Individuality of Isolated Bangla Characters

Chayan Halder
Department of Computer Science,
West Bengal State University,
Barasat, Kolkata 700126, WB, India
Email: chayan.halderz@gmail.com

Kaushik Roy
Department of Computer Science,
West Bengal State University,
Barasat, Kolkata 700126, WB, India
Email: kaushik.mrg@gmail.com

Abstract*—Writer Identification and Verification, a behavioural bio-metric study has gained a renewed interest in research work for its promising prospect in real life applications. In current times, to the best of our knowledge there is no complete system of Writer Identification/Verification on Indic scripts including Bangla. In this paper a scheme has been proposed for individuality of handwriting and writer identification which we plane to use in our general unconstrained writer identification/verification. The scheme has great prospect not only in Writer Identification but also in Writer Verification, Graphological Analysis and in various field of Handwriting Forensic. As there is no such standard Bangla writer database, for the proposed work a database consisting of total 31950 characters and vowel modifiers from 90 writers with 5 sets from each writer has been developed. Standard and robust features like 64 and 400 dimensional has been used for the evaluation of the collected data. The LIBLINEAR and MLP classifiers of WEKA tool has been used for analysis of the characters and vowel modifiers. An analysis of each characters and vowel modifiers reveals that the character *GA* ("গ") has the highest individuality of 55.85% and modifier *REF* ("/") has lowest individuality of 11.05%. The individuality has been calculated by finding out the writer identification accuracy for individual character. Writer identification accuracy of 99.75% has been achieved for 90 writers with 5-fold cross validation.

Keywords*—*Individuality of Handwriting, Writer Identification, Bangla Handwriting Analysis, WEKA, MLP, LIBLINEAR

I. Introduction

Writer identification belongs to the category of behavioural bio-metric. Bio-metric templates are extracted from handwritings for the identification purpose. Identification is performed by comparing the measured bio-metric template of an unknown individual to the previously measured templates of known individuals stored in the database.

The work on Writer Identification and Verification is an active area of research in document processing for more than two decades. An automated system of Writer Identification and Verification has the potential to be used in forensic science, banking, graphology etc. As the handwritten characters carries additional information about the personality and characteristics of the writer compared to electronic or printed numerals, there exists a high possibility to authenticate/identify the writer. Every individual has a certain degree of stability in their handwriting which makes it possible to identify and verify the writer. Lots of works has been reported in literature on writer identification of non-Indic scripts (mainly in Roman script) [1]–[10]. A texture analysis based approach has been used by Said et al. [2] for the development of a text independent writer identification system on Roman script. 1000 test documents from 40 writers have been used for the purpose. They have achieved an accuracy of 96.0%. Marti et al. [3] have also worked on Roman script and used 100 pages of text written by 20 writers as the data for their work. 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 to achieve identification accuracy of 87.8% and 90.7% respectively. Srihari et al. [4] have develop text dependent approach to find out individuality of handwriting on Roman script. 1568 writers have been used 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. The edge-based directional feature has been used by Bulacu et al. [5] on 500 documents from 250 writers for their experiment on Roman script. They have achieved accuracy of 75% for their work. Siddiqi and Vincent [8] have used 50 documents from same number of writers for their work on Roman script. A local approach has been used by them to extract writer specific characteristics. Bayesian classifier has been used in their work to achieve an identification accuracy of 94%. Bulacu et al. [10] have implemented text-independent Arabic writer identification using the combination of textural and allographic features. The IFN/ENIT dataset [11] has been used for their work. Firstly they have extracted the textural features and then generated the probability distribution function to calculate the distance measure using nearest neighbour classifier. To calculate the allographic features they have generated a codebook containing 400 allographs from 61 writers. They have used the similarity of allographs as another feature.

The database has been collected from 350 writers with 5 samples per writer (each sample contains 2 lines (about 9 words)). The best accuracies that have been seen in experiments are 88% in top-1 and 99% in top-10.

Indic scripts needs a lot of attention because there are many scripts and also the structural complexity of the scripts are higher than the non-indic scripts like Roman, Arabic etc. Though few works are available in this area like [12]–[17]. Chanda et al. [12] have used the directional chain-code and curvature feature for their work on Oriya script. SVM classifier has been used for the work and they have achieved an accuracy of 94% on writer identification. In the work of Sarkar and Garain [14] on Bangla characters, they have used gradient based contour encoding feature and 192 bit feature vector. 20 writers and 3 sample document for each writer has been

978-1-4799-6052-1/14/$31.00 ©2014 IEEE

used in this respect. The K-means clustering has been used to get an accuracy of 40% for identification. Chanda et al. [15] have worked on text independent writer identification using the Bangla characters. They have used 400 dimensional gradient features with SVM classifier to achieve an accuracy of 95.19%. In the work of Biswas and Das [16], they have used Radon transform projection profile as the feature for writer identification of Bangla handwritten documents. They have worked on their own BESUS Database consisting of 55 writers with four sample documents on two different topics for each writer. 83.63% accuracy has been achieved by them. In the work of Halder et al. [17] 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. As the aim of the proposed work is to develop the database and to analyse the individuality of isolated Bangla handwritten characters along with writer identification, we have used standard and robust features like 64 and 400 dimensional for the evaluation of our work.

The subsequent part of the paper is organized as follows: section II describes the properties of Bangla Script. In section III Data Collection and Pre-processing is described. Individuality of Bangla Handwriting are discussed in section IV. Feature Extraction is described in Section V. Section VI describes about WEKA tool followed by Results in section VII. At last we concluded in the section VIII.

II. Properties of Bangla Script

In terms of population Bangla is the second most popular language in India and sixth most popular language in the world. It is an ancient Indo-Aryans language [18]. Bangla script has been used in texts of Bangla, Assamese and Manipuri languages and is the national language of Bangladesh. It is the official language of the state West Bengal and Tripura in India. Bangla has 11 vowels and 39 consonants and lots of complex and compound shapes that can be formed by combination of vowels and consonants with consonant(s). In the Fig. 1 example of Bangla alphabets along with some complex and compound characters are shown. From the figure we can observe that the character shapes are quite complex in nature than the conventional Roman script. There exists head lines which are some times absent for certain characters or the length gets smaller for some characters like: *UNGO* ("ঙ") here denoted as "UN" and *GA* ("গ") respectively. Also when they form complex and compound shapes the resultant shapes become totally different from the source character shapes, also these shapes can be written in different techniques.

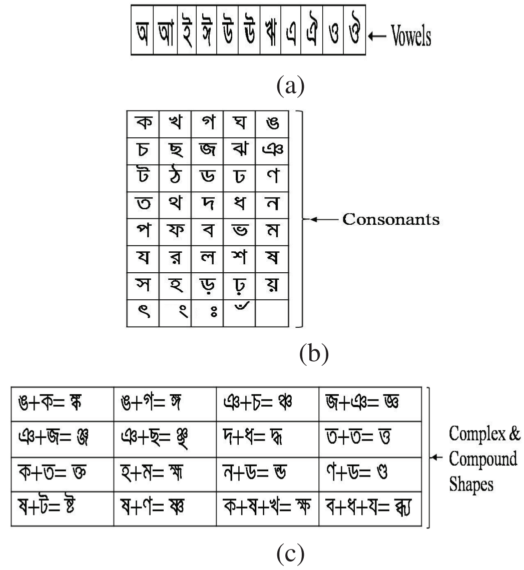


Fig. 1. Example of (a) Bangla Vowels (b) Bangla Consonants (c) Some Bangla Complex and Compound characters

III. Data Collection and Pre-Processing

As the main interest is to analyse the individuality of Bangla handwriting and to identify the writers (writer identification), a standard database is essential in this respect. But no such standard database is currently available. So, to create a standard database for Bangla handwriting analysis, we have designed a sample document consisting of all Bangla isolated characters and vowel modifiers. Each participant (writer) was asked to copy-out the printed characters in the particular box area of the sample document of 5 set in different time. Data has been collected from more than 200 writers of different professions including students. Out of 200 writers for the current work full 5 sets of data from 90 writers has been used. A total of 26367 Bangla alphabets, 5170 numerals and 5170 vowel modifiers have been used for the work. An example of our designed character sample collection document form is shown in Fig. 2. There exists no restrictions for writers regarding the type of pen and ink they use. We digitized them using a flat-bed scanner in gray mode at 300/600 dpi and stored in TIF format. For our current work the 300 dpi images are used and 600 dpi images are stored for future works.

During the Pre-processing, the digitized images are firstly binarized using a global binarization method [18]. Then a simple and fast character extraction technique is used to extract the characters from the digitized forms. Firstly, the global binarization of the whole document has been carried out. Then maximum run length has been computed on horizontal and vertical histogram. Then the horizontal and vertical lines are detected. After that those lines are deleted to get an image which contains only the suggestive characters and the original handwritten characters. Then the suggestive characters are removed and bounding box for each handwritten character has been calculated to get the isolated characters. These information have been stored for further processing as gray toned isolated character images and verified manually.

IV. Individuality of Bangla Handwriting

Handwriting of individuals not only varies between writers but also from his/her own handwriting but the intra-writer variation (the variation within a person's own handwriting samples) is quite less than that of the inter-writer variation (the variation between the handwriting samples of two different people). So, while analysing handwriting there are two main point of concern: identification of inter-writer variation and at the same time minimizing the effect of intra-writer variation. These variations can be seen from Fig. 3 (a) where 10

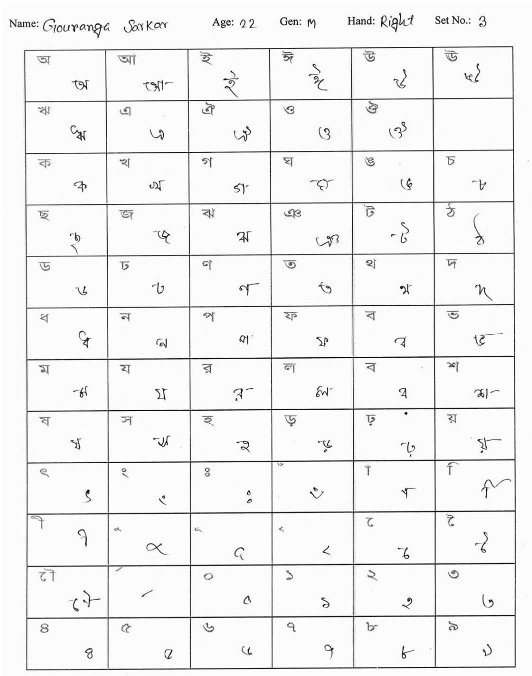


Fig. 2. Sample data collection form used for collection of Bangla Handwritten isolated characters and Vowel modifiers

characters of 5 different writers extracted from the database has been shown. From Fig. 3 (a) it can be easily found that the character *GA* ("গ") has more inter writer variability then other characters. It can also be seen that the variability differs from character to character for example *REF* ("/") has a little inter-writer variation. To emphasize more on that and also to check intra-writer variation 5 characters among the above characters each with 5 instances of a single writer have been extracted from the database and shown in Fig. 3 (b). From Fig. 3 (b) it can be seen that the character *REF* ("/") has a very little variation whether its from same writer or from different writer but the character *GA* ("গ") has more variation when its written from different writer compared to when written by same writer.

To analyse the individuality of characters graphically, the characters of different writers of each category has been superimposed into single individual image. This is done to analyse the inter and intra writer variability visually for individual characters. The steps are follows:

Firstly, the bounding box of isolated gray character images are calculated after global binarization and then they are normalized to 128x128 pixels. Next, the normalized 128x128 binary images are projected into a white 128x128 image.In the projection technique, for each object pixel of the original image, corresponding pixel of the white 128x128 image has been decremented by 1. Thus, we get gray images of the characters as individual image capturing the writing variation (the individuality of the characters). If the character is not individual then there is less variation among the characters so the character will have more black parts (as the same point will be decremented repeatedly) which can be seen in the image as example of concentrated areas and rest part as white where none of the writer has written. This procedure is repeated for every character category once writer wise and once for all writers. For example see Fig. 4 (a) where three samples of 6 handwritten characters: *GA* ("গ"), *BORGIAJA* ("জ") here denoted as "JA", *PA* ("প"), *PANCH* ("৫") denoted as "5", *SUNNO* ("০") denoted as "0" and *REF* ("/") from three different writers extracted from the database have been shown. In Fig 4 (b) three samples of superimposed images of the above characters from the same writers taking ten instances from each writer and in 4 (c) three different samples of superimposed images from randomly selected ten different writers of the database taking one instance from each writer have been shown.

Analysing Fig 4 (b) and (c), it can be easily observed that for a single writer the characters are almost same as there is very less variation in own writing but in case of multiple writers there are presence of gray portion due to the variation in writing. This is because positioning, size, shape and writing pattern etc. varies with writers. But these variations also vary with characters, which can be seen from the image of the modifier *REF* ("/") of 4 (c), which is almost consistent over the writing, it means that the variation between writers are less compared to character *GA* ("গ") which has more gray portions. So if we consider modifier *REF* ("/") and character *GA* ("গ") for writer identification the modifier will give less identification rate than the character as character *GA* ("গ") has more regions that are different for different writers. So it is clear that the modifier *REF* ("/") is less individual than character *GA* ("গ"). Numeral *5* ("৫") has individuality information only on the right side as it has more gray part compared to its left side. This is because generally there are less variation in circular portions compared to other specially start and end portion and hence the individuality of numeral *0* ("০") is also low. The description shows that some character has individuality property which can be used for writer identification/verification.

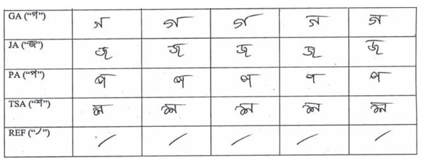
V. Feature Extraction

To analyse the individuality along with writer identification and to evaluate the quality of the database 400 and 64 dimensional feature extraction technique has been used [18]. For the analysis of individuality of Bangla characters the 400 dimensional feature and for writer identification 64 dimensional feature has been used. The 400 dimensional feature has given some encouraging results for different Indic script numeral recognition like the work of Pal et al. [19]. The 64 dimensional feature has been widely used in this type of pattern recognition problems because of the ease of use as it has low computational complexity compared to 400 dimensional feature.

To obtain 400 dimensional features firstly, the input gray image is binarized and then normalized into 73x73 pixels. Then 2x2 mean filtering has been applied five times to get the gray-scale image. After that the image is segmented into 9x9 blocks and Roberts filter has been used to get the gradient of the image. At last histogram has been computed and down sampled into 5x5 blocks. For the down sampling we have applied a Gaussian filter and obtained 400 (5x5x16) dimensional feature. For more details in this see [18].

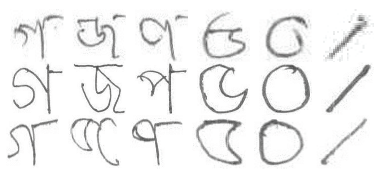
For the 64 dimensional feature extraction firstly contour points of the two-tone image has been computed. Initially all the contour points are divided into 7x7 blocks one contour point in each block. Then direction code for each block

(a)

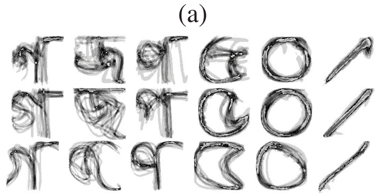


(b)

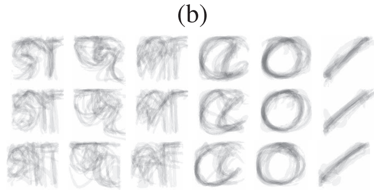
Fig. 3. (a) Some Bangla handwritten characters of five different writers, (b) Five instances of five characters from a single writer.



(a)



(b)



(c)

Fig. 4. (a) Example of 3 sets of isolated characters from different writers, (b) Example of 3 sets of superimposed images from same writer, (c) Example of 3 sets of superimposed images from different writers, for the characters shown in (a)

containing contour point has been computed. After down sampling the initial 7x7 blocks into 4x4 blocks 64 (4x4x4) directions code features are being obtained. For normalizing the features (between 0 and 1) the maximum value of the histogram peaks in each direction from all the blocks has been computed. Each of the above features is then divided by the maximum value of their respective direction to get the feature value between 0 and 1. For more details in this see [18].

VI. WEKA

WEKA is one of the widely used tools in the area of machine learning [20]. 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 computation.

A. LIBLINEAR

The LIBLINEAR is 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 data-set. 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 [21].

B. MLP

The Multi Layer Perceptron (MLP) is a layered feed forward artificial neural network [18]. The number of hidden layers and the number of neurons in a hidden layer should be determined during training process [18]. For this work, the 400 and 64 dimensional features are used for individuality calculation of each character where only the 64 dimensional feature is used for the 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 default values for this work like the learning rate has been set to 0.3 and momentum to 0.2.

VII. Results

The present work has been carried out on total 31950 characters and vowel modifiers from documents written by 90 writers taking 5 samples each. In this work 5-fold cross validation scheme for individuality computation of each character and also for writer identification has been used. To compute the individuality of each character, actually the writer identification accuracy of each character for all the writers has been computed. This makes sense as the identification accuracy of writers for any particular character is the measure of the uniqueness of the character for the given set of writers i.e. the individuality of that particular character. After that, writer identification has been computed on all characters. The results are discussed below:

A. Result of Individuality of characters

The individuality for all the characters and vowel modifiers has been calculated using 400 dimensional feature. The results of both the LIBLINEAR and MLP classifiers for the individuality of each character on all the writers are shown in Table

I. From the table it can be observed that the character *GA* ("গ") is most individual with writer identification accuracy of 55.85% and 53.90% for the MLP and LIBLINEAR classifiers respectively, followed by character *JA* ("জ") with accuracy of 53.17% in case of MLP classifier and *PA* ("প") with accuracy of 52.20% for LIBLINEAR classifier. The least individual is the vowel modifier *REF* ("/") with accuracy of 11.05% and 9.77% for MLP and LIBLINEAR classifiers respectively, preceded by *0* ("০") with the accuracy of 15.89% and 13.20% respectively among all the alphabets, numerals and vowel modifiers. It is also the least individual among all numerals. If only alphabets are considered then the least individual alphabet is *BISORGO* ("ঃ") here denoted as "BI" with accuracy of 20.73% and 16.59% for the MLP and LIBLINEAR classifiers respectively. For both the classifiers the vowel modifier *DIRGHOU* ("ূ") here denoted as "-UU" has the highest individuality with accuracy of 42.54% and 46.70% for LIBLINEAR and MLP classifiers respectively among all vowel modifiers. The numerals *DUI* ("২") denoted as "2" for MLP and *5* ("৫") for LIBLINEAR have the highest individuality among all numerals with accuracy of 39.22% and 35.37% respectively.

If we go back to Fig. 4 it can be seen that at (b) *GA* ("গ") has less gray portions for same writer that is the intra writer variation is less than that of *JA* ("জ") and *PA* ("প") but in (c) *GA* ("গ") has more gray portion for different writers i.e. the inter writer variation is more than *JA* ("জ") and *PA* ("প"). For this reason *GA* ("গ") has maximum individuality for writer identification than others as it reduces the intra writer variability and maximize the inter writer variability. In case of *JA* ("জ") and *PA* ("প") though they have higher intra writer variability than *5* ("৫"), the individuality is higher as they have very high inter writer variability than *5* ("৫") as evident from 4 (b) and (c).

B. Result of Writer Identification

For writer identification purpose the 64 dimensional feature has been calculated for feature extraction. The main constraint is the system available which is unable to work with very large feature set is one of the reasons that we have chosen 64 dimensional feature over 400 dimensional feature in this respect. For the same reason also the MLP classifier has not been used. Though the work on both has been in progress but for the time being the results are not available. The LIBLINEAR classifier has been used for writer identification. We have achieved an accuracy of 99.75% for 90 writers but it increases to 100% when the numbers of writers are 87.

C. Comparison of Results

To the best of our knowledge there is no such works in literature on this topic in Bangla except Sarkar and Garain [14] has worked on similar area with 60 documents from 20 writers and Halder et al. [17] on Bangla Numerals. In Table II a comparative study of different writer identification method on Bangla script has been shown. Sarkar and Garain have achieved highest accuracy of 18.33% and 13% for individuality of characters and writer identification respectively compared to 55.85% and 99.75% for 90 writers and 100% for 87 writers for the current work. Although the database size and type is totally different from Sarkar and Garain [14]. As the current database and feature set are similar to the work of Halder et al. [17] the results on individuality of numerals are almost same for LIBLINEAR classifier but for MLP the highest individual numeral is *2* ("২") rather than *5* ("৫") also the writer identification accuracy is higher.

TABLE I. INDIVIDUALITY OF EACH BANGLA CHARACTER

Chars	Classification Accuracy		Chars	Classification Accuracy	
	LIBLINEAR(%)	MLP (%)		LIBLINEAR (%)	MLP (%)
A ("অ")	33.41	39.02	*MA* ("ম")	41.71	41.46
AA ("আ")	40.00	39.51	*YA* ("য")	39.27	42.68
I ("ই")	34.39	37.56	*RA* ("র")	36.10	40.98
II ("ঈ")	37.56	44.88	*LA* ("ল")	44.63	46.83
U ("উ")	28.54	30.73	*TSA* ("শ")	51.34	51.34
UU ("ঊ")	36.43	37.90	*MSA* ("ষ")	35.37	37.56
RI ("ঋ")	45.34	45.10	*SS* ("স")	45.61	51.22
E ("এ")	35.54	41.67	*HA* ("হ")	38.78	39.76
OI ("ঐ")	39.71	39.95	*DRA* ("ড়")	34.15	33.90
O ("ও")	30.07	33.99	*ADA* ("ঢ়")	29.02	33.41
OU ("ঔ")	30.32	30.07	*Y* ("য়")	38.05	42.44
K ("ক")	43.41	44.15	*KT* ("ৎ")	22.20	27.07
KH ("খ")	51.71	51.95	*AN* ("ং")	39.51	40.73
***GA* ("গ")**	**53.90**	**55.85**	***BI* ("ঃ")**	**16.59**	**20.73**
GH ("ঘ")	45.85	48.05	*CN* ("ঁ")	26.23	29.41
UN ("ঙ")	44.99	47.19	*-AA* ("া")	29.02	30.24
C ("চ")	26.34	31.46	*-I* ("ি")	24.88	27.07
CH ("ছ")	43.66	46.83	*-II* ("ী")	27.14	27.63
***JA* ("জ")**	50.73	**53.17**	*-U* ("ু")	34.96	37.41
JH ("ঝ")	45.61	42.68	***-UU* ("ূ")**	**42.54**	**46.70**
EN ("ঞ")	37.80	38.78	*-RI* ("ৃ")	21.76	23.72
T ("ট")	33.90	38.05	*-E* ("ে")	23.47	26.16
TTA ("ঠ")	26.65	27.14	*-OI* ("ৈ")	28.36	31.54
DDA ("ড")	30.07	32.27	*-OU* ("ৌ")	31.54	34.72
DDH ("ঢ")	24.88	25.12	***REF* ("/")**	**9.77**	**11.05**
MNA ("ণ")	37.80	46.83	***0* ("০")**	**13.20**	**15.89**
TA ("ত")	23.90	26.34	*1* ("১")	28.68	29.41
THA ("থ")	40.00	41.46	***2* ("২")**	29.90	**39.22**
DA ("দ")	31.95	35.12	*3* ("৩")	20.05	25.18
DHA ("ধ")	39.02	42.44	*4* ("৪")	25.12	28.54
NA ("ন")	36.34	40.73	***5* ("৫")**	**35.37**	34.63
***PA* ("প")**	**52.20**	52.93	*6* ("৬")	33.66	34.88
PHA ("ফ")	44.88	41.95	*7* ("৭")	26.59	28.29
BA ("ব")	34.63	32.44	*8* ("৮")	27.32	28.54
BHA ("ভ")	32.44	32.68	*9* ("৯")	33.17	35.85

TABLE II. COMPARISON OF WRITER IDENTIFICATION METHODS

Method	Writers (data details)	Samples /Writer	Highest Individuality Rate (%)	Writer Identification Rate (%)
Sarkar [14] *and Garain*	20 (Alpha-) (numeric)	3	18.33	13.00
Halder [17] *et al.*	90 (Numerals)	5	35.90	96.50
Current Method	90 (Alpha) (-numeric)	5	55.85	99.75

VIII. CONCLUSION

In this paper a scheme for analysis of Individuality of handwriting based on isolated Bangla characters and writer identification has been proposed using 400 and 64 dimensional feature with LIBLINEAR and MLP classifier. We have achieved writer identification rate of 99.75% for 90 writers. It is a part of our work on writer identification on unconstrained Bangla handwriting. We plan to verify writers based on individuality property proposed here. The main emphasis was on data collection and analysis of individuality of isolated characters. In future we plan to increase the size of the database both in terms of writer as well as the number of samples and also make the database open. We plan to use combination of features, classifiers and use some script dependent features in future.

Acknowledgment

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

References

[1] R. Plamondon and G. Lorette, "Automatic signature verification and writer identification - the state of the art," *Pattern Recognition*, vol. 22, no. 2, pp. 107–131, 1989.

[2] H. E. S. Said, T. N. Tan, and K. D. Baker, "Personal identification based on handwriting," *Pattern Recognition*, vol. 33, no. 1, pp. 149–160, 2000.

[3] U. V. Marti, R. Messerli, and H. Bunke, "Writer identication using text line based features," pp. 101–105, 2001.

[4] S. N. Srihari, S. H. Cha, H. Arora, and S. Lee, "Individuality of handwriting," *Journal of Forensic Science*, vol. 47, no. 4, pp. pp. 1–17, 2002.

[5] M. Bulacu, L. Schomaker, and L. Vuurpijl, "Writer identification using edge-based directional features," pp. pp. 937–941, 2003.

[6] A. Bensefia, T. Paquet, and L. Heutte, "Information retrieval based writer identification," pp. pp. 946–950, 2003.

[7] S. N. Srihari, M. J. Beal, K. Bandi, and V. Shah, "A statistical model for writer verification," pp. 1105–1109, 2005.

[8] I. Siddiqi and N. Vincent, "Writer identification in handwritten documents," pp. 108–112, 2007.

[9] M. Bulacu and L. Schomaker, "Text-independent writer identification and verification using textural and allographic features," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 701–717, 2007.

[10] M. Bulacu, L. Schomaker, and A. Brink, "Text-independent writer identification and verification on offline arabic handwriting," pp. 769–773, 2007.

[11] M. Pechwitz, S. Maddouri, V. Margner, N. Ellouze, and H. Amiri, "Ifn/enit-database of handwritten arabic words," pp. 129–136, 2002.

[12] S. Chanda, K. Franke, and U. Pal, "Text independent writer identification for oriya script," pp. 369–373, 2012.

[13] P. Purkait, R. Kumar, and B. Chanda, "Writer identification for handwritten telugu documents using directional morphological features," pp. 658–663, 2010.

[14] A. Sarkar and U. Garain, "Individuality of handwriting: a study on handwriting in an indic script," pp. 188–191, 2006.

[15] S. Chanda, K. Franke, U. Pal, and T. Wakabayashi, "Text independent writer identification for bengali script," pp. 2005–2008, 2010.

[16] S. Biswas and A. K. Das, "Writer identication of bangla handwritings by radon transform projection prole," pp. 215–219, 2012.

[17] C. Halder, J. Paul, and K. Roy, "Individuality of bangla numerals," pp. 264–268, 2012.

[18] K. Roy and U. Pal, *On the Development of an OCR System for Indian Postal Automation*. 2011.

[19] U. Pal, T. Wakabayashi, N. Sharma, and F. Kimura, "Handwritten numeral recognition of six popular indian scripts," pp. 749–753, 2007.

[20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[21] 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*, vol. 9, pp. 1871–1874, 2008.