Journal of Theoretical and Applied Information Technology
15th December 2019. Vol.97. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HANDWRITTEN BENGALI CHARACTER RECOGNITION THROUGH GEOMETRY BASED FEATURE EXTRACTION

[1]MOSHIUR RAHMAN, [2]IQBAL MAHMUD, [3]MD. PALASH UDDIN, [4]MASUD IBN AFJAL, [5]MD. AHSAN HABIB, [6]FAISAL KABIR

[1&2]Software Engineer, Samsung R&D Institute, Bangladesh

[3&5] PhD Candidate, Deakin University, Australia and Assistant Professor, Hajee Mohammad Danesh Science and Technology University (HSTU), Department of Computer Science and Engineering, Dinajpur, Bangladesh

[4]Assistant Professor, Hajee Mohammad Danesh Science and Technology University (HSTU), Department of Computer Science and Engineering, Dinajpur, Bangladesh

[6]B. Sc. Student (Session 2014), Hajee Mohammad Danesh Science and Technology University (HSTU), Dinajpur, Bangladesh

E-mail: [1]mrmohim74@gmail.com, [2]i.mahmud494@gmail.com, [3]palash_cse@hstu.ac.bd,[4]masud@hstu.ac.bd, [5]ahsan.habib@hstu.ac.bd, [6]faisalkabir0010@gmail.com

ABSTRACT

Unlike English characters, one of the major drawbacks in recognizing handwritten Bengali script is the massive amount of characters in Bengali language and their complex shapes. There are 50 complex shaped characters in Bengali alphabet set and working with this huge amount of characters with an appropriate set of feature is a tough problem to solve. Moreover, the ambiguity and precision error are common in handwritten words. In addition, among the huge amount of complex shaped characters, some are very similar in shape those possess severe difficulty to recognize handwritten Bengali characters. Bearing in mind the complexity of the problem, an efficient approach for recognizing handwritten Bengali alphabet is proposed in this work. This proposed approach for identifying Bengali characters is based on character geometry-oriented feature extraction for different handwritten characters. In this paper, different image processing steps are used including image acquisition, digitization , preprocessing, segmentation and feature extraction for tackling the difficulty. Most importantly, the geometry based feature extraction method has been employed to extract the effective features from the Bengali characters for the classification purposes. Then, the classification result was measured for SVM and Artificial Neural Network (ANN) based classifiers on self-generated training and testing data sets which contain 2500 different samples of 50 characters in the Bengali character-set. The proposed technique produces an average recognition rate of 84.56% using SVM and 74.47% using ANN.

Keywords: *Bengali alphabets, image segmentation, feature extraction, support vector machine, artificial neural network*

1. INTRODUCTION

Recently, there has been much interest and anticipation in automatic character recognition. Between handwritten and printed forms, handwritten character recognition is more challenging research area in computer vision and pattern recognition. Handwritten characters written by different persons are not identical and differ in both size and shape. Several variations in writing styles of individual character make the recognition task difficult. The similarities in distinct character shapes, the overlaps and the interconnections of the adjacent characters further complicate the problem [1]-[12]. To cope up with the difficulties, a typical handwritten character recognition system consists of two major steps [13]-[16]:

i. effective feature extraction from the character set, and
ii. employment of proper learning tool(s) to classify individual character

In addition, handwritten character identification is one of the artificial intelligence

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

tasks that fall into the scientific discipline called pattern recognition. This is one of the complicated tasks for the complicated logic and theories with it which requires more effort to improve the accuracy and performance of the system. A system for the identification of handwritten characters can be considered very useful. Because it is stress-free to make adaptations on a digital document rather than editing the content of a document written on a paper. For that reason, a number of various classification methods are used for online and offline character recognition.

Recently, several feature extraction and classification techniques have also been employed for Bengali character recognition. For instance, the feature extraction techniques include Chain Code, Zone Based Centroid, Background Directional Distribution and Distance Profile Features applied to the preprocessed images [17], [18]. However, the geometry based feature extraction [17] can be one of the methods that can be used to collect character features from the Bengali characters. Consequently, the geometry based feature extraction has been adopted for Bengali character recognition in this paper. In addition, several approaches based on artificial neural network (ANN) and support vector machine (SVM) are also examined for handwritten Bengali character recognition [19], [20] for the following reasons. The artificial neural network (ANN) has been successful in character pattern identification which does not undergo any of mathematical algorithms [21]. On the other hand, Support vector machine (SVM) is one of the appropriate techniques that analyzes data and recognizes patterns for classification and regression tasks [16].

As mentioned earlier, a model has been proposed for recognition of Bengali handwritten characters through the extraction of geometry based features of Bengali alphabets using ANN and SVM. Although the accuracy of printed Bengali character recognition has been reached near 100%, there is still low performance of existing handwritten Bengali character recognition systems. Thus, this work uses line classifier to extract the geometric based features of the characters to detect and classify the handwritten text. In dealing with the problem of recognition of character patterns of varying shapes and sizes, the geometric based features are used to achieve high recognition performance. Some image processing techniques are used for removing the background noise. To this end, the proposed model aims to evaluate the performance of the line based feature set for effective recognition of isolated handwritten Bengali basic characters.

The rest of this paper is organized into the following sections. Section 2 is a description and investigation of the existing character recognition models. In Section 3, we discuss the proposed Bengali character recognition model with the detail explanation of the constituent steps. In Section 4, we focus on the experimental setup and result analysis of the proposed model whereas Section 5 summarizes the explanations and concludes the paper.

2. RELATED WORK

The available systems for handwritten character recognition are not perfect in all aspects. Most of the developed systems cannot detect exactly and they may fail in the critical points and the accuracy is not satisfactory. Existing systems cannot recognize Bengali handwritten character from the image properly [22]. Among Indian scripts, first research work on handwritten Devnagari characters was reported in [23] in 1977. Ragha and Sasikumar [24] extracted moments features from Gabor wavelets for Kannada handwritten character recognition. Rajput and Mishra [25] used replacement of the recognized characters with standard fonts through backpropagation algorithm to extract the features and classify these features using ANN for recognition of Devnagari handwriting. Rajashekararadhya reported a feature extraction technique based on distance metric, zone metric and Neural network for the recognition of Telugu and Kannada numerals [26]. Arora and Bhattacharjee [27] used shadow features and chain code histogram features for recognition of handwritten non-compound Devnagari characters using combination of multilayer perceptron (MLP) and minimum edit distance and later on a two stage classification approach was reported by them [28]. John, Pramod and Kannan [29] reported a technique using chain code and image centroid for feature extraction. Sigappi, Palanivel and Ramalingam [30] used profile features for retrieval of handwritten Tamil documents. A fuzzy approach used in recognition of handwritten Malayalam characters and state space point distribution parameters were utilized by Lajish [31]. Kumar and Ravichandran [32] were used a collection of structural features for recognition of handwritten Tamil characters. Sangame, Ramteke and Benne [33] proposed an invariant moments feature extraction technique for recognition of handwritten Kannada vowels. Shanthi and Duraishwamy [34] used variation of zonal pixel densities was considered for feature extraction and classify these features using SVM to recognize handwritten Tamil

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

characters. Rahiman [35] took HLH intensity patterns for recognize Isolated handwritten Malayalam characters. Attempt of Suresh and Arumugam [36] to recognize handwritten Tamil characters was based on fuzzy approach. A Technique used by Sureshkumar and Ravichandran [37] for both recognition and conversion of handwritten Tamil characters was based on spatial space detection. Raju [38] proposed a system for handwritten Malayalam character recognition by using zero-crossing wavelet coefficients. Bhattacharya, Ghosh and Parui [39] used K-means clustering in a two stage recognition approach for handwritten Tamil characters.

On the other side, mentionable research works on Bengali handwritten character recognition was begun in early 1990. Chaudhuri, Majumder and Parui [40] proposed a recognition scheme for Bengali handwritten numerals based on matching of character skeleton. An analytical scheme involving stroke features, center of gravity and histogram features were used for handwritten Bengali cursive word recognition by Bhattacharya and Nigam [41]. Biswas and Bhattacharya [42] adopted bilinear interpolation technique in a HMM based approach using Dirichlet distributions for online handwritten Bengali character recognition. Dutta and Chaudhuri [43] used curvature features for recognition of Bengali alpha-numeric character. A technique of directional chain code histogram features of contour points in association with water reservoir principle to derive a lexicon driven method was used by Pal, Roy and Kimura [44] for recognition of unconstrained Bengali handwritten words. Rahman and Fairhurst [45] proposed a multistage technique involving some major structural features for handwritten Bengali character recognition. Basu and Das [46] shows their maximum recognition rate 75.05% to considered a feature set comprising of 76 elements (16 centroid features, 36 longest-run features and 24 shadow features) along with MLP classifier for recognition purpose. A technique based on direction code for recognition of online handwritten characters of Bengali reported by Bhattacharya and Gupta [47]. Bhoumick, Bhattacharya and Parui [48] proposed an approach for recognition of Bengali handwritten characters via an MLP based scheme. Bhattacharya, Shridhar, Parui, Sen and Chaudhuri [49] contributed by reporting the generation of a database of handwritten basic characters of Bengali language and also by developing an appropriate handwritten character recognition scheme for Bengali alpha-numeric using a two stage classifier basing on rectangular grid technique. In this paper, we propose a Bengali handwritten character recognition model using SVM and ANN that focuses on the extraction of geometry based features from the self-generated dataset.

3. PROPOSED MODEL

The proposed model for Bengali handwritten character recognition constitutes five main steps: image acquisition, preprocessing, segmentation, feature extraction, and classification and recognition as illustrated in Fig. 1. The model first takes the self-generated dataset images to apply preprocessing techniques and normalizing the characters and then it converts the images into binary form to ease the analysis of the behavior of characters. After that, it extracts different line types that form a particular character and it also focusses on the positional features of the same and stored them in the feature matrix. The feature matrix is fed to a machine leaning classifier for training purpose. Finally, it tests different characters for accuracy from the trained model. To this end, the complete architecture of the proposed Bengali handwritten character recognition model is illustrated in Fig. 2.

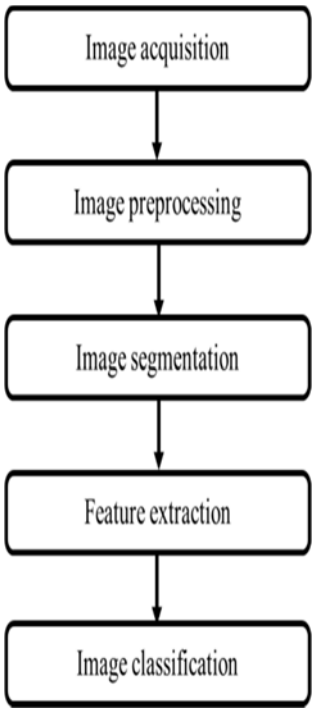

Figure 1: Flow diagram of proposed model

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

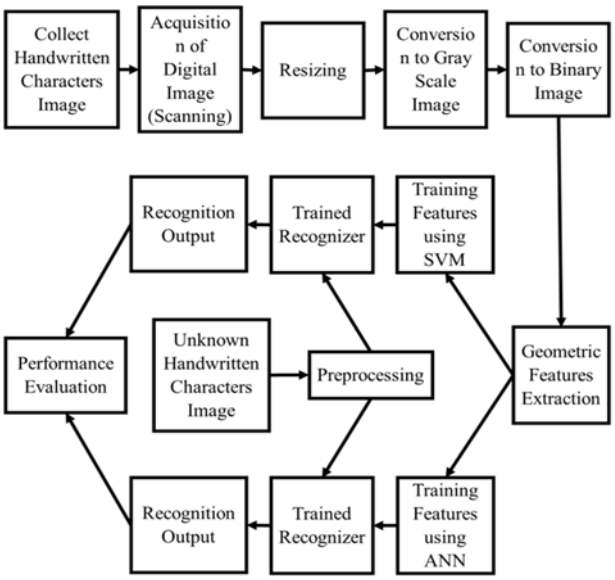

Figure 2: Architecture of proposed model

The operations involved in the steps of the proposed model are discussed in detail as follows:

3.1 Dataset Preparation

In this work, we prepared a comparatively large handwritten dataset for 50 Bengali isolated characters containing 11 vowels and 39 consonants as shown in Fig. 3. We have collected our dataset from around 25 persons from different ages and education levels. The generated dataset size is 2500 handwritten images where each character has 50 different samples. The prepared dataset contains wide variation of discrete characters because of different persons writing styles and some of these character images are very composite shaped and closely interrelated with others.

Figure 3: Bengali handwritten characters in the prepared dataset used in recognition

3.2 Preprocessing

The preprocessing is a sequence of operations performed on the scanned handwritten characters image. It basically enhances the image interpretation which is suitable for segmentation. The preface of pre-processing is to segment the interesting pattern from the background. The techniques used to enhance the image are described below:

3.2.1 Noise Removal

The noise of written characters may be introduced due to any writing mistakes or disturbance. Some scanning devices also introduced noises like complete loops, disconnected lines, bumps and gaps in line of characters [50]. The noise removing is mandatory to reorganization purpose.

3.2.2 Normalization

Normalization is an important part of handwritten characters recognition in preprocessing phase which challenges to remove variations in images. In this case, image will not be able to change identity of character [51]. Normalization provides the appropriate shapes to the images so that features of the character images can be compared. Basically, it deals with sizes of the images. In this proposed model, the image size 165×165 is used.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2.3 Binarization

After applying normalization, the RGB image is converted into gray level image and gray level image is converted into binary form to ease the analysis of the behavior characters [52]. Some sample conversions are shown in Fig. 4.

Figure 4: RGB to gray scale and gray scale to binary conversion

Binarization transformed gray scaled image into binary image where all 0 displays black pixels and all 1 displays white pixels. Global thresholding picks one threshold value for the whole document image based on an estimation of the background level from the intensity histogram of the image. In this model, Otsu's method is applied to binarize the images. In Otsu's method [53], we exhaustively search for the threshold that reduces the intra-class variance (the variance within the class), defined as a weighted sum of variances of the two classes:

$$\sigma_{\omega}^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t)$$

Where, weights ω_0 and ω_1 are the probabilities of the two classes separated by a threshold t, σ_0^2 and σ_1^2 are variances of these two classes. Then, dilation of edges in the binarized image is done using Sobel technique.

3.3 Image Segmentation

Image Segmentation is an image partition procedure into its fundamental parts or objects. Generally, autonomous segmentation is one of the toughest tasks in image processing. A heavy segmentation procedure brings the process in a long way toward a successful solution of imaging problems which require objects to be identified independently. Alternatively, weak or irregular segmentation algorithms almost always guarantee eventual failure. As a general rule, the more accurate segmentation, the more likely recognition is to succeed.

3.4 Feature Extraction

The proposed model extracts different line types (geometrical features) that form a particular character. It also focuses on the positional features of the characters. As mentioned earlier, the proposed employment of geometry-based feature extraction technique explained below was finally tested using a SVM and ANN.

3.4.1 Universe of Discourse

The universe of discourse is the aggregate of the individual objects which exist, that is are independently side by side in the collection of experiences to which the deliverer and interpreter of a set of symbols have agreed to refer and to consider. The features extracted from the character image include the locations of different line segments in the character image. For this situation, the universe of discourse is selected. So that, every character image should be independent of its image size and resized it into 165×165, which is shown in Fig. 5.

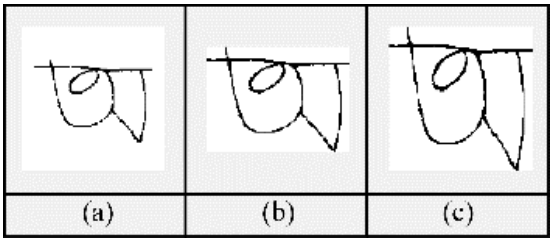

Figure 5: (a) Original image (b) Universe of discourse (c) Resized to 165×165

3.4.2 Zoning

After universe of discourse is selected, the image is divided into 9 equal-sized windows as shown in Fig. 6 and the feature is done on individual windows. For getting more information about the details of character skeleton, feature extraction was applied to specific zones rather than the whole image. Similarly, if zoning is used, then positions of different line segments in a character skeleton become a feature. As in almost cases, a particular line segment of a character occurs in a particular zone and the entire skeleton zone should be traversed into those line segments. For this persistence, certain pixels in the character skeleton were defined as starters, minor starters and intersections.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

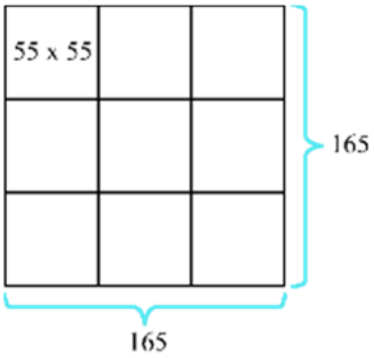

Figure 6: Divided windows of equal size

3.4.3 Starters and Minor Starters

The pixels which have only one neighbor in the character skeleton are called starters. All the starters in the particular zone is selected and is stored in a list before character traversal starts. The starters of Bengali 'অ' character found are given in the Fig. 7(a). When the current pixel under consideration has more than two neighbors, minor starters are created. They are found along the course of traversal along the character skeleton. The minor starters of Bengali 'অ' character found are given in the Fig. 7(c).

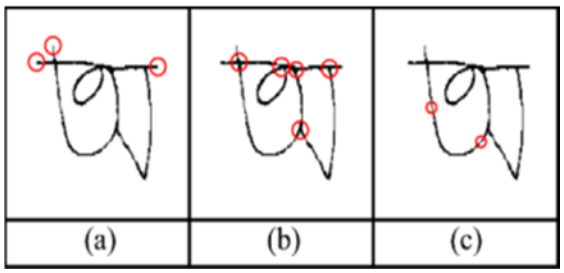

Figure 7: (a) Starters are rounded (b) Intersections are rounded (c) Minor starters are rounded

There are two conditions that may occur, intersections and non-intersections. If the current pixel is an intersection, then all the unvisited neighbors are populated in the minor starters list and the current line segment will end there. On the other hand, the non-intersection situations can occur when the current pixel under consideration has more than two neighbors but still it's not an intersection. In these situations, the current direction of traversal depend on the location of the previous pixel and if any of the unvisited pixels in the neighborhood is in this direction of traversal, then it is measured as the next pixel and all other pixels are populated in the minor starters list. Alternatively, if none of the pixels is not in the current direction of traversal, then all the pixels in the neighborhood are populated in the minor starters list and the current segment is ended there.

3.4.4 Intersections

The intersection is necessary but insufficient criterion for a pixel that it should have more than one neighbor. The intersections of Bengali 'অ' character found are given in the Fig. 7(b). For each pixel, a new property called true neighbors is defined and it is classified as an intersection or not based on the number of true neighbors. For this reason, the neighboring pixels are classified into two classes: direct pixels and diagonal pixels. All pixels in the neighborhood of the pixel under consideration in the horizontal and vertical directions is called direct pixels and the remaining pixels in the neighborhood which are in a diagonal direction to the pixel under consideration is called diagonal pixels. Then, for calculating the number of true neighbors for the pixel under consideration, it needs to be classified further based on the number of neighbors it has in the character skeleton. Pixels under consideration are classified as those with various neighbors. Such as,

- For 3 neighbors: The pixel under consideration cannot be an intersection, if any one of the direct pixels is adjacent to anyone of the diagonal pixels. On the other hand, intersection exists if none of the neighboring pixels is adjacent to each other.
- For 4 neighbors: The pixel under consideration cannot be considered as an intersection, if each and every direct pixel has an adjacent diagonal pixel or vice-versa.
- For 5 or more neighbors: The pixel under consideration considered as an intersection, if the pixel has five or more neighbors.

Formerly, all the intersections in the image are identified and stored in a list.

3.4.5 Feature Vector

Finally, the feature vector is formed for every zone based on the line type of each segment. In the proposed model, every zone has a feature vector with a length of 9 which is used to train SVM and ANN to recognize the characters. The contents of each zone feature vector are divided into two categories.

i. A number of -
   - Horizontal lines
   - Vertical lines
   - Right diagonal lines
   - Left diagonal lines

ii. Normalized -
   - Length of all horizontal lines
   - Length of all vertical lines
   - Length of all right diagonal lines
   - Length of all left diagonal lines

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- Area of the Skeleton

Where, the number of any particular line type is normalized by using the following method,

$$Value = 1 - \frac{Number\ of\ lines * 2}{10}$$

And the normalized length of any particular line type is determined by using the following method,

$$Length = \frac{Total\ pixels\ in\ that\ line\ type}{Total\ zone\ pixels}$$

Here, explained 9 features are extracted individually for each zone. So, if there are *N* zones, then there will be 9**N* elements in feature vector for each zone. Finally, certain features were extracted based on the regional properties. Namely,

- Euler Number: Euler number is defined as the difference between number of objects and number of holes in the image.
- Eccentricity: Eccentricity is defined as the eccentricity of the smallest ellipse that fits the skeleton of the image.
- Regional Area: Regional area is defined as the ratio of the number of the pixels in the skeleton to the total number of pixels in the image.

3.5 Classification

Characters are classified by classifier, which works as decision making to classify from one category to another category of classes of characters. The performance of a classifier depends on proper features. Among various types of classification techniques, we used SVM and ANN classifiers to achieve the best possible result in our proposed model. The training features from the characters are extracted using the feature extraction technique as mentioned in the above section. The SVM and ANN are provided 108 feature values from the character features.

In this proposed model, we used multilayer perceptron (MLP) which is a feed forward artificial neural network that maps sets of inputs to desired outputs. This network is used with three layers including a hidden layer for different types of features sets which consist of 108 features for each character image. In ANN, each node in network is a neuron with a nonlinear activation function except input nodes. In this technique, supervised learning is used which is called Backpropagation to train a network. Results of this experiment are obtained using the feature extraction technique for recognition of Bengali characters. In this proposed model, we used isolated characters so there is no need of segmenting the characters. We have used 1250 number of inputs for which the total feature set of geometric feature extraction method and the hidden layer are not fixed. We worked on the values 76-101 to get optimal results. On the other hand, SVM is a powerful discriminating binary classifier which models the decision boundary between two classes as a separating hyper plane. This hyper plane tries to split, one class consists of the target training vector (labeled as +1), and the other class consists of the training vectors from an impostor (background) population (labeled as -1). Using the labeled training vectors, SVM optimizer finds a separating hyper plane that maximizes the margin of separation between these two classes which is shown in Fig. 8.

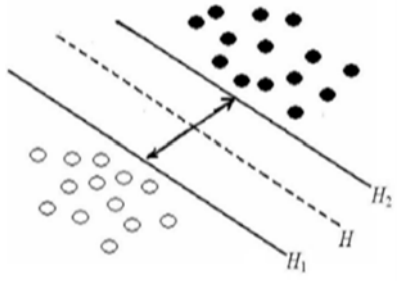

Figure 8: The optimal plane of SVM in linearly separable condition

4. EXPERIMENTAL RESULT ANALYSIS

For the recognition process of the proposed Bengali handwritten character recognition model, we used two different classifiers ANN and SVM separately to recognize the characters. The recognition results have gained from handwritten Bengali characters dataset on 2500 sample images individually. The proposed model has been operated on PNG format, where every image was in 165×165 resolution. Table 1 shows the detail of the dataset.

Table 1: SVM and ANN learning dataset

Purpose	Number of images	Sample of each alphabet	Resolution	Pixel value	Image type
Training	1250	25	165×165	uint8	PNG
Testing	1250	25	165×165	uint8	PNG

For both SVM and ANN classification 50% of the character dataset have been occupied for training and 50% characters have been used in testing section. In the proposed model, we obtained the maximum accuracy 84.47% by using the SVM classifier with 10 fold cross validation and through

Journal of Theoretical and Applied Information Technology
15th December 2019. Vol.97. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the ANN classifier the recognition rate was 74.56%. The confusion matrix in Fig. 9 and Fig. 10 show the classification result for testing set using SVM and ANN classifier respectively. Here, class 1 to class 50 describes the classes of character 'অ' to the class of character 'ঃ' respectively. The correctness and the errors can be easily determined from the confusion matrices. The maximum recognition rate is calculated with mean diagonal of the confusion matrix. However, Table 2 shows the accuracy of SVM and ANN classifiers based on our proposed model for the testing set. From the result analysis, it can be concluded that the SVM worked satisfactory with the geometry based features for the Bengali handwritten character recognition as it produces the highest accuracy.

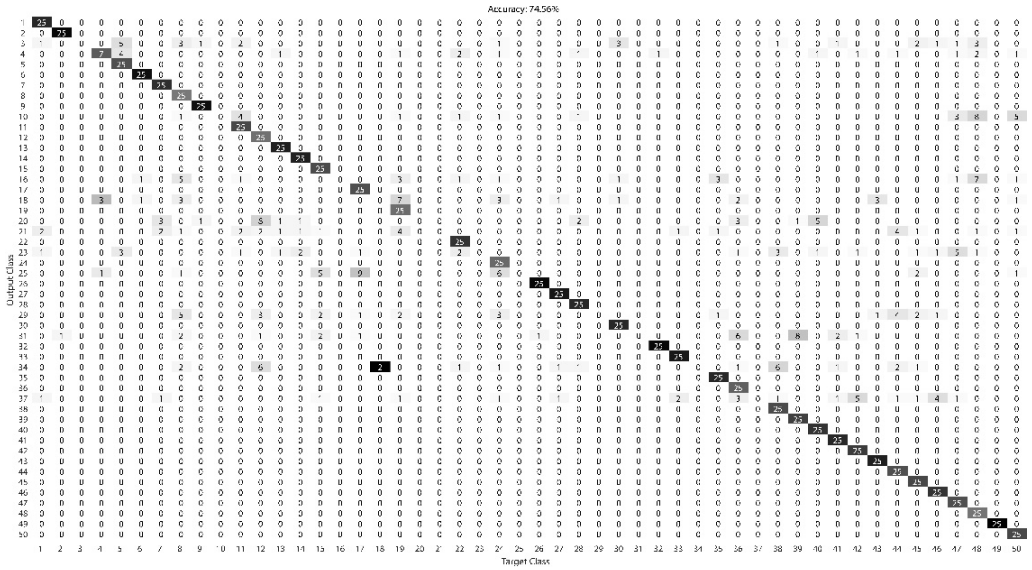

Figure 9: Confusion matrix using ANN

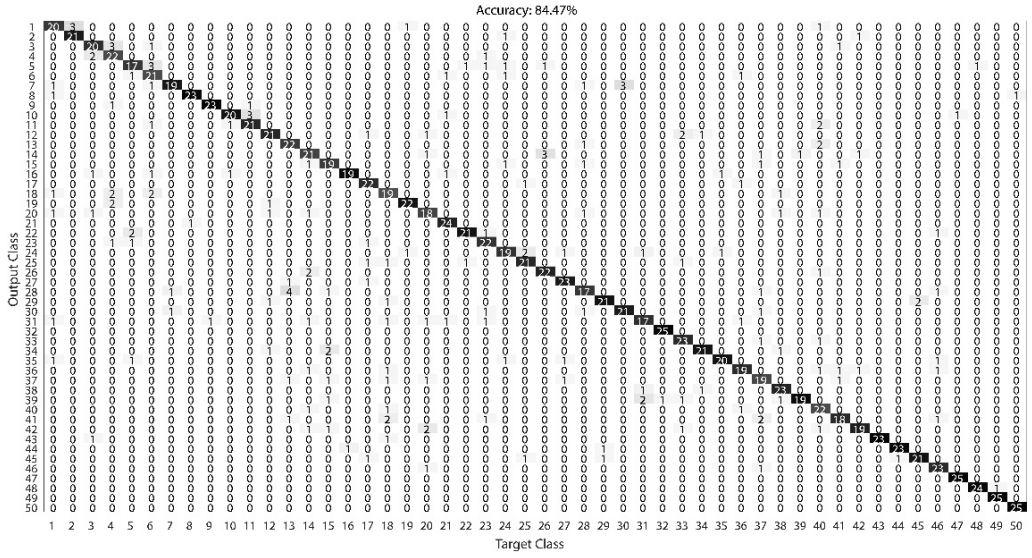

Figure 10: Confusion matrix using SVM

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Experimental result

Characters	Train	Test	Recognized	
			SVM	ANN
অ	25	25	20	25
আ	25	25	21	25
ই	25	25	20	0
ঈ	25	25	22	7
উ	25	25	17	25
ঊ	25	25	21	25
ঋ	25	25	19	25
এ	25	25	23	25
ঐ	25	25	23	25
ও	25	25	20	0
ঔ	25	25	21	25
ক	25	25	21	25
খ	25	25	22	25
গ	25	25	21	25
ঘ	25	25	19	25
ঙ	25	25	19	0
চ	25	25	22	25
ছ	25	25	19	0
জ	25	25	22	25
ঝ	25	25	18	0
ঞ	25	25	24	0
ট	25	25	21	25
ঠ	25	25	22	0
ড	25	25	19	25
ঢ	25	25	21	0
ণ	25	25	22	25
ত	25	25	23	25
থ	25	25	17	25
দ	25	25	21	0
ধ	25	25	21	25
ন	25	25	17	0
প	25	25	25	25
ফ	25	25	23	25
ব	25	25	21	0
ভ	25	25	20	25
ম	25	25	19	25
য	25	25	19	0
র	25	25	23	25
ল	25	25	19	25
শ	25	25	22	25
ষ	25	25	18	25

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

স	25	25	19	25
হ	25	25	23	25
ড়	25	25	23	25
ঢ়	25	25	21	25
য়	25	25	23	25
□	25	25	25	25
ং	25	25	24	25
ঃ	25	25	25	25
ঁ	25	25	25	25
Overall Accuracy			84.47%	74.56%

5. CONCLUSION

An efficient model is proposed in this work for recognizing handwritten Bengali characters by using geometric features of the character image. A detailed experimental result is shown in Table 2, where we can see that the proposed model works with 84.47% accuracy for SVM and 74.56% accuracy for ANN. In Fig. 9 and Fig. 10 confusion matrices are included with the recognition accuracy using a 2500 dataset (1250 for training and 1250 for testing). It is now an integral part of computer science to recognize handwritten characters as many documents present are written in hand and using handwritten character recognition we can further exploit this field such as summarizing handwritten text or finding keywords from handwritten text. From results of the experiments, this work gives a noticeable reduction in the number of most discriminating regions as well as significant increment of the recognition accuracy. The results have shown great promise in this approach. Therefore, it opens up a new frontier for more successful handwritten character recognition systems. Also, it presents with future scope for researchers to improve its performance by using different feature-set or employing a more powerful variant of support vector machine method present in the literature.

REFERENCES:

[1] M. Asgari, F. Pirahansiah, M. Shahverdy and M. Fartash, "Using an Ant Colony Optimization Algorithm for Image Edge Detection as a Threshold Segmentation for OCR System", *Journal of Theoretical and Applied Information Technology*, Vol. 95, No. 21, pp. 5654-5664, 2017.

[2] N. H. Abbas, K. N. Yasen, K. H. A. Faraj, L. F. A. Razak and F. L. Malallah, "Offline Handwritten Signature Recognition using Histogram Orientation Gradient and Support Vector Machine", *Journal of Theoretical and Applied Information Technology*, Vol. 96, No. 8, pp. 2075-2084, 2018.

[3] S. M. Ismail and S. N. H. S. Abdullah,"Geometrical-Matrix Feature Extraction for On-Line Handwritten Characters Recognition", *Journal of Theoretical and Applied Information Technology*, Vol. 49, No. 1, pp. 86-93, 2013.

[4] S. D. Kulik, "Neural Network Model of Artificial Intelligence for Handwriting Recognition", *Journal of Theoretical and Applied Information Technology*, Vol. 73, No. 2, pp. 202-211, 2015.

[5] L. Naika R., R. Dinesh and Santoshnaik, "Handwritten Electronic Components Recognition: An Approach Based on HOG + SVM", *Journal of Theoretical and Applied Information Technology*, Vol. 96, No. 13, pp. 4020-4028, 2018.

[6] M. Abaynarh, H. Fadili and L. Zenkouar, "Enhanced Feature Extraction of Handwritten Characters and Recognition using Artificial Neural Networks", *Journal of Theoretical and Applied Information Technology*, Vol. 72, No. 3, pp. 355-365, 2015.

[7] N. V. Rao, A. S. C. S. Sastry, A. S. N. Chakravarthy and P. Kalyanchakravarthi, "Optical Character Recognition Technique Algorithms", *Journal of Theoretical and Applied Information Technology*, Vol. 83, No. 2, pp. 275 - 282, 2016.

[8] A. James, K. Sujala, and C. Saravanan, "A Novel Hybrid Approach for Feature Extraction in Malayalam Handwritten Character Recognition", *Journal of Theoretical and Applied Information*

Technology, Vol. 96, No. 13, pp. 4191-4202, 2018.

[9] S. H. S. Al-Kilidar and L. E. George, "Texture Recognition using Co-occurrence Matrix Features and Neural Network" *Journal of Theoretical and Applied Information Technology*, Vol. 95, No. 21, pp. 5949 - 5961, 2017.

[10] S. V. Rajashekararadhya and P. Vanajaranjan, "Efficient Zone Based Feature Extraction Algorithm for Handwritten Numeral Recognition of Four Popular South-Indian Scripts", *Journal of Theoretical and Applied Information Technology*, Vol. 4, No. 12, pp. 1171-1181, 2008.

[11] Y. P. Singh, V. S. Yadav, A. Gupta and A. Khare, "Bi Directional Associative Memory Neural Network Method in the Character Recognition", *Journal of Theoretical and Applied Information Technology*, Vol. 5. No. 4, pp. 382-386, 2009.

[12] 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, 2014.

[13] L. F. C. Pessoa and P. Maragos. "Neural networks with hybrid morphological/rank/linear nodes: a unifying framework with applications to handwritten character recognition", *Pattern Recognition*, vol. 33, no. 6, pp. 945-960, 2000.

[14] A. Bellili, M. Gilloux and P. Gallinari. "An MLP-SVM combination architecture for offline handwritten digit recognition", *Document Analysis and Recognition*, vol. 5, no. 4, pp. 244-252, 2003.

[15] J. Dong, A. Krzyak and C. Y. Suen. "An improved handwritten Chinese character recognition system using support vector machine", *Pattern Recognition Letters*, vol. 26, no. 12, pp. 1849-1856, 2005.

[16] G. Vamvakas, B. Gatos and S. J. Perantonis. "Handwritten character recognition through two-stage foreground subsampling", *Pattern Recognition*, vol. 43, no. 8, pp. 2807-2816, 2010.

[17] D. D. Gaurav and R. Ramesh. "A Feature Extraction Technique Based On Character Geometry for Character Recognition", *Computing Research Repository*, 2012.

[18] A. Singh and K. A. Maring. "Handwritten Devanagari Character Recognition using SVM and ANN", International *Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 8, 2015.

[19] R. Azim, W. Rahman and M. F. Karim. "Bangla Hand Written Character Recognition Using Support Vector Machine", *International Journal of Engineering Works*, vol. 3, no. 6, pp. 36-46, 2016.

[20] Md. M. Rahman, M. A. H. Akhand, S. Islam and P. C. Shill. "Bangla Handwritten Character Recognition using Convolutional Neural Network", *I. J. Image, Graphics and Signal Processing*, vol. 7, no. 8, pp. 42-49, 2015.

[21] JM.H. M. Jayamaha and HMM. Naleer. "Feature Extraction Technique Based Character Recognition Using Artificial Neural Network For Sinhala Characters", *Enriching the Novel Scientific Research for the Development of the Nation*, 2016.

[22] T. K. Bhowmik, U. Bhattacharya, and S. K. Parui. "Recognition of Bangla Handwritten Characters Using an MLP Classifier Based on Stroke Features", *International Conference on Neural Information Processing*, pp. 814-819, 2004.

[23] I. K. Sethi and B. Chatterjee. "Machine Recognition of constrained Hand printed Devnagari", *Pattern recognition*, vol. 9, no. 2, pp. 69-75, 1977.

[24] L. Ragha and M. Sasikumar. "Using moments features from Gabor directional images for Kannada handwriting character recognition", *International Conference and Workshop on Emerging Trends in Technology*, pp. 53-58, 2010.

[25] K. Y. Rajput and S. Mishra. "Recognition and Editing of Devnagari Handwriting Using Neural Network", *International Conference on SPIT-IEEE Colloquium*, Vol. 1.M.

[26] S. V. Rajashekararadhya and P. V. Ranjan. "Neural Network Based Handwritten Numeral Recognition of Kannada and Telugu Scripts", *TENCON IEEE Region 10 Conference*, 2008.

[27] S. Arora, D. Bhattacharjee, M. Nasipuri, D. K. Basu, and M. Kundu. "Recognition of Non-Compound Handwritten Devnagari Characters using a Combination of MLP and Minimum Edit Distance". *International Journal of Computer Science and Security*, vol. 29, pp. 641-662, 1997.

[28] S. Arora, D. Bhatcharjee, M. Nasipuri, and L. Malik. "A two stage classification approach for handwritten Devnagari characters", *International Conference on Computational Intelligence and Multimedia Applications*, 2007.

[29] J. John, K. Pramod, and K. Balakrishnan. "Offline handwritten Malayalam Character Recognition based on Chain code histogram", *International Conference on Emerging Trends in Electrical and Computer Technology*, 2011.

[30] A. Sigappi, S. Palanivel, and V. Ramalingam. "Handwritten Document Retrieval System for Tamil Language", *International Journal of Computer Application*, vol. 31, no. 4, pp. 42-47, 2011.

[31] V. L. Lajish. "Handwritten Character Recognition using Perpetual Fuzzy Zoning and Class Modular Neural Networks", *International Conference on Innovations in IT*, 2007.

[32] C. S. Kumar and T. Ravichandran, "Handwritten Tamil Character Recognition using RCS Algorithms", *International Journal of Computer Applications*, vol. 8, 2010.

[33] S. K. Sangame, R. J. Ramteke and R. Benne. "Recognition of isolated handwritten Kannada vowels", *Advances in Computational Research*, vol. 1, pp. 52-55, 2009.

[34] N. Shanthi and K. Duraiswamy. "A Novel SVM-based Handwritten Tamil Character Recognition System", *Pattern Analysis and Applications*, vol. 13, no. 2, pp. 173-180, 2010.

[35] M. A. Rahiman, A. Shajan, A. Elizabeth, M. Divya, G. M. Kumar, and M. Rajasree. "Isolated handwritten Malayalam character recognition using HLH intensity patterns", *International Conference on Machine Learning and Computing*, 2007.

[36] R. Suresh, S. Arumugam, and L. Ganesan. "Fuzzy Approach to Recognize Handwritten Tamil Characters", *International Conference on Computational Intelligence and Multimedia Applications*, 1999.

[37] C. Sureshkumar and T. Ravichandran. "Recognition and Conversion of Handwritten Tamil Character", *International Journal of Research and Reviews in Computer Science*, vol. 1, 2010.

[38] G. Raju. "Recognition of unconstrained handwritten Malayalam characters using zero-crossing of wavelet coefficients," *International Conference on Advanced Computing and Communications*, 2006.

[39] U. Bhattacharya, S. Ghosh, and S. K. Parui. "A Two Stage Recognition Scheme for Handwritten Tamil Characters", *International Conference on Document Analysis and Recognition*, 2007.

[40] B. B. Chaudhuri, D. D. Majumder, and S. K. Parui. "A Procedure for Recognition of Connected hand written Numerals", *International Journal of Systems Sciences*, vol. 13, pp. 1019-1029, 1982.

[41] U. Bhattacharya, A. Nigam, Y. Rawat, and S. Parui. "An analytic scheme for online handwritten Bangla cursive word recognition", *International Conference on Frontiers in Handwriting Recognition*, 2008.

[42] C. Biswas, U. Bhattacharya and S. Parui. "HMM Based Online Handwritten Bangla Character Recognition using Dirichlet Distributions", *International Conference on Frontiers in Handwriting Recognition*, 2012.

[43] A. Dutta and S. Chaudhury. "Bengali alpha-numeric character recognition using curvature features", *Pattern Recognition*, vol. 26, no. 12, pp. 1757-1770, 1993.

[44] U. Pal, K. Roy, and F. Kimura. "A Lexicon Driven Method for Unconstrained

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Bangla Handwritten Word Recognition", *International Workshop on Frontiers in Handwriting Recognition*, 2006.

[45] A. F. R. Rahman, R. Rahman, and M. C. Fairhurst. "Recognition of handwritten Bengali characters: a novel multistage approach", *Pattern Recognition*, vol. 35, no. 5, pp. 997-1006, 2002.

[46] S. Basu, N. Das, R. Sarkar, M. Kundu, M. Nasipuri, and D. K. Basu. "Handwritten Bangla Alphabet Recognition using an MLP Based Classifier", *National Conference on Computer Processing of Bangla*, 2005.

[47] U. Bhattacharya, B. K. Gupta, and S. K. Parui. "Direction code based features for recognition of online handwritten characters of Bangla", *International Conference on Document Analysis and Recognition*, 2007.

[48] T. K. Bhowmick, U. Bhattacharya, and S. K. Parui. "Handwritten Characters Using an MLP Classifier Based on Stroke Features", *International Conference on Neural Information Processing*, 2004.

[49] U. Bhattacharya, M. Shridhar, S. K. Parui, P. K. Sen, and B. B. Chaudhuri. "Offline recognition of handwritten Bangla characters: an efficient two-stage approach", *Pattern Analysis and Applications*, vol. 15, no. 4, pp. 445-458, 2012.

[50] Fan, Kuo-Chin, Yuan-Kai Wang, and Tsann-Ran Lay. "Marginal noise removal of document images", *Pattern Recognition*, vol. 35, no. 11, pp. 2593-2611, 2002.

[51] Liu, Cheng-Lin, and K. Marukawa. "Pseudo two-dimensional shape normalization methods for handwritten Chinese character recognition", *Pattern Recognition*, vol. 38, no. 12, pp. 2242-2255, 2005.

[52] Yang, Feng, Z. Ma, and M. Xie. "A novel binarization approach for license plate", *IEEE Conference in Industrial Electronics and Applications*, 2006.

[53] Wikipedia. "Otsu's method". [Online]. Available:https://en.m.wikipedia.org/wiki/Otsu%27s_method.