Journal of Network and Innovative Computing
ISSN 2160-2174 Volume 1 (2013) pp. 33-42

Individuality of Isolated Bangla Numerals

Chayan Halder[1] **and Kaushik Roy**[2]

[1]Department of Computer Science, West Bengal State University,
Barasat, Kolkata 700126, WB, India
chayan.halderz@gmail.com

[2]Department of Computer Science, West Bengal State University,
Barasat, Kolkata 700126, WB, India
kaushik.mrg@gmail.com

Abstract: **Writer Identification and Verification is a study in the field of Computer Vision and Pattern Recognition. The identification and verification of writer through the analysis of handwriting has significant role for many critical judicial decisions. The forensic experts often encounters this type of problem to identify or verify writer from given documents or part of documents. In todays high-tech world an automated system with the ability to identify and verify writers can play vital role in the judicial system. In current times, to the best of our knowledge there is no such automated system of Writer Identification and Verification on Indic script like Bangla, Oriya, Telugu etc. Analysis of individuality of handwritten isolated Bangla numerals are presented here. It has a great prospect not only in Writer Identification and Writer Verification but also in Graphological Analysis and also in various fields of Forensic Science based on handwritten documents etc. As there is no such standard Bangla writer database, we have collected data samples consisting of total 4500 numerals from 90 writers with 5 sets from each writer. After collecting and extracting characters from filled in forms, 64 and 400 dimensional feature vectors are computed on numerals based on directional chain code and gradient of the images. In our experiment we have used LIBLINEAR and MLP classifier of WEKA environment. We have computed and analyzed the Individuality of each numeral and observed that the numeral *PANCH* (5) is the most individual than other numerals except when MLP classifier is used for classification on 400 dimensional feature set where numeral *DUI* (2) is the most individual. It has also been observed that numeral *SUNNO* (0) has the least individuality. We have also done the writer identification with all the numerals and using 64 dimensional feature we have obtained 97.07% accuracy for LIBLINEAR classifier and 94.62% accuracy for MLP classifier with all writers. For 400 dimensional feature with LIBLINEAR classifier 96.5% writer identification accuracy has been achieved.**

Keywords: Individuality of Handwriting, Writer Identification, Bangla Handwriting, WEKA, MLP, LIBLINEAR.

I. Introduction

Its been a decade or two that writer Identification and verification is an active area of research in document level analysis. It has such potential that it can be used in forensic science, banking, graphology etc. As the handwritten numerals carries additional information about the personality and characteristics of the writer compared to electronic or printed numerals, there exists a high possibility to authenticate and identify the writer. Writer identification rests on the hypothesis that there exists a certain degree of stability in the writing style of an individual which makes it possible to identify the writer even from his/her handwritten numerals. Our objective is to verify this hypothesis using our experiment, testing and results. The task of writer Identification focuses on extracting the characteristic attributes like character shape, size, and slant angle etc. from Bangla handwriting which can be done by averaging out the variation in handwriting of different individuals. These attributes are used by the expert document analyzers to quantitatively establish individuality of a writing style.

There exist various pieces of works in literature on writer identification/verification of non-Indic scripts [1, 2, 3, 4, 5, 8, 6, 10, 9, 7, 15, 11]. In [2] Said et al. have proposed a multi-channel Gabor filter and gray scale co-occurrence matrix (GSCM) based approach for writer identification on non-uniform skewed handwritten images. 150 documents from 10 writers have been used for their work. They have also used two classification techniques viz Weighted Euclidian Distance (WED) and K-Nearest Neighbor (K-NN) classifiers. An accuracy of 96.0% has been achieved by them. For their work the WED classifier has given better performance than the K-NN. In [3] they have used 1000 test documents from 40 writers to develop a text independent writer identification system. They have used a texture analysis based approach and achieved an accuracy of 96.0%. Marti et al. in [4] have used 100 pages of text written by 20 writers as the data for their experiment. They have computed twelve features based on visible characteristics of the writings. K-nearest-neighbour classifier and Feed forward Neural network are being used for their experiment. They have achieved Identification accuracy of 87.8% and 90.7% respectively. Srihari et al. in [5] have used 1568 writers for their database and each writer is being asked to copy out the sample document three times. They have extracted Macro and Micro features from total text document, paragraphs, separated words and

even from characters. An accuracy of 98% has been achieved by them. In [6] on Individuality calculation of numerals they have used 1000 writers each with 3 sets of samples, thus having total 3000 samples each with 10 numerals. The WMR features composed of 72 local features and 2 global features: aspect ratio and stroke ratio of an entire character, along with clusterization algorithm on the resultant feature has been used for their work purpose. They have used manual clustering and the Bhattacharya Distance to measure the individuality of numerals for verification purpose. In their work the numeral 5 has the most individuality for both verification and identification. The 79.6% and 78.4% accuracy for verification and identification respectively has been achieved by them for all accumulated numerals. In another work of [7], on statistical model for writer verification, they have used the same database, Macro and Micro features like [5]. In this work they have been focused on extracting the characteristics from the questioned and known documents and computing corresponding differences. Also they have computed likelihoods for two classes assuming statistical independence of the distances i.e. the conditional probabilities for the differences, has been estimated using parametric probability densities either Gaussian or Gamma. Log Likelihood Ratio (LLR) for the same and Cumulative Distribution Functions (CDFs) of the LLRs has been used to calibrate the LLR values into a nine-point scale. They have achieved an accuracy of 94.6% for same writers and 97.6% for different writers. Bensefia et al. in [8] have used 88 sample documents from 88 writers. Two types of testings have been performed for that work. They have achieved an accuracy of 97.7% for their work. In the work of [9], Bensefia et al. have used two type of database for their work. The first one has been collected from 88 writers database of [8] and the second one has been collected from 39 writers from original correspondence of Emile Zola.

Schomaker et al. in [10] have used 500 documents from 250 writers for their experiment. The edge-based directional features are being used for identification procedure. Schomaker et al. [12] have used Quill feature, based on directional ink trace width measurement on pixel of contours for their work on Writer Identification. They have used a total of four data sets two of which are medieval handwriting datasets namely Dutch charter dataset and the English diverse dataset; the other two databases of contemporary handwritings are IAM [13] and Firemaker [14] datasets. The Quill feature $p(\varphi, w)$ has been consists of few simple parts that together form a powerful method for writer identification like: contour tracing, angle measurements, width measurements and calculation of a probability distribution. They have also tried a complex variant of the feature Quill-Hinge feature. After the feature extraction they have chosen the nearest-neighbor (instance-based) classifier for classification. They have achieved 97% accuracy for their approach. Siddiqi et al. [15] have used 50 documents from same number of writers for their work. They have used a local approach, based on the extraction of characteristics that are specific to a writer. Bayesian classifier has been used in their work and an identification accuracy of 94% has been achieved.

Indic scripts like Bangla, Oriya, Telugu and Kannada etc. are very popular though there exist very few works on these Indic scripts. Some works of Pal et al. on Oriya script [16]; Chanda et al. on Telugu script [17]; Hangarge et al. on Kannada script [18]; Garain et al., Pal et al., Das et al. and Halder et al. [19, 20, 21, 22] on Bangla script are among the very few available works in literature on Indic script. U.Pal et al. [16] have used the directional chain-code and curvature feature for their work. SVM classifier has been used for the work and they have achieved an accuracy of 94% on writer identification. Hangarge et al. in their work [18] on Kannada script have used 3 different kinds of features like Discrete Cosine Transform (DCT), Gabor Filtering and Gray Level Co-Occurrence Matrix (GLCM). 20 writer have been used for their datasets. The K-NN classifier has been used for the work and they have achieved an accuracy of 77.0% for DCT, 88.5% for Gabor Filtering and 79.5% for GLCM. They have presented that the Gabor Filtering has more potential in this respect for Kannada script than DCT and GLCM.

In the work of Garain et al. [19] on Bangla characters, they have used 60 documents from 20 writers. Gradient based contour encoding feature and 192 bit feature vector have been used in this respect. The K-means clustering has been used to get an accuracy of 40% for identification. Pal et al. have worked on the database of 204 documents from 102 writers for text independent writer identification using the Bangla characters [20]. They have used 400 dimensional gradient features with SVM classifier to achieve an accuracy of 95.19%. In the work of Das et al. [21] they have used their own Database consisting of 55 writers. Each writer has four sample documents on two different topics. Radon transform projection profile has been used as the feature for their work. In the work of Halder et al. [22] for individuality calculation on Bangla numerals, the numbers of writers and database are same as the present work database. In this work the highest individuality accuracy has been achieved for the numeral 5 and for writer identification 96.5% accuracy has been achieved.

To the best of our knowledge there exists no work in literature on the Individuality of Bangla numerals. Only the work of Garain et al. [19]; that is on individuality of handwriting based on Bangla characters is comparable with our work, if only the individuality of numerals is considered. Here for our work the 400 dimensional feature based on gradient has been used. A comparative analysis of different writer identification schemes has been described in Table 1.

The subsequent part of the paper is organized as follows: section II describes the properties of Bangla numerals. In section III data collection is described. Pre-processing strategies and feature extraction are discussed in section IV and V respectively. Section VI describes about WEKA tool followed by results in section VII. At last we concluded in the section VIII.

II. Properties of Bangla Numerals

Bangla, the second most popular language in India and the sixth most popular language in the world, is an ancient Indo-Aryans language [23]. Bangla script alphabet is used in texts of Bangla, Assamese and Manipuri languages. Bangla is the national language of Bangladesh and also the official language of the state West Bengal and Tripura in India. Bangla has 11 vowels and 39 consonants and a number of complex

Table 1: Details of various schemes on writer identifications

System	Data (writer)	Script /Language	Feature	Classi-fiers	Results (%)
Said et al. [2] [3]	150 (10) [2], 1000 (40) [3]	English	Texture features using Gabor filter and gray scale co-occurrence matrix	Weighted Euclidian Distance (WED) and K-Nearest Neighbor (K-NN)	96.00
Marti et al. [4]	100 (20)	English	visible characteristics	K-nearest-neighbour classifier and Neural network	90.70
Bensefia et al. [8] [9]	88 (88) ZOLA-BASE, PSI-DataBase	French	Textual and grapheme based	Multiple sequential invariant clustering and Cosine similarity	97.70 [8], 93.30 [9]
Schomaker et al. [10] [11] [12]	500 (250) IAM database, Firemaker set, Unipen database	Dutch, English	Edge based directional PDFs , Quill, Quill-Hinge	K-NN, neural network, SVM	75.00 [10], 87.00 [11], 97.00 on IAM [12]
Siddiqi et al. [15]	The IAM database, RIMES database	English , French	Writer Specific local features	Bayesian classifier	94.00
Pal et al. in [16] [20]	200 (100) [16], 208 (104) [20]	Oriya, Bangla	Directional chain-code and curvature feature	SVM classifier	94.00 [16], 95.19 [20]
Hangarge et al. [18]	400 (20)	Kannada	Discrete Cosine Transform, Gray Scale Co-Occurrence Matrices (GSCM)	K-NN	DCT: 77.00, Gabor Energy: 88.50, GLCM: 79.50
Das et al. [21]	BESUS Database	Bangla	Radon transform projection profile	Euclidean distance	83.63
Srihari et al. [5] [6]	3000 (1000)	English	Macro and Micro features [5], WMR features [6]	Multi-Layer Perceptron [5] Bhattacharya Distance [6]	98.00[5], 78.40 [6]
Garain et al. [19]	60 (20)	ISI database Bangla	Gradient based contour encoding	K-means clustering, Euclidean distance	13.00, individuality 18.33
Halder et al. [22]	450 (90)	Bangla	Gradient feature	LIBLINEAR classifier	96.50 individuality: 35.90

numeral with its uniqueness can able to identify a writer with certain amount of accuracy. For this very reason we worked on the Individuality of Bangla numerals.

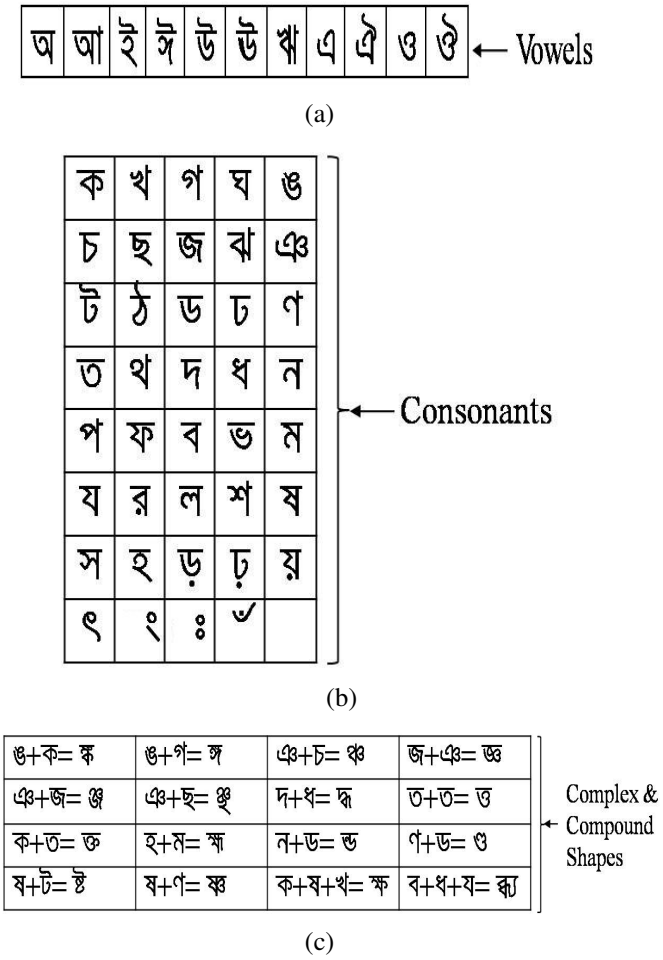


Figure. 1: Example of (a) Bangla Vowels (b) Bangla Consonants (c) Some Bangla Complex and Compound characters

and compound shapes that can be formed by combination of vowels and consonants with consonent(s). In the Figure 1 example of Bangla alphabets and some complex and compound characters are shown. To get an idea of Bangla numerals and their variability in handwriting, five sets of handwritten Bangla numerals of different writers are shown in Figure 2. Each person writes differently from other and each person write differently from himself/herself but intuitively, the intra-writer variation (the variation within a person's handwriting samples) is quite less compared to the inter-writer variation (the variation between the handwriting samples of two different people). There are two main point of concern while comparing handwritings: the variability of the handwriting of the same individual and the variability of the handwriting from one individual to another. These two variations can be seen when several individuals are asked to write the same numerals many times (in our work it is 5 times). For example if we consider Figure 2 which exhibits the numeral 5 and 0 of different writers, where the variation is much more in between same numerals than in Figure 3 when a single writer writes the numerals. From the figure it can be noted that the variation in writing for numeral 5 is more than the numeral 0. It also describes that for every writer there exist a certain amount of uniqueness for every character they write. For this reason using this uniqueness of each numeral the identification of writer can be possible. This means each

III. Data Collection

As we are interested in computing individuality of handwriting, we have designed a sample document consisting of all Bangla alphabets (vowel and consonants), numerals and vowel modifiers. A total number of 120 writers, predominantly students, were asked to copy-out the printed characters in the particular box area of the sample document form. Total 5 documents are being given to each writer for data collection. Most of the writers are in between age group 17-25. We also have writers in between age group 30-45 and even 50-60. Most of the writers are right handed. Out of 120 writers until now we have managed to collect full 5 sets of data from 90 writers (used for current work purpose) and for remaining writers till now we have collected fewer numbers of sets. Each set contains 10 Bangla numerals and 51 Bangla alphabets and 10 Bangla vowel modifiers. We have a total of 26367 Bangla alphabets, 5170 numerals and 5170 vowel modifiers. An example of our designed character sample collection document form is shown in Figure 4. There exists no boundary for writers regarding the type of pen and ink they use. We scanned these documents using a flatbed scanner

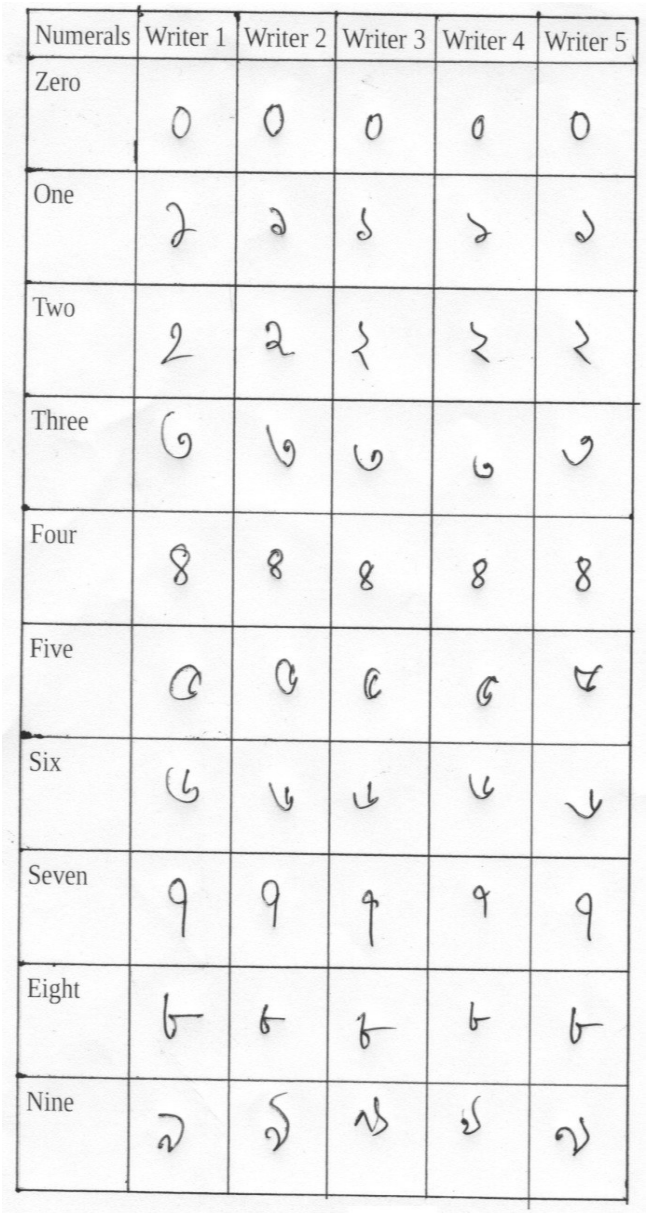


Figure. 2: Sample of all Bangla handwritten numerals from 5 writers

for digitization. The images are in gray tone and digitized at 300/600 dpi and stored in Tagged Information File Format (TIFF). In this work we have used only the numerals for the writer identification and individuality. For our current work we have used the 300 dpi gray toned TIFF format image.

IV. Pre-Processing

The digitized images have been binarized using a global binirization method [23]. Then a character extraction technique has been used to extract the characters from the digitized form and they are being stored in gray mode.

A. Character extraction Technique

This technique is used for extraction of each individual character from the document form of handwritten characters. The steps are:

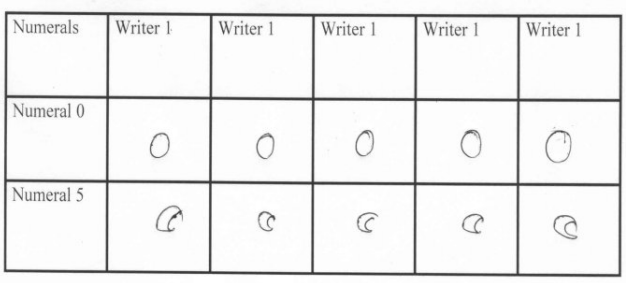


Figure. 3: Sample of 5 instances of numeral 0 and 5 from same writer.

Firstly, the global binirization of the whole document has been carried out. Then maximum run length has been computed on horizontal and vertical histogram of that document form. Using the maximum run length of horizontal and vertical histogram, we have identified the horizontal lines and vertical lines of the document form. After the identification of vertical and horizontal lines we have deleted those lines from the document form image to get an image which contains only the suggestive characters and the original handwritten characters. Then using the horizontal and vertical line information we have calculated the top corner point values of each block. These points have been used to remove the suggestive characters of each block. After this, bounded box for each handwritten character has been calculated and the gray values of characters have been stored for further processing.

V. Feature Extraction

In this work both 400 and 64 dimensional feature extraction techniques [23] have been used for individuality calculation and writer identification [23]. The 400 dimensional feature has given some encouraging results for different Indic script numeral recognition like the work of U. Pal et al. [24]. The 64 dimensional feature has been widely used in this type of pattern recognition problems as at the time of classification the 64 feature will take lesser resources and time. These are some of reasons we have opted for the 400 and 64 dimensional features to work with.

A. 400 Dimensional Feature Extraction

The following steps have been applied to obtain the 400 dimensional feature.

Step 1 At first the binirization of the input gray image is done.

Step 2 The normalization of the binary image is done. Here we normalize the image into 73 x 73 pixels.

Step 3 The binary image is then converted into a gray-scale image applying a 2 x 2 mean filtering 5 times.

Step 4 The gray-scale image is normalized so that the mean gray scale becomes zero with maximum

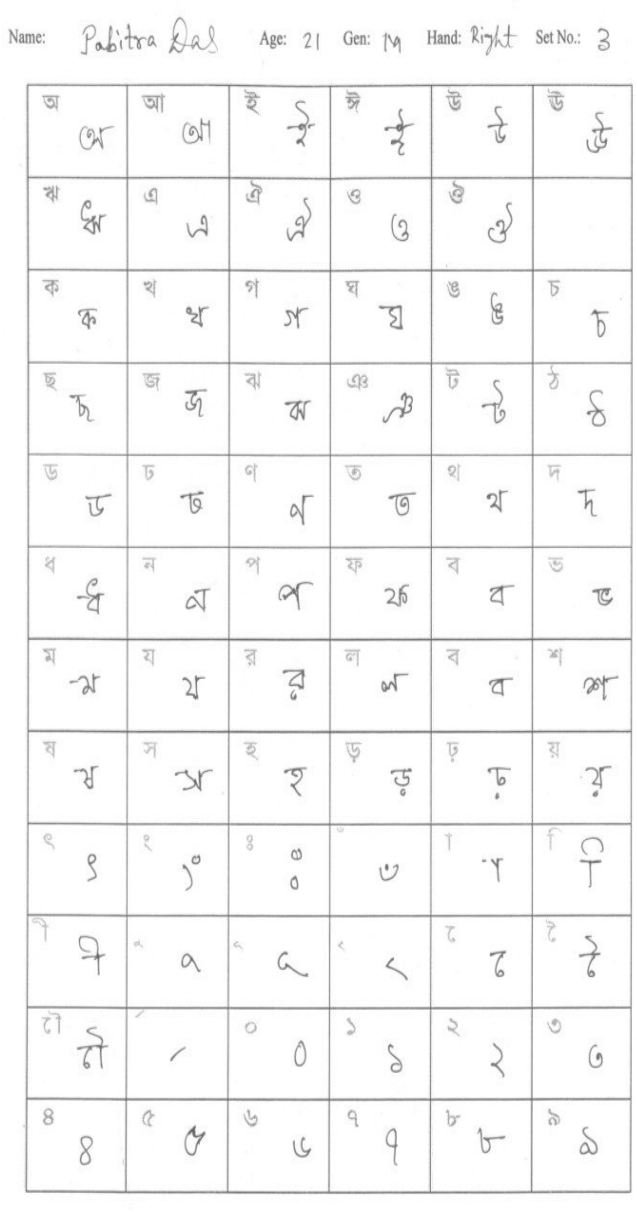


Figure. 4: Sample data collection form used for collection of Bangla Handwritten isolated characters and Vowel modifiers.

value 1.

Step 5 Normalized image is then segmented into 9x9 blocks.

Step 6 A Roberts filter is then applied on the image to obtain gradient image. The arc tangent of the gradient (strength of gradient) is quantised into 16 directions and the strength of the gradient is accumulated with each of the quantized direction. By strength of Gradient (f(x, y)) we mean

$f(x,y) = \sqrt{(\Delta u)^2 + (\Delta v)^2}$ and

by direction of gradient $(\theta(x,y))$ we mean,

$\theta(x,y) = tan^{-1}\Delta v/\Delta u$, where

$\Delta u = g(x+1, y+1) - g(x,y)$, and

$\Delta v = g(x+1, y) - g(x, y+1)$, and

$g(x,y)$ is a gray scale point (x,y).

Step 7 Histograms of the values of 16 quantized directions are computed in each of 9 x 9 blocks.

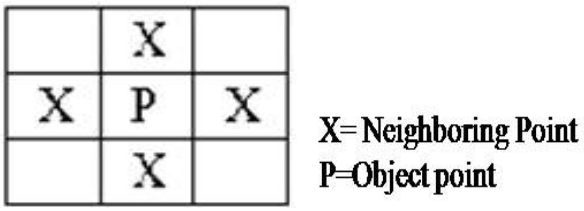
Step 8 9x9 blocks is down sampled into 5x5 by a Gaussian filter. Thus, we get 5x5x16 = 400 dimensional feature.

B. 64 Dimensional Feature Extraction

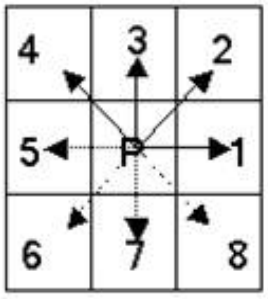
For the 64 dimensional feature extraction we first find contour points of the given two-tone image. Each object point in the image is to be surrounded by a 3 x 3 window. If any one of the four neighboring points of an object point within the window as shown in Figure 5(a) is a background point then this object point (P) is considered as a contour point. Otherwise it is a non-contour point. At first we form the bounding box covering the image of a sample numeral as shown in Figure 6(a-c). To get 64 dimensional features, this bounding box is divided into 7 x 7 blocks, as shown in Figure 6(c). In each of these blocks the direction code for each contour point is noted and frequency of each of four direction codes is computed. Here we use direction code of four directions only [directions 1 (horizontal), 2 (45 degree slanted), 3 (vertical) and 4 (135 degree slanted)]. As shown in Figure 5(b). We assume that direction code of direction 1 and 5 are same. Also, we assume that pair wise direction code 2 and 6, 3 and 7, 4 and 8 are same. Thus, in each block, we get an array of four integer values representing the frequencies of direction code in these four directions as shown in Figure 6(d). These frequencies are used as features. After down sampling the initial 7 x 7 blocks into 4 x4 blocks 64 (4x4x4) directions code features are obtained shown in Figure 6(e).
To normalize the above features (between 0 and 1) we have computed the maximum value of the histogram peaks in each direction from all the blocks. We then divided each of the above features by the maximum value of their respective direction to get the feature value between 0 and 1 [23].

VI. WEKA

WEKA is one of the widely used tools for data analysis and recognition in the area of machine learning [25]. The built in tools can be called from own Java code or using the weka.jar file of the package or directly from GUI interface. It contains tools for various applications like data pre-processing, classification, clustering, regression, association rules, visualization etc. For the current work the LIBLINEAR (Library for Large Linear Classification) and MLP (Multi-Layer Perceptron) have been used for classification. We have used these two classifiers as in [26] the LIBLINEAR and MLP have given some encouraging results for numeral recognition and MLP is one of the widely used classifier in this type of patter recognition problem. The LIBLINEAR is faster in terms of convergence criteria [26] and have already given some promising results in [22].



(a)



(b)

Figure. 5: (a) A 3x3 window. (b) Here, for a point P the direction code for its neighboring eight points are shown [23]

A. LIBLINEAR

The LIBLINEAR is a suitable linear classifier for most cases when the amount of data with instances or features to be classified is large enough. The convergence rate is much faster in comparison with other classifiers of WEKA for our dataset. We have used the L2-Loss Support Vector Machine (dual) as the SVM Type parameter of the LIBLINEAR. Both the Bias and Cost parameters are 1.0. The EPS (the tolerance of the termination criterion) is 0.01. For more details see [27].

B. MLP

The MLP is a layered feed forward network, pictorially represented with a directed acyclic graph. Each node represents an artificial neuron of the MLP, and the labels in each directed arc denote the strength of synaptic connection between two neurons and the direction of the signal flow in the MLP [23]. For pattern classification, the number of neurons in the input layer of an MLP is determined by the number of features selected for representing the relevant patterns in the feature space and the output layer is chosen by the number of classes in which the input data belongs. The neurons in hidden and output layers compute the sigmoid function on the sum of the products of input values and weight values of the corresponding connections to each neuron. Training process of an MLP involves tuning the strengths of its synaptic connections so that it can respond appropriately to every input taken from the training set. The number of hidden layers and the number of neurons in a hidden layer should be determined during training process [23]. For this work, the 400 and 64 dimensional features are used for individuality calculation and writer identification. The number of neurons for the hidden layer is chosen automatically (the default value) by the MLP classifier of the WEKA tool.
Most of the parameters for MPL classifier of WEKA tool are set to its default values for this work like the learning rate has been set to 0.3 and momentum to 0.2.

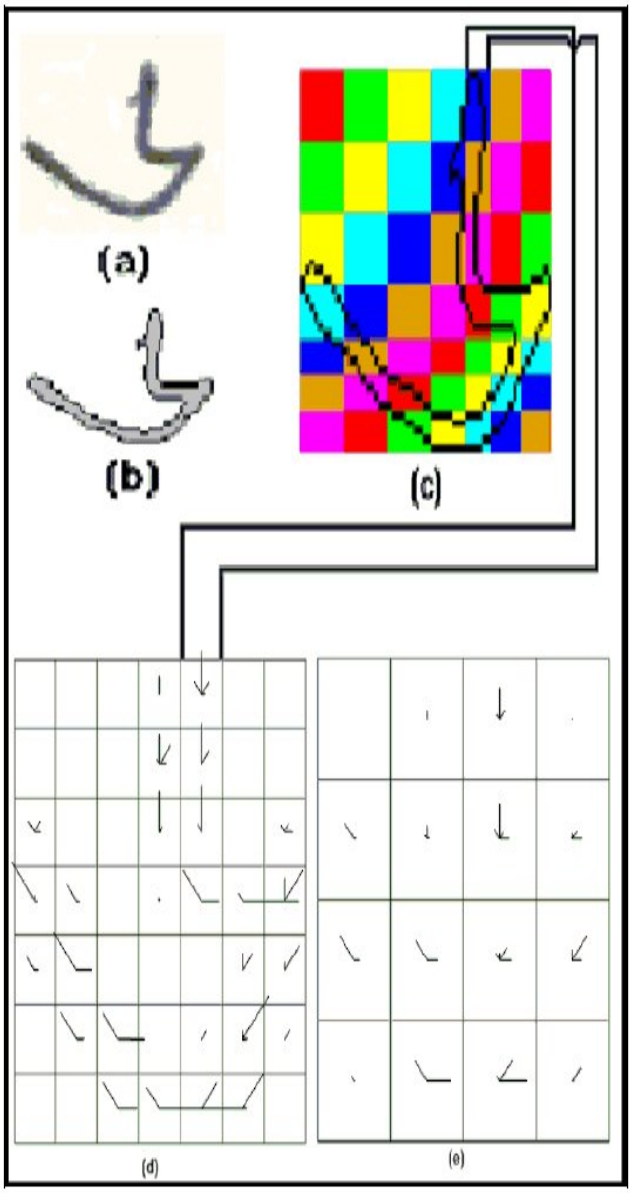


Figure. 6: Example of 64 Dimensional feature extraction (a) An image of a Bangla numeral. (b) Contour of the numeral (c) Minimum bounding box divided into 7 x 7 segmented blocks, (d) Block-wise direction code histogram of contour points, (e) Direction code Histogram of contour points of 4 x 4 blocks obtained through down sampling from 7 x 7 blocks [23]

VII. Results

The present work has been carried out on 4500 numerals from 450 documents written by 90 writers. We have used 5-fold cross validation scheme for computing individuality of numerals and also for writer identification. We have computed the individuality of each numeral for all the writers. Both 400 and 64 dimensional features have been used for the present work. To compute the individuality of each numeral, actually the writer identification accuracy of each numeral for all the writers has been computed. This makes sense as the identification accuracy of writers for any particular numeral is the measure of the uniqueness of the numeral for the given set of writers i.e. the individuality of that particular numeral. Then writer identification has been computed on all numerals. The results are discussed below:

A. Result on Individuality of Numerals

The computed individuality of numerals using 400 dimensional and 64 dimensional features have been described in Table 2 and Table 3 respectively. Results computed from both LIBLINEAR and MLP classifiers have been shown in the tables.

1) Result of 400 dimensional feature for Individuality:

The corresponding chart of the Table 2 has been showed in Figure 7. From the table and the figure it can be observed that the numeral *PANCH* (5) is most individual followed by numerals *CHOY* (6), *NOY* (9) and *DUI* (2) for the classifier LIBLINEAR. But in case of MLP classifier with the same 400 dimensional feature the most individual numeral is *DUI* (2) followed by *NOY* (9), *CHOY* (6) and *PANCH* (5).Though for both the classifiers numeral *SUNNO* (0) has the least individuality. The corresponding values of individuality for *PANCH* (5) and *SUNNO* (0) are 35.37% and 13.20% respectively in case of LIBLINEAR classifier with 400 dimensional feature. The individuality value of numeral *DUI* (2) for MLP classifier is 39.22% and for *SUNNO* (0) the value is 15.89% with the same feature.

2) Result of 64 dimensional feature for Individuality:

Figure 8 shows the corresponding chart of the Table 3. Both the table and the chart shows that the numeral *PANCH* (5) is most individual followed by numerals *CHOY* (6), *NOY* (9) and *DUI* (2) for both LIBLINEAR and MLP classifiers. Only change that can be observed is that for LIBLINEAR *CHOY* (6) is more individual than *NOY* (9) where for MLP it is in reverse order. For both the classifiers the least individual numeral is *SUNNO* (0). The individuality values for *PANCH* (5) and SUNNO (0) are 42.44% and 18.09% respectively for LIBLINEAR classifier and for MLP classifier the values are 43.41% and 15.89% in case of 64 dimensional feature.
From all the results of individuality it can be concluded that the numeral 5 is most consistent among different features and classifiers. Analyzing all the numerals that are shown in Figure 2, it can be said that numeral 0 has more similarity among different writers. But if we consider numeral 5 then the variation in writing is much higher among the writers. This is because the starting stroke angle, shape, writing pattern of numeral 5 is different among the writers and also the writing complexity of the numeral itself is higher. These are some of the reasons that numeral 5 is most stable in terms of individuality for different situations.

B. Result of Writer Identification

Using all the numerals for writer identification an accuracy of 96.5% has been achieved for 400 dimensional feature with LIBLINEAR classifier. We have achieved 97.07% and 94.62% accuracies for 64 dimensional feature with LIBLINEAR and MLP classifiers respectively. We have also experimented by varying the number of writer sets in this respect. The details of different identification results for different writer sets in case of 400 dimensional feature have been presented in Table 4. The results of 64 dimensional feature for LIBLINEAR and MLP classifiers have been showed in Table 5.

1) Result of 400 dimensional feature for Writer Identification:

In case of 400 dimensional feature, we have achieved an accuracy of 100% upto 20 writers for writer identification. The accuracy has been dropped to 98% for 40 writers. Strangely the drop rate in the accuracy has been quite higher for 50 writers and this rate continues upto 65 writers. From 70 writers the accuracy has been increased afterward. We have tried and tested several times but the results that we have got remained unchanged. Figure 9 has been used to describe using line graph, the variation in results that have been achieved for different writer sets.

2) Result of 64 dimensional feature for Writer Identification:

From Table 5, it can be observed that for 64 dimensional feature the writer identification accuracies are 98.97% and 98.04% less for both the classifiers in comparison with 400 dimensional feature if only 20 writers are considered. Although when the numbers of writers have been increased upto 45 writers the accuracies for both classifiers have also increased. From 45 writers to 70 writers the accuracy has been decreased for both the classifiers. But after 70 writers the accuracy has been increased for LIBLINEAR classifier but it has been decreased for MLP. From 85 to 90 writers the reverse has been occurred. The variations of writer identification results for LIBLINEAR and MLP classifiers have been presented in Figure 10 with comparative line graphs of LIBLINEAR and MLP classifiers.

C. Comparison of Results

The work of Garain et al. [19] on 60 documents from 20 writers; that is on individuality of handwriting based on Bangla characters can be compared with our work, if we consider only numerals individuality. In their work the percentage of individuality for numeral 4 is around 5.60% for numeral 5 it is around 11.9% and for 7 it is around 18.5% where for those numerals in our work the corresponding accuracies in case of 400 dimensional feature are 25.12%, 35.37% and for 7 it is 26.59% respectively if only LIBLINEAR classifier is considered. For MLP classifier with the same feature the values are

28.54%, 34.63% and 28.29% respectively. For 64 dimensional feature with LIBLINEAR classifier the individuality values have been increased to 27.56%, 42.44% and 27.32% respectively. For the same feature with MLP classifier the values are 25.85%, 43.41% and 26.59%. Though the features and classifiers are different along with different database but the results are much higher for the present work in this respect.

The other work which can be compared with this present work is the work of Srihari et al. [6]. Both works have been conducted on individuality of numerals. Though the scripts and the structure of the numerals are different as for their work the Roman script has been used where in our case the Bangla script has been used. There exists a little similarity in the result of these two works that the numeral 5 has the most individuality than all the other numerals.

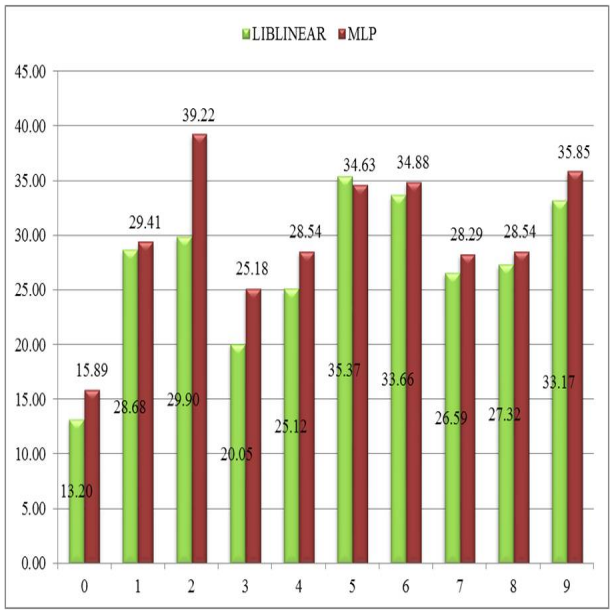


Figure. 7: Individuality of Bangla Numerals on 400 Dimensional Feature

Table 2: Individuality of Bangla numerals for 400 dimensional feature

Numerals	Classification Accuracy	
	LIBLINEAR (%)	MLP (%)
0	13.20	15.89
1	28.68	29.41
2	29.90	39.22
3	20.05	2 5.18
4	25.12	28.54
5	35.37	34.63
6	33.66	34.88
7	26.59	28.29
8	27.32	28.54
9	33.17	35.85

VIII. Conclusion

In this paper we have presented numeral based Individuality of handwriting and writer identification based on isolated Bangla numerals. The main emphasis has been on data collection and evaluation of the individuality of numerals to evaluate the quality of the dataset. The research has been done on all numerals from 90 writers with 5 sets each. After extraction of the isolated numerals from dataset images both 64 and 400 dimensional features have been applied and the LIBLINEAR and MLP classifiers of the WEKA tool have been used for the present work. For the present work we have achieved the highest individuality of 43.41% for numeral *PANCH* (5) and highest writer identification accuracy of 97.07% for all writers.

In future we intended to increase the number of writers. The results in this proposed work are promising and we will try to increase the data and also try different classifiers and combination of classifiers for our future works. We are also interested in finding out different feature extraction techniques for our future works.

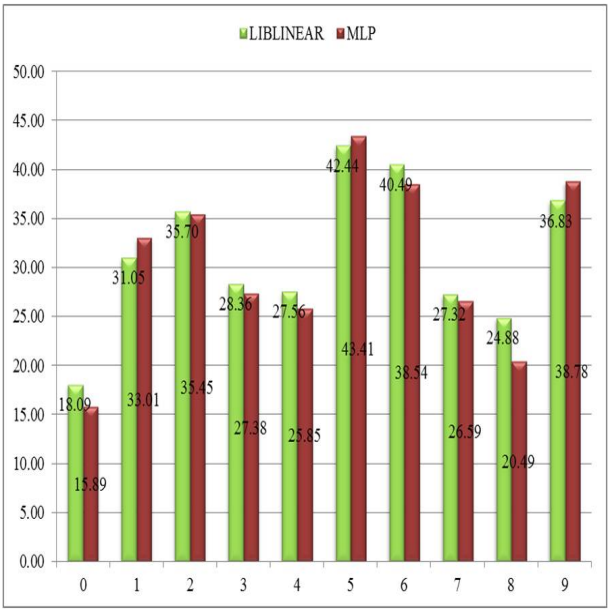


Figure. 8: Individuality of Bangla Numerals on 64 Dimensional Feature

Table 3: Individuality of each Bangla Numeral for 64 dimensional feature

Numerals	Classification Accuracy	
	LIBLINEAR (%)	MLP (%)
0	13.20	15.89
1	28.68	29.41
2	29.90	39.22
3	20.05	2 5.18
4	25.12	28.54
5	35.37	34.63
6	33.66	34.88
7	26.59	28.29
8	27.32	28.54
9	33.17	35.85

Acknowledgement

One of the authors would like to thank Department of Science and Technology (DST) for support in the form of INSPIRE fellowship.

Table 4: Writer Identification using 400 dimensional feature and LIBLINEAR classifier

Number of Writers	Accuracy (%)
20	100.00
40	98.00
50	96.00
55	96.00
60	95.33
65	95.08
70	96.29
90	96.50

Table 5: Writer identification using 64 dimensional feature on different classifiers

Number of Writers	**Identification Rate (%)**	
	LIBLINEAR	MLP
20	98.97	98.04
45	99.10	98.65
70	96.41	95.81
85	97.55	94.12
90	97.07	94.62

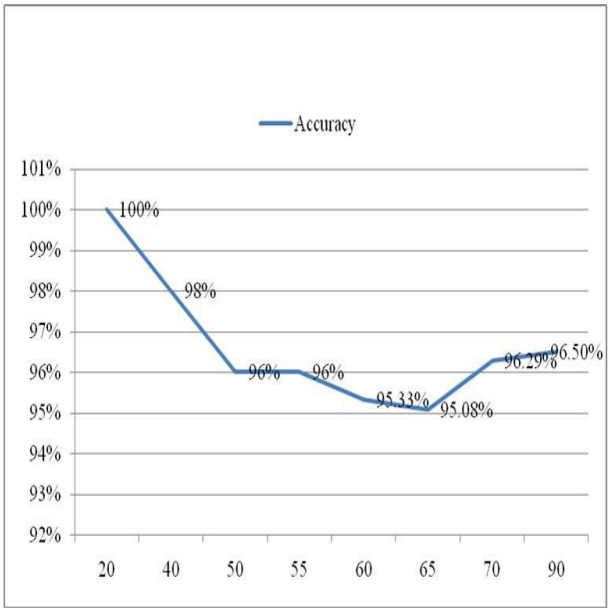


Figure. 9: The variability of writer identification rate using 400 dimensional feature and LIBLINEAR classifier for different writers.

References

[1] R. Plamondon and G. Lorette, Automatic signature verification and writer identification the state of the art, *Pattern Recognition*, 22 (2), pp. 107-131, 1989.

[2] H. E. S. Said, G. S. Peake, T. N. Tan and K. D. Baker, Writer Identification from Non-uniformly Skewed Handwriting Images, In *Proceedings of 9th British Machine Vision Conference (BMVC)*, pp. 478-487, 1998.

[3] H. E. S. Said, T. N. Tan and K. D. Baker, Personal Identification Based on Handwriting, *Pattern Recognition*, 33(1), pp. 149-160, 2000.

[4] U. V. Marti, R. Messerli and H. Bunke, Writer Identication using Text Line Based Features, In *Proceedings of 6th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 101-105, 2001.

[5] S. N. Srihari, S. H. Cha, H. Arora and S. Lee. Individuality of Handwriting, *Journal of Forensic Science*, 47(4), pp. 1-17, 2002.

[6] S. N. Srihari, C. Tomai, S. Lee and B. Zhang, Individuality of Numerals, In *Proceedings of 7th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1096-1100, 2003.

[7] S. N. Srihari, M. J. Beal, K. Bandi and V. Shah, A Statistical Model for Writer Verification, In *Proceedings of 8th International Conference on Document Analysis and Recognition (ICDAR)*, 2, pp. 1105-1109, 2005.

[8] A. Bensefia, A. Nosary, T. Paquet and L. Heutte, Writer Identification By Writers Invariants, In *Proceedings of 8th International Workshop on Frontiers in Handwriting Recognition (IWFHR)*, pp. 274-279, 2002.

[9] A. Bensefia, T. Paquet and L. Heutte, Information Retrieval Based Writer Identification, In *Proceedings of 7th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 946-950, 2003.

[10] M. Bulacu, L. Schomaker and L. Vuurpijl, Writer Identification using Edge-Based Directional Features, In *Proceedings of 7th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 937-941, 2003.

[11] M. Bulacu and L. Schomaker, Text-Independent Writer Identification and Verification Using Textural and Allographic Features, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(4), pp. 701-717, 2007.

[12] A. A. Brink, J. Smit, M. L. Bulacu and L. R. B. Schomaker, Writer identification using directional ink-trace width measurements, *IEEE Trans. Pattern Recognition*, 45(1), pp. 162-171, 2012.

[13] U. Marti and H. Bunke, The IAM-Database: An English Sentence Database for Offline Handwriting Recognition, *International Journal on Document Analysis and Recognition*, 5(1), pp. 39-46, 2002.

[14] L. Schomaker and L. Vuurpijl, Forensic Writer Identification: A Benchmark Data Set and a Comparison of Two Systems, technical report, *Nijmegen: NICI*, 2000.

[15] I. Siddiqi and N. Vincent, Writer Identification in Handwritten Documents, In *Proceedings of 9th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 108-112, 2007.

[16] S. Chanda, K. Franke, U. Pal, Text Independent Writer Identification for Oriya Script, In *Proceedings of 10th International Association for Pattern Recognition (IAPR) International Workshop on DAS*, pp. 369-373, 2012

[17] P. Purkait, R. Kumar, B. Chanda, Writer Identification for Handwritten Telugu Documents Using Directional Morphological Features, In *Proceedings of 12th*

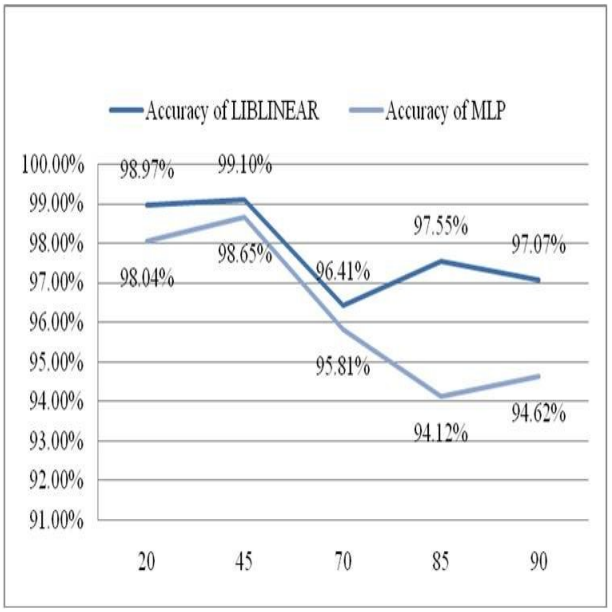


Figure. 10: The comparative results of writer identification rate using 64 dimensional feature with LIBLINEAR and MLP classifiers for different writers.

International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 658-663, 2010.

[18] B. V. Dhandra, M. B. Vijayalaxmi, G. Mukarambi, M. Hangarge, Writer Identification by Texture Analysis Based on Kannada Handwriting, *International Journal of Communication Network Security*, 1(4), pp. 80-85, 2012.

[19] A. Sarkar, U. Garain Individuality of Handwriting: a Study on Handwriting in an Indic Script, In *Proceedings of International Conference on Recent Trends in Information Systems (ReTIS)*, pp. 188-191, 2006.

[20] S. Chanda, K.Franke, U.Pal and T.Wakabayashi, Text Independent Writer Identification for Bengali Script, In *Proceedings of International Conference on Pattern Recognition (ICPR)*, pp. 2005-2008, 2010.

[21] S. Biswas and A. K. Das, Writer Identication of Bangla handwritings by Radon Transform Projection Prole, In *Proceedings of 10th International Association for Pattern Recognition (IAPR) International Workshop on DAS*, pp. 215-219, 2012.

[22] C. Halder, J. Paul and K. Roy, Individuality of Bangla Numerals, In *Proceedings of 12th Intelligent Systems Design and Applications (ISDA)*, pp. 264-268, 2012.

[23] K. Roy and U. Pal, On the Development of an OCR System for Indian Postal Automation, *LAP LAMBERT Academic Publishing*, Germany, ISBN: 978-38-443-1403-8, 2011.

[24] U. Pal, T. Wakabayashi, N. Sharma1 and F. Kimura, Handwritten Numeral Recognition of Six Popular Indian Scripts, In *Proceedings of 9th International Conference on Document Analysis and Recognition (ICDAR)*, 2, pp. 749-753, 2007.

[25] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, The WEKA Data Mining Software: An Update, *SIGKDD Explorations*, 11(1), pp. 10-18, 2009.

[26] C. Halder, J. Paul and K. Roy, Comparison of the Classifiers in Bangla Handwritten Numeral Recognition, In *Proceedings of 12th International Conference on Radar, Communication and Computing (ICRCC)*, pp. 399-404, 2012.

[27] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang and C. J. Lin, LIBLINEAR: A library for large linear classification, *Journal of Machine Learning Research*, 9, pp. 1871-1874, 2008.

Author Biographies

Chayan Halder was born in the year 1987 at Kolkata, West Bengal, India. He has done his Bachelors in computer science from BRS College Barrackpore under Calcutta University in 2008 and completed his Masters in computer science from West Bengal State University Barasat in 2010. He has worked as a Project linked personnel in Indian Statistical Institute (ISI) Kolkata, in 2011. Right now he is pursuing his PhD from West Bengal State University as an INSPIRE Fellow of Department of Science and Technology Delhi, since Nov. 2011. His current area of research includes Document processing, Handwriting Analysis etc. He has 4 conference publications in the area of Computer vision and pattern recognition.

Kaushik Roy Kaushik Roy received his B.E from Assam University and his M. E. and PhD from Jadavpur University (J.U) in Computer Science and Engineering in 1998, 2002 and 2008, respectively. He has worked as a Project linked personnel in I.S.I, Kolkata, from 2003 to 2005, and worked as a Scientific Officer in Centre for Development of Advanced Computing, Kolkata in 2006. From 2006 to 2009 he has worked as lecturer in West Bengal University of Technology, Salt lake, Kolkata and Now he is working as Associate Professor and Head in Department of Computer Science, West Bengal State University, Barasat, since 2009. He has more than 50 publication in various International and National Journals and Conferences proceedings to his credit. His areas of teaching and research includes are Pattern Recognition, Medical Image Processing, Artificial Intelligence, Handwriting Analysis and Online Handwriting Recognition.