Convolutional Neural Network based Handwritten Bengali and Bengali-English Mixed Numeral Recognition

M. A. H. Akhand, Mahtab Ahmed
Dept. of Computer Science and Engineering, Khulna University of Engineering & Technology (KUET)
Khulna-9203, Bangladesh
Email: {akhand, mahtab}@cse.kuet.ac.bd

M. M. Hafizur Rahman
Dept. of Computer Science, KICT, International Islamic University Malaysia (IIUM)
Selangor, Malaysia
Email: hafizur@iium.edu.my

Abstract—Recognition of handwritten numerals has gained much interest in recent years due to its various potential applications. Bengali is the fifth ranked among the spoken languages of the world. However, due to inherent difficulties of Bengali numeral recognition, a very few study on handwritten Bengali numeral recognition is found with respect to other major languages. The existing Bengali numeral recognition methods used distinct feature extraction techniques and various classification tools. Recently, convolutional neural network (CNN) is found efficient for image classification with its distinct features. In this paper, we have investigated a CNN based Bengali handwritten numeral recognition scheme. Since English numerals are frequently used with Bengali numerals, handwritten Bengali-English mixed numerals are also investigated in this study. The proposed scheme uses moderate pre-processing technique to generate patterns from images of handwritten numerals and then employs CNN to classify individual numerals. It does not employ any feature extraction method like other related works. The proposed method showed satisfactory recognition accuracy on the benchmark data set and outperformed other prominent existing methods for both Bengali and Bengali-English mixed cases.

Index Terms—Image Pre-processing, Convolutional Neural Network, Bengali Numeral, Handwritten Numeral Recognition.

I. INTRODUCTION

Recognition of handwritten numerals has gained much interest in recent years due to its various potential applications. These include postal system automation, passports & travel document analysis, automatic bank cheque processing, and even for number plate identification [1, 2]. Research on recognition of unconstrained handwritten numerals has made impressive progress in Roman, Chinese, and Arabic script [3, 4]. On the other hand, recognition of handwritten Bengali numeral is largely neglected although it is a major language in Indian subcontinent having fifth ranked in the world and is the main language of Bangladesh. Some Bengali numerals have very similar shape, and even in the printed form few differs very little than other numerals. Therefore, Bengali numeral recognition is a challenging task.

An interesting aspect of Bengali documents is that English entries are commonly available in those. These include the text books written in Bengali scripts often have entries in English especially the numerals; Bangladeshi currencies contain both Bengali and English numerals to represent values; handwritten Bengali-English mixed numerals are frequently found in Bangladesh while writing postal code, bank cheque, age, number plate, mobile number and other tabular form documents. Moreover, often people casually enter one or more English numerals which results a mixed-script situation. There is a similarity in writing style of several Bengali numerals (e.g., '০', '২','৪' and '৭') with English numerals (e.g., '0','2', '8' and '9') that makes Bengali-English mixed numeral recognition more challenging. The reason behind the uses of this mixed case is that English is used in parallel with Bengali in official works and education systems. Therefore, Bengali-English mixed numeral recognition system is challenging and important for practical applications. Although several remarkable works are available for Bengali and English handwritten numeral recognition separately, a very few works are available for mixed Bengali-English numeral recognition whose performance are not at satisfactory level.

The rest of the paper is organized as follows. Section II reviews several related works and explains motivation of the present study. Section III explains proposed recognition scheme using convolutional neural network (CNN) which contains dataset preparation, pre-processing and classification. Section IV presents

experimental results of the proposed method and comparison of performance with other related works. Finally, a brief conclusion of the work is given in Section V.

II. RELATED WORKS

A few notable works are available for Bengali handwritten numeral recognition with respect to other popular Indian subcontinent scripts such as Devanagari [5-7]. Bashar et al. [8] investigated a digit recognition system based on windowing and histogram techniques. Windowing technique is used to extract uniform features from scanned image files and then histogram is produced from the generated features. Finally, recognition of the digit is performed on the basis of generated histogram.

Pal et al. [3] introduced a new technique based on the concept of water overflow from the reservoir for feature extraction and then employed binary tree classifier for handwritten Bengali numeral recognition. Basu et al. [4] used Dempster-Shafer (DS) technique to combine the classification decisions obtained from two MLP based classifiers for handwritten Bengali numeral recognition using two different feature sets. Feature sets they investigated are called shadow feature and centroid feature. Khan et al. [9] employed an evolutionary approach to train artificial neural network for Bengali handwritten numeral. At first, they used boundary extraction on a numeral image in a single window by horizontal-vertical scanning and scaled the image into fixed sized matrix. Then, Multi-Layer Perceptron (MLP) are evolved for recognition.

Wen et al. [10] proposed a handwritten Bengali numeral recognition system for automatic letter sorting machine. They used Support Vector Machine (SVM) classifier combined with extensive feature extractor using Principal Component Analysis (PCA) and kernel PCA (KPCA). Das et al. [11] also used SVM for classification but used different techniques for feature selection. Seven different sets of variable sized local regions are generated using different region selection strategies in their work. A genetic algorithm (GA) based region sampling strategy has been employed to select an optimal subset of local regions containing high discriminating information about the pattern shapes from the above mentioned set.

Bhattacharya and Chaudhuri [12] presented a multistage cascaded recognition scheme using wavelet-based multi-resolution representations and MLP classifiers. The scheme first computes features using wavelet-filtered image at different resolutions. The scheme has two recognition stages and the first stage involves a cascade of three MLP classifiers. If a decision about the possible class of an input numeral image cannot be reached by any of the MLPs of the first stage, then certain estimates of its class conditional probabilities obtained from these classifiers are fed to another MLP of the second stage. In this second stage, the numeral image is either classified or rejected according to a precision index.

Recently, Wen and He [13] proposed a kernel and Bayesian Discriminant based method to recognize handwritten Bengali numeral. Most recently, Nasir and Uddin [14] proposed a hybrid system for recognition of handwritten Bengali numeral for the automated postal system, which performed feature extraction using k-means clustering, Baye's theorem and maximum of a Posteriori, then the recognition is performed using SVM.

On the other hand, according to best of our knowledge, a very few studies are available for Bengali-English mixed handwritten numeral recognition. Mustafi et al. [15] proposed a system which extract topological and structural features of the handwritten Bengali and English numerals using water overflow from the reservoir. For recognition, a MLP is trained with the extracted feature values using Back-propagation (BP) algorithm. They considered 16 classes, considering similarity between Bengali numerals 'o', '২', '৪', '৭' with English numerals '0', '2', '8', '9', respectively, in same class.

Vazda et al. [16] investigated mixed Bengali-English numeral recognition in Indian postal document automation. At first they extracted numeral images from the destination address block section of a postal document. They normalized the images into 28×28 pixel and train a 16 class MLP using these pixel values. To improve recognition accuracy, other two MLPs with 10 classes are considered for Bengali and English numeral recognition. In recognition stage, at first the six digit pin-code is checked with 16 class MLP. If majority of the numerals are recognized as Bengali then the entire pin-code is considered to be written in Bengali and recognized through 10 classes Bengali classifier. Similarly, 10 class English classifier is used if the majority of the numerals are recognized as English by the 16-class classifier. Bhattacharya and Chaudhuri [12] also investigated handwritten Bengali-English mixed numeral recognition along with Bengali. They tested Bengali-English mixed numerals in two different MLPs: a 16 class MLP as like previous studies and 20 class MLP considering 10 classes for each of Bengali and English.

The objective of this study is to develop a recognition scheme for handwritten Bengali and Bengali-English mixed numerals which is capable of providing high recognition accuracy. Toward this goal, we pre-processed the handwritten images in a moderate way to generate patterns and then CNN is employed for classification of Bengali and Bengali-English mixed handwritten numerals. Recently, CNN is found efficient for image classification with its distinct features; such as, it automatically provides some degree of translation invariance [17]. We investigated both 16 class and 20 class classifiers for mixed case. Experimental studies reveal that the proposed CNN based method shows satisfactory classification accuracy and outperformed other existing methods.

III. BENGALI AND BENGALI-ENGLISH MIXED HANDWRITTEN NUMERAL RECOGNITION USING CNN

This section explains proposed scheme in detail which has two major steps: pre-processing of raw images of

Bengali Numeral	Samples of ISI Bengali numeral images					English Numeral	Samples of MNIST English numeral images				
০						0					
১						1					
২						2					
৩						3					
৪						4					
৫						5					
৬						6					
৭						7					
৮						8					
৯						9					

Fig.1. Samples of handwritten Bengali and English numerals.

numerals and classification using CNN. The following subsection gives brief description of each step. At first it explains dataset preparation for better understanding.

A. Handwritten Numeral Image Data and Pre-processing

For Bengali handwritten numerals, the benchmark image dataset maintained by CVPR unit, ISI, Kolkata [18] is considered in this study. Several recent studies used this dataset or in a modified form [12, 19]. The samples of CVPR dataset are the scanned images from pin codes used on postal mail pieces. The digits are from people of different age and sex groups as well as having different levels of education. The dataset has been provided in training and test images. The test set contains total 4000 images having 400 samples for each of 10 digits. On the other hand, the training set contains total 19392 images having around 1900 images of each individual digit. In this study, all 4000 test images and 18000 training images (1800 images from each digit) are considered and pre-processed. On the other hand, we have considered the well-known MNIST database for English numerals. All 60000 training (around 6000 samples for each numeral) and 10000 test (1000 samples for each numeral) samples of MNIST are considered in this study.

In Bengali-English mixed case, English numerals of MNIST database are combined with ISI Bengali numerals. In 20 class classifier, total 78000(=18000 + 60000) and 14000(=4000+10000) samples are used for training and test cases, respectively, combining Bengali and English

numeral samples. On the other hand, 16 class classifier considered all 10 classes of Bengali samples and 6 classes of English samples excluding the 4 classes which are similar to Bengali (i.e., 0, 2, 8, and 9). Therefore, in 16 class case training and test samples are 54000 (=18000+6000×6) and 10000 (=4000+1000×6), respectively. Figure 1 shows few sample images of each numeral.

Pre-processing is performed on the images into common form that makes it appropriate to feed into classifiers. The original images are in different sizes, resolutions and shapes. Matlab R2015a is used to pre-process the images into same dimension and format. For Bengali numeral, ISI images are transformed into binary image. At first, an image is transformed into binary image with automatic thresholding of Matlab. This step removes background as well as improves intensity of written black color. Since black color is used for writing on white paper (background), the binary image files contains more white point (having value 1) than black (having value 0). To reduce computational overhead, images are converted through foreground numeral black to white and background changed to black. Written digit may be a portion in the scanned image that is easily visible from the foreground-background interchanged image. An image has been cropped to the actual writing portion removing black lines from all four sides (i.e., left, right, top and bottom). Finally, images are resized into 28×28 dimension to maintain appropriate and equal inputs for all

the numerals. To capture pattern values of resized images, the double type matrix is considered (instead of binary in the previous stages) so that best possible quality in the resized images is retained. On the other hand, MNIST images are available in 28×28 pixels therefore resizing is not required for the images. For better understanding of pre-processing, outcome of stepwise transformation on three selected Bengali numeral images from "০" to "২" are presented in Figure 2.

Steps	Stepwise pre-processing outcome of Bengali numeral samples		
	০	১	২
Step 1			
	Originnal image in tif format		
Step 2			
	Binary image with automatic thresholding		
Step 3			
	Forground and background interchanged		
Step 4			
	Cropped to original writing portion		
Step 5			
	Resized into 28×28 dimension		

Fig.2. Stepwise outcomes in pre-processing on sample handwritten images from three Bengali numerals.

B. Classification using CNN

CNNs [20] are multi-layer neural networks particularly designed to work with two-dimensional data, such as images. It adds the new dimension in image classification systems and recognizing visual patterns directly from pixel images with minimal pre-processing. In CNNs, information generally propagates through the multiple layers of the network where features of the observed data at each layer is obtained by applying digital filtering technique. Since handwritten numeral classification is a high-dimensional complex task, CNNs

are being popular for this task. CNNs use two main processes: convolution and subsampling. General architecture of a CNN consists of input layer, multiple convolution-subsampling layers, hidden layer and output layer.

In convolution process, a convolved feature map (CFM) is generated using a small sized filter (called kernel) from input feature map (IFM) which is a previous layer feature map [21]. A kernel is nothing but a set of weights and a bias. Small portion of the IFM is termed as local receptive field (LRF) and a particular LRF with the kernel will give a particular point in the CFM. All the LRFs of an IFM with the same kernel will give a complete CFM. Thus, the weights and bias of a kernel is shared in the convolution process. A common form of convolution operation to get a CFM from an IFM through kernel (K) is shown in Eq. (1).

$$CFM_{x,y} = f\left(b + \sum_{r=1}^{K_H} \sum_{c=1}^{K_W} K_{r,c} * IFM_{x+r,y+c}\right) \quad (1)$$

Here, $f(.)$ is the activation function, b is bias value of the kernel, K_H and K_W denote the size of the kernel as $K_H \times K_W$ matrix. It is useful to apply the kernel everywhere in the image. This makes sense, because if the weights and bias are such that the hidden neuron can pick out a vertical edge in a particular local receptive field then it is also likely to be useful at other places in the image. That's why CNNs are well adapted to the translation invariance of images. While distinct kernels may produce distinct CFMs from the same IFM; operations of multiple kernels are composed to produce a CFM for multiple IFMs. It is worth mentionable that original input image is the IFM of first convolution operation to produce first set of CFMs.

In CNN, a sub-sampling layer is followed by each convolutional layer which simplifies the information produced by the convolutional operation and produces a feature map, may called sub-sampled feature map (SFM). In general, sub-sampling operation condense the CFM retaining its important feature points. General form of sub-sampling operation is shown in Eq. (2).

$$SFM_{x,y} = down\left(\sum_{r=0}^{R-1} \sum_{c=0}^{C-1} CFM_{xR-1+r,yC-1+c}\right) \quad (2)$$

Where R and C denote the size of the pooling area as $R \times C$ matrix of CFM; down(.) represents a subsampling operation on a pooling area. It is also possible to pass the value of Eq. (2) through an activation function after applying multiplicative coefficient and additive bias, respectively [22]. The size of SFM becomes 2-times smaller with respect to CMF in both spatial dimensions when $R \times C$ is 2×2. In case of local averaging in down (.) operation, the 4 pixels in a 2×2 area of CFM are taken and their average value is considered as a single point in the SFM. On the other hand, in max-subsampling, a point in SFM is the maximum value of a 2×2 pooling region of the CFM.

In CNN, after multiple convolution-subsampling operations, a hidden layer is considered before output layer. Nodes of both hidden and output layers are fully
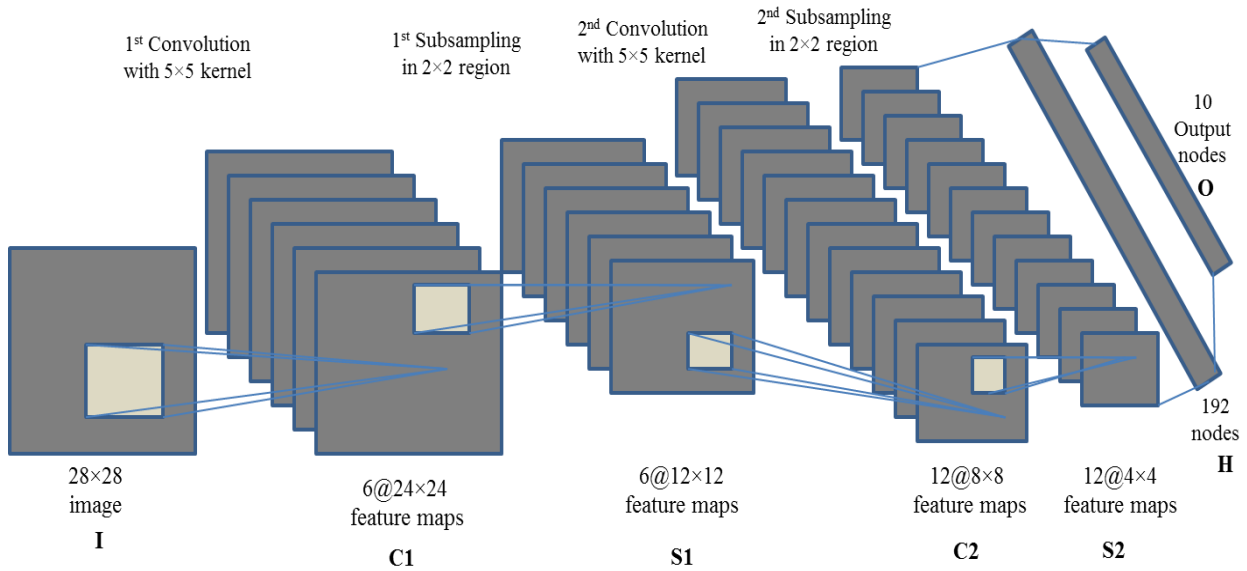
Fig.3. Structure of CNN considered in this study.

connected; but hidden layer nodes may be just linear representation of the previous SFM values or connected though weights. The output of a particular output node is the weighted sum of hidden layer values passing through an activation function. In the output layer, errors are measured by comparing desired output with the actual output. The training of CNN is performed to minimize the error (E):

$$E = \frac{1}{2}\frac{1}{PO}\sum_{p=1}^{P}\sum_{o=1}^{O}\big(d_o(p) - y_o(p)\big)^2 \qquad (3)$$

where P is the total number of patterns; O is the total output nodes of the problem; d_o and y_o are the desired and actual output of a node for a particular pattern p. In training, the kernel values with bias in different convolution layers and weights of hidden-output layers are updated. Therefore, learning parameters are very small with respect to the fully connected multi-layer neural network. A modified version of Back-Propagation is used to train a CNN and description regarding this is available in [20, 22].

Figure 3 shows CNN structure of this study for classification of handwritten numerals that holds two convolutional layers (C1 and C2) with kernel size of 5×5 and two subsampling layers (S1 and S2) with 2×2 local averaging area. In the input layer (I), 28×28 pixels are considered as 784 linear nodes on which convolution operation are to be performed. In the first convolution operation, the input image I and six kernels are convolved to produce 24×24 sized six CFMs in C1. In the first sub-sampling operation, the six CFMs of C1 are subjected to 2×2 local averaging and produces 12×12 sized six SFMs in S1. The second convolution layer (i.e., C2) contains 12 feature maps. In the convolution operation, six different kernels are applied on each of the SFM of S1 to produce an 8×8 sized CFM in C2 and; therefore, total 72 (=12$\times6$) kernels are operated to produce 12 CFMs. The second sub-sampling operation is similar to first sub-sampling

operation and produces 12 SFMs and each of them has a size of 4×4. The values of these 12 SFMs ($12\times4\times4 = 192$) are placed linearly as hidden layer (H) with 192 nodes. Finally, nodes of hidden layer are connected to the 10 output nodes for the numeral set. Each output node represents a particular digit and the desired value of the node was defined as 1 (and other 9 output nodes value as 0) for the input set of the pattern.

The first convolution layer contains total of 156 (= $(5\times5+1) \times 6$) parameters for six kernels whose values to be updated during training. Similarly, total training parameters for 12 CFMs are 1812 (= $(6\times5\times5+1) \times12$) for the second convolutional layer. Finally, all 1920 (=192×10) weights of fully connected output layer are also updated during training. The training procedure is repeated for the defined number of epochs or until the error is minimized up to a certain level.

C. Significance of the Proposed Recognition System

There are several significant differences between the proposed scheme and the traditional methods for recognition of Bengali related handwritten numeral recognition. Firstly, the proposed method is simpler than existing ones; involved moderated pre-processing of images; and training CNN with the patterns from processed images. Proposed method did not conceive any feature selection scheme. On the other hand, traditional methods use different feature selection schemes along with different pre-processing techniques and classification with different machine learning tools. Automatic thresholding level in transformation of original image into binary format helped to improve pattern generation quality of unclear images. Therefore, it is observed in Figure 2 that pre-processed outcomes of images for numerals "০ " and "১ " belong to same quality for numeral "২ ". On the other hand, image cropped to original writing portion (after foreground and background interchange) helped to optimally fit the writing and hence

improved pattern generation from the image. So that original writing portions for numerals "১ " and "২ " are found same as "০ " in Fig. 2.

IV. EXPERIMENTAL STUDIES

We have applied CNN on the resized and normalized grayscale image files without any feature extraction technique. The experiment has been conducted on HP pro desktop machine (CPU: Intel Core i7 @ 3.60 GHz and RAM: 8.00 GB) in Window 7 (64bit) environment using Matlab R2015a. Experimental results using the proposed recognition scheme have been collected based on the samples of the prepared dataset discussed earlier. Due to large sized training set, batch wise training was performed in this study; and experiments conducted with different batch sizes. Weights of the CNN are updated once for a batch of images. Number of batch size (BS) is considered as a user defined parameter and experiments are conducted for four different BS of 25, 50, 75 and 100 to observe the effect of batch size on the performance. We have considered learning rate as 1.0 as previous study [17] identified such value for better result. In the following subsections, experimental results for Bengali and Bengali-English mixed numeral recognition are presented and discussed accordingly.

A. Bengali Numeral Recognition

Figure 4 portrays the effect of batch size on training error (E) calculated using Eq. (3). For any batch size, error was reduced rapidly in initial iterations (e.g., up to 100) and after that error minimization with iteration was not significant. However, lower number of BS seems to minimize error rapidly; for BS = 10, the error (E) is lower than that of BS values 50 or 75. For smaller BS, the CNN adjusted consequently with fewer number of patterns and therefore conceiving the opportunity of better adjustment and hence minimizes error rapidly.
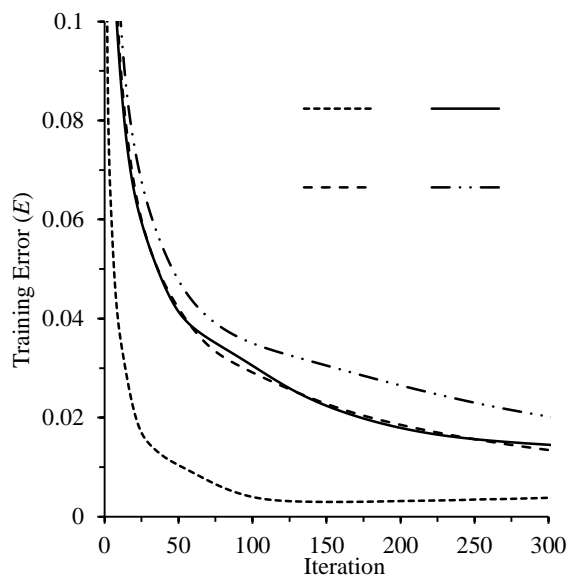
Figure 5 depicts the training set and test set classification (i.e., recognition) accuracy at different iterations for different batch sizes. It is observed that recognition accuracy is improved with iteration for both training and test sets rapidly at lower iteration values (e.g., up to 100). Training set recognition accuracy was also improved for higher iteration values coinciding minimization of E, since CNN are trained with bit values of the patterns from training set images. On the other hand, test set accuracy did not coincide with training set accuracy because its patterns were unseen by CNN during training. It is notable that the accuracy on test set is more desirable which indicates the generalization ability of a system.
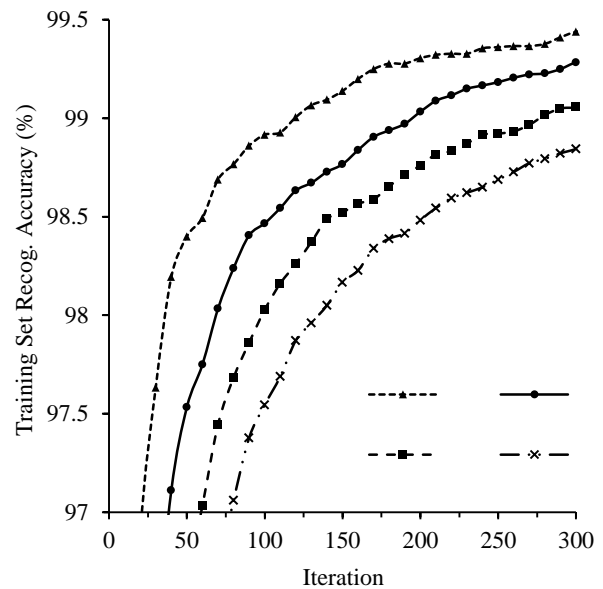
(a) Effect of iteration and batch size (BS) on training set recognition accuracy (%).
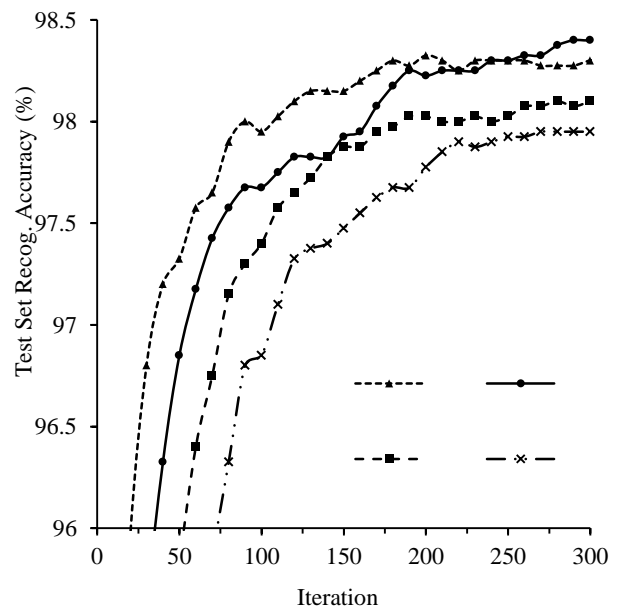
(b) Effect of iteration and batch size (BS) on test set recognition accuracy (%).

Fig.5. Training and test set recognition accuracy of Bengali numeral recognition for different iteration and batch sizes

Fig.4. Effect of Batch Size (BS) on Training Error (*E*).

Table 1 compares required training time, training error, training set recognition accuracy and test set recognition accuracy after 300 iterations for different BS values. With minimum training error, the best training set recognition accuracy (i.e., 99.44%) was for BS = 25. On the other hand, the best test set recognition accuracy (i.e., 98.40%) was achieved for BS = 50; although at this stage training set recognition accuracy was worse than that for BS = 25. Moreover, larger BS required less time than smaller BS values. To complete 300 iterations, total training time was 123.37 and 146.42 minutes for BS values 50 and 25, respectively. In general, an iteration required 28.80, 24.45, 22.54 and 21.86 seconds for BS values 25, 50, 75 and 100, respectively. From the table it is notable that for fixed 300 iterations, training with BS = 25 requires remarkable time than for BS = 100. For larger BS, CNN was updated once for a large number of training patterns and took relatively smaller time to train. Finally, batch size has a remarkable effect on training time as well as training and test set accuracy.

Table 1. Performance evaluation of different Batch Size (BS) after 300 iterations.

Batch Size	Time in Minutes	Training Error (E)	Training Set Rec. Accuracy	Test Set Rec. Accuracy
25	146.42	0.003796	99.44%	98.30%
50	123.37	0.014506	99.28%	98.40%
75	113.35	0.013449	99.06%	98.10%
100	109.77	0.020205	98.84%	97.95%

We have observed recognition accuracy of the system for various fixed number of iterations and best test set recognition accuracy was 98.45% (misclassifying 62 cases out of total 4000 test patterns) at iteration 360 for BS = 50. At that point, the method misclassified 114 cases out of 18000 training patterns showing accuracy rate of 99.37%. Table 2 shows the confusion matrix of test set samples at that point. From the table it is observed that the proposed method worst performed for the numerals "১" and "৯" truly classifying 383 and 385 cases respectively, out of 400 test cases of each one. Among the Bengali numerals, these two numerals seem to be most similar even in printed form. Numeral "১" recognized as "৯" in three cases; on the other hand "৯" recognized as "১" in 11 cases. Similarly, in the Bengali handwritten numeral script, "৫" and "৬" looks similar; therefore in seven cases system misclassified those as one another. It is notable that diverse writing styles enhance confusion in several numerals. But the proposed method shows best performance for numerals "২" and"৪", truly classifying all 400 test samples of each numeral. Table 3 shows some handwritten numeral images from total 62 misclassified images. It is observed from the table that due to large variation in writing styles, such images are difficult to correctly recognize even by human. All other images are also found ambiguous and misclassification by the system is acceptable.

Table 2. Confusion matrix for test samples of Bengali handwritten numerals. Total samples are 4000 having 400 samples for each numeral.

Bengali Num.	Total samples of a particular numeral classified as									
	০	১	২	৩	৪	৫	৬	৭	৮	৯
০	397	1	0	0	0	0	1	1	0	0
১	0	383	0	0	6	0	6	2	0	3
২	0	0	400	0	0	0	0	0	0	0
৩	0	0	0	396	0	0	3	0	1	0
৪	0	0	0	0	400	0	0	0	0	0
৫	1	1	1	0	2	389	4	2	0	0
৬	0	0	0	6	0	3	391	0	0	0
৭	0	0	1	0	1	0	0	398	0	0
৮	0	0	0	0	0	0	1	0	399	0
৯	0	11	0	1	0	0	0	3	0	385

Table 3. Sample handwritten numerals those are misclassified by CNN based classifier.
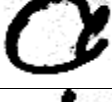
Handwritten Numeral Image	Image Classified as	Image in Category
	২	১
	৯	১
	০	৩
	৬	৫
	০	৫
	৬	৩
	১	৯

Table 4 compares the test set recognition accuracy of the proposed method with other prominent works of Bengali handwritten numeral recognition. The table also presents distinct features of individual methods in the form of feature selection, classification technique and dataset for brief overview of the methods. It is notable that proposed method did not employ any feature selection technique; whereas, existing methods use single or two stage of feature selections. The dataset used in our

study consists of sufficient number of training and test patterns. Without feature selection, proposed scheme is shown to outperform the existing methods. According to the table, proposed scheme achieved test set accuracy of 98.45%, on the other hand, the test set accuracy on the same dataset are 95.10% and 98.20% for the works of [4] and [12], respectively. The system also outperformed on the training set accuracy. The recognition scheme of [12] correctly recognizes 99.14% training samples; whereas the present scheme correctly recognizes 99.37%. It is notable that the work of [12] is the best performed existing method and used 173920 patterns for training, manipulating 17392 original training patterns. On the other hand, system of this study used only 18000 patterns for training and seem to be efficient. Besides recognition performance, the proposed system without feature selection is simpler than other existing methods. At a glance, proposed scheme trained with patterns from moderated pre-processing based images revealed as a good recognition system for Bengali handwritten numeral.

Table 4. A comparative description of proposed CNN based Bengali handwritten numeral recognition with contemporary methods

Work reference and Year	Feature Selection	Classification	Dataset; Training and Test Samples	Test Set Recog. Accuracy
Pal et al. [3], 2006	Water overflow from reservoir	Binary decision tree	Self-prepared; 12,000	92.80%
Basu et al. [4], 2005	Shadow feature and Centroid feature	MLPs with Dempster-Shafer technique	Samples from CVPR, ISI, India [18]; 6,000 and 2,000	95.10%
Wen et al. [10], 2007	Principal component analysis (PCA) and Kernel PCA	SVM	Dhaka automatic letter sorting machine; 16000 and 10000	95.05%
Das et al. [11], 2012	Genetic algorithm	SVM	CMATERdb 3.1.1 [19]; 6,000 and 2,000	97.70%
Bhattacharya and Chaudhuri [12], 2009	Wavelet filter at different resolutions	Four MLPs in two stages (three + one)	CVPR, ISI [18]; 173,920 and 4,000	98.20%
Wen and He [13], 2012	Eigenvalues and eigenvectors	Kernel and Bayesian discriminant (KBD)	Dhaka automatic letter sorting machine; 45,000 and 15,000	96.91%
Nasir and Uddin [14], 2013	K-means clustering Bayes' theorem and Maximum a Posteriori	SVM	Self-prepared: 300	96.80%
Proposed Scheme	No	CNN	CVPR, ISI [18]; 18,000 and 4,000	98.45%

Table 5. Confusion matrix produced for test samples of Bengali-English mixed numerals for 20 classes.

| Numeral	Total samples of a particular numeral classified as																			
	০	১	২	৩	৪	৫	৬	৭	৮	৯	0	1	2	3	4	5	6	7	8	9
০	395	1	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0
১	0	376	1	0	1	2	1	1	0	15	0	0	0	0	0	3	0	0	0	0
২	0	1	393	0	0	2	0	1	0	1	0	0	2	0	0	0	0	0	0	0
৩	1	0	0	392	0	4	1	0	2	0	0	0	0	0	0	0	0	0	0	0
৪	0	2	0	0	396	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
৫	1	0	0	0	4	387	1	2	3	0	0	0	0	0	0	0	1	0	0	1
৬	0	1	0	4	0	5	389	0	0	1	0	0	0	0	0	0	0	0	0	0
৭	0	0	0	0	3	0	0	396	0	0	0	0	0	0	1	0	0	0	0	0
৮	0	1	0	0	0	0	2	0	396	0	0	0	0	0	0	0	1	0	0	0
৯	0	23	0	1	1	0	2	4	0	369	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	978	0	0	0	0	0	0	1	1	0
1	0	0	0	0	0	0	0	0	0	0	0	1132	1	0	0	0	1	0	1	0
2	0	0	0	0	0	0	0	0	0	0	1	0	1020	2	1	0	4	3	1	0
3	0	0	0	0	0	0	0	0	0	0	0	0	2	1004	0	2	0	1	1	0
4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	971	0	3	0	1	6
5	0	1	0	0	0	0	0	0	0	0	1	0	0	7	0	880	2	0	1	0
6	0	0	0	0	0	1	0	0	0	0	4	2	0	0	1	0	947	0	3	0
7	0	0	0	0	0	0	0	0	1	0	0	4	3	1	0	0	0	1014	1	4
8	0	0	0	0	0	0	0	0	0	0	2	0	1	0	3	2	0	3	961	2
9	0	0	0	0	0	0	0	0	0	0	1	3	0	1	8	3	0	1	6	986

B. Bengali-English Mixed Numeral Recognition

Figure 6 portrays the training and test set accuracies for both 16 class and 20 class classifier varying iteration from 10 to 300. It is observed from the figure that accuracies increased rapidly in all four cases in initial training period e.g., up to 150 iterations. It is also notable from the figure that accuracy for 16 class is always better than 20 class case in both training and test sets.

Table 5 shows the confusion matrix of test set samples after fixed 1000 iterations for 20 class case. It is observed from the table that total 111 patterns of ISI Bengali dataset are misclassified. On the other hand, total 107 patterns out of MNIST 10000 patterns are misclassified showing better recognition accuracy than Bengali numerals. MNIST patterns are already preprocessed and easily distinguishable from Bengali numerals; therefore, only three patterns are misclassified as Bengali numerals. The proposed method shows worst performance for "9" from MNIST, misclassifying total 23 samples but all with MNIST numerals. On the other hand, it is also observed from the table that the proposed method worst performed for the numeral "২" and 369 cases it classified truly out of 400 test cases, misclassifying 23 cases as "১". Finally, in case of 20 class classifier, the proposed scheme misclassified 218 test samples (out of 14000 cases) and only 543 training samples (out of 78000 samples) showing recognition accuracy of 98.44% and 99.30%, respectively.

Table 6 compares the outcome of the proposed method with other works of handwritten Bengali-English mixed numeral recognition for both 16 and 20 class cases. Ref. [15] and [16] only considered 16 class case. Table 6 also presents distinct features of individual methods. It is notable that proposed method did not employ any feature selection technique whereas some existing methods uses one or two feature selection methods. Without feature selection, proposed scheme is shown to outperform the existing methods. In case of 16 class classifier, proposed scheme achieved test set recognition accuracy of 98.71% after 670 iteration with BS=100. On the other hand, the test set recognition accuracy is 98.47%, 87.24% and

92.10% for the works of [12], [15] and [16] respectively. In case of 20 class classifier, the test set accuracy of [12] is only 69.24%, whereas proposed CNN based scheme achieved test set accuracy of 98.44%. It is notable that 20 class system is more applicable than 16 class system. And therefore, the proposed method is more acceptable than that of other contemporary methods.
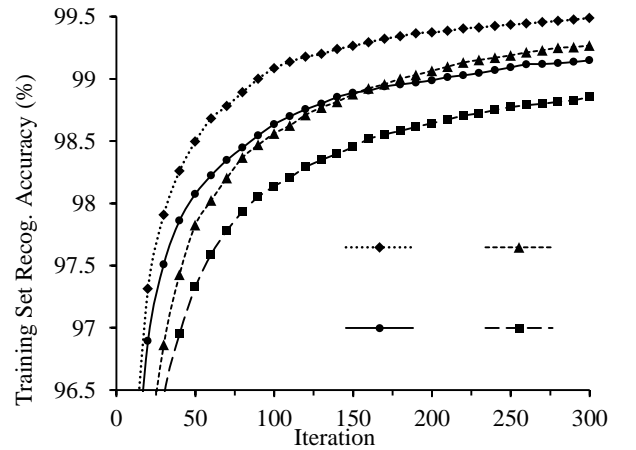
(a) Effect of iteration and batch size (BS) on training set accuracy (%)
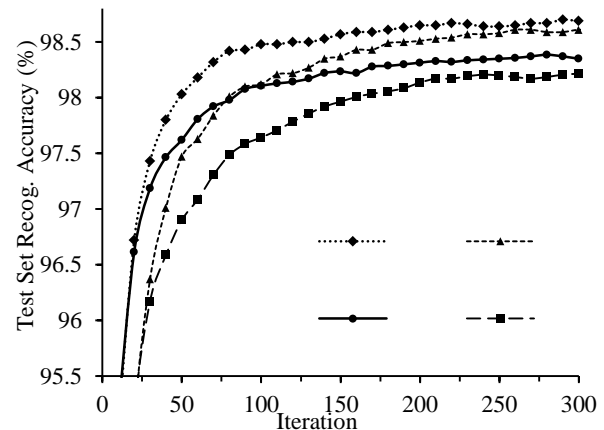
(b) Effect iteration and of batch size (BS) on test set accuracy (%)

Fig.6. Training and test sets recognition accuracy of Bengali-English mixed numeral recognition for iteration and different BS.

Table 6. A comparative description of proposed CNN based Bengali-English mixed recognition with contemporary methods.

Work reference and Year	Feature Selection	Classification	Class	Dataset; Training and Test Samples	Test Set Rec. Accuracy
Mustafi et al. [15], 2004	Water overflow from the reservoir	MLPs	16	Self-prepared; 2,400 and 2,800	87.24%
Vazda et al. [16], 2009	No	MLPs	16	Self-prepared; 8,690 and 6,406	92.10%
Bhattacharya and Chaudhuri [12], 2009	Wavelet filter at different resolutions	MLPs in two stages	16	CVPR, ISI and MNIST [18]; 86,000 and 14,000	98.47%
			20	CVPR, ISI and MNIST [18]; 108,000 and 14,000	69.24%
Proposed Scheme	No	CNN	16	CVPR, ISI and MNIST [18]; 54,000 and 10,000	98.71%
			20	CVPR, ISI and MNIST [18]; 78,000 and 14,000	98.44%

V. Conclusion

CNN has the ability to recognize visual patterns directly from image pixels with minimal pre-processing. This is why, in this paper, a CNN structure is investigated without any feature selection for handwritten Bengali and Bengali-English mixed numeral recognition. A moderated pre-processing is employed to improve recognition accuracy. The proposed recognition method has been tested on a large sized handwritten benchmark numeral dataset and outcome has been compared with existing prominent methods. The illustrated results in the previous section reveal that the training set accuracy and test set accuracy of the proposed method is significantly higher than those of other methods. Therefore, the proposed method outperformed the existing methods with respect to both training and test set accuracy.

In this study, the most common CNN architecture is considered and is achieved acceptable performance. Different CNN architectures varying number of convolutional and subsampling layers along with different kernel sizes may give better performance and remain as a scope for future study.

Acknowledgment

References

[1] R. Plamondon and S. N. Srihari, "On-line and off-line handwritten recognition: A comprehensive survey," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, pp. 62-84, 2000.

[2] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document Recognition,"*Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278–2324, November 1998.

[3] U. Pal, C. B. B. Chaudhuri and A. Belaid, "A System for Bangla Handwritten Numeral Recognition," *IETE Journal of Research, Institution of Electronics and Telecommunication Engineers*, vol. 52, no. 1, pp. 27-34, 2006.

[4] S. Basu, R. Sarkar, N. Das, M. Kundu, M. Nasipuri and D. K. Basu, "Handwritten Bangla Digit Recognition Using Classifier Combination Through DS Technique," *LNCS*, vol. 3776, pp. 236–241, 2005

[5] R. Kumar and K. K. Ravulakollu, "Offline Handwritten Devnagari Digit Recognition", *ARPN Journal of Engineering and Applied Sciences*, vol. 9, no.2, pp. 109-115, Feb 2014.

[6] R. Kumar and K. K. Ravulakollu, "Handwritten Devnagari Digit Recognition: Benchmarking on New Dataset", *Journal of Theoretical and Applied Information Technology*, vol. 60, no.3, pp. 543-555, Feb 2014.

[7] P. Singh, A. Verma and N. S. Chaudhari, "Devanagri Handwritten Numeral Recognition using Feature Selection Approach", *I.J. Intelligent Systems and Applications*, MECS Press, vol. 6, no.12, pp. 40-47, Nov 2014.

[8] M. R. Bashar, M. A. F. M. R. Hasan, M. A. Hossain and D. Das, "Handwritten Bangla Numerical Digit Recognition using Histogram Technique," *Asian Journal of Information Technology*, vol. 3, pp. 611-615, 2004.

[9] M. M. R. Khan, S. M. A. Rahman and M. M. Alam, "Bangla Handwritten Digits Recognition using Evolutionary Artificial Neural Networks*" in Proc. of the 7th International Conference on Computer and Information Technology (ICCIT 2004),* 26-28 December, 2004, Dhaka, Bangladesh.

[10] Y. Wen, Y. Lu and P. Shi, "Handwritten Bangla numeral recognition system and its application to postal automation," *Pattern Recognition*, vol. 40, pp. 99-107, 2007.

[11] N. Das, R.Sarkar, S. Basu, M. Kundu, M. Nasipuri and D. K. Basu, "A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application," *Applied Soft Computing*, vol. 12, pp. 1592-1606, 2012.

[12] U. Bhattacharya and B. B. Chaudhuri, Handwritten numeral databases of Indian scripts and multistage recognition of mixed numerals, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 444-457, 2009.

[13] Y. Wen and L. He, "A classifier for Bangla handwritten numeral recognition," *Expert Systems with Applications*, vol. 39, pp. 948-953, 2012.

[14] M. K. Nasir and M. S. Uddin, "Hand Written Bangla Numerals Recognition for Automated Postal System," *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 8, no. 6, pp. 43-48, 2013.

[15] J. Mustafi, K. Roy, and U. Pal, "A System for Handwritten Bangla and English Numeral Recognition," *National Conference on Advanced Image Processing and Networking,* India, 2004.

[16] S. Vazda et al., "Automation of Indian Postal Documents written in Bangla and English," *International Journal of Pattern Recognition and Artificial Intelligence*, World Scientific Publishsing, pp. 1599-1632, 2009.

[17] Md. Mahbubar Rahman et al. "Bangla Handwritten Character Recognition using Convolutional Neural Network," *I.J. Image, Graphics and Signal Processing*, vol. 7, no. 8, pp. 42-49, 2015.

[18] Off-Line Handwritten Bangla Numeral Database, Available: http://www.isical.ac.in/~ujjwal/, accessed July 12, 2015.

[19] CMATERdb 3.1.1: Handwritten Bangla Numeral Database, Available:http://code.google.com/p/cmaterdb/, accessed July 12, 2015.

[20] T. Liu et al., "Implementation of Training Convolutional Neural Networks," *arXiv preprint arXiv:1506.01195*, 2015.

[21] Feature extraction using convolution, UFLDL Tutorial. Available: http://deeplearning.stanford.edu/, accessed November 12, 2015.

[22] J. Bouvrie, "Notes on Convolutional Neural Networks," Cogprints. Available: http://cogprints.org/5869/, accessed November 12, 2015.

Authors' Profiles

M. A. H. Akhand received his B.Sc. degree in Electrical and Electronic Engineering from Khulna University of Engineering and Technology (KUET), Bangladesh in 1999, the M.E. degree in Human and Artificial Intelligent Systems in 2006, and the Doctoral degree in System Design Engineering in 2009 from University of Fukui, Japan. He joined as a lecturer at the Department of Computer Science and Engineering at KUET in 2001, and is now a Professor. He is also head of Computational Intelligence Research Group of this department. He is a member of Institution of Engineers, Bangladesh (IEB). His research interest includes artificial neural networks, evolutionary computation, bioinformatics, swarm intelligence and other bio-inspired computing techniques.

Mahtab Ahmed received his B.Sc. degree in Computer Science and Engineering (CSE) from Khulna University of Engineering and Technology (KUET), Bangladesh in 2014. At present he is serving as a Lecturer in the Dept. of CSE, KUET. His research interest includes pattern recognition, speech and image processing and advanced machine learning.

M. M. Hafizur Rahman received his B.Sc. degree in Electrical and Electronic Engineering from Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh, in 1996. He received his M.Sc. and Ph.D. degree in Information Science from the Japan Advanced Institute of Science and Technology (JAIST) in 2003 and 2006, respectively. Dr. Rahman is now an assistant professor in the Dept. of Computer Science, KICT, IIUM, Malaysia. Prior to join in the IIUM, he was an associate professor in the Dept. of CSE, KUET, Khulna, Bangladesh. He was also a visiting researcher in the School of Information Science at JAIST and a JSPS postdoctoral research fellow at Research Center for Advanced Computing Infrastructure, JAIST & Graduate School of Information Science (GSIS), Tohoku University, Japan in 2008 and 2009 & 2010-2011, respectively. His current research include parallel and distributed computer architecture, hierarchical interconnection networks and optical switching networks. Dr. Rahman is member of IEICE of Japan and IEB of Bangladesh.