2020 11th International Conference on Electrical and Computer Engineering (ICECE) | 978-1-6654-2254-3/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICECE51571.2020.9393070

BANGLA SIGN DIGITS: A DATASET FOR REAL TIME HAND GESTURE RECOGNITION

Dardina Tasmere, Boshir Ahmed, Md Mehedi Hasan
Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh
Email: dardina.ruet@gmail.com , boshir78@gmail.com , mmehedihasann@gmail.com

Abstract— With current widespread computer vision study, research with a particular culture and aesthetic contexts lead to enhanced performance with improved user experience. Despite a large number of deaf and dumb people, Bangla Sign Language has been less-focused on the research area. This paper proposes the building of identifying Bangla Sign Language using hand gestures to reduce the isolation of the hearing-impairment world. For the last little advancement in deep learning, Convolution Neural Network (CNN) has been used to classify Bangla Sign Language with higher accuracy. Our methodology overcomes all constraints from existing research of Bangla hand gesture recognition. In this paper, we presented a novel real-time method for Bangla sign digits (0-9) that focuses on three sections: image acquisition, preprocessing, and lastly, recognition of Bangla sign digits. For training, we have created a dataset of 1674 images from different people from various conditions. The proposed CNN model architecture classifies all Bangla sign digits with an overall accuracy of 97.63% on real-time video that ensures a potential gesture model that contributed significantly in the domain of Bangla sign number recognition compared to previous researches.

Keywords— A dataset for Bangla Sign digits, Bangla Sign Digits Recognition, Deep Convolution Neural Network, Real Time Recognition.

I. INTRODUCTION

Only communication can spell all aspects of living a life. Technology is helping aid this process. Disability is a vital issue and cannot be disregarded. Sign language stands dominantly for hearing and speech impaired people to communicate. Sign languages differ from country to country and culture to culture. Like other countries, Bangladesh also has its sign language named Bangla Sign Language (BaSL). According to the Bangladesh Bureau of Statistics (BBS), the estimated total percentage of hearing (13%) and speech impairment (9%) is 22% by the distribution of disabled by type of disability [1]. Though nowadays different developed countries have focused on the learning process of physically disabled people Bangladesh is still far away from the experience of such technical development. To fill up the gap of naturalness and reaching out to the problem of impaired people, sign language recognition has become a more iterative topic in the computer vision field.

Using advanced technological infrastructure, large numbers of research projects are dedicated to translating ASL (American Sign Language) in real-time. Following a few years, deep learning techniques have proven a flawless performance [2] to suppress traditional systems in computer vision. Such as Peijun Bao et al. [2] applied a deep convolution neural network composed with an accuracy of 97.2%. At the same time, recognition of gestures with both written words and audible speech was proposed [3] by image matching with an overall accuracy of 85%-90%. Compared with different sign language recognition, Bangla sign language recognition has drawn the leading fascination of researchers. Before the beginning, K-Nearest Neighbors (KNN) algorithm was used to classify the Bangla alphabets [4]. Beyond hand detection, Bangla hand signs are directly classified by Shayhan Chowdhury et al. [5] using Support Vector Machine (SVM). In [6] neural network is introduced to classify Bangla sign language. Follow this NN trend, in [7] deep convolutional neural network has been used on the validation dataset to accurately recognize Bangla signs with an accuracy of 84.68%. For improvement in this research area, they used a multi-layered Convolutional Neural Network (CNN). CNN can automate the structure formulation process. [8]. Very recent research conducted by Shahjalal Ahmed et al. on Bangla sign digits gained a mean accuracy of 92% over validation data [9] that outperforms all other previous research for Bangla digits. Very recent research was conducted on Bangla sign language, by hand skeletal features and CNN to recognize and classify hand images [10].

To facilitate deaf and dumb people of Bangladesh in many aspects such as education, medical as well as safety and social purpose, this paper proposed a real-time hand gesture recognition system on our new dataset, “Bangla Sign Digits’’. Fig. 1 presents Bangla sign digits.

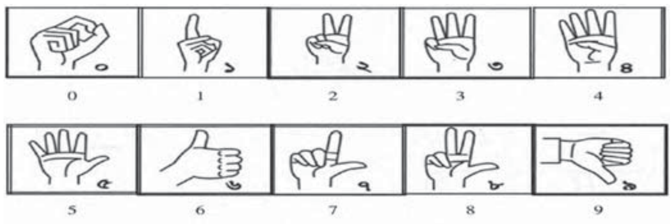

Fig. 1 Bangla sign digits

I. DATASET DESCRIPTION

Dataset plays a crucial role in the significance of research by posing challenging problems. Despite having some works with Bangla sign language, there is not enough dataset available to work. The existing dataset for Bangla sign digits is a lack image quantity. Development in this field has become complicated for lacking proper datasets. For increasing classification accuracy, a large dataset plays a significant role. In this research paper, a Bangla sign digits dataset has been created. [16] [https://github.com/MehediCSERUET/Bangla-Sign-Digits]. Images of different people (male and female) have been captured to originate the dataset. The training images for Bangla sign digits have been taken by the webcam captured within a specific time frame. The dataset assimilates 1066 images composed for one-handed Bangla sign digits images for training, 271 images for validation, and 337 images for

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

testing. The number of data for each class is shown in Table I.

Table I. Number of Images for Each Class

Class	No. of Training Data	No. of Validation Data	No. of Test Data
0	163	41	51
1	166	42	52
2	163	41	51
3	48	12	16
4	64	16	20
5	86	22	27
6	112	29	36
7	85	22	27
8	83	21	26
9	96	25	31
Total	1066	271	337

In Fig. 2 Data sample images of each class are shown.

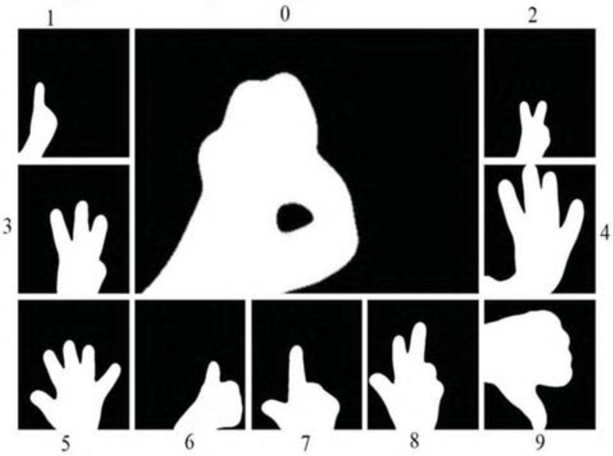

Fig. 2 Data sample of each class (The number around each image denotes the class label)

II. PROPOSED FRAMEWORK

In this present section of the paper, the detailed system for the recognition of Bangla sign digits has been elaborated. Our proposed system detects hand area from a real-time video extracts hand features after preprocessing, and trains a deep CNN model to recognize correct gestures. Fig. 3 illustrates the flowchart of the proposed real-time Bangla sign number recognition system.

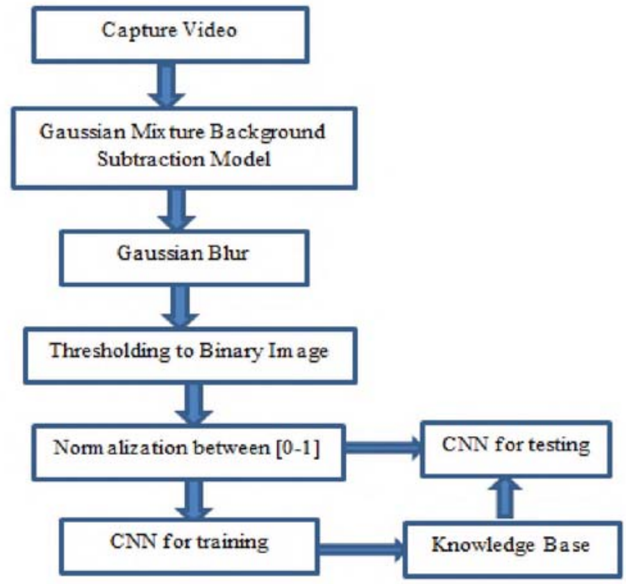

Fig. 3 Methodology of the proposed framework

A. *Image Acquisition*

Though a small amount of tasks has been done for Bangla hand gestures, it is limited to a variable environment having environmental noise. The proposed system gives the facility to capture a hand from a video. For each position of the hand, images have been captured from video frames. All hand images have been stored in .png format.

B. *Preprocessing*

Without enough preprocessing, it is burdensome to expect to get a tantalizing result no matter how good the data we have.

a) Gaussain mixture-based background algorithm Detection of moving objects is an essential process in visual surveillance.It contributes a better adaption capability to varying scenes due illumination changes[13]. In order to adapt to possible changes the training set should be updated. A appropriate time conversion period T has chosen. At time T, $X_T = \{x^t, \ldots\ldots\ldots, x^{t-T}\}$ For every current sample, updated training dataset, X_T and reestimates the density, $\hat{p}(\vec{x} \mid X_T, BG)$ Although, among from the past samples there could be values that belong to foreground objects. The estimation is denotaed as, $\hat{p}(\vec{x} \mid X_T, BG + FG)$

We use a GMM with M components:

$$\hat{p}(\vec{x} \mid XT, BG + FG) = \sum_{m=1}^{M} \hat{\Pi}_m N(\vec{x} \mid; \vec{\hat{\mu}}_m ; \vec{\sigma}^2{}_m I) \quad (1)$$

Where $\vec{\hat{\mu}}_1, \ldots, \vec{\hat{\mu}}_M$ are the estimates of the averages and $\vec{\sigma}^2 1, \ldots . \vec{\sigma}^2 M$ are the estimates of the variances that define the Gaussian components [14]. For a new sample the ownership $o^{(t)}{}_m$ is set to 1 for the 'close' component with largest $\hat{\Pi}_m$ and the others are set to zero. If there are no 'close' components a new component is generated with

$$\hat{\Pi}_{M+1} = \alpha, \hat{\Pi}_{M+1} = \sigma_0 \quad (2)$$

and $\hat{\Pi}_{M+1} = \sigma_0$ (3)

where σ_0 is some appropriate initial variance [11]. If the maximum components is reached we remove the component with smallest $\widehat{\Pi}_m$ [13,14]. For a new static object, a secondary stabile cluster will be created. Since the old background is occluded the weight $\widehat{\Pi}_{B+1}$ of the new cluster will be non-stop rising. If the object stand static for a long period, its weight becomes greater than C_f and it can be reviewed to be background portion [14].

b) Gaussian Blur

Gaussian blurring uses normal distribution along with convolution function. The convolution function creates a Gaussian distribution of pixel rate, weave some of the volatility. Separate pixel will give a new value for every pixel to a weighted mean of its neighboring pixels, with more weight given to the closer ones than to those remote pixels.

c) Binary Image Conversion

This filtering approach facilitates the recognition process. It is relatively easier to deal with binary image than multiple color channels. Thresholding transforms images into binary images. Pixels below threshold value (128) are converted to 0 and pixels above the threshold value are converted to 1. We resize a full image to size 64X 64 before feeding it to our network.

d) Normalization

Normalization is the process of making all the data fall under the same range. It helps because if there were different ranges of data, the optimizer algorithm would take more time to converge and even hurt the accuracy of the model. We used equation 1 to normalize our images.

$$X_{new} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (4)$$

C. Proposed CNN Model Architecture

In neural networks, a convolutional neural network is one of the main categories to do image analysis. Convolutional Neural Network is one of the algorithms that have been constructed with the growth in computer vision with deep learning. Roughly convolutional neural network contains three layers: convolutional layer, pooling layer, and fully connected layer. Convolutional layers start working with the input. Output information is then sent on to the next layer. The outputs of clusters of neurons are combined into a single neuron in the next layer. Fully connected layers attach each neuron in one layer to each one neuron in the next layer.

The performance of the CNN model relies on the architecture design with a proper choice of convolution layer and numbers of the neuron. There is no particular guidance for choosing the number of convolution layers and neurons [15]. For Bangla sign digits recognition, the proposed CNN model architecture consists of four convolution layers, four pooling layers, and four fully connected layers, and a dropout layer with a 40% drop out in between the last two fully connected layers. The proposed CNN model is shown in Fig. 4.

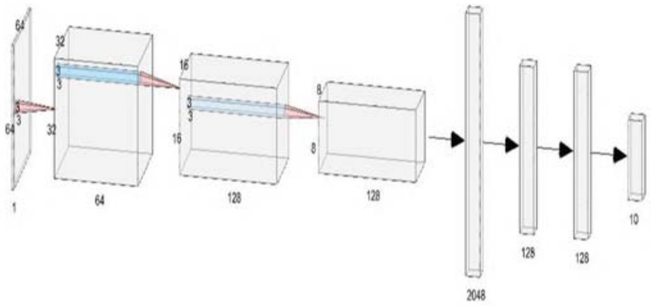

Fig. 4 Architecture of the proposed CNN model

III. Experimental Result Analysis

This part of the study broadly analyzes experimental results of the proposed framework of recognizing Bangla sign digits from a video. The experiment was executed on Intel core i5-7300HQ Processor, 8GB RAM, and NVIDIA Geforce GTX 1050 4GB using python 3.7 environments. It is important to mention that the proposed framework is trained with 1066 frames and tested on 337 frames.

We trained the model on our dataset and to understand the progress of neural networks and direction of network learning, we draw the accuracy and loss curve. The accuracy and loss curve of the model are given below in Fig. 5 and Fig. 6 for 50 epochs.

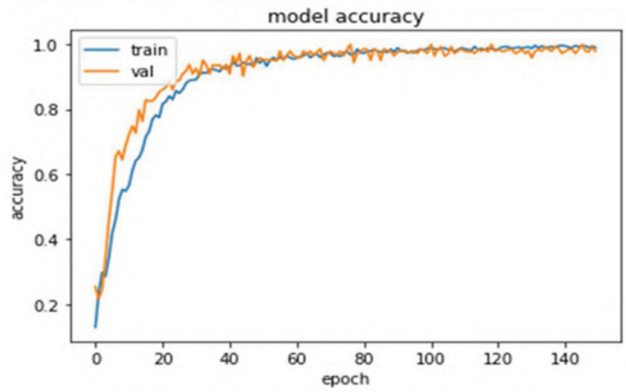

Fig. 5 Accuracy curve of the proposed CNN model

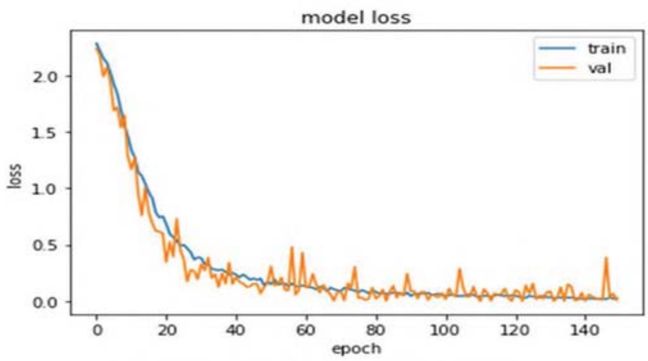

Fig. 6 Loss curve of the proposed CNN model

The accuracy and loss curve shows high training accuracy with lower overfitting. The accuracy of the proposed architecture is 97.63%. The performance of the proposed architecture for Bangla sign language recognition from input video in terms of accuracy, precision, recall, f1-score, and support are given below in Table II.

TABLE II. PERFORMANCE OF THE PROPOSED ARCHITECTURE

Class	Accuracy	Precision	Recall	F1-Score	Support
0	1.0	0.98	1.00	0.99	51
1	0.9807	0.98	0.98	0.98	52
2	0.9803	1.00	0.98	0.99	51
3	0.9375	0.94	0.94	0.94	16
4	0.9	0.95	0.90	0.92	20
5	0.9630	0.93	0.96	0.95	27
6	0.9722	0.97	0.97	0.97	36
7	1.0	0.96	1.00	0.98	27
8	0.9615	1.00	0.96	0.98	26
9	1.0	1.00	1.00	1.00	31

In Fig. 6 confusion matrix of the classification performance of Bangla hand sign language is enumerated. Columns and rows are comparable to the true and predicated label respectively [15].

Fig. 6 Confusion matrix of the proposed CNN

Real time Bangla sign digits recognition system using proposed methodology is sketched out in Fig. 7.

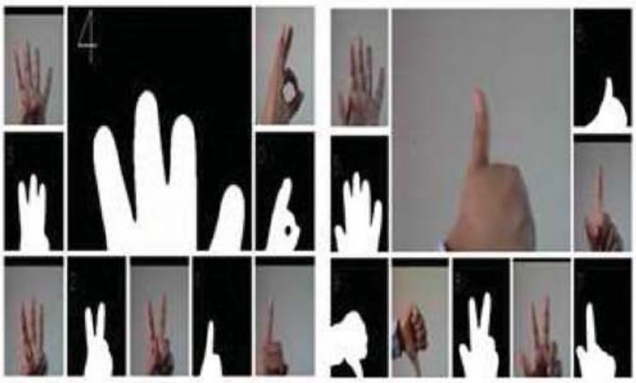

Fig.7 Real Time example using the proposed methodology

IV. CONCLUSION

In this study, a real-time hand gesture recognition system has been proposed lessen the communication troublesome of hearing-impaired people. Our proposed methodology incorporates four convolution layers together with four pooling layers including four fully connected layers and a dropout layer with a 40% dropout. To keep pace with the apparent advancement of convolutional neural networks, our proposed system comes up with an accuracy of 97.63%. Such a recognition system compared to earlier proposed models in this field gives a prominent performance. We developed a new dataset of 1674 images of Bangla sign digits. The development of this dataset might reduce the limitation of future research in this field. Our next target is to build a wide-ranging dataset and emphasizing to implement a more realistic real-time Bangla sign language recognition system.

REFERENCES

[1] Bangladesh Bureau of Statistics, "Disability in Bangladesh: Prevalence and Pattern, Population Monograph : Volume-5," *Bangladesh Bureau of Statistics, Statistics and Informatics Division (SID) Ministry of Planning, Ministry of Planning*, November 2015.

[2] P. Bao, A. I. Maqueda, C. R. Del-Blanco, and N. García, "Tiny hand gesture recognition without localization via a deep convolutional network," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 3, pp. 251–257, 2017.

[3] A. Akoum and N. A. Mawla, "Hand Gesture Recognition Approach for ASL Language Using Hand Extraction Algorithm," *Journal of Software Engineering and Applications*, vol. 08, no. 08, pp. 419–430, 2015.

[4] 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)*, pp. 192–197. IEEE, 2014.

[5] M. A. Uddin, Md Azher 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)*, pp. 1-4. IEEE,2016.

[6] A.M. Jarman, S. Arshad, N. Alam and M.J.Islam, "An Automated Bengali Sign Language Recognition System Based on Fingertip Finder Algorithm", *International Journal of Electronics & Informatics*, Vol. 4, No. 1, pp. 1-10, July 2015.

[7] M. Hossen, A. Govindaiah, S. Sultana, and A. Bhuiyan, "Bengali Sign Language Recognition Using Deep Convolutional Neural Network," *2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 2018.

[8] M. S. Islam, S. S. S. Mousumi, A. S. A. Rabby, and S. A. Hossain, "A Simple and Mighty Arrowhead Detection Technique of Bangla Sign Language Characters with CNN," *Communications in Computer and Information Science Recent Trends in Image Processing and Pattern Recognition*, pp. 429–437, 2019.

[9] S. Ahmed, M. Islam, J. Hassan, M.U. Ahmed, B.J. Ferdosi, S. Saha, and M. Shopon, "Hand Sign to Bangla Speech: A Deep Learning in Vision based system for Recognizing Hand Sign Digits and Generating Bangla Speech", *International Conference on Sustainable Computing in Science, Technology and Management*, February 26-28, 2019.

[10] 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, IEEE.

[11] K. E. Aziz, A. Wadud, S. Sultana, M. A. Hussain, and A. Bhuiyan, "Bengali Sign Language Recognition using dynamic skin calibration and geometric hashing," *2017 6th International Conference on Informatics, Electronics and Vision & 2017 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMHT)*, 2017.

[12] A.M. Rafi, N. Nawal, N.S.N. Bayev, L. Nima, D.C. Shahnaz, S.A. Fattah, "Image-based Bengali Sign Language Alphabet Recognition for Deaf and Dumb Community",*IEEE Global Humanitarian Technology Conference (GHTC)*,October,2019.

[13] Y.-M. Song, S. Noh, J. Yu, C.-W. Park, and B.-G. Lee, "Background subtraction based on Gaussian mixture models using color and depth information," *The 2014 International Conference on Control, Automation and Information Sciences (ICCAIS 2014)*, 2014.

[14] Z. Zivkovic and F. V. D. Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006.

[15] 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.

[16] Bangla Sign Digits: *https://github.com/MehediCSERUET/Bangla-Sign-Digits*