[13]	4800	97.7%
KNN[23]	7200	96.6%
Proposed model	37970	99.53%

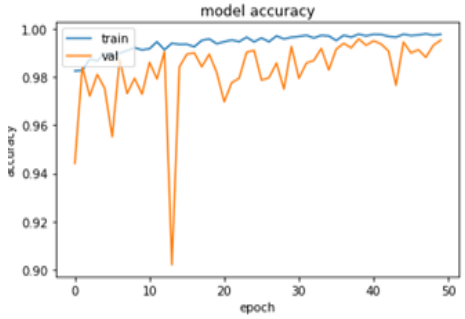

Fig. 5. Accuracy curve of the proposed model

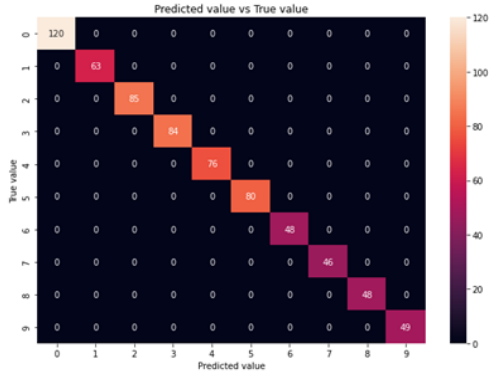

Fig. 6. Confusion matrix for numerals

V. Conclusions And Future Work

We developed a dataset that contains all of the basic characters and digits are presented in Bangla sign language. CNN outperforms other models in recognizing Bangla sign language according to our research. We built the first phase of an automatic interpreter as a result of our research, which turns the static hand sign image into spoken language with improved accuracy. We used the enlarged dataset to test the outcomes of the suggested model in our study. We were able to attain excellent accuracy in BdSL identification as a result of this. This dataset could help to fill in the gaps in the future study.We plan to expand this dataset in the future to include additional numerical combinations. In addition, we will create a dataset on Bengali words. Finally, we'll test our model against that dataset to check if it can distinguish specific Bangla words.

Acknowledgment

We would like to express our special thanks to Md. Hafizur Rahman for providing their dataset for our research.

References

[1] M. S. Islalm, M. M. Rahman, M. H. Rahman, M. Arifuzzaman, R. Sassi and M. Aktaruzzaman, "Recognition Bangla Sign Language using Convolutional Neural Network," 2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2019, pp. 1-6, doi: 10.1109/3ICT.2019.8910301

[2] Patwary MJ, Wang XZ. Sensitivity analysis on initial classifier accuracy in fuzziness based semi-supervised learning. Information Sciences. 2019 Jul 1;490:93-112.

[3] 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. 1-7, doi: 10.1109/GHTC46095.2019.9033031.

[4] 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.369373,doi:10.1109/ICIEV.2018.8640962

[5] Patwary MJ, Wang XZ, Yan D. Impact of fuzziness measures on the performance of semi-supervised learning. International Journal of Fuzzy Systems. 2019 Jul;21(5):1430-42.

[6] Liu J, Patwary MJ, Sun X, Tao K. An experimental study on symbolic extreme learning machine. International Journal of Machine Learning and Cybernetics. 2019 Apr;10(4):787-97.

[7] Patwary MJ, Liu JN, Dai H., Recent advances of statistics in computational intelligence (RASCI), International Journal of Machine Learning and Cybernetics, 2018 January Volume 9, Issue 1, pp. 1–3

[8] Rabby, Akm Shahariar Azad Haque, Sadeka Islam, Md. Sanzidul Abujar, Sheikh Hossain, Syed. (2018). BornoNet: Bangla Handwritten Characters Recognition Using Convolutional Neural Network. Procedia Computer Science. 143. 528-535. 10.1016/j.procs.2018.10.426.

[9] S. Hossain, D. Sarma, T. Mittra, M. N. Alam, I. Saha and F. T. Johora, "Bengali Hand Sign Gestures Recognition using Convolutional Neural Network," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp.636-641,doi:10.1109/ICIRCA48905.2020.9183357.

[10] F. Yasir, P. W. C. Prasad, A. Alsadoon, A. Elchouemi and S. Sreedharan, "Bangla Sign Language recognition using convolutional neural network," 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), 2017,pp.49-53, doi: 10.1109/ICICICT1.2017.8342533.

[11] F. Zhan, "Hand Gesture Recognition with Convolution Neural Networks," 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), 2019, pp. 295-298,doi:10.1109/IRI.2019.00054.

[12] P. Haque, B. Das and N. N. Kaspy, "Two-Handed Bangla Sign Language Recognition Using Principal Component Analysis (PCA) And KNN Algorithm," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019, pp. 1-4, doi: 10.1109/ECACE.2019.867918

[13] 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.

[14] R. Yasir and R. A. Khan, "Two-handed hand gesture recognition for bangla sign language using lda and ann," in 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), 2014, pp.1–5

[15] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in International Conference on Machine Learning, 2015, pp. 448–456

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processingsystems,2012

[17] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," 2018 International Interdisciplinary PhD Workshop (IIPhDW), 2018, pp. 117-122, doi: 10.1109/IIPHDW.2018.8388338.

[18] 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.

[19] M. A. Rahaman, M. Jasim, M. H. Ali, and M. Hasanuzzaman, "Realtime computer vision-based bengali sign language recognition," in 17th International Conference on Computer and Information Technology (ICCIT), 2014, pp. 192–197

[20] Patwary MJ, Parvin S, Akter S. Significant HOG-histogram of oriented gradient feature selection for human detection. International Journal of Computer Applications. 2015 Jan 1;132(17).

[21] Jahin D, Emu IJ, Akter S, Patwary MJ, Bhuiyan MA, Miraz MH. A Novel Oversampling Technique to Solve Class Imbalance Problem: A Case Study of Students' Grades Evaluation. In2021 International Conference on Computing, Networking, Telecommunications Engineering Sciences Applications (CoNTESA) 2021 Dec 9 (pp. 69-75). IEEE.

[22] Faisal MF, Saqlain MN, Bhuiyan MA, Miraz MH, Patwary MJ. Credit Approval System Using Machine Learning: Challenges and Future Directions. In2021 International Conference on Computing, Networking, Telecommunications Engineering Sciences Applications (CoNTESA) 2021 Dec 9 (pp. 76-82). IEEE