Bangladeshi Sign Language Recognition Employing Neural Network Ensemble

Bikash Chandra Karmokar[1], Kazi Md. Rokibul Alam[2] and Md. Kibria Siddiquee[3]
Department of Computer Science and Engineering, Khulna University of Engineering and Technology
Khulna-9203, Bangladesh
bikash_kuet@yahoo.com[1], rokibcse@yahoo.com[2], shakil_2k7@live.com[3]

ABSTRACT
This paper proposes a Bangladeshi sign language recognizer (BdSLR), an initiative to recognize sign language of Bangladeshi deaf and mute (D&M) people. Although all over the world, the D&M people are a part of the community, the communication between the general and the D&M people becomes tough when interaction is required. Moreover in different races, the D&M people use different sign languages. In this regard BdSLR has been developed that can interpret Bangladeshi sign language into Bengali text and vice versa. In BdSLR, the inputs of Bangladeshi sign language have been taken by webcam and later on recognized by efficient Neural Network Ensemble (NNE). Without any major modification, BdSLR can be used as an interpreter for the sign languages of other races. In the Proposed BdSLR the use of an efficient NNE technique converges the training time faster and recognizes with good generalization ability.

General Terms
Pattern Recognition, Sign language

Keywords
Neural network ensemble, Negative correlation learning, Feature extraction, Bangladeshi sign language

1. INTRODUCTION
Deaf and mute (D&M) people suffer from hearing and speech impairment and use sign language to express their feelings. In the social activities, the communication between the D&M and the general people is hard because usually the sign language is not understandable by the general people. Only few general people who have learned the sign language can understand and translate it for the general ones. The D&M people also cannot understand what the general people say as well as the lip reading [1] too. Moreover for interaction, most of the D&M people do not prefer to write the normal text as their sign language structure is different from it [2]. Thus because of communication gap, the D&M people are ignored in the society in many cases.

There are various sign languages all over the world, namely American Sign Language (ASL), French Sign Language, British Sign Language (BSL), Japanese Sign Language (JSL) etc. All of these were developed independently [3] and are unlike. Similarly Bangladeshi Sign Language (BdSL) is also different from others. Researchers use techniques such as fuzzy logic [4], neural network (NN) [5], PCA [6], Hidden Markov Model (HMM) [7] etc to recognize sign language. In this regard, the existing approaches usually take pre-captured images of sign language as their inputs. Input sources like using a webcam has been rarely used in this purpose. An approach like this can help the D&M people in social activities, i.e. it can be used to translate the sign language of the D&M people to its corresponding normal text that is understandable by the general people and vice versa.

There are lots of D&M people around us who use BdSL for their interactions. Up to now, some works have been done to recognize handwritten Bengali characters [8] as well as Bengali OCR [9]. Besides, the approach presented in [10] has recognized Bengali characters written by touchless fashion i.e. by using webcam. However there is no remarkable progress of recognizing BdSL captured by webcam. If the expressions of D&M people who use BdSL can be captured by webcam and then interpreted, their participation in the social activities can be promoted. In this regard BdSLR has been developed which is an intelligent system. It captures the BdSL of D&M people and capable to translate the captured sign to its corresponding Bengali text and vice versa. Thereby it can improve the communication facilities for the D&M people who use BdSL.

The outline of this paper is as follows: Section II describes existing sign language recognition techniques. Section III presents the proposed BdSLR. Experimental studies have been discussed in Section IV. Finally, concluding remarks are explained in Section V.

2. EXISTING WORKS
The recognition of sign language has begun to appear at the end of 1990. A primary effort was made by using some electrochemical devices to recognize it. The device was used to determine hand gesture parameter like hand's location, angle, position etc. This approach is known as glove-based system. But this approach compels the signer to wear a cumbersome device. It also encounters problem with accuracy and efficiency of the recognition system [11].

The system developed in [12] analyzes video clips of different gestures of sign language taken as input and gives a regular language expression as an audio output. Here actual frame rate of the animation is too quick for interpreting the sign language for which the frame rate was decreased manually.

The system named as "Intelligent Assistant" for human machine interface [13] was developed to communicate the D&M people. For capturing sound the system had used Microsoft's Voice Command and Control Engine along with microphone and converts it into text. But it could not perform efficiently in noisy environments.

Comparing with the above mentioned systems, proposal in [14] is more complicated. Here the signs were shown by

wearing a glove containing different dots in each finger. In real time the signs were taken as inputs. The program then analyzed the dots of the graphics in the image file to understand what the sign had been shown. Then it recognizes the sign in regular language. The technique was based on clustering of the dots. Pre-defined tables are used for mapping the result of this clustering. This system was only limited to Bengali numbers ranging from 1 to 10, which actually does not require any intelligent system to understand [11].

The use of NN is another approach to recognize signs. The mechanism proposed in [15] provided a system that had combined Radial Basis Function and Bayesian Classifier to classify a hybrid vocabulary of state and dynamic hand gestures. Another system proposed in [16] also used NN which classifies JSL. Recently another approach proposed in [11] used intelligent classification technique to recognize BdSL where there is no proper description about the test case and it used a slow learning algorithm for training.

3. PROPOSED BANGLADESHI SIGN LANGUAGE RECOGNIZER (BdSLR)

This paper proposes a Bangladeshi Sign Language recognizer (BdSLR) that learns and recognizes BdSL for D&M people. According to the manual of BdSL [18], there are approximately 5000 sets of gestures for alphabets, numbers and common words. The input of BdSLR is the images of sign of BdSL taken by the webcam captured within a specific time frame. The proposed BdSLR combines feature extraction [19] and neural network ensemble (NNE) for training that yields a better performance. A NNE consists of a set of independent trained NN whose predictions are combined by various methods. Bagging, Boosting, Negative correlation learning (NCL) [17], DECORATE [20] etc are familiar techniques of NNE.

There are numerous reasons for using NNE; some of them are given below:

- Intuition evinces that in almost every cases, NNE offers better generalization ability when compared with a single monolithic NN in the same problem domain.
- In spite of under-fitting and over-fitting [21] problems, NNE works well in comparison with an individual NN, as NNE involves data re-sampling.
- In many cases, additional information is available that cannot be given to the same NN tackling the core information. The additional information can be trained separately and then combined [22].

In order to develop BdSLR among various NNE techniques, NCL [17] has been employed. The NCL technique along with efficient feature extraction enables BdSLR to attain faster training and recognition performance.

Fig. 1: Captured BdSL sign represents 'ক'.

3.1 Major Stages of BdSLR

Major stages of the proposed scheme proceed as follows:

3.1.1 Pre-processing

This process starts through the detection of skin colour of hand in front of a webcam to capture the images of BdSL. The skin colour of the hand has been kept unique in the environment to ensure uniform detection. A captured image of BdSL which is the representation of Bengali alphabet 'ক' has been shown in Fig. 1.

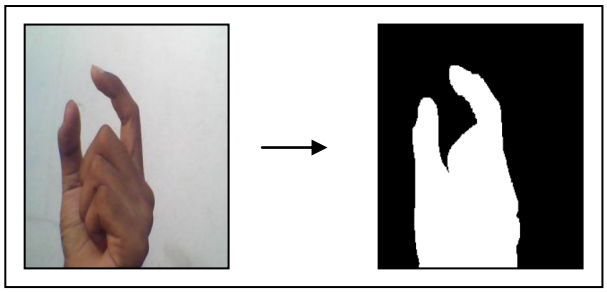
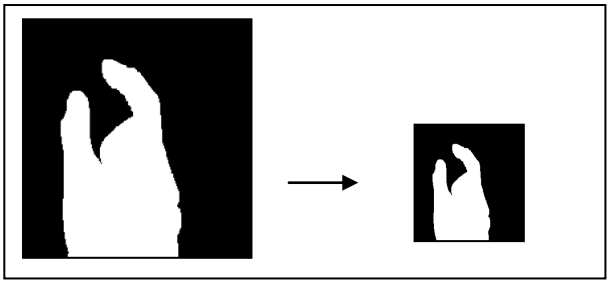


Fig. 2: Conversion into threshold image.



(a) Original size **(b) Normalized size**

Fig. 3: Normalization.

3.1.2 Image processing

At first captured image has been converted into its threshold value where skin colour is shown in white pixels and background is shown in black pixels which has been shown in Fig. 2. Then the converted image has been normalized to 30x33 scale pixels which has been shown in Fig. 3. Then feature extraction method has been applied and NCL [17] algorithm has been used to train these images.

3.1.3 Operational Steps of BdSLR

The block diagram of the proposed BdSLR has been shown in Fig. 4. Its operational steps are as follows:

- Capturing images of BdSL by using a webcam.
- Converting the images into threshold images.
- Applying normalization process to get the reduced size of pixels.
- Extracting the features of the images.
- Matrix representation of the extracted images.
- Training of these obtained data i.e. patterns by using NCL algorithm up to an expected error level.
- Testing i.e. recognizing BdSL sign by taking inputs later on.

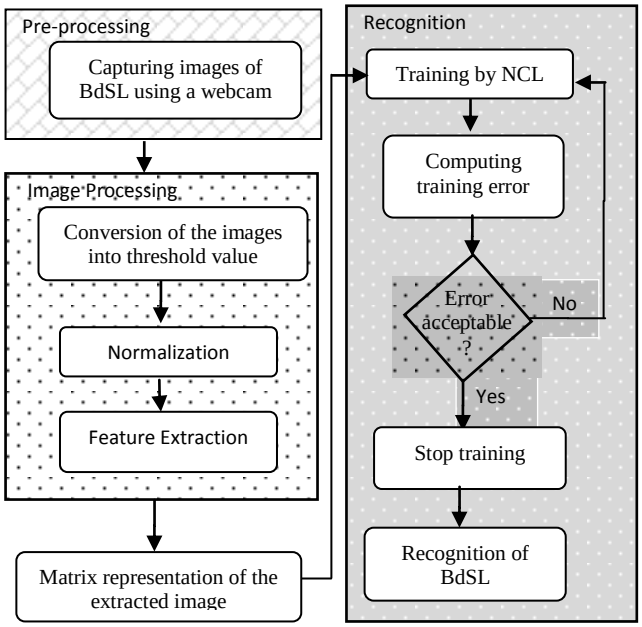


Fig. 4: Block diagram of the proposed BdSLR.

4. EXPERIMENTAL STUDIES

4.1 Experimental Setup

In the experiment, we have used 47 signs of BdSL as input to train where each sign has 5 samples that make 47 x 5 = 235 samples of pre-defined images of same pixel size. Then BdSLR can recognize any of 47 of captured BdSL. In the training of BdSLR, the NCL [17] learns through adjusting its weights which is a supervised learning algorithm. For example for BdSL signs "অ", "ক", "১", "২" etc, we have sufficiently trained BdSLR by NCL in an iterative process that adjust the weights in the mode of pattern-by-pattern updating considering its penalty function. In BdSLR, the chosen architecture of individual NN involved in NCL has 30x33 i.e. 990 nodes (after normalization) in input layer and one hidden layer with flexible no. of nodes. Its output layer contains 47 nodes and the training continues until the error reaches to a certain error level.

The output of NCL is a simple averaging of outputs of a set of NNs which is given by

$$F(n)=\frac{1}{M}\sum_{i=1}^{M}F_i(n)$$

where M is the number of the individual NNs in the NNE, $F_i(n)$ is the output of NN *i* on the nth training pattern, and $F(n)$ is the output of the NNE on the *n*th training pattern. NCL uses a correlation penalty term into the error function of each individual NN in the NNE so that all the NNs can be trained simultaneously and interactively on the same training data set. The error function E_i for NN *i* in NCL is defined by

$$E_i=\frac{1}{N}\sum_{n=1}^{N}E_i(n)=\frac{1}{N}\sum_{n=1}^{N}\frac{1}{2}(F_i(n)-d(n))^2+\frac{1}{N}\sum_{n=1}^{N}\lambda p_i(n)$$

where $E_i(n)$ is the value of the error function of NN *i* at presentation of the nth training pattern. Here, the first term in the right side of the above equation is the empirical risk function of NN *i*. The second term *pi* is a correlation penalty function which has the form

$$p_i(n)=(F_i(n)-F(n))\sum_{j\neq i}(F_j(n)-F(n)).$$

Here the parameter $0 \le \lambda \le 1$ is used to adjust the strength of the penalty.

4.2 Performance of Recognition

As mentioned earlier, BdSLR has been trained by 235 samples. Here, training has been continued until the error rate has reached to 0.02. As shown in Fig. 5, the error rate decreases as the number of training cycles increases and the curve becomes steady after 2500 training cycles. To gain the better performance, we have chosen 3000 training cycles for our experiment.

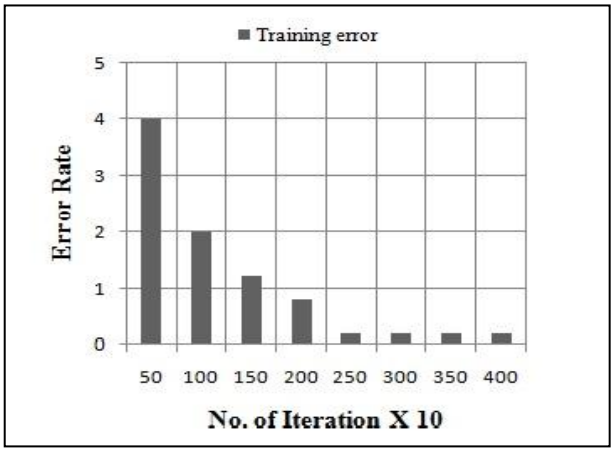


Fig. 5 Training error.

Table I shows the performance of the proposed BdSLR system. Considering λ = an arbitrary value and λ = 0, NCL has been used to develop BdSLR. Here for λ = 0, individual NNs are trained independently which is as same as standard back-propagation (BP) algorithm. The table shows that NCL produces better recognition accuracy in comparison with BP in all perspectives. Moreover, NCL learns faster than BP.

Table I, Recognition Performance of BdSLR

No. of outputs = 47	**Iterations**	**Training time (s)**		**Accuracy (%)**			
				NCL		**BP**	
No. of samples 47x5 = 235		**NCL**	**BP**	**Training**	**Test**	**Training**	**Test**
Input pixel size = 30x33	3000	152	175	98	95	86	68
		148	183	96	93	80	65
		156	179	97	91	83	71
	Average	152	179	97	93	83	68

4.3 Performance with different NNs involved in NCL

For the pursuit of better generalization, NCL has been used with different NN architectures to develop BdSLR. In the

domain of the same data space, NCL has been designed considering five and ten individual NNs. Table II shows the comparison of performance of BdSLR among these different NCL architectures. As training data space is large, NCL consisting of ten NNs with feature extraction shows the best accuracy.

Table II, Performance Comparison of BdSLR With Different Architectures

Network Type	Accuracy (%)
NCL with 05 NNs	81
NCL with 10 NNs	88
NCL with 10 NNs with feature extraction	93

4.4 Discussion

In this experiment the capturing process of the images by detecting hand gesture using the webcam is little bit cumbersome and tedious. Better performance can be achieved by considering the following steps:

- There should not be any colour in the experimental environment which conflicts with human skin colour.
- Camera pixels should be high enough for a better quality of images.
- Adequate light should be provided while capturing the images.

5. CONCLUSION

In comparison with existing works, the BdSLR system proposed in this paper is an advanced initiative to recognize BdSL. D&M people accustomed with BdSL can be benefitted by the interpretation facility of BdSLR while engaging in social vital activities. To implement BdSLR, we have exploited feature extraction [19] along with NCL [17] algorithm for training that is capable enough to perform a good recognition. BdSLR has been implemented involving different numbers of individual NNs in NCL. The experimental results show that NCL with feature extraction yields good recognition accuracy approximately 93%. BdSLR can be enhanced to learn other sign languages. A future plan of improvement is to employ BdSLR to recognize BdSL in real time.

6. REFERENCES

[1] Beatrice de Gelder, Jean Vroomen and Lucienne van der Heide, “Face recognition and lip reading in autism”, European journal of cognitive psychology, 1991, Vol. 3(1), pp. 69-86.

[2] Fudickar, S. and Nurzynska, K., “A user friendly sign language chatr” in proceedings of the conference ICL 2007.

[3] Padden Carol A. and Tom Humphries, “Deaf in America”, in Harvard University Press, 1988.

[4] Holden, E.-J., R. Owens and G. Roy, “Adaptive fuzzy expert system for sign recognition”, in proceedings of the international conference on signal and image processing, Las Vegas, USA, 2000, pp.141-146.

[5] M.B. Waldron, and S. Kim, "Isolated ASL Sign Recognition System for Deaf Persons," in IEEE Transactions on Rehabilitation Engineering, Vol. 3, No.3, 1995, pp. 261-271.

[6] H. Birk, T. B. Moeslund, and C. B. Madsen, "Real-time Recognition of Hand Alphabet Gesture Using Principal Component Analysis”, in Proceeding of Scandinavian Conference on Image Analysis, Finland, 1997.

[7] Vogler, C. and D. Metaxas, "Adapting Hidden Markov Models for ASL Recognition by Using three-Dimensional Computer Vision Methods," in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics SMC97, IEEE Computer Society: Orlando, Florida. 1997, pp. 156-161.

[8] Ahmed Shah Mashiyat, Ahmed Shah Mehadi, Kamrul Hasan Talukder, “Bangla off-line Handwritten Character Recognition Using Superimposed Matrices”, ICCIT, 2004.

[9] U. Pal, “On the development of an optical character recognition (OCR) system for printed Bangla script”, Ph.D. Thesis, Indian Statistical Institute, 1997.

[10] Bikash Chandra Karmokar, Kazi Md. Rokibul Alam and Md. Kibria Siddiquee, “An Intelligent Approach to Recognize Touchless Written Bengali Characters” International Conference on Informatics, Electronics & Vision (ICIEV), 2012, Dhaka, Bangladesh.

[11] Foez M. Rahim, Tamnun E Mursalin, Nasrin Sultana, “Intelligent Sign Language Verification System – Using Image Processing, Clustering and Neural Network Concepts” International Journal of Engineering Computer Science and Mathematics, ISSN:0976-6146, Vol.1 No.1 January-June 2010.

[12] D. Shahriar Hossain Pavel, Tanvir Mustafiz, Asif Iqbal Sarkar, M. Rokonuzzaman, “Geometrical Model Based Hand Gesture Recognition for Interpreting Bengali Sign Language Using Computer Vision”, ICCIT, 2003.

[13] Adnan Eshaque, Tarek Hamid, Shamima Rahman, M. Rokonuzzaman, "A Novel Concept of 3D Animation Based 'Intelligent Assistant' for Deaf People: for Understanding Bengali Expressions", ICCIT, 2002.

[14] Sohalia Rahman, Naureen Fatema, M.Rokonuzzaman, "Intelligent Assistants for Speech Impaired People", ICCIT, 2002.

[15] Sandberg, “Gesture Recognition using Neural Networks”, 1995.

[16] K. Murakami and H. Taguchi, “Gesture Recognition using recurrent neural networks”, in proceedings of CH191 Human factors in Computing Systems, 1991.

[17] Y. Liu, X.Yao, “Ensemble learning via negative correlation”, Neural Networks 12 (1999) pp. 1399–1404.

[18] Centre for Disability in Development (CDT), "Manual on Sign Supported BangIa," in Computer Vision and Image Understanding, 1-50, 2002.

[19] Isabelle Guyon and Andr´e Elissee, “An Introduction to Feature Extraction”, Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer, 2006.

[20] Melville and Mooney, “Creating Diverse Ensemble Classifiers to Reduce Supervision”, PhD thesis, Department of Computer Sciences, University of Texas at Austin, November 2005.

[21] Hafiz T. Hassan, Muhammad U. Khalid and Kashif Imran, “Intelligent Object and Pattern Recognition using

Ensembles in Back Propagation Neural Network", International Journal of Electrical & Computer Sciences (IJECS-IJENS) Vol. 10, No: 06.

[22] Robi Polikar, "Ensemble based systems in decision making", Article IEEE Circuits and Systems Magazines, 2006.