

# Bangla numerical sign language recognition using convolutional neural networks

F. M. Javed Mehedi Shamrat<sup>1</sup>, Sovon Chakraborty<sup>2</sup>, Md. Masum Billah<sup>3</sup>, Moumita Kabir<sup>4</sup>, Nazmus Shakib Shadin<sup>5</sup>, Silvia Sanjana<sup>6</sup>

<sup>1,3</sup>Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh

<sup>2,4</sup>Department of Computer Science and Engineering, European University of Bangladesh, Dhaka, Bangladesh

<sup>5,6</sup>Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh

---

## Article Info

### Article history:

Received Mar 31, 2021

Revised May 1, 2021

Accepted Jun 17, 2021

---

### Keywords:

Bandlet transmission

Bengali sign language

Deep CNN

Digit recognition

Image processing

---

## ABSTRACT

The amount of deaf and mute individuals on the earth is rising at an alarming rate. Bangladesh has about 2.6 million people who are unable to interact with the community using language. Hearing-impaired citizens in Bangladesh use Bangladeshi sign language (BSL) as a means of communication. In this article, we propose a new method for Bengali sign language recognition based on deep convolutional neural networks. Our framework employs convolutional neural networks (CNN) to learn from the images in our dataset and interpret hand signs from input images. Checking their collections of ten indications (we used ten sets of images with 31 distinct signs) for a total of 310 images. The proposed system takes snapshots from a video by using a webcam with applying a computer vision-based approach. After that, it compares those photos to a previously trained dataset generated with CNN and displays the Bengali numbers (০-৯). After estimating the model on our dataset, we obtained an overall accuracy of 99.8%. We want to strengthen things as far as we can to make silent contact with the majority of society as simple as probable.

*This is an open access article under the [CC BY-SA](#) license.*



---

## Corresponding Author:

F. M. Javed Mehedi Shamrat

Department of Software Engineering

Daffodil International University

102/1, Sukrabad, Mirpur Road, Dhaka 1207, Bangladesh

Email: javedmehedicom@gmail.com

---

## 1. INTRODUCTION

Sign language recognition technologies are used to identify indications of numbers, alphabets, phrases, or some other signs, such as traffic signal hand movements. Some scientists are focused on real-time symbol recognition, while others are focusing on static pictures. Artificial neural network-based approaches for real-time American sign language (ASL) term [1] and alphabet [2] recognition have recently been used. Hand sign recognition with the Microsoft sensor system [3] is used in the report to detect signals for two real-world applications: arithmetic calculation and the rock-paper-scissors game, with a mean accuracy of over 90% for ASL. Deep learning has been used to identify signs in Indian sign languages [4], [5], Arabic sign language [6], and other languages.

In the modern era, deep learning and machine learning [7]-[11] besides Bengali sign language (BSL) recognition studies have gotten a lot of attention, and various approaches for implementing a Bengali sign language recognition scheme have been suggested. Karmokar *et al.* [12] introduced a system of BSL

recognition using a neural network in 2012, with a 93 percent accuracy. They used a dataset of distinct skin tones but the same context, allowing for fast identification. Rahaman *et al.* suggested a computer-based Bengali sign language recognition system in 2014 [13]. Rahim *et al.* [14] applied the backpropagation approach of ANN to identify indications of certain popular Bengali terms in 2015. Fingertip finder algorithms were used for BSL by numerous researchers in 2015 and 2016 [15]-[17]. For the identification of Bengali Sign Language, many methods based on ANN [18] have been suggested. The precision of the majority of the techniques seems to be promising. They do, however, have drawbacks due to the usage of limited datasets, a controlled backdrop or scenario, and in certain situations, major mistakes in lighting effects or skin tone, which they monitor to prevent more complications. The approaches that do utilize neural networks [12] use a lot of preprocessing, which isn't necessarily suitable for real-time applications. Using a CNN, the proposed system recognizes only the Bangla numeric digits (CNN). It sensed the symbol with only one hand. It is split into two parts: the trained and the signs detection part.

The following section follows the same structure. This portion includes the most recent developments in Bengali Signs Identification. Section ii describes the research method for modeling the whole system. The iii section looks at the results of the system that has been established. Section iv ends with a theory, flaws, and proposals for potential study.

## 2. METHODOLOGY

This proposed system takes the image from a camera and then preprocesses the images. To identify the image structure firstly the system applies bandlet transformation on images, then it applies the logarithm replace technique to control the extra light effects on images. To fixed up the low-resolution image, the system also used the D-LBP technique. Then it measures the picture dimension [19]. As dimension calculation is done, it segments the skin color from the picture and transforms it into a binary image, removes the broad blobs from the image, and uses deep learning techniques to construct a qualified dataset [19]. Classifies with the qualified dataset after the recognizer section and identifies the numerical digit of the Bangla. In this paper, we used deep learning methods for Bengali numerical sign detection. In Figure 1 we have displayed the entire proposed model diagram.

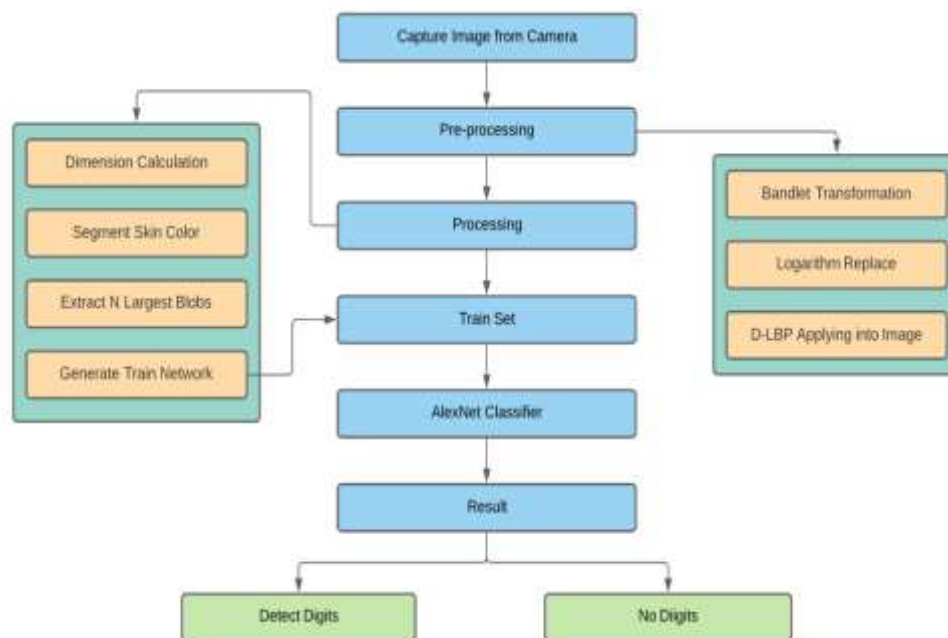


Figure 1. Proposed model diagram

### 2.1. Data collection

We collect images from the video frame using the webcam. To capture the image, we used a Xiaomi Vidlok W77 webcam. We use Ryzen 3900X Core i3 3.90 (base speed) GHz PC with 32 GB RAM for running the system. The dataset consists of ten hand signs (a total of 310 pictures), with each signing class consisting

of 31 images captured at various lengths, orientations, and intensities. Bengali numerical sign language is depicted by the photos. We used the data for the train and test our model to recognition Bengali numerical signs. In Figure 2 we showed the dataset images sample.



Figure 2. Datasets images samples

**2.2. Data preprocessing**

**2.2.1. Bandlet transformation**

A 3x3 transformation matrix is needed for perspective conversion. Well after the transition, straight lines would remain straight. We'll need four points from the input image and four points from the output image to find this transition matrix. Three of the four points could not be in sync with one another. Then use the method cv2.getPerspectiveTransform to locate the transition matrix. For this 3x3 transition matrix, add cv2.warpPerspective. In Figure 3 we exhibited the output of bandlet transformation after implementation.

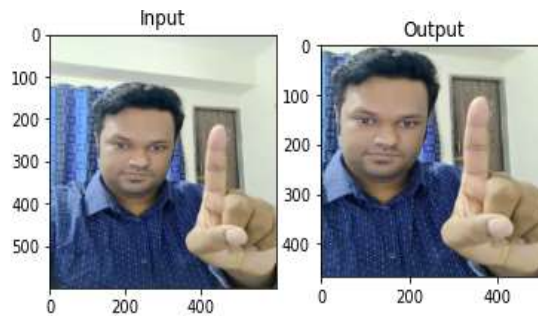


Figure 3. The output of bandlet transformation

**2.2.2. Logarithm replace**

Each pixel's value is replaced with its logarithm value in a log transformation. As shown in (1) can be used to describe log transformations.

$$S = C \log (r + 1) \tag{1}$$

The output and input image pixel values are s and r, respectively, and c is a constant. Since the input image has a pixel intensity of 0, log (0) equals infinity, and each of the input image's pixel values is given a sum of one. As a result, 1 is applied to the minimum amount to render it at least 1.

Expands the image's dark pixels thus compressing the image's lighter pixels. The combination of maximum and minimum amplitude values is referred to as dynamic range. Lower values are removed where

the image's dynamic spectrum exceeds that of the viewing system. We use log transform to solve this issue. The log transformation compresses the dynamic range first, then upscales the picture to the display device's dynamic range. Lower values are boosted in this way, resulting in a far more detailed picture. In Figure 4 we showed the image after applying logarithm replace.

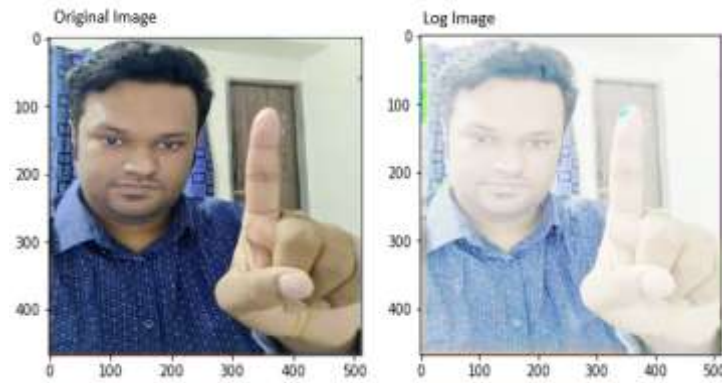


Figure 4. After applying logarithm replacement

**2.2.3. D-LBP**

Two complementary measures: local spatial patterns and grayscale comparison, according to the LBP operator, will characterize two-dimensional surface textures. The original LBP operator produces labels for image pixels by thresholding the 3x3 neighborhood of each pixel with the center value and treating the result as a binary integer. The texture descriptor is then the histogram of these 28 = 256 different marks. This operator, when coupled with a simple local contrast measure, achieved excellent unsupervised texture segmentation results. The LBP operator is denoted by the following notation: Ru<sup>2</sup>; LBPP; The subscript indicates that the operator is being used in the (P, R) neighborhood. The superscript u<sup>2</sup> denotes using only standardized patterns and labeling the residual patterns with a single dot. After obtaining the LBP labeled picture fl (x, y), the LBP histogram can be identified as (1) and in Figure 5 we showed the output of D-LBP after implementation.

$$H_i = \sum_{x,y} I\{f_1(x,y) = i\}, i = 0, \dots, n - 1 \tag{2}$$

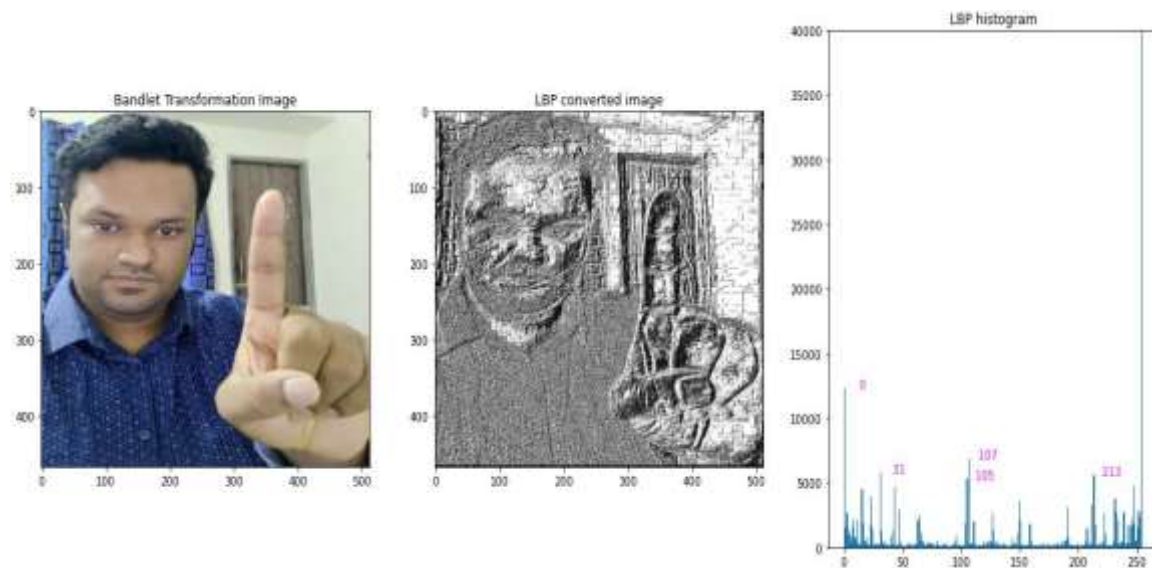


Figure 5. After applying D-LBP on the image

## 2.3. Proposed convolution neural network (CNN) model

### 2.3.1. Dimension calculation

Joint Photographic Experts Group (JPEG) is one of the most common image compression methods. The headers (first few bytes) in the majority of file formats provide valuable details about the file. JPEG headers, for example, provide details such as height, weight, color depth (grayscale or RGB), and so on. We find the resolution of a JPEG image without utilizing any external libraries in this software by reading these headers.

### 2.3.2. Segment skin color

This method is only used to detect the color of a human being, such as a hand or a face [20]. It reads the RGB picture and then measures the image's proportions. It senses the skin color after transforming the RGB picture to YCbCr [20], [21]. In Figure 6, the skin color is segmented from a different background.



Figure 6. Skin color segmentation

### 2.3.3. Extract N largest blobs

This role assists in the identification of broad blobs of an item in a binary picture. It takes the blob region after getting all of the blob properties from the binary picture. It also detects the largest entity from the chosen boundaries and establishes boundaries in the original picture. This method employs a custom feature to remove the N largest blobs from the binary picture seen in Figure 7.

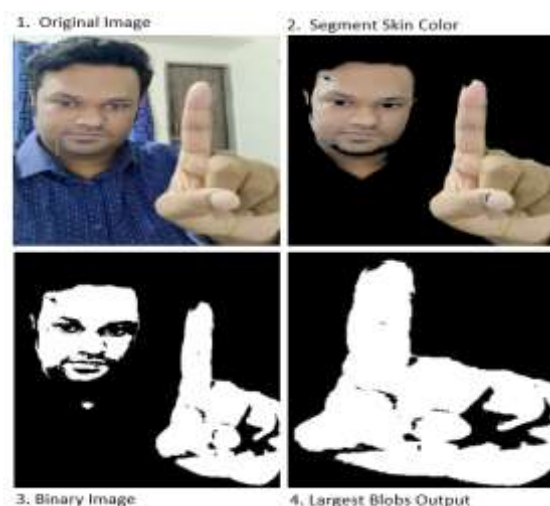


Figure 7. Extract the largest blobs from the binary image

### 2.3.4. Generate train network

This method even used a webcam to build the qualified network but preprocessed the webcam images utilizing the same process as mentioned above. This implies that the trained dataset was generated using the largest binary blob of an image. To begin, multiple photos of the same sign where needed. For each



symbol, approximately 31 photos were taken [22]. To create the divisions, the suggested method stores various sign photos into separate folders. This machine was given 31 photos to create a qualified network to improve accuracy. After taking pictures, this method generates a qualified dataset utilizing a convolutional neural network [23] methodology. For the educated network, the proposed framework used the Alex net of the deep learning process. This framework, in particular, employs the convolutional network's Transfer Learning process. For the convolutional layer data, it resized the binary picture blob into a [28x28]. Then, using a random selection process, it generated a qualified dataset and a research dataset. The term "qualified dataset" refers to a classification database, whereas "test dataset" refers to a classification database that is used in conjunction with the trained dataset. For the qualified dataset, the machine used 750 images and for the evaluation dataset, it used 250 images. Initial learn rate = 0.0001, Max Epochs = 20, and Mini batch size = 64 were used.

#### 2.4. Classifying trained network

This system loads the training dataset at first, then the sign has been taken when the figure window is opened. This system can detect Bangla numerical signs only using one hand. The proposed system has used the webcam image for recognition of the sign. When completed the largest binary blob image then it does classify with the custom training dataset that was first created. The proposed model used AlexNet classifier to detect the signs. Finally, it gives the outcome that is shown in Figure 8.

When the figure portal is raised, this device loads the training dataset first, then the symbol. Only one hand is intended to identify a Bangla numerical symbol. For the identification of the symbol, the proposed system used a webcam picture. It begins by preprocessing the picture with the technique mentioned in the 2.1, 2.2, and 2.3 sections above. When the largest binary blob picture is over, it can be classified using the custom training dataset that was generated at the beginning. Finally, it produces the result (০, ১, ২, ৩, ৪, ৫, ৬, ৭, ৮, ৯) indicated in Figure 8.



Figure 8. Bengali numerical sign language recognition

### 3. RESULTS AND DISCUSSION

To develop the area of sign language interpretation, the proposed model has used a variety of approaches. We developed the proposed model to improve the technique on Bengali numerical sign detection. The accuracy of each sign's identification is calculated by this method for experimental results. We measured the performance of two models using precision, recall, true negative rate, and accuracy after completing the training and testing phase. The formulas that we used are as shown in (3)-(4):

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{5}$$

$$TN Rate = \frac{TN}{TN+FP} \tag{6}$$

Table 1 shows how the system uses 25 tests on hand signs for each digit, with 9 signs correctly recognized at the maximum peak and 1 sign recognition level is comparatively low. The precision is achieved 100% for each sign (○-৯), besides Recall also gain 100% for 9 signs, except one Bengali sign (৩). For all signs, True negative gain 0%, it carried a huge impact on overall accuracy. Finally, our model for the Bengali numerical sign’s language detection achieved overall 99.8% accuracy. Figure 9 depicted the value of every test result (Precision, Recall, True Negative, Accuracy) for each sign. Figure 10 shows the accuracy evaluation of ten separate numerical Bengali signs.

Table 1. Outcomes for each numerical Bengali sign

Bangla Sign	Precision	Recall	True Negative	Accuracy
○	100%	100%	0%	100%
১	100%	100%	0%	100%
২	100%	100%	0%	100%
৩	100%	98%	0%	98%
৪	100%	100%	0%	100%
৫	100%	100%	0%	100%
৬	100%	100%	0%	100%
৭	100%	100%	0%	100%
৮	100%	100%	0%	100%
৯	100%	100%	0%	100%

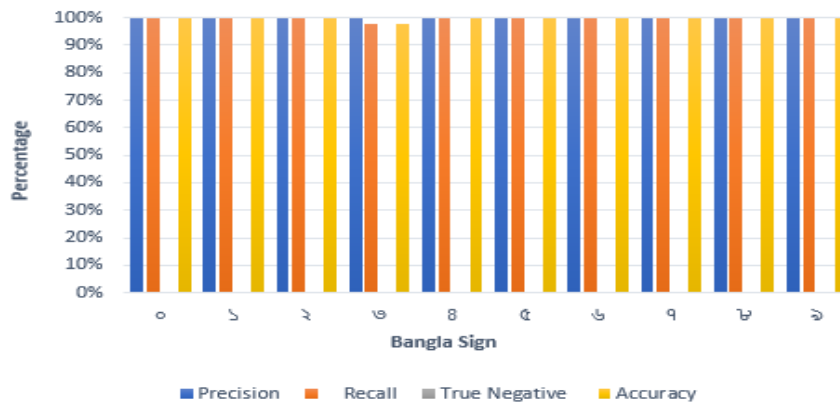


Figure 9. Precision, recall, true negative, accuracy graph for each sign

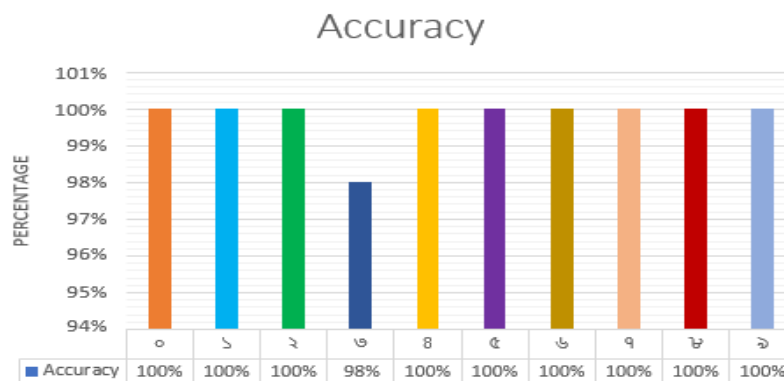


Figure 10. Comparing the accuracy of ten different Bengali numerical signs

In Table 2 we compared some previous works, which are on Bengali and different language signs detection. The methodologies they have been used, article publishing year, and accuracy of their methodologies, we depicted in Table 2. They have many similarities with our proposed system, a few methodologies they have been used including CNN, artificial neural network, support vector machine, Convex hull method, 3D convolutional neural network (CNN), key maximum curvature points, stroke sub-segment vectors, equivolumetric partition, Gabor filter, kernel PCA, morphological operation, SVM. Most of the models changed the images into grayscale and directly apply some skin detection algorithms to detect the shape of the hand. On the other hand, our proposed model preprocesses the images as well to get the highest accuracy on test and training phases, then we apply correspondingly dimension calculation, segment skin color, extract N largest blobs, and trained our network at the processing phase. Finally, we applied the AlexNet classifier to classify the images. Fortunately, our model achieved the highest accuracy (99.8%) among the existing related works.

Table 2. Comparative analysis on previous related works

	Year	Methods	Overall Accuracy
G. A. Raouf et al. [24]	2018	CNN	92.88%
Chowdhury et al. [25]	2017	Artificial neural network, Support vector machine, Convex hull method	84.11%
El-badawy et al. [26]	2017	3D Convolutional Neural Network (CNN)	98%
Kaimal [27]	2017	Key Maximum Curvature Points, Stroke Sub-segment Vectors, Equivolumetric Partition	85.4%
Uddin et al. [28]	2016	Gabor filter, Kernel PCA, Morphological operation, SVM	99.5%
Our Proposed Model	2021	CNN	99.8%

#### 4. CONCLUSION

Machine learning and deep learning doing a great role in recent times in image processing, the medical sector and various purposes. To identify Bengali numerical signs from the input images, our current model takes around a minute. Our aim is to render this process as smooth, fast, and real-time as possible. The dataset that we have used for training, testing, and validation contains the Bengali sign (০-৯). We intend to complete our dataset by including all Bengali alphabets. We haven't incorporated gesture recognition in our research, nor have we included the recognition of terms and full sentences, so we're excited to embed this into our model. We intend to integrate our framework into a portable computer and IoT media, specifically smartphones, and operate it as an application after implementing both of these methods. It would be conveniently available to all people in this manner, assisting us in the incorporation of such technologies into our culture.

#### REFERENCES

- [1] N Sarawate, M Chan LEU, and Cemil OZ, "A real-time American Sign Language word recognition system based on neural networks and a probabilistic model," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 23, no. 1, pp 2107-2123, 2013, doi: 10.3906/elk-1303-167.
- [2] B Garcia and S Alarcon Viesca, "Real-time American Sign Language Recognition with Convolutional Neural Networks," *Convolutional Neural Networks for Visual Recognition*, vol. 2, pp. 225-232, 2012.
- [3] Z Z. Ren, J. Yuan, J. Meng and Z. Zhang, "Robust Part-Based Hand Gesture Recognition Using Kinect Sensor," in *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1110-1120, Aug. 2013, doi: 10.1109/TMM.2013.2246148.
- [4] S. Padmavathi, M. S. Saipreethy, and V. Valliammai, "Indian Sign Language Character Recognition using Neural Networks," *IJCA Special Issue on Recent Trends in Pattern Recognition and Image Analysis RTPRIA*, vol. 1, pp: 40-45, 2013.
- [5] N. Priyadharsini and N. Rajeswari, "Sign Language Recognition Using Convolutional Neural Networks," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 5, no. 6, pp. 2321-8169, 2017.
- [6] O Al-Jarrah and Alaa Halawani, "Recognition of gestures in Arabic Sign Language using neural networks," *Artificial Intelligence and Soft Computing*, vol. 133, no. 1-2, pp. 28-30, 2006, doi: 10.1016/S0004-3702(01)00141-2.
- [7] F. M. Javed Mehedi Shamrat, Z. Tasnim, P. Ghosh, A. Majumder and M. Z. Hasan, "Personalization of Job Circular Announcement to Applicants Using Decision Tree Classification Algorithm," *2020 IEEE International Conference for Innovation in Technology (INOCON)*, Bangluru, India, 2020, pp. 1-5, doi: 10.1109/INOCON50539.2020.9298253.



- [8] F. M. Javed Mehedi Shamrat, P. Ghosh, M. H. Sadek, M. A. Kazi and S. Shultana, "Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease," *2020 IEEE International Conference for Innovation in Technology (INOCON)*, Bangluru, India, 2020, pp. 1-7, doi: 10.1109/INOCON50539.2020.9298026.
- [9] P. Ghosh, F. M. Javed Mehedi Shamrat, S. Shultana, S. Afrin, A. A. Anjum and A. A. Khan, "Optimization of Prediction Method of Chronic Kidney Disease Using Machine Learning Algorithm," *2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, Bangkok, Thailand, 2020, pp. 1-6, doi: 10.1109/iSAI-NLP51646.2020.9376787.
- [10] P. Ghosh *et al.*, "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques," in *IEEE Access*, vol. 9, pp. 19304-19326, 2021, doi: 10.1109/ACCESS.2021.3053759.
- [11] F.M. Javed Mehedi Shamrat, Asaduzzaman, A. K. M. Sazzadur Rahman, R. T. H. Tusher, and Zarrin Tasnim, "A Comparative Analysis of Parkinson Disease Prediction Using Machine Learning Approaches," *International Journal of Scientific & Technology Research*, vol. 8, no. 11, pp: 2576-2580, 2019.
- [12] B. C. Karmokar, K. M. R. Alam, and M. K. Siddiquee, "Bangladeshi sign language recognition employing neural network ensemble," *International Journal of Computer Applications (IJCA)*, vol. 58, no. 16, pp. 43-46, November 2012, doi: 10.5120/9370-3846.
- [13] 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)*, 2014, pp. 192-197, doi: 10.1109/ICCITechn.2014.7073150.
- [14] Md. Abdur Rahim, Tanzillah Wahid, and Md. Khaled Ben Islam, "Visual Recognition of Bengali Sign Language using Artificial Neural Network," *International Journal of Computer Applications*, vol. 94, no. 17, pp. 1-5, 2014, doi: 10.5120/16448-0572.
- [15] A. M. Jarman, S. Arshad, N. Alam, and M. J. Islam, "An automated bengalisign language recognition system basedon fingertip finder algorithm," *International Journal of Electronics & Informatics*, vol. 4, no. 1, pp. 1-10, 2015.
- [16] M. Hasan, T. H. Sajib and M. Dey, "A machine learning based approach for the detection and recognition of Bangla sign language," *2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec)*, 2016, pp. 1-5, doi: 10.1109/MEDITEC.2016.7835387.
- [17] S. T. Ahmed and M. A. H. Akhand, "Bangladeshi Sign Language Recognition using fingertip position," *2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec)*, 2016, pp. 1-5, doi: 10.1109/MEDITEC.2016.7835364.
- [18] Golam Kayas, Sakhawat Hossain, and Himeland MahrajHasan, "Automatic Recognition of Bangla Sign Language Using Artificial Neural Networks (ANNS) For Deaf And Dumb to Bridge The Communication Gap," *International Journal of Recent Advances in Multidisciplinary Research*, vol. 3, no. 7, pp.1649-1654, 2016.
- [19] M. T. Hoque, M. Rifat-Ut-Tauwab, M. F. Kabir, F. Sarker, M. N. Huda and K. Abdullah-Al-Mamun, "Automated Bangla sign language translation system: Prospects, limitations and applications," *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, 2016, pp. 856-862, doi: 10.1109/ICIEV.2016.7760123.
- [20] K. B. Shaik, P. Ganesan, V. Kalist, B. S. Sathish, and J. M. M. Jenitha, "Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space," *Procedia Computer Science*, vol. 57, pp. 41-48, 2015, doi: 10.1016/j.procs.2015.07.362.
- [21] V. Bhamre, R. Sreemathy and H. Dhumal, "Vision based hand gesture recognition using eccentric approach for human computer interaction," *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2014, pp. 949-953, doi: 10.1109/ICACCI.2014.6968545.
- [22] G. A. Rao, K. Syamala, P. V. V. Kishore and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition," *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*, 2018, pp. 194-197, doi: 10.1109/SPACES.2018.8316344.
- [23] Jonathan Long, Evan Shelhamer, Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, pp. 3431-3440.
- [24] M. J. Hossein and M. Sabbir Ejaz, "Recognition of Bengali Sign Language using Novel Deep Convolutional Neural Network," *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, 2020, pp. 1-5, doi: 10.1109/STI50764.2020.9350418.
- [25] A. R. Chowdhury, A. Biswas, S. M. F. Hasan, T. M. Rahman and J. Uddin, "Bengali Sign language to text conversion using artificial neural network and support vector machine," *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, 2017, pp. 1-4, doi: 10.1109/EICT.2017.8275248.
- [26] M. ElBadawy, A. S. Elons, H. A. Shedeed and M. F. Tolba, "Arabic sign language recognition with 3D convolutional neural networks," *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 2017, pp. 66-71, doi: 10.1109/INTELCIS.2017.8260028.
- [27] M. M. Hasan, A. Y. Srizon, A. Sayeed and M. A. M. Hasan, "Classification of Sign Language Characters by Applying a Deep Convolutional Neural Network," *2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*, 2020, pp. 434-438, doi: 10.1109/ICAICT51780.2020.9333456.
- [28] 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 (ICISSET)*, 2016, pp. 1-4, doi: 10.1109/ICISSET.2016.7856479.