Front. Comput. Sci., 2020, 14(3): 143302
https://doi.org/10.1007/s11704-018-7253-3

RESEARCH ARTICLE

Bangla language modeling algorithm for automatic recognition of hand-sign-spelled Bangla sign language

Muhammad Aminur RAHAMAN (✉), Mahmood JASIM, Md. Haider ALI, Md. HASANUZZAMAN

Department of Computer Science and Engineering, University of Dhaka, Dhaka-1000, Bangladesh

Abstract Because of using traditional hand-sign segmentation and classification algorithm, many diversities of Bangla language including joint-letters, dependent vowels etc. and representing 51 Bangla written characters by using only 36 hand-signs, continuous hand-sign-spelled Bangla sign language (BdSL) recognition is challenging. This paper presents a Bangla language modeling algorithm for automatic recognition of hand-sign-spelled Bangla sign language which consists of two phases. First phase is designed for hand-sign classification and the second phase is designed for Bangla language modeling algorithm (BLMA) for automatic recognition of hand-sign-spelled Bangla sign language. In first phase, we have proposed two step classifiers for hand-sign classification using normalized outer boundary vector (NOBV) and window-grid vector (WGV) by calculating maximum inter correlation coefficient (ICC) between test feature vector and pre-trained feature vectors. At first, the system classifies hand-signs using NOBV. If classification score does not satisfy specific threshold then another classifier based on WGV is used. The system is trained using 5,200 images and tested using another $(5,200 \times 6)$ images of 52 hand-signs from 10 signers in 6 different challenging environments achieving mean accuracy of 95.83% for classification with the computational cost of 39.972 milliseconds per frame. In the Second Phase, we have proposed Bangla language modeling algorithm (BLMA) which discovers all "hidden characters" based on "recognized characters" from 52 hand-signs of BdSL to make any Bangla words, composite numerals and sentences in BdSL with no training, only based on the result of first phase. To the best of our knowledge, the proposed system is the first system in BdSL designed on automatic recognition of hand-sign-spelled BdSL for large lexicon. The system is tested for BLMA using hand-sign-spelled 500 words, 100 composite numerals and 80 sentences in BdSL achieving mean accuracy of 93.50%, 95.50% and 90.50% respectively.

Keywords Bangla sign language (BdSL), hand-sign, classification, Bangla language modeling rules (BLMR), Bangla language modeling algorithm (BLMA)

1 Introduction

Like the spoken language, sign language (SL) is a separate language with its own grammar and rules which is used by speech and/or hearing impaired people to communicate with non-sign people and themselves. SL is a part of the cultural, social, historical and religious heritage. Approximately 7% of the world's populations use SL as their first language [1, 2]. Almost 2.6 million sign people are living in Bangladesh [3] and they use Bangla sign language (BdSL) to communicate.

The recognition of continuous, natural signing in BdSL is challenging, in terms of both video analysis and linguistics. Nowadays a significant goal is to achieve the real-time SL recognition in naturalistic scenarios where occlusions, illumination changes and cluttered background are handled. Computer vision based image processing methods present a backward compatible, user friendly and robust solution to the

Received July 18, 2017; accepted August 24, 2018

E-mail: aminur.wg@gmail.com

SL recognition problem in real time [4]. So, the demand of computer vision based real-time continuous SL recognition research is increasing rapidly.

The Bangla sign language dictionary [5] uses 36 (6 vowels and 30 consonants as shown in Fig. 1(a) and Fig. 1(b)) two-handed Bangla sign alphabet from 51 Bangla written alphabet based on pronunciation, and 10 basic numerals (0 to 9) as shown in Fig. 1(c). But, most of this word and sentence level signs are gestures. In BdSL, about 5,000 set of gestures [6] are used to express sign words and sentences which are mostly impossible to memorize for a human.

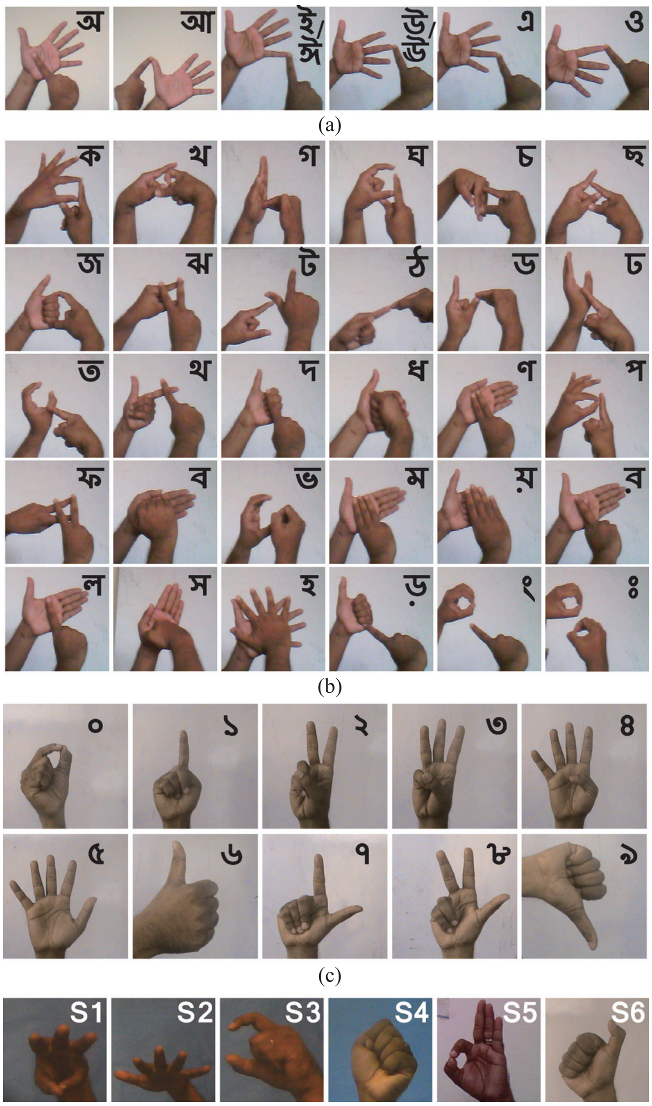


Fig. 1 Example postures of 52 hand-signs in BdSL. (a) Example postures of BdSL vowel signs; (b) example postures of BdSL consonant signs; (c) example postures of BdSL numeral signs; (d) example postures of special signs

To establish communication between sign and non-sign people, there is a need to develop BLMA, abbreviation for Bangla language modeling algorithm for automatic recognition of hand-sign-spelled BdSL in real-time by discovering "hidden caracters" that are not in BdSL (described indetails in Section 3) using only Bangla alphabet (36 letters) and basic numerals (0 to 9) which will be able to make any words, composite numerals and sentences by hand-sign-spelling. Memorizing only 36 alphabet and 10 numeral signs is possible for anyone.

The proposed system is the extension of our previous systm [7, 8]. Our previous systms [7] was developed only for hand-signs segmentation and classification using fuzzy rule based RGB (FRB-RGB) model and window-grid vector (WGV) analysis. But the proposed system is designed for automatic recognition of hand-sign-spelled BdSL in real-time which contains two phases as shown in Fig. 2. In first phase, the system is designed for hand-sign classification e.g., individual sign letters classification trained with 52 hand-signs (6 vowels+30 consonants+10 numerals+6 special signs) in BdSL as shown in Fig. 1 using our previously used NOBV or vector contours (VC) [8] and WGV [7] in combination. Li et al. [9] developed a static gesture recognition system based on high-level features which was tested by hand digit gestures of 0–9 accurately. Dong et al. [10] proposed a descriptor named holons visual representation (HVR) which was a derivative mutational self-contained combination of global and local information. Our previous system [8] using NOBV, e.g., vector contours (VC), could not recognize among the hand-signs where outer contours are similar but inner shapes are different. So, in this paper, we have combined the rotation, translation and scale invariant feature vector WGV [7] which includes not only outer contour but also the inner shape to overcome the limitation of our previous system [8]. In the proposed system, we have significantly improved the classification module with the proposed two-steps classifier based on NOBV [8] and WGV [7]. Dong et al. [11] proposed a discriminative light unsupervised learning network (DLUN) to counter the image classification challenge. Carcia-Ceja and Brena [12] proposed an improved three-stage classifier for activity recognition and Lee et al. [13] proposed a person-specific saliency system for the recognition of dynamic gestures using two-stage classifiers based on different features. But our proposed two-steps classifier is simple and time efficient. By combining the two features NOBV and WGV in our proposed system, we have increased the recognition accuracy than previous systems in cluttered and dynamic background with illumination variation. Here, we have used 6 special signs as shown in Fig. 1(d) to implement the second phase. In second phase, the system is designed as BLMA for

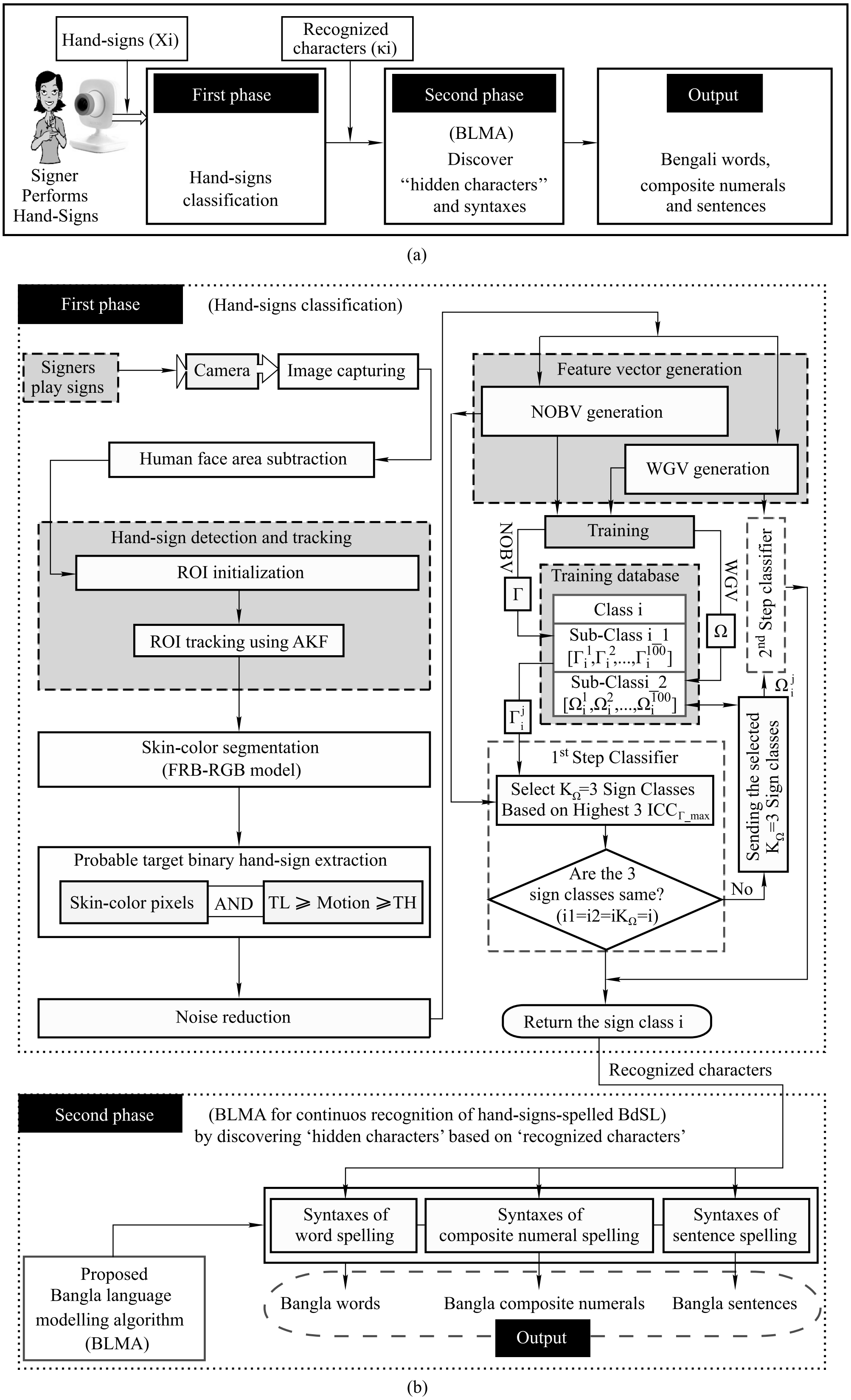


Fig. 2 Architecture of the proposed system, (a) block diagram and (b) details view

automatic recognition of hand-sign-spelled words, composite numerals and sentences for large lexicon in BdSL with no training, only based on the result of first phase (hand-signs classification).

BLMA is mainly a part of the natural language understanding (NLU). Mills et al. [14, 15] and Santoni and Pourabbas [16] worked on NLU. Because of using only 36 hand-signs to represent 51 Bangla written characters and other diversities of Bengla language including joint-letters, dependent vowels etc., implementation of BLMA is challenging. In this paper, the implementation of the Second Phase for the proposed BLMA is the unique contribution for hand-sign-spelled Bangla sign language recognition (BdSLR).

The moving hand-sign detection and tracking in cluttered and dynamic background is challenging and most essential part in vision based hand-signs classification. More recently, various approaches have been proposed to detect the moving objects [17, 18]. Dong et al. [19] used E-GrabCut for video object segmentation. Their another research [20] was done for nature image segmentation. Li et al. [21] proposed and developed an approach to segment streaming video in data noises and/or corruptions affected environment. A real-time moving hand-signs detection system was developed by Chen et al. [22], using motion, skin-color and edge detection. Alon et al. [23] proposed hand-signs detection performed in front of moving and cluttered background by combining only skin-color and motion cues. Another system developed by Asaari et al. [24], presents an efficient method for hand detection and tracking using integration of adaptive Kalman filter (AKF) and Eigen hand method using skin-color and motion cues as the main tracking features. The system developed by Khaled et al. [25] isolates the moving hand from the whole image by subtracting the static background with continuous update. Instead of ROI processing, these systems process whole image to isolate hand-sign which increases the computational cost (CC). Skin-color based hand segmentation and detection is easy and invariant to different types of hand postures, scale, translation, and rotation changes [26]. But skin-color based segmentation method does not perform well under various illumination conditions, cluttered and dynamic backgrounds [27]. For this reason, in this paper, we propose a solution to detect and track the hand-signs. After initialization of the ROI [27] adaptive Kalman filter (AKF) [24, 28–30] is applied to track the ROIs considering all the hand-signs are performed within the ROIs with cluttered and dynamic background. Then the system extracts hand-signs as binary image by segmented skin-color using a robust fuzzy rule based RGB (FRB-RGB) model from the ROI with specific motion which is describe in details in our previous system [7].

Notable researches on Bangla sign language recognition (BdSLR) [27, 31–34] have been done in the last few years. But most of the BdSLR systems were developed for only sign alphabet and/or numbers recognition. As in the case of several computer vision tasks and deep learning has also recently irrupted in SL recognition, achieving outstanding results [35, 36]. Asadi-Aghbolaghi et al. [37] collected and reviewed all deep learning methods for gesture recognition including their highlighting features, and advantages and challenges. Liu et al. [38] and Varol et al. [39] used 3D filters in the convolutional layers of their deep learning model. Zhu et al. [40] used pyramidal 3d convolutional networks for large-scale isolated gesture recognition using RGB depth data. Wang et al. [41] used depth data of 2D networks for gesture recognition including dynamic depth image, dynamic depth normal image and dynamic depth motion normal image. Xu [42] developed a real-time hand gesture recognition and human-computer interaction system using convolutional neural network (CNN) classifier which are the state-of-the-art in these research area.

Our previous system [43] was developed for Bangla sign words recognition using 18 sign words achieving recognition accuracy of 90.11%. But the system was applicable for only word gestures recognition instead of continuous hand-sign-spelled words recognition. For large lexicon based BdSL recognition, the system needs to train for each word gestures. Another system was developed by Park et al. [44] for Korean finger-spelling recognition with similar ideas. Kane and Khanna [45] developed a system for finger-spelling recognition using depth sensors. The system was tested against one-handed American sign language (ASL), NTU hand digit and two-handed Indian sign language (ISL) with 94.1% accuracy. But these finger-spelling recognition systems are mainly character sign recognition systems. The syntaxes to make words and/or sentences using hand-signs-spelling were not used in these systems. Fang et al. [46] developed a continuous Chinese sign language (ChSL) recognition system with large vocabulary. The system was tested using 5113 Chinese signs/sentences and obtained an average accuracy of 91.9%. Liwicki and Everingham [47] developed an automatic recognition of finger-spelled words in British sign language (BSL). The system was tested using 1,000 low quality webcam videos of 100 words achieving with 98.9% accuracy. But the system is signer dependent and does not perform well in cluttered background. More recently, Koller et al. [48] developed ASL recognition using statistical approach handling multiple signers for a large vocabulary. The system performs well with large vocabulary databases signer independently. But the system was designed and tested for ASL gesture recognition in plain background, not for finger-spelled words. Some systems were developed for large lexicon finger-spelled or hand-sign-spelled words recognition in various SL such as [47] but this can not be used for Bangla language as the structure of Bangla words and/or sentences is much different from other languages. To the best of our knowledge, the proposed system is the first system in BdSL designed on automatic recognition of hand-sign-spelled BdSL for large lexicon using BLMA which is able to make any Bangla words, composite numerals and sentences from only 52 hand-signs.

The main contributions of the proposed system can be summarized as (1) Hand-signs classification (a two-steps classification technique is proposed based on WGV and NOBV for hand-signs classification instead of traditional classifier to achieve high accuracy and reduced computational cost); and (2) We have proposed Bangla language modeling algorithm (BLMA) to interpret the hand-sign-spelled BdSL into Bangla words, composite numerals and sentences for large lexicon.

The rest of the paper is organized as follows. Section 2 describes the hand-signs classification. Section 3 describes the proposed Bangla language modeling algorithm (BLMA) for automatic recognition of hand-sign-spelled BdSL by discovering 'hidden characters' that are not in BdSL. Section 4 presents the experimental results with discussion. Finally, the paper is concluded in Section 5.

2 First phase: hand-signs classification

In this section, we describe the first phase of the proposed system in which the system classifies the hand-signs of individual characters in BdSL. Figure 2 presents the architecture of the proposed system. Although the first phase is already implemented in our previous systems [7, 8, 27], we have improved only the classifier in this phase. The system captures image sequence denoted by $I^m_{rgb}(x, y)$ by using a USB or CCD camera, where, m represents the sequence number. After ROI generation by detecting hand-signs using Haar classifier, the system tracks the ROI using AKF [24, 28–30]. Face area subtraction, Hand-signs detection and ROI generation, skin-color segmentation using FRB-RGB model, probable binary hand-signs extraction, and noise removal are described in details in our previous system [7]. In this paper, we have proposed and implemented a two-step classifier based on NOBV and WGV. NOBV was used in our previous system [8] in terms of vector contour (VC) using complex number "$a + ib$" representation. WGV generation process is described in detail in our previous system [7]. After extracting the feature vectors NOBV and WGV, the system is trained for each sign class i ($i = 1, 2, 3, \ldots, n$ where, $n = 52$ is the number of sign classes) using the NOBV (Γ^j_i) and WGV (Ω^j_i) as Eq. (1) and Eq. (2) where, $j = 1, 2, 3, \ldots, 100$ respectively from 10 different signers. The resulted training images for NOBV and WGV for each sign class, i are $(10 \times 10) + (10 \times 10) = 200$.

$$\begin{aligned}\Gamma^j_i = {} & [\Gamma^1_1, \Gamma^2_1, \Gamma^3_1, \ldots, \Gamma^{100}_1][\Gamma^1_2, \Gamma^2_2, \Gamma^3_2, \ldots, \Gamma^{100}_2] \\ & [\Gamma^1_3, \Gamma^2_3, \Gamma^3_3, \ldots, \Gamma^{100}_3] \cdots [\Gamma^1_n, \Gamma^2_n, \Gamma^3_n, \ldots, \Gamma^{100}_n],\end{aligned} \tag{1}$$

$$\begin{aligned}\Omega^j_i = {} & [\Omega^1_1, \Omega^2_1, \Omega^3_1, \ldots, \Omega^{100}_1][\Omega^1_2, \Omega^2_2, \Omega^3_2, \ldots, \Omega^{100}_2] \\ & [\Omega^1_3, \Omega^2_3, \Omega^3_3, \ldots, \Omega^{100}_3] \cdots [\Omega^1_n, \Omega^2_n, \Omega^3_n, \ldots, \Omega^{100}_n].\end{aligned} \tag{2}$$

In the training phase, the system combines two feature vectors NOBV(Γ^j_i) and WGV (Ω^j_i) for each sign class, i, which is represented by Eq. (3).

$$\xi^j_i = [\Gamma^j_i, \Omega^j_i]. \tag{3}$$

Here, NOBV (Γ^j_i) and WGV (Ω^j_i) are stored in two separate sub-classes (sub-class i_1 and sub-class i_2 respectively) within each sign class, i. The structure of the combined feature vector $\xi^j_i = [\Gamma^j_i, \Omega^j_i]$ is shown in Fig. 3.

The combined feature vector ξ^j_i of the 52 input hand-signs (অ, আ, ই, উ, এ, ও, ক, খ, গ, ঘ, চ, ছ, জ, ঝ, ট, ঠ, ড, ঢ, ত, থ,

Class 1	Class 2	...	Class n
Sub-Class 1_1 $[\Gamma^1_1, \Gamma^2_1, \Gamma^3_1, \ldots, \Gamma^{100}_1]$	Sub-Class 2_1 $[\Gamma^1_2, \Gamma^2_2, \Gamma^3_2, \ldots, \Gamma^{100}_2]$	...	Sub-Class n_1 $[\Gamma^1_n, \Gamma^2_n, \Gamma^3_n, \ldots, \Gamma^{100}_n]$
Sub-Class 1_2 $[\Omega^1_1, \Omega^2_1, \Omega^3_1, \ldots, \Omega^{100}_1]$	Sub-Class 2_2 $[\Omega^1_2, \Omega^2_2, \Omega^3_2, \ldots, \Omega^{100}_2]$		Sub-Class n_2 $[\Omega^1_n, \Omega^2_n, \Omega^3_n, \ldots, \Omega^{100}_n]$

Fig. 3 Example structures of the combined feature vector

দ, ধ, ণ, প, ফ, ব, ভ, ম, য়, র, ল, স, হ, ড়, ং, ঃ, ০, ১, ২, ৩, ৪, ৫, ৬, ৭, ৮, ৯, S1, S2, S3, S4, S5, S6) are to be assigned to the sign class labels $i = 1, 2, 3, \ldots, 52$ respectively in the training database. Where, each class label contains $j = 100$ NOBVs and $j = 100$ WGVs.

The size of each combined feature vector $\xi_i^j = [\Gamma_i^j, \Omega_i^j]$ for a single sign is $[K_\Gamma + M_\Omega] = [50 + 25] = 75$. Hence the size of $\xi_i^j = [\Gamma_i^j, \Omega_i^j]$ for a single sign class, i will be $[50 + 25] \times 100 \times 1 = 7500$ where, j=100 and i=1. The resulted size of the combined feature vector ξ_i^j for 52 hand-signs is $[K_\Gamma + M_\Omega] \times J \times i = [50 + 25] \times 100 \times 52 = 390,000$ where according to the previous system [27] the size of the combined feature vector considered as $(M \times N) \times j \times i = (150 \times 150) \times 100 \times 52 = 117,000,000$ which was more than 300 times larger than the proposed system. As a result, the proposed system is capable of minimizing the CC.

After training the system, the proposed two-steps classification technique is applied to recognize BdSL by comparing with pre-trained feature vectors of hand-sign. At the first step classifier selects K_Ω most frequent sign classes (where $7 \geqslant K_\Omega \geqslant 3$ for better performance decided by observing the graph as presented in Section 4) based on calculating the K_Ω most similarity between pre-trained NOBV (Γ_i^j) and test NOBV (Γ) using Eq. (4).

$$ICC_{\Gamma\_\max}(K_\Omega) = \begin{cases} Classi1, & \text{if} \left(\frac{ICC_\Gamma(\eta)}{|\Gamma_1^j||\Gamma|}\right) \geqslant TH_{ICC}, \\ Classi2, & \text{if} \left(\frac{ICC_\Gamma(\eta)}{|\Gamma_2^j||\Gamma|}\right) \geqslant TH_{ICC}, \\ Classi3, & \text{if} \left(\frac{ICC_\Gamma(\eta)}{|\Gamma_3^j||\Gamma|}\right) \geqslant TH_{ICC}, \\ \vdots & \vdots \\ ClassiK_\Omega, & \text{if} \left(\frac{ICC_\Gamma(\eta)}{|\Gamma_n^j||\Gamma|}\right) \geqslant TH_{ICC}, \end{cases} \tag{4}$$

where $ICC_{\Gamma\_\max}(K_\Omega)$ represents the K_Ω most frequent sign classes based on the K_Ω most $ICC_{\Gamma\_\max}$ between pre-trained Γ_i^j and test Γ. In this system we have used $K_\Omega = 3$ for the best performance which is decided from the observation graph as presented in Section 4. So that the system selects 3 sign classes from the pre-rained database of Γ_i^j using the Eq. (4). $\left(\frac{ICC_\Gamma(\eta)}{|\Gamma_i^j||\Gamma|}\right)$ measures the maximum Inter-Correlation Coefficient (ICC) between pre-trained Γ_i^j and test Γ with the value among 0 to 1; and returns the sign class i. Where, $|\Gamma_i^j|$ and $|\Gamma|$ represent the normalized length of pre-trained Γ_i^j and test Γ respectively. $ICC_\Gamma(\eta)$ is the ICC between pre-trained Γ_i^j and test Γ which is calculated using Eq. (5) [8]. Where, $\Gamma^{(\eta)}$ represents a outer boundary vector point received from test Γ by cycle shift by its vector point γ^η on "η" of elements.

$$ICC_\Gamma(\eta) = (\Gamma_i^j, \Gamma^{(\eta)}) \tag{5}$$

If the $ICC_{\Gamma\_\max}$ satisfies the maximum similarity of the threshold ($TH_{ICC} = 0.85$) value ($ICC_{\Gamma\_\max} \geqslant 0.85$) then the system returns the sign classes iK_Ω. Note that the threshold value ($TH_{ICC} = 0.85$) is selected by observing the graph as shown in Fig. 4 using selected offline sample of hand-signs. From Fig. 4, the similarity rate is increased with respect to the threshold (TH_{ICC}) but the number of candidate (NOC) selection is decreased and the number of recognition drop (NORD) is increased from the value of $TH_{ICC} > 8.5$. After selecting the K_Ω most sign classes, the system checks the selected sign classes whether they are same classes or not. If selected sign classes are same ($Classi1 = Classi2 = ClassiK_\Omega = i$) then the system return the sign class i and recognizes the hand-sign labeled with the sign class i. Otherwise, confusion is raised and the system sends the selected different K_Ω most sign classes ($i1, i2, \ldots, iK_\Omega$) into the 2nd step classifier. The 2nd step classifier selects one of the sign classes among the K_Ω most sign classes ($i1, i2, \ldots, iK_\Omega$) based on maximum ICC between pre-trained Ω_i^j and test Ω using Eq. (6).

$$ICC_{\Omega\_\max} = \left(\frac{ICC_\Omega(\eta)}{|\Omega_i^j||\Omega|}\right), \tag{6}$$

where, $ICC_{\Omega\_\max}$ measures the maximum ICC between pre-trained WGV (Ω_i^j) and test WGV (Ω) with the value among 0 to 1. $|\Omega_i^j|$ and $|\Omega|$ represent the mean value of pre-trained Ω_i^j and test Ω respectively. $ICC_\Omega(\eta)$ is the ICC between pre-trained Ω_i^j and test Ω which is calculated using Eq. (7). Where, $\Omega^{(\eta)}$ represents the ηth Window-Grid (ω^η) of the WGM for each hand-sign received from test Ω_i^j by cycle shift by its Window-Grid (ω^η) on "η" of elements.

$$ICC_\Omega(\eta) = (\Omega_i^j, \Omega^{(\eta)}). \tag{7}$$

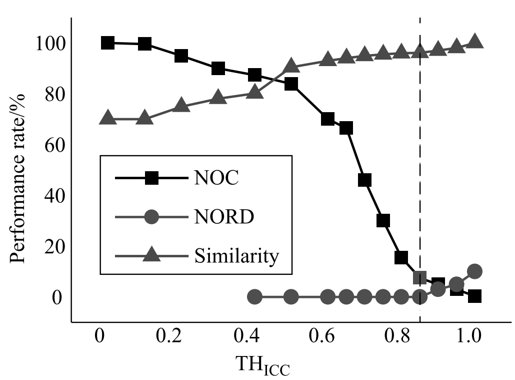


Fig. 4 Performance rates for different threshold (TH_{ICC}) values

If the $ICC_{\Omega\_\max}$ satisfies the maximum similarity of the threshold ($TH_{ICC} = 0.85$) value ($ICC_{\Omega\_\max} \geqslant 0.85$) then the system returns the sign class i from the sign classes $i1, i2, \ldots, iK_\Omega$. Most of the hand-signs classifications are done using only 1st step classifier based on only NOBV. The 2nd step classifier is used only for the hand-signs classifications based on WGV in which the NOBVs of those had-signs are similar.

3 Second phase: Bangla language modeling algorithm (BLMA)

In this section, we propose Bangla language modeling algorithm (BLMA) for automatic recognition of hand-sign-spelled BdSL by discovering "hidden characters" from "recognized characters". Bangla "hidden characters" are those characters which are not present in BdSL and "recognized characters" are those characters which are already recognized from BdSL using the first phase (hand-sign classification) of our proposed system. As presented in Algorithm 1, for each character, word, composite numeral and sentence states the input is κ_i which are recognized from each hand-sign X_i; and X_m is the position of κ_i in the text box. /.../ represents pronunciation of Bangla alphabet using Roman alphabet to make compound words [49], and (...) represents meaning of the Bangla compound words in English.

Modeling of Bangla language is different from other written language. Some vowels which are used with consonants to make Bangla words are changed as opposed to English. For example, the syntax of Bangla word "আমার"/aAmAr/ is "আ/aA/+ম/m/+ আ/A/+র/r/" which uses four letters in which the 4th vowel "আ"/aA/ is replaced with "া"/A/ but the English version of that word is simply "My" Another example for the syntax of Bangla word "জ্বর"/Jbr/ is "জ/J/+'্'/LK/+ ব/b/+র/r/" in which first three letters "জ/J/+'্'/LK/+ ব/b/" are replaced with "জ্ব" but in the English version of that word is "Fever" there is no need to joint two letters. As a single hand-sign represents multiple characters, so selecting one of the characters among multiple characters corresponding the hand-sign is another complex issue. To make the sentences in Bangla language, there is no predefined traditional syntax such as "Subject+Verb+Object" as like as syntax of English language. Some of the Bangla sentences are made without verb such as "আমার নাম রহমান" /aAmAr nAm rhmAn/ or "আঙ্গুর ফল টক"/aANggUr Tk/. Most of the cases, Bangla "verb" is placed in the last part of the sentences. Use of different punctuation marks is another complex issue. For this reason, the implementation of BLMA is more challenging but most essential.

In this system, 52 characters are recognized from BdSL (Fig. 1) which are categorized as seven categories listed in Table 1. But only 36 hand-signs are used from 51 written characters based on pronunciation [5]. To make complete Bangla words, composite numerals and sentences, we need to discover another (51-36=15) fifteen "hidden characters" which are 6 hidden vowels {ঈ/ii/, ঊ/uu/, ঐ/ei/, ঔ/oi/, ঋ/ri/, ঌ/li/} and 9 hidden consonants {ঙ/Ng/, ঞ/iy/, ন/n/, য/j/,

---

Algorithm 1 Bangla language modeling algorithm (BLMA)

Input: $\kappa_i \in (T1$ =Table 1), X_m = Position of κ_i, $T2$ =Table 2, $T3$ = Table 3, $T4$ = Table 4, $T5$ = Table 5;

Output: Recognized hand-signs-spelled BdSL.

Initialization: $TempWord = \phi$; $TempChar = \phi$; $SentenceFlag = 0$; $QuestionFlag = 0$; $Rule = \phi$;
if ($\kappa_i \in \{ct1, ct2, ct3, ct5\}$ AND ($X_{m-1}(\kappa_i) == \kappa_{space}$ OR $X_{m-1}(\kappa_i) == \phi$)) **then**
  Print κ_i;
  $TempChar = \kappa_i$;
else
  if ($\kappa_i \in T2.input$ AND $X_{m-2}(\kappa_i) \in ct3$) **then**
    Print $T2.output$ \\ Table 2: use of vowels after consonants
    TempChar=$T2.output$
  else if ($\kappa_i \in T3.input$ AND $Rule == T3.rule$) **then**
    Print $T3.output$ \\ finding 'hidden character' from 'recognized character' using Table 3
    $TempChar = T3.output$
  else
    Print κ_i \\ If no rules are matched then print all κ_i including Bangla numerals
    $TempChar = \kappa_i$
  end if
end if
if ($TempChar \neq \kappa_{space}$ AND $TempChar \neq \kappa_{punctuation}$) **then**
  $TempWord = TempWord \cup TempChar$ \\ for Bangla word forming
else
  if ($TempWord \in T4.input$) **then**
    $TempWord = T4.output$
  end if
  Print $TempWord$
  if ($SentenceFlag == 0$ AND $TempWord \in T5$) **then**
    $QuestionFlag = 1$ \\ is the first word of the sentence QWod?
  else
    $QuestionFlag = 0$
  end if
  if ($TempChar == \kappa_{punctuation}$ **then**
    if ($TempWord \in T5$ OR $QuestionFlag = 1$) **then**
      $TempChar = \kappa_{punctuation}$ ='?' \\ $\kappa_{punctuation}$ is changed as '?'
    end if
    Print $TempChar$
    $SentenceFlag = 0$ \\ starting new sentence
  else
    $QuestionFlag = 1$ \\ in between sentence
  end if
end if

শ/Shh/, ষ/SHh/, ৎ/Tt/, ঁ/CB/}. We ignore the "hidden characters" {ঌ/li/ and ঁ/CB/} which are not used available now in Bangla language [50]. Bangla language uses ten dependent vowels known as "kar" (listed as output of Table 2), (jfola) ্য/j/, (rfola) ্র/RR/, (ref) র্/rr/ and about 285 joint characters [51, 52] to make words also defined as "hidden characters". To discover the all of "hidden characters" for making Bangla words and sentences by using Algorithm 1, we have modeled the Bangla language as shown in Fig. 5 which consists of 17 sub-models. We have also listed all alphabets, vowels, all syntax and rules, example of correct words and sentences of Bangla language into different tables represented as Tables 1–5 respectively to implement the Algorithm 1.

The syntax of a Bangla word is defined by $\kappa_{space} \ldots \kappa_{space}$ or $\kappa_{space} \ldots \kappa_{punctuation}$ and the syntax of a Bangla sentence is defined by $\kappa_{punctuation} \ldots \kappa_{punctuation}$ or $\kappa_{space} \ldots \kappa_{punctuation}$. Punctuation ($\kappa_{punctuation}$) may be '।'/./, ','/,/(comma), ';'/;/(semicolon), '!'/!/(Exclamation mark), '?'/?/(Question Mark) etc. [52]. In this paper, we have consider only two punctuation marks '।'./(Full stop) and '?'/?/(Question

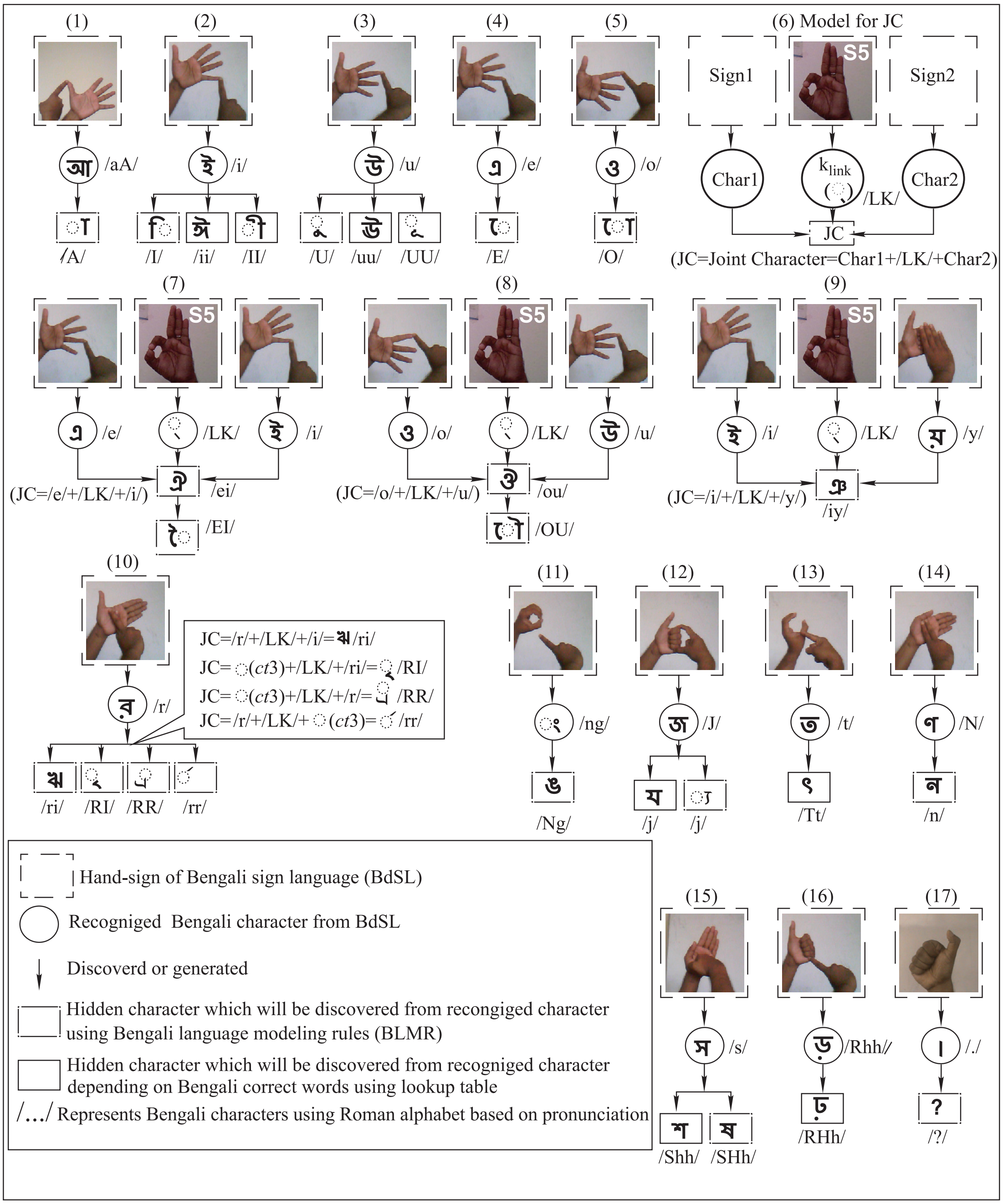


Fig. 5 Bangla language modeling for discovering "hidden characters" from "recognized characters"

Table 1 List of Bangla 'recognized characters' from corresponding BdSL

Categories	Characters from BdSLR	BdSL
$ct1$	[অ/a/]	অ of Fig. 1(a)
$ct2$[1]	[আ/aA/, ই/e/, উ/u/, এ/e/, ও/o/]	[আ-ও] of Fig. 1(a)
$ct3$[2]	[ক/k/, খ/Khh/, গ/g/, ঘ/Ghh/, চ/c/, ছ/Chh/, জ/J/, ঝ/Jhh/, ট/T/, ঠ/THh/, ড/D/, ঢ/DHh/, ত/t/, থ/Thh/, দ/d/, ধ/Dhh/, ণ/N/, প/p/, ফ/Phh/, ব/b/, ভ/Bhh/, ম/m/, য/y/, র/r/, ল/l/, স/s/, হ/h/, ড়/Rhh/]	[ক-ড়] of Fig. 1(b)
$ct4$	[ং/ng/, ঃ//Nh]	[ং, ঃ] of Fig. 1(b)
$ct5$	[১/1/, ২/2/, ৩/3/, ৪/4/, ৫/5/, ৬/6/, ৭/7/, ৮/8/, ৯/9/]	[১-৯] of Fig. 1(c)
$ct6$	[০/z/, ০ ০ ০/3z/, ০ ০ ০ ০ ০/5z/, ০ ০ ০ ০ ০ ০ ০/7z/]	০ of Fig. 1(c) and [S1-S3] of Fig. 1(d)
$ct7$	[κ_{space} = ' '//, κ_{link} = ্/LK/, $\kappa_{punctuation}$ = ।/./[3]]	[S4-S6] of Fig. 1(d)

[1] here we have categorized only those vowels which are originated from BdSLR. The complete list of dependent and in dependent vowels (except অ/a/ which has not dependent vowel form) presented in Table 2

[2] here we have categorized only those consonants which are originated from BdSLR. Another consonants which are generated from Algorithm 1 are categorized as $ct3\kappa$=[ঙ/Ng/, ঞ/iy/, ন/n/, য/j/, শ/Shh/, ষ/SHh/, ঢ়/RHh/, ৎ/Tt/, (jfola) ্য/j/, (rfola) ্র/RR/, (ref) র্/rr/] according to Fig. 5

[3] $\kappa_{punctuation}$ may be ।/./ or ?/?/ depending on the syntax of Bangla sentences

Table 2 List of Bangla vowels with corresponding dependent vowels as output depending on BLMR for vowels

Index	Input	Output	Placement	Index	Input	Output	Placement
1	আ /a/	া /A/	right of ct3	6	Table 3.output.rule1 (ঐ /ei/)	ৈ /EI/	left of ct3
2	ই /i/	ি /I/	left of ct3	7	Table 3.output.rule2 (ঔ /ou/)	ৌ /OU/	around the ct3
3	উ /u/	ু /U/	base of ct3	8	Table 3.output.rule3 (ঋ /ri/)	ৃ /RI/	base of ct3
4	এ /e/	ে /E/	left of ct3	9	ঈ /ii/ (Table 4.output for ই /i/)	ী /II/	right of ct3
5	ও /o/	ো /O/	around the ct3	10	ঊ /uu/ (Table 4.output for উ /u/)	ূ /UU/	base of ct3

Table 3.output.rulei indicates the output of Table 3 for corresponding ith rule

Table 4.output indicates the output of Table 4 depending on the corresponding correct words

Table 3 List of Bangla language modeling rules (BLMR) to implement BLMA

Index	Rule	Output
1	if (κ_i== ই/i/ AND $X_{m-2}(\kappa_i)$ == এ/e/ AND $X_{m-1}(\kappa_i)$==κ_{link})	ঐ/ei/[1]
2	if (κ_i == উ/u/ AND $X_{m-2}(\kappa_i)$ == ও/o/ AND $X_{m-1}(\kappa_i)$ == κ_{link})	ঔ/ou/[2]
3	if (κ_i == ই/i/ AND $X_{m-2}(\kappa_i)$ ==র/r/ AND $X_{m-1}(\kappa_i)$ == κ_{link})	ঋ/ri/[3]
4	if ($\kappa_i \in$ { ই/i/, উ/u/, ও/o/} AND $X_{m-2}(\kappa_i) \in ct3$ AND $X_{m-1}(\kappa_i)$ == κ_{link})	κ_i [4]
5	if ($\kappa_i \in ct3$ AND $X_{m-2}(\kappa_i)$ == র/r/ AND $X_{m-1}(\kappa_i)$ == κ_{link})	র্ (ref)/rr/[5]
6	if (κ_i == র/r/ AND $X_{m-2}(\kappa_i) \in ct3$ AND $X_{m-1}(\kappa_i)$ == κ_{link})	্র(rfola)/RR/[6]
7	if (κ_i == য/y/ AND $X_{m-2}(\kappa_i)$ == ই/i/ AND $X_{m-1}(\kappa_i)$ == κ_{link})	ঞ/iy/[7]
8	if (κ_i == স/s/ AND $Rule$ == RSH [8]	ষ/SHh/
9	if (κ_i == স/s/ AND $TempWord \in$ {পুর/pUr/, ভা/BhhA/, তির/tIr/} AND $X_{m+1}(\kappa_i)$ == κ_{link} AND $X_{m+2}(\kappa_i) \in$ { স/k/, ক/Khh/, খ/p/, প/Phh/})	ষ/s/
10	if (κ_i == স/s/ AND $X_{m-1}(\kappa_i) \in ct5$)	শ/Shh/
11	if (κ_i == ণ/N/ AND $Rule$ == RN [9]	ণ/N/
12	if (κ_i == ণ/N/ AND $Rule \neq RN$)	ন/n/
13	if (κ_i == জ/J/ AND $X_{m-2}(\kappa_i) \in$ {$ct3$, JC, ন/n/, য/j/} AND $X_{m-1}(\kappa_i)$ == κ_{link})	্য (jfola)/j/[10]
14	if (κ_i == ং/ng/ AND $X_{m+1}(\kappa_i)$ == κ_{link} AND $X_{m+1}(\kappa_i) \in${ ক/k/, খ/Khh/, গ/g/, ঘ/Ghh/, ম/m/})	ঙ/Ng/
15	if($\kappa_i \in ct3$ AND $X_{m-1}(\kappa_i)$ == κ_{link} AND $X_{m-2}(\kappa_i) \in \{ct3, ct3x^{11}, JC\}$)	JC[12]

[1] Delete ($\kappa_i, X_{m-1}(\kappa_i), X_{m-2}(\kappa_i)$) AND κ_i = ঐ/ei/

[2] Delete ($\kappa_i, X_{m-1}(\kappa_i), X_{m-2}(\kappa_i)$) AND κ_i = ঔ/ou/

[3] Delete ($\kappa_i, X_{m-1}(\kappa_i), X_{m-2}(\kappa_i)$) AND κ_i = ঋ/ri/

[4] Delete ($X_{m-1}(\kappa_i)$) AND $\kappa_i = \kappa_i$

[5] Delete ($X_{m-1}(\kappa_i), X_{m-2}(\kappa_i)$) AND Link κ_i = র্ (ref)/rr/ on the top of $\kappa_i \in ct3$

[6] Delete (κ_i = র/r/, $X_{m-1}(\kappa_i)$) AND Link κ_i = ্র (rfola)/RR/ to the base of ($X_{m-2}(\kappa_i) \in ct3$)

[7] Delete ($\kappa_i, X_{m-1}(\kappa_i), X_{m-2}(\kappa_i)$) AND κ_i = ঞ/iy/

[8] $RSHh$ = Rules of using ষ/SHh/(ষত্ব/SHhtb bIDhhAn/) [52]; Example:(i) if (κ_i == বিধান/s/ AND $X_{m-1}(\kappa_i)$ == স/ri/) then output= ঋ/SHh/; Example:(ii) if (κ_i == ষ/s/ AND $TempWord \in$ { স/atI/, অতি/aBhhI/, অতি/anU/, অনু/sU/} then output= সু/SHh/; Example: (iii) if (κ_i == ষ/s/ AND $TempWord \in$ { স/nI/, নি/dU/, দু/bhI/, বহি/aAbI/, আবি/ctU/, চতু/pRRAdu/} AND $X_{m+1}(\kappa_i)$ == κ_{link} AND $X_{m+2}(\kappa_i) \in$ { প্রাদু/k/, ক/Khh/, খ/p/, প/Phh/}) then output= ফ/SHh/)

[9] RN = Rules of using ষ/N/ (ণ ণত্ব /Ntb bIDhhAn/) [52]; Example: if (κ_i == বিধান/N/ AND $X_{m-1}(\kappa_i)$ == {ণ/ri/, ঋ/r/, র/SHh/}) then output=ষ/N/)

[10] Delete ($\kappa_i, X_{m-1}(\kappa_i)$) AND Link κ_i = ণ (jfola)/j/ to the right of $X_{m-2}(\kappa_i)$

[11] $ct3x$ ={্য/Ng/, ঙ/iy/, ঞ/n/, ন/Shh/, শ/SHh/}$\in ct3\kappa$

[12] JC =Joint Character= Delete ($\kappa_i, X_{m-1}(\kappa_i), X_{m-2}(\kappa_i)$) AND Link ($X_{m-2}(\kappa_i), X_{m-1}(\kappa_i)$) [51]

Table 4 Lookup table for sample incorrect Bangla words and corresponding correct words

Index	Input (incorrect word)	Output (correct word)	Index	Input (incorrect word)	Output (correct word)
1	ইদ /id/	ঈদ /iid/	30	ধরেণ /DhhrEN/	ধরেন /DhhrEn/
2	ইগল /igl/	ঈগল /iigl/	31	জাবেণ /JAbEN/	যাবেন /jAbEn/
3	ইর্ষা /irrSHhA/	ঈর্ষা /iirrSHhA/	32	দুর্ণাম /dUrrNAm/	দুর্নাম /dUrrnAm/
4	ইস্বরি /isbrI/	ঈশ্বরী /iiShhbrII/	33	ব্রাক্ষ্মণ /bRRAkSHhmN/	ব্রাক্ষ্মন /bRRAkSHhmn/
5	পক্ষি /pkSHhI/	পক্ষী /pkSHhII/	34	জিনিষ /JInISHh/	জিনিস /JInIs/
6	স্রি /stRRI/	স্ত্রী /stRRII/	35	ষোড়স /SHhORhhs/	ষোড়শ /SHhORhhShh/
7	বিণা /bINA/	বীণা /bIINA/	36	গ্রিষ /gRRISHh/	গ্রিস /gRRIs/
8	উর্ধ্ব /urrDhhb/	ঊর্ধ্ব /uurrDhhb/	37	মিষর /mISHhr/	মিশর /mIShhr/
9	উষা /uSHhA/	ঊষা /uuSHhA/	38	পুলিস /pUlIs/	পুলিশ /pUlIShh/
10	উহ্য /uhj/	ঊহ্য /uuhj/	39	স্টেসন /sTEsn/	স্টেশন /sTEShhn/
11	মুক /mUk/	মূক /mUUk/	40	সির /sIr/	শির /ShhIr/
12	মুর্খ /mUrrKhh/	মূর্খ /mUUrrKhh/	41	সাক /sAk/	শাক /ShhAk/
13	মুমুর্ষ /mUmUrrSHh/	মুমূর্ষ /mUmUUrrSHh/	42	সাসন /sAsn/	শাসন /ShhAsn/
14	বিমুড় /bImURhh/	বিমূঢ় / bImUURHh/	43	সিত /sIt/	শীত /ShhIIt/
15	দৃড় /dRIRhh/	দৃঢ় /dRIRHh/	44	সরিক /srIk/	শরীক /ShhrIIk/
16	দৃড়তা /dRIRhhtA/	দৃঢ়তা /dRIRHhtA/	45	বিসেস /bIsEs/	বিশেষ /bIShhESHh/
17	গাড /gARhh/	গাঢ় /gARHh/	46	সরির /srIr/	শরীর /ShhrIIr/
18	রুড়িং /rURhhI/	রূঢ়ি /rUURHhI/	47	উজ্যল /uJjl/	উজ্জল /uJJl/
19	রুড় /rURhh/	রূঢ় /rUURHh/	48	প্রতিজোগি /pRRtIJOgI/	প্রতিযোগী /pRRtIjOgII/
20	রাড় /rARhh/	রাঢ় /rARHh/	49	জুদ্ধ /JUdDhh/	যুদ্ধ /jUdDhh/
21	আষাড় /aASHhARhh/	আষাঢ় / aASHhARHh/	50	জদি /JdI/	যদি /jdi/
22	হঠাত /hTHhAt/	হঠাৎ / hTHhATt/	51	জম /Jm/	যম /jm/
23	ইষত /iSHht/	ঈষৎ /iSHhTt/	52	জখন /JKhhn/	যখন /jKhhn/
24	উতসাহ /utsAh/	উৎসাহ /uTtsAh/	53	জাবজ্যিবন /JAbJjIbn/	যাবজ্জীবন /jAbJJIIbn/
25	উতসব /utsb/	উৎসব /uTtsb/	54	জাতনা /JAtnA/	যাতনা /jAtnA/
26	তড়িত /tRhhIt/	তড়িৎ /tRhhITt/	55	জস /Js/	যশ /jShh/
27	ততকালিন /ttkAlIn/	তৎকালীন /tTtkAlIIn/	56	জমজ /JmJ/	যমজ /jmJ/
28	ততপর /ttpr/	তৎপর /tTtpr/	57	জাবত /JAbt/	যাবৎ /jAbTt/
29	ততসম /ttsm/	তৎসম /tTtsm/	58	জুগ /JUg/	যুগ /jUg/

Table 5 List of question-word used in interrogative sentence

কি /kI/ (what), কোথায় /kOThhy/ (where), কখন /kKhhn/(when), কখোন /kKhhOn/(when), কে /kE/(who), কেমন /kEmn/(how), কোনটি /kOnTI/(which), কয়টি /kyTI/(how many)...

Mark) from a single sign S6 for simplicity. If a sentence contains any Question Word (QWord) presented in Table 5 then the sentence is treated as interrogative and the $\kappa_{punctuation}$ is set to '?'/?/(Question Mark); otherwise $\kappa_{punctuation}$ is set to '।'/./(Full stop). A single word may represent a sentence in Bangla language such as খাব। /KhhAb./(I shall eat.) or ক? ?/kE?/(Who?) etc.

All syntax and rules for the BLMA are listed in Tables 2 and 3 which are generated from [51–54] as input according to the BLM as shown in Fig. 5. After applying this Bangla lanuage modeling rules (BLMR), the system may generate some incorrect words (where BLMR is not applicable) which will be corrected according to the lookup Table 4. Here, the special signs S1, S2 and S3 of ct6 represent "000", "10000" and "1000000" respectively [55] which are used to make composite numeral signs as shown in Figs. 6(c)–6(f). We have categorized the three special signs S4, S5 and S6 of ct7 defined in Table 1. Example of hand-sign-spelled words, composite numerals and sentences in BdSL are shown in Fig. 6.

4 Experimental result and discussion

4.1 Experimental setup

The proposed system uses a built-in webcam (USB 2.0 UVC HD Webcam) of ASUS ZenBook UX305CA series for capturing the image sequence. The system uses an ASUS ZenBook UX305CA with Intel Core m7 (Intel(R) Core(TM) m7-6y75 CPU 1.20GHz 1.51GHz) processor and 8GB RAM. The system uses EmguCV (C# of Microsoft®Visual Studio®2008 and OpenCV wrapper) [56] in 64-bit operating system of MS Windwos10®.

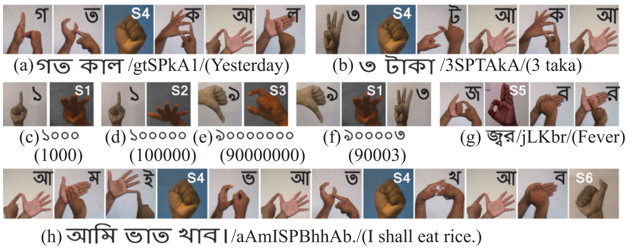


Fig. 6 Example of hand-sign-spelled word, composite numeral and sentence in BdSL

In this experiment, the proposed system is trained using 5,200 images for 52 hand-signs. 100 images are captured for each hand-sign from 10 different signers to train the system for hand-signs classification phase. The system asks 10 different skin-colored signers to perform those signs, where four are female and six are male. 10 images of each hand-sign are captured from each signer for training the system by generating the NOBV (Γ_i^j) and WGV (Ω_i^l). Figure 1 presents the example set of 52 hand-signs training dataset. Note that the training data of the proposed system (first phase and second phase as shown in Fig. 2) consists of only the hand-signs of single characters (not words, composite numerals and sentences).

For testing the system, six sets of images have been used for 52 hand-signs in BdSL. The total number of images have been used by the system for testing are $(5200 \times 6) = 31200$. For each set of images, 10 new signers who did not take part in training phase, participate to perform hand-signs, where $10 \times 10 = 100$ samples for each hand-sign are prepared for testing in the first phase of the system (hand-signs classification phase). Test data are prepared in the following six different environments:

- Environment-1(E1): plain background with proper lighting;
- Environment-2(E2): illumination variation environment with plain background;
- Environment-3(E3): cluttered and static background where skin-color static objects are present;
- Environment-4(E4): illumination variation environment with cluttered background;
- Environment-5(E5): cluttered and dynamic background where other persons are moving behind the signer; and
- Environment-6(E6): cluttered and dynamic background with illumination variation environment where other persons are moving behind the signer.

The second phase of the proposed system is tested using 500 video clips of 500 hand-sign-spelled words, 100 video clips of 100 hand-sign-spelled composite numerals and 80 video clips of 80 hand-sign-spelled sentences in BdSL from the 10 signers. Each video clip contains 10 of each hand-sign-spelled BdSL (words, composite numerals and sentences) from 10 signers. For preparing the sets of testing video clips for the Second Phase, the signs are generated randomly in different environment with different background from the 10 different signers. We allow other moving objects or persons behind the signer only who performs the hand-signs.

For the system training and testing, we plot the accuracy versus CC graphs (as shown in Fig. 7) to fine-tune the values of K_Γ of NOBV (Γ_i^j); the normalizing size of the clipping binary images $I_{norm}^m(x, y)$; the size of each WGM for WGV (Ω_i^j) generation; and the different K_Ω values of 1st step classifier using the selected off-line samples of hand-signs. From the observations of the graphs as shown in Fig. 7, we decide that the $K_\Gamma = 50, I_{norm}^m(x, y) = 150 \times 150, WGM = 5 \times 5$, and $K_\Omega = 3$ are set for achieving the high accuracy with reduced CC.

The system uses accuracy and computational cost for performance measurement. Accuracy is calculated using Eq. (8) [57] and tabulated in Table 6. The Computational Cost (CC) is calculated by measuring the time to capture an image, detect and segment hand-signs from captured image, extract features and match it with the training feature vectors in milliseconds per frame (ms/f).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100, \qquad (8)$$

where the values of TP (true positive), FP (false positive), FN (false negative), and TN (true negative) are generated from separate confusion matrices for six environments (E1, E2, E3, E4, E5 and E6). The confusion matrices are not shown here for the save of simplicity and space.

4.2 Result of hand-signs classification (first phase)

The result of hand-sign detection and skin-color segmentation are not present here due to these being the contributions of our previous system [7]. Here, we have presented the experimental result of hand-signs classification using our proposed two-step classifier.

Table 6 presents the summarized results of 52 hand-signs of BdSL alphabet, numerals and special signs recognition considering six environments (E1, E2, E3, E4, E5 and E6). From the test results shown in Table 6, it is evident that the 52 hand-signs are recognized and distinguished properly in cluttered and dynamic background with illumination varia-

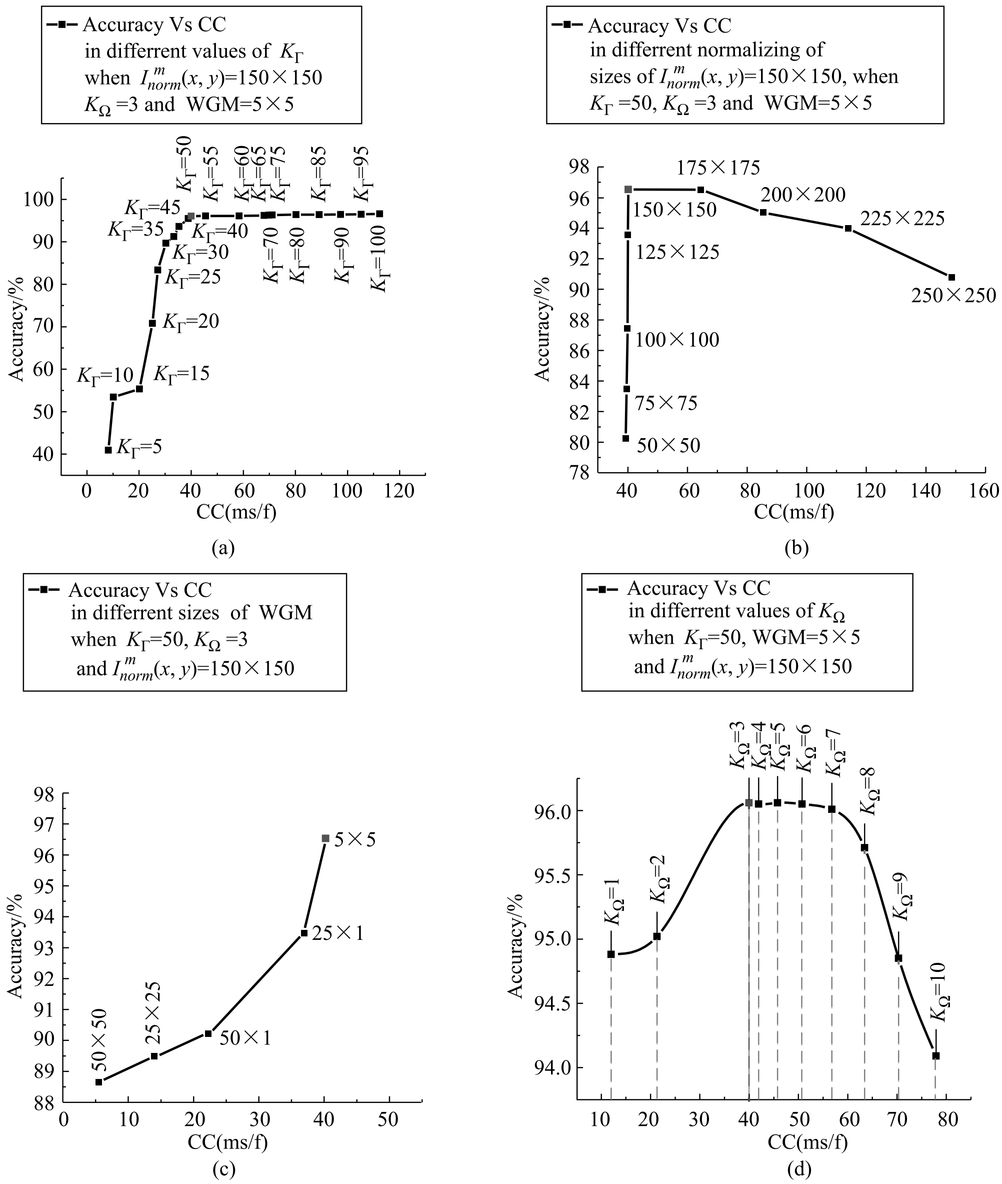


Fig. 7 The graphs of accuracy verses CC for (a) different K_Γ value of normalized outer boundary vector (Γ^j_i), (b) different normalizing sizes of $I^m_{norm}(x, y)$, (c) different sizes of each WGM, and (d) different K_Ω values of 1st step classifier

tion environment achieving the mean accuracy of 97.12% for Environment-1, 96.73% for Environment-2, 96.21% for Environment-3, 95.83% for Environment-4, 95.17% for Environment-5, 93.94% for Environment-6 and 95.83% for overall system with the computational cost of 39.97 ms/f. For each environment, the performance of the recognition rate of hand-signs “ৱ” and “ল” are decreased to less than about 92%. Because, the skin color image of hand-signs “ৱ” and “ল” are distinguishable but the binary hand-signs of them are very similar to each other as shown in Fig. 8. In several environments, the performances may decrease due to wrong performing of the hand-signs in front of fixed camera. In particular, performing the hand-signs “ও”, “থ”, “জ” and “৯” is most difficult in front of fixed camera. For each environment, the performance of one-handed numeral signs and special signs recognition in BdSL is higher than the performance of two-handed BdSL alphabet signs recognition. However, the test results are satisfactory for all environments, in which Environment-1 (E1) is the best case and Environment-6 (E6) is the worst case. Our experimental results show that the proposed system recognizes hand-signs under various illumination and backgrounds successfully. Signer dependency is tackled with NOBV and WGV feature vectors. Signer independence is proved by the testing result of the proposed sys-

Table 6 Results of BdSL Hand-signs classification accuracy and mean computational costs in six different environments

Hand-signs	Accuracy /%							Computational cost (ms/f)
	E1	E2	E3	E4	E5	E6	mean	
অ/a/	99	98	98	97	95	93	96.67	39.972
আ/aA/	100	100	98	98	97	95	98	38.873
ই/i/	98	97	97	97	95	94	96.33	39.953
উ/u/	96	97	97	97	96	95	96.33	39.995
এ/a/	97	97	96	97	97	95	96.5	39.985
ও/o/	96	97	95	96	95	94	95.5	40.882
ক/k/	96	96	96	96	96	94	95.67	40.775
খ/Khh/	96	95	95	95	95	93	94.83	40.885
গ/g/	100	99	97	97	98	96	97.83	39.775
ঘ/Ghh/	97	96	95	95	94	94	95.17	40.985
চ/c/	95	95	95	94	95	93	94.5	39.034
ছ/Chh/	95	95	94	94	94	93	94.17	40.755
জ/J/	95	95	95	95	93	93	94.33	40.755
ঝ/Jhh/	95	95	94	94	93	93	94	40.789
ট/T/	100	100	98	97	96	95	97.67	38.887
ঠ/THh/	100	99	98	98	97	95	97.83	39.342
ড/D/	97	96	96	95	95	94	95.5	39.458
ঢ/DHh/	95	94	94	94	94	92	93.83	39.459
ত/t/	95	95	95	95	95	94	94.83	40.257
থ/Thh/	96	96	95	95	94	93	94.83	40.125
দ/d/	95	95	95	95	94	93	94.5	40.155
ধ/Dhh/	95	95	95	95	94	92	94.33	39.348
ণ/N/	94	94	94	94	94	92	93.67	39.345
প/p/	95	95	94	94	95	92	94.17	41.312
ফ/Phh/	95	95	94	94	93	93	94	40.255
ব/b/	95	94	94	94	94	93	94	40.245
ভ/Bhh/	98	97	96	95	95	93	95.67	39.988
ম/m/	96	96	96	95	94	93	95	38.985
য/y/	96	96	95	95	94	91	94.5	38.987
র/r/	92	91	92	91	89	87	90.33	39.985
ল/l/	92	92	92	91	91	89	91.17	39.907
স/s/	97	95	95	94	93	92	94.33	40.415
হ/h/	98	97	96	96	95	93	95.83	40.235
ড়/Rhh/	97	96	96	95	95	95	95.67	40.125
◌ং/ng/	98	97	97	96	95	93	96	38.998
◌ঃ/Nh/	97	97	96	96	95	92	95.5	39.985
০/0/	100	100	99	99	98	97	98.83	39.972
১/1/	100	100	100	99	98	97	99	40.565
২/2/	100	100	98	98	97	96	98.17	39.987
৩/3/	100	100	99	98	98	98	98.83	40.997
৪/4/	100	99	100	98	97	97	98.5	38.987
৫/5/	100	99	99	99	97	97	98.5	39.895
৬/6/	99	99	99	98	97	97	98.17	39.887
৭/7/	99	98	99	98	96	95	97.5	40.128
৮/8/	99	99	98	98	97	95	97.67	40.654
৯/9/	98	97	98	96	96	95	96.67	41.354
S1	98	97	96	95	95	94	95.83	38.447
S2	98	98	96	96	95	94	96.17	38.456
S3	97	97	96	96	96	95	96.17	39.734
S4	97	97	96	96	95	95	96	40.231
S5	98	98	97	96	96	95	96.67	41.173
S6	99	98	98	97	96	95	97.17	38.842
mean:	97.12	96.73	96.21	95.83	95.17	93.94	95.83	39.972

tem with 10 different signers who did not take part in training.

Before applying the proposed two-steps classification technique, we have tested the hand-signs classification using four cases: Case-1, Case-2, Case-3 and Case-4. We have plotted the test results (accuracy versus computational cost (CC)) of the four cases in the observation graph as shown in Fig. 9 using the selected off-line samples of hand-signs to prove that why our proposed two step classification technique is best case. In Case-1, Hand-signs are classified based on only NOBV (Γ_i^j) with least accuracy (94.88%) but the highest classification speed. In Case2, hand-signs are classified based on only WGV (Ω_i^j) with lowest accuracy (94.78%) and more CC than Case-1. In Case-3, hand-signs are classified based NOBV (Γ_i^j) and WGV (Ω_i^j) as a single feature vector defined as $\xi_i^j = [\Gamma_i^j, \Omega_i^j]$ by a single step classification technique which obtains highest accuracy (96.08%) but the CC is increased significantly. In Case-4, hand-signs are classified by the proposed two-steps classification technique based on NOBV and WGV respectively. If the classification score of the 1st step classifier based on NOBV does not satisfy the specific condition then the system calls the 2nd step classifier based on WGV to classify the hand-signs. By applying the Case-4, the system obtains high accuracy (96.06%) about same as Case-3 but the CC is minimized significantly than the Case-3. From Fig. 9, we conclude that the proposed two-steps classification technique (Case-4) is the best case.

The previous system [27] process $i \times j \times M \times N = 52 \times 100 \times 150 \times 150 = 117,000,000$ pixel vector of 5,200 hand-signs to recognize a single hand-sign but the proposed system needs to match with maximum $[(j \times i) \times (K_\Gamma + (M_\Omega \times K_\Omega))] = [(100 \times 52) \times (50 + (25 \times 3))] = 650,000$ vector elements of 5,200 hand-signs, even if the selected K_Ω-sign classes are same then the system needs to match with only $[(100 \times 52) \times (50 + (25 \times 0))] = 260,000$ vector elements and also two classes are same but one class is different among the selected K_Ω-sign classes by the 1st classifier, the system needs to match with $[(100 \times 52) \times (50 + (25 \times 2))] = 520,000$ vector elements of 5,200 hand-signs which reduces the CC.

4.3 Comparative analysis of different system for hand-signs classification

The comparative analysis of the test results of the proposed system (BdSLR) with existing reputed BdSL classification systems developed by Rahaman et al. [8] using contour matching (CM) algorithm (denoted by "CM"); Rahaman et al. [7] using WGV (denoted by "HSSCS"); Jasim et al. [4] using Haar like feature based cascaded classifier and KNN

Fig. 8 Similarity of binary hand-signs “র” with “ল”. (a) Distinguishable skin color images of “র” and “ল”; (b) indistinguishable binary hand postures of “র” and “ল”

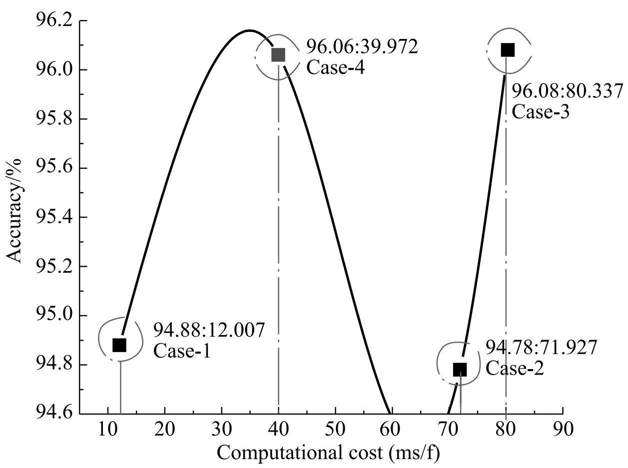


Fig. 9 Performance analysis graph (accuracy vs computational cost) of hand-signs classification in several four cases

Classifier (denoted by “Haar-KNN”), Xu [42] using convolutional neural network (CNN) classifier (denoted by “CNN”); Rahaman et al. [27] using KNN (denoted by “KNN”) and Karmokar et al. [31] using Neural network ensemble (NNE) (denoted by “NNE”) is shown in Fig. 10. We have tested these systems (not only classifiers but also whole system including hand-signs detection and segmentation, feature extraction, training and classification) using the same dataset of our proposed system (BdSLR). We have calculated the mean accuracies and CCs from the several confusion matrices for the several systems. The system's mean accuracy verses CC are plotted in the graph as shown in Fig. 10 where different points indicate the overall performance of different systems which represents that the proposed system (BdSLR) is approximately two times faster than the previous systems “Haar-KNN” [4], “CNN” [42] and “KNN” [27] and about three times faster than the “NNE” [31] while maintaining higher accuracy. The proposed system is more than five times slower than the previous system “CM” [8] and “HSSCS” [7] but the accuracies are increased significantly than the system “CM” in different challenging environments (as shown in Fig. 11) and also archives higher accuracy than “HSSCS”. We have compromised computational cost but not accuracy for maintaining the robustness of the system. From Fig. 10, we claim that the test results of the proposed system (BdSLR) show better performance than existing reputed hand-sign classification systems.

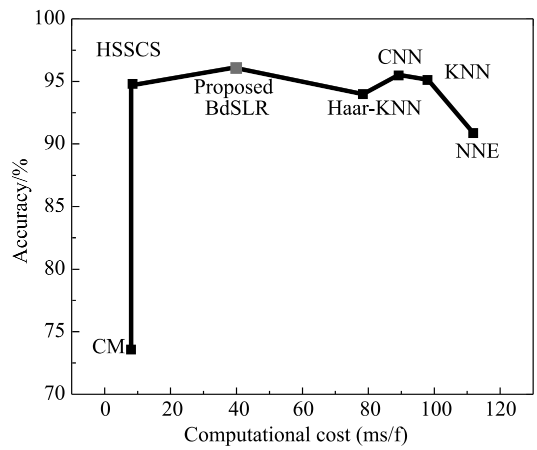


Fig. 10 Comparative analysis of different systems

4.4 Result of hand-sign-spelled BdSL recognition using BLMA (second phase)

Table 7 shows the accuracy of BLMA for discovering “hidden characters” based on corresponding “recognized characters” from BdSL. Here, we have used 100 of occurrences of each “recognized characters” in different words to discover corresponding “hidden characters” using selected offline samples. In the experiment, the system achieves the accuracy of 100% for most of the rules except discovering the “hidden characters” “ন /n/, ষ /SHH/, শ /SHh/, and জ /j/” because of exceptionality of Bangla words where the BLMRs of BLMA are not applicable and absence of alternative correct words in the lookup Table 4. From the test results shown in Table 7, it is evident that the proposed BLMA works properly by discovering all “hidden characters” based on “recognized characters” from hand-sign-spelled BdSL.

Table 7 The accuracy of BLMA for discovering “hidden characters” based on corresponding “recognized characters”

Hidden characters	Recognized characters	Number of occarence	Accuracy /%
◌া/A/	আ/a/	100	100
◌ি/I/	ই/i/	100	100
◌ু/U/	উ/u/	100	100
◌ে/E/	এ/e/	100	100
◌ো/O/	ও/o/	100	100
ঈ/ii/	ই/i/	100	100
◌ী/II/	ই/i/	100	100
ঊ/uu/	উ/u/	100	100
◌ূ/UU/	উ/u/	100	100
ঐ/ei/	এ+◌্+ই/eLKi/	100	100
◌ৈ/EI/	ct3+এ+◌্+ই/. . . eLKi/	100	100
ঔ/ou/	ও+◌্+উ/oLKu/	100	100
◌ৌ/OU/	ct3+ও+◌্+উ/. . . oLKu/	100	100
ঋ/ri/	র+◌্+ই/rLKi/	100	100
◌ৃ/RI/	ct3+◌্+র+◌্+ই/. . . rLKi/	100	100
্র/RR/	ct3+◌্+র/. . . Lkr/	100	100
র্/rr/	র+◌্+ct3/rLK. . . /	100	100
্য/iy/	ই+◌্+য়/iLKy/	100	100
JC	Char1+LK+Char2	100	100
ঙ/Ng/	◌ং/ng/	100	100
য/j/	জ/J/	100	96
্য/j/	জ/J/	100	100
ৎ/Tt/	ত/t/	100	100
ন/n/	ণ/N/	100	95
শ/Shh/	স/s/	100	95
ষ/SHH/	স/s/	100	98
ঢ়/RHh/	ড়/Rhh/	100	100
?/?/	।/./	100	100

In the experiment, we have recorded $(10 \times 1) = 10$ video clip observations from 10 signers on each hand-sign-spelled BdSL (word, composite numeral and sentence) using the 52 hand-signs for selected 500 words, 100 composite numerals and 80 sentences in BdSL. The system achieves the mean accuracy of 93.50% for hand-sign-spelled words, 95.50% for composite numerals and 90.50% for sentences recognition. Tables 8–10 present the sample results of hand-sign-spelled words, composite numerals and sentences recognition in BdSL respectively. From the test results shown in Tables 8–10, it is evident that the proposed BLMA works properly by discovering “hidden characters” based on “recognized characters” from 52 hand-signs and making Bangla words, composite numerals and sentences. In this case, if one of the hand-sign of the words or composite numerals or sentences is classified wrongly in spelling then the whole words or composite numerals or sentences are recognized as wrong which affects the overall performance of the system. Higher classification rate of basic numeral (0–9) signs in BdSL increases the recognition rate of hand-sign-spelled composite numerals recognition rate as shown in Table 9. Figure 11 shows the example snapshots of successful hand-signs classification and automatic recognition of hand-sign-spelled words, composite numerals and sentences recognition outputs in several environments.

Table 8 Example results[r1] of accuracy for automatic recognition of hand-sign-spelled words in BdSL

Bangla words	Hand-sign spelling	Accuracy /%
ঈদ /iid/(Eid)	ই+দ	95
মূক /mUUk/(Dumb)	ম+উ+ক	93
শীত /ShhIIt/(Winter)	স+ই+ত	89
সঙ্ঘ /sNgGhh/(Organization)	স+◌ং+S5+ঘ	93
মৃত্যু /mRItjU/(Death)	ম+S5+র+ S5+ই+ত+S5+জ+উ	87
হঠাৎ /hTHhaTt/(Sudden)	হ+ঠ+আ+ত	95
রাজ্য /rAJj/(State)	র+আ+জ+S5+জ	88
ডাক্তার /DAktAr/(Doctor)	ড+আ+ক+S5+ত+আ+র	87
জ্বর /Jbr/(Fever)	জ্++S5+ব+র	87
দুঃখিত /dUNhKhhIt/(Sorry)	দ+উ+◌ঃ+খ+ই+ত	89
ভালো /BhhAlO/(Good)	ভ+আ+ল+ও	91
বাথরুম /bAThhrUm/(Toilet)	ব+আ+থ+র+উ+ম	88
বাসা /bAsA/(House)	ব+আ+স+আ	89
মা /mA/(Mother)	ম+আ	92
বাবা /bAbA/(Father)	ব+আ+ব+আ	90
কষ্ট /kSHhT/(Trouble)	ক+স+S5+ট	88
ঘুম /GhhUm/(Sleep)	ঘ+উ+ম	97
খাবার /KhhAbAr/(Food)	খ+আ+ব+আ+র	80
খুশি /KhhUShhI/(Happy)	খ+উ+শ+ই	94
ভাত /BhhAt/(Rice)	ভ+আ+ত	93
খাব /KhhAb/(Eat)	খ+আ+ব	87
আগামি কাল /aAgAmI kAl/ (Tomorrow)	আ+গ+আ+ম+ই+S4+ক+আ+ল	87
গত কাল /gt kAl/(Yesterday)	গ+ত+S4+ক+আ+ল	87
৩ টাকা /3 TAkA/(3 Taka)	৩+S4+ট+আ+ক+আ	97
মাথা /mAThhA/(Head)	ম+আ+থ+আ	89
দধি /dDhhI/(Ke r)	দ+ধ+ই	91
আম /aAm/(Mango)	আ+ম	97

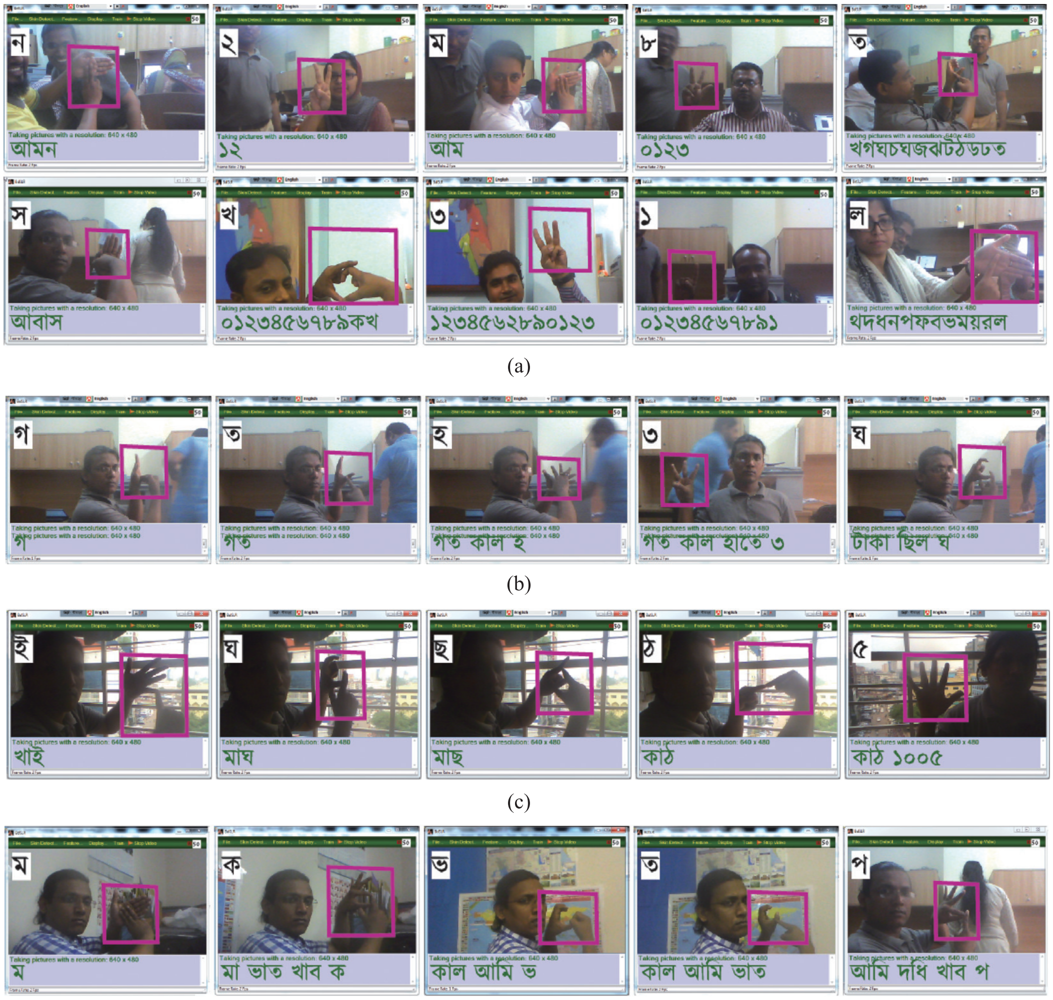
[r1] only 27 sample results are presented among 500 words

Table 9 Example results[r2] of accuracy for automatic recognition of hand-sign-spelled composite numerals in BdSL

Bangla composite numerals	Hand-sign spelling	Accuracy/%
১০০ (100)	১+০+০	92
১০০০ (1000)	১+S1	97
১০০০০০ (100000)	১+S2	97
৯০০০০০০০ (90000000)	৯+S3	96
১০১ (101)	১+০+১	97
৫৪৭ (547)	৫+৪+৭	96
৯০০০০৩ (900003)	৯+S1+০+৩	94

[r2] only 7 sample results are presented among 100

By using AKF to track the ROI and using the proposed FRB-RGB model to segment the skin-color and the binary



(a) (b) (c) (d)

Fig. 11 Example snapshots of the output of the proposed system in different challenging environments. (a) Example hand-signs recognition with 10 different signers for 10 different signs; (b) example hand-signs-spelling recognition (word and/or sentence macking) with partial occlusion behind another persons are moving; (c) example hand-signs-spelled words recognition with illumination variation in outdoor scene; (d) example hand-signs-spelled sentence recognition with cluttered background

hand-sign extraction from the ROI based on segmented skin-color pixels with specific motion, the proposed system achieves the ability to recognize the hand-signs performed in cluttered and dynamic background where other persons and skin-color objects are moving behind. It makes robustness of the system possible in real-time. The proposed system works successfully in different challenging environments, such as when signs are performed by different signers (Fig. 11(a)), illumination is varied (Fig. 11(c)), hand-signs are performed in cluttered background or hand-signs overlap with some skin-color region (Fig. 11(d)), and hand-signs are on the face (Fig. 11(a)). The proposed system can classify hand-signs and recognize hand-sign-spelled words and/or sentences in dynamic backgrounds where other persons are moving behind (as shown in Fig. 11(b)).

5 Conclusion

This paper presents a Bangla language modeling algorithm for automatic recognition of hand-sign-spelled Bangla sign language that interprets hand-sign-spelled BdSL into Bangla written words, composite numerals and sentences. In the First Phase of the system, proposed two-steps classifiers implements hand-signs classification which is tested for 52 Bangla hand-signs classification considering four cases: using NOBV based classifier, WGV based classifier, combined feature vector using NOBV and WGV based classifier and then the proposed two step classifiers. For the proposed two step classifiers, at first the system classifies the hand-signs using NOBV and if the classification score is unsatisfied then the system uses another classifier based on WGV. For

Table 10 Example results[r3] of accuracy for automatic recognition of hand-sign-spelled sentences in BdSL

Bangla sentences	Hand-sign spelling	Accuracy/%
ডাক্তার ডাক। /DAktAr DAk./ (Please call Doctor.)	ড+আ+ক+S5+ত+আ+র+S4+ড+আ+ক+ S6	87
আমাকে ৯০০০ টাকা দাও।/aAmAkE 9000 TAkA dAo./ (Please give me 9000 taka)	আ+ম+আ+ক+এ+S4+৯+ S1+S4+ট+আ+ক+আ+S4+ দ+আ+ও+S6	93
জ্বর জ্বর লাগছে। /Jbr Jbr lAgChhE./ (I feel fever.)	জ+S5+ব+র+S4+জ+S5+ ব+র+S4+ল+আ+গ+ছ+এ+S6	87
বাথরুমে যাব। /bAThhrUmE jAb./ (I feel to toilet.)	ব+আ+থ+র+উ+ম+এ+S4+জ+আ+ব+S6	90
মা ঘুম পাচ্ছে। /mA GhhUm pAcChhE./ (Mother, I want to sleep.)	ম+আ+S4+ঘ+উ+ম+S4+প+আ+চ+S5+ ছ+এ+S6	87
বাবা ঘুরতে যাব। /bAbA GhhUrtE jAb./ (Father, I want to go out.)	ব+আ+ব+আ+S4+ঘ+উ+র+ত+এ+S4+জ+আ+ব+S6	90
আমি ভাত খাব। /aAmI BhhAt KhhAb./ (I shall eat rice.)	আ+ম+ই+S4+ভ+আ+ত+S4+থ+আ+ব+S6	86
তোমার নাম কি? /tOmAr nAm kI?/ (What is your name?)	ত+ও+ম+আ+র+S4+ণ+আ+ম+S4+ক+ই+S6	86
কোথায় যাবে? /kOThhAy jAbE?/ (Where shall you go?)	ক+ও+থ+আ+য়+S4+জ+আ+ব+এ+S6	86

[r3] only 9 sample results are presented among 80 sentences

all classifiers, the hand-signs classification is done based on maximum Inter Correlation Coefficient (ICC) between test feature vector and pre-trained feature vectors. For the classifier perspective analysis, the system achieved mean accuracy of 94.88% for NOBV based classifier, 94.78% for WGV based classifier, 96.08% for combined feature vector using NOBV and WGV based classifier, and 96.06% for proposed two step classifiers based on NOBV and WGV with the computational cost of 12.007, 71.927, 80.337 and 39.972 ms/f respectively as shown in Fig. 9. The analytical result shows that the performance of the proposed two step classifier based on NOBV and WGV is better than other cases. The proposed system is tested considering in six different challenging environments (E1, E2, E3, E4, E5 and E6) presented in Table 6 achieving mean accuracy of 97.12% for E1, 96.73% for E2, 96.21% for E3, 95.83% for E4, 95.17% for E5, 93.94% for E6 and 95.83% for overall system with the computational cost of 39.97 ms/f. The experimental results prove that the system is capable of recognizing any hand-signs of any sign language in any environment, if it is trained properly. The proposed system is faster and simpler along with keeping high accuracy than other related systems as shown in Fig. 10. In the Second Phase of the system, proposed BLMA is used to make Bangla written words, composite numerals and sentences by discovering the "hidden characters" based on "recognized characters" from 52 hand-signs. Finally, the system is tested for classifying hand-sign-spelled of 500 words, 100 composite numerals and 80 sentences in BdSL using BLMA. For this experiment the system achieves mean accuracy of 93.50% for words, 95.50% for composite numerals and 90.50% for sentences recognition in BdSL. These experimental results prove that the proposed BLMA works properly with acceptable result. However the system sometimes fails to distinguish properly two similar binary signs such as "র" and "ল" in which the color images of them are distinguishable but the binary images of them are very similar as shown in Fig. 8. The system may fail to segment the hand-signs, if any skin-color objects with similar motion of hand-signs are presented in the ROI. These limitations will be overcome by future development of the system. The BLMA needs to further develop in future work including all punctuation marks and all kind of joint letter representation to interpret hand-sign-spelled words and sentences into Bangla written language successfully. However this research will provide a starting point to the researchers into the field of hand-sign-spelled BdSL recognition. The system can be applied as an interpreter for communication between sign and non-sign people and it can also be used for human–omputer/machine interaction or robot control.

Acknowledgements This research was partially supported and funded by the Information and Communication Technology (ICT) Division, Ministry of Posts, Telecommunications and IT, Government of the People's Republic of Bangladesh.

References

1. Sutton V. Introduction on deafness, sign language & sign-writing. See Signwritingorg Website, 2016

2. WFD. Sign language. See World Federation of the Deaf Website, 2016
3. Dey S. Bangladesh sign language day. See DPI-AP Website, 2014
4. Jasim M, Zhang T, Hasanuzzaman M. A real-time computer vision-based static and dynamic hand gesture recognition system. International Journal on Image and Graphics, 2014, 14(01n02): 145006
5. Majumder M, Hossain M, Mamtaz M, Ahmed M, Khan H, Ali M H, Khaled O, Iqbal M. Bengali Sign Language Dictionary. 2nd ed. Dhaka, Bangladesh: National Centre for Special Education and Ministry of Social Welfare in Cooperation with The Norwegian Association of The Deaf and Bangladersh National Federation of The Deaf, 1994
6. Begum S, Hasanuzzaman M. Computer vision-based bangladeshi sign language recognition system. In: Proceedings of the 12th International Conference on Computer and Information Technology. 2009, 414–419
7. Rahaman M A, Jasim M, Zhang T, Ali M H, Hasanuzzaman M. A real-time hand-signs segmentation and classification system using fuzzy rule based RGB model and grid-pattern analysis. Frontiers of Computer Science, 2018, 12(6): 1258–1260
8. Rahaman M A, Jasim M, Zhang T, Ali M H, Hasanuzzaman M. Real-time bengali and chinese numeral signs recognition using contour matching. In: Proceedings of the IEEE International Conference on Robotics and Biomimetics. 2015, 1215–1220
9. Li J, Wang J, Ju Z. A novel hand gesture recognition based on high-level features. International Journal of Humanoid Robotics, 2018, 15(2): 1750022
10. Dong L, Liang Y, Kong G, Zhang Q, Cao X, Izquierdo E. Holons visual representation for image retrieval. IEEE Transactions on Multimedia, 2016, 18(4): 714–725
11. Dong L, He L, Zhang Q. Discriminative light unsupervised learning network for image representation and classification. In: Proceedings of the 23rd ACM International Conference on Multimedia. 2015, 1235–1238
12. Garcia-Ceja E, Brena R F. An improved three-stage classifier for activity recognition. International Journal of Pattern and Recognition and Artifical Intelligence, 2018, 32(1): 1860003
13. Lee G, Mallipecldi R, Lee M. Trajectory-based vehicle tracking at low frame rates. Expert System with Application, 2017, 80: 46–57
14. Mills M T, Bourbakis N G. A comparative survey on NLP/U methodologies for processing multi-documents. International Journal on Artificial Intelligence Tools, 2012, 21(4): 1250017
15. Mills M, Psarologou A, Bourbakis N. Modeling natural language sentences into SPN graphs. In: Proceedings of the 25th IEEE International Conference on Tools with Artificial Intelligence. 2013, 889–896
16. Santoni D, Pourabbas E. Automatic detection of words associations in texts based on joint distribution of words occurrences. Computational Intelligence, 2016, 32(4): 535–560
17. Lee G, Mallipeddi R, Lee M. Trajectory-based vehicle tracking at low frame rates. Expert Systems with Applications, 2017, 80: 46–57
18. Liberman Y, Perry A. A visual tracking scheme for accurate object retrieval in low frame rate videos. International Journal on Artificial Intelligence Tools, 2016, 25(5): 1640003
19. Dong L, Feng N, Mao M, He L, Wang J. E-grabcut: an economic method of iterative video object extraction. Frontiers of Computer Science, 2017, 11(4): 649–660
20. Dong L, Feng N, Zhang Q. LSI: semantic label inference for nature image segmentation. Pattern Recognition, 2016, 59: 282–291
21. Li C, Lin L, Zuo W, Wang W, Tang J. An approach to streaming video segmentation with sub-optimal low-rank decomposition. IEEE Transactions on Image Processing, 2016, 25(4): 1947–1960
22. Chen F S, Fu C M, Huang C L. Hand gesture recognition using a real-time tracking method and hidden markov models. Image Vision Computer, 2003, 21(8): 745–758
23. Alon J, Athitsos V, Yuan Q, Sclaroff S. A unified framework for gesture recognition and spatiotemporal gesture segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(9): 1685–1699
24. Asaari M S M, Rosdi B A, Suandi S A. Adaptive kalman filter incorporated eigenhand (akfie) for real-time hand tracking system. Multimedia Tools and Applications, 2015, 74(21): 9231–9257
25. Khaled H, Sayed S G, Saad E S M, Ali H. Hand gesture recognition using modified 1 and background subtraction algorithms. Mathematical Problems in Engineering, 2015, (Article ID: 741068): 1–8
26. Gurav R M, Kadbe P K. Vision based hand gesture recognition with haar classifier and adaboost algorithm. International Journal of Latest Trends in Engineering and Technology, 2015, 5(2): 155–160
27. Rahaman M A, Jasim M, Ali M H, Hasanuzzaman M. Realtime computer vision-based bengali sign language recognition. In: Proceedings of the 17th International Conference on Computer and Information Technology. 2014, 192–197
28. Weng S K, Kuo C M, Tu S K. Video object tracking using adaptive kalman filter. Journal of Visual Communication and Image Representation, 2006, 17(6): 1190–1208
29. Li N, Liu L, Xu D. Corner feature based object tracking using adaptive kalman filter. In: Proceedings of the 9th International Conference on Signal Processing. 2008, 1432–1435
30. Luo Y, Celenk M. A new adaptive kalman filtering method for block-based motion estimation. In: Proceedings of the 15th International Conference on Systems, Signals and Image Processing. 2008, 89–92
31. Karmokar B C, Alam K M R, Siddiquee M K. Bangladeshi sign language recognition employing neural network ensemble. International Journal of Computer Applications, 2012, 58(16): 43–46
32. Jarman A M, Arshad S, Alam N, Islam M J. An automated bengali sign language recognition system based on fingertip finder algorithm. International Journal of Electronics and Informatics, 2015, 4(1): 1–10
33. Yasir F, Prasad P W C, Alsadoon A, Elchouemi A. Sift based approach on bangla sign language recognition. In: Proceedings of the 8th IEEE International Workshop on Computational Intelligence and Applications. 2015, 35–39
34. Ayshee T F, Raka S A, Hasib Q R, Hossain M, Rahman R M. Fuzzy rule-based hand gesture recognition for bengali characters. In: Proceedings of the IEEE International Advance Computing Conference. 2014, 484–489
35. Wang, Qiao Y, Tang X. Action recognition with trajectorypooled deep-convolutional descriptors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015, 4305–4314
36. Feichtenhofer, Pinz A, Wildes R. Spatiotemporal residual networks for video action recognition. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. 2016, 3476–3484
37. Asadi-Aghbolaghi M, Clapes A, Bellantonio M, Escalante H J, Ponce-

López V, Baró X, Guyon I, Kasaei S, Escalera S. A survey on deep learning based approaches for action and gesture recognition in image sequences. In: Proceedings of the 12th IEEE Conference on Automatic Face and Gesture Recognition (FG 2017). 2017, 476–483

38. Liu Z, Zhang C, Tian Y. 3D-based deep convolutional neural network for action recognition with depth sequences. Image and Vision Computing, 2016, 55(2): 93–100
39. Varol G, Laptev I, Schmid C. Long-term temporal convolutions for action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(6): 1510–1517
40. Zhu G, Zhang L, Mei L, Shao J, Song J, Shen P. Large-scale isolated gesture recognition using pyramidal 3D convolutional networks. In: Proceedings of the 23rd International Conference on Pattern Recognition. 2016, 19–24
41. Wang P, Li W, Liu S, Gao Z, Tang C, Ogunbona P. Largescale isolated gesture recognition using convolutional neural networks. In: Proceedings of the 23rd International Conference on Pattern Recognition. 2016, 7–12
42. Xu P. A real-time hand gesture recognition and humancomputer interaction system. 2017, arXiv preprint arXiv:1704.07296
43. Rahaman M A, Jasim M, Ali M H, Hasanuzzaman M. Computer vision based bengali sign words recognition using contour analysis. In: Proceedings of the 18th International Conference on Computer and Information Technology. 2015, 335–340
44. Park A, Yun S, Kim J, Min S, Jung K. Real-time visionbased korean finger spelling recognition system. International Journal of Computer, Electrical, Automation, Control and Information Engineering, 2008, 2(8): 2623–2628
45. Kane L, Khanna P. A framework for live and cross platform fingerspelling recognition using modified shape matrix variants on depth silhouettes. Computer Vision and Image Understanding, 2015, 141: 138–151
46. Fang G, Gao W, Zhao D. Large-vocabulary continuous sign language recognition based on transition-movement models. IEEE Transactions on Systems, Man and Cybernetics, Part A, 2007, 37(1): 1–9
47. Liwicki S, Everingham M. Automatic recognition of fingerspelled words in British sign language. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2009, 50–57
48. Koller O, Forster V J, Ney V H. Continuous sign language recognition: towards large vocabulary statistical recognition systems handling multiple signers. Computer Vision and Image Understanding, 2015, 141: 108–125
49. Rameshar S. Sadaran Vhasabiggan and Bangla Vhasa Dhaka, Bangladesh: Ananda Press, 1996
50. Mukhopaddhay A. Adorsho Bangla banan: ekti prostabona. See Galpersamay Website, 2017
51. Kibria G. Bangla joint letter. See Daffodilvarsity.edu Website, 2013
52. Chowdhuri M, Chowdhuri M H. Bangla Bhashar Bakoron (Grammar of Bengali language). 2nd ed. Dhaka, Bangladesh: Jatio Shikkhakromo Opatthopustok Board, 2014
53. Choudhury J. Bangla Banan Abhidhan. 3rd ed. Dhaka, Bangladesh: Bangla Academy, 2008
54. Ishaaque A. Samakalin Bangla Bhashar Abhidhan, 2nd ed. Dhaka, Bangladesh: Bangla Academy, 2003
55. CDD. Manual on Sign Supported Bangla. Dhaka, Bangladesh: Center for Disability in Development (CDD), 2002
56. Sourceforge.net. EmguCV. See Sourceforge Website, 2015
57. Nagarajan S, Subashini T S. Static hand gesture recognition for sign language alphabets using edge oriented histogram and multi class SVM. International Journal of Computer Applications, 2013, 82(4): 28–35

Muhammad Aminur Rahaman received his BSc and MSc degree in Computer Science & Engineering from the Department of Computer Science & Engineering, Islamic University, Bangladesh in 2003 and 2004, respectively. He is a candidate for PhD degree under supervision of Prof. Dr. Md. Hasanuzzaman and Prof. Dr. Md. Haider Ali in the Department of Computer Science & Engineering, University of Dhaka, Bangladesh. He is a founder Director of Worldgaon (Pvt.) Limited which is a one of the famous software development company of Bangladesh. His current research interests including computer vision, sign language recognition and human-computer interaction. He is a member of IEEE.

Mahmood Jasim received his BSc and MSc degree in Computer Science & Engineering from the Department of Computer Science & Engineering, University of Dhaka, Bangladesh in 2011 and 2013, respectively. He joined the Department of Computer science and engineering as a lecturer in 2014. Currently he is pursuing his PhD degree in the University of Massachusetts Amherst. His current research interests include human-computer interaction, image processing, computer vision and artificial intelligence.

Md. Haider Ali received PhD degree from the Department of Electronics & Information Engineering, Toyohashi University of Technology, Japan in 2001. Prof. Ali has completed his Bachelor and Master degree from the Department of Applied Physics & Electronics (presently EEE) University of Dhaka 1984 and 1985 respectively. He is a professor of the Department of Computer Science & Engineering, University of Dhaka since June 2007. His current research interests include human face recognition and expression detection,

post surgical expression simulation, soft-tissue deformation modeling, polygonal mesh simplification, narrow band video transmission/video conferencing, etc.

Md. Hasanuzzaman received PhD degree from the Department of Informatics, National Institute of Informatics (NII), The Graduate University for Advanced Studies, Japan. He graduated (with Honors) in 1993 from the Department of Applied Physics & Electronics, University of Dhaka, Bangladesh. He completed Masters of Science (MSc) in Computer Science in 1994 from the University of Dhaka. He joined as a Lecturer in the Department of Computer Science & Engineering, University of Dhaka, Bangladesh in 2000. Since March 2013, he has been serving as a professor in the Department of Computer Science & Engineering, University of Dhaka, Bangladesh. His current research interests include human-computer interaction, image processing, computer vision and artificial intelligence.