2017 3rd International Conference on Electrical Information and Communication Technology (EICT), 7-9 December 2017, Khulna, Bangladesh

A Feature Fusion Based Optical Character Recognition of Bangla Characters Using Support Vector Machine

Mst. Tasnim Pervin[1], Shyla Afroge[2], Aminul Huq[3]
[1, 2, 3]Dept. of Computer Science and Engineering (Rajshahi University of Engineering and Technology, Bangladesh)
Email: tasnimpervin@gmail.com[1], shyla.ruet@gmail.com[2], aminul.huq11@gmail.com[3].

Abstract— **This paper introduces a feature fusion based optical Bangla character recognition approach using Support Vector Machine (SVM) including the feature extraction method named Zoning and Gabor Filter. Optical character recognition (OCR) is mostly dependent on the features extracted from characters which highly influences the classification and recognition accuracy. This paper emphasizes on the feature fusion of two different feature vectors obtained by Zoning and Gabor filter of suitable frequencies and orientations along with finding the recognition rate by passing the features into a well-developed classifier, Support Vector Machine (SVM) for classification. Here, a comparison has also been made among the recognition accuracy by individual features and by feature fusion which reveals that feature fusion based method performs better (92.99%) than a single feature extraction method (68.15% for Zoning, 89.73% for Gabor filter) during classification.**

Keywords—Feature extraction, Zoning, Gabor filter, Feature fusion, SVM classifier..

I. INTRODUCTION

Optical character recognition (OCR) plays a vital role in document image analysis which has become a challenging research since recent years [1]. Advance automation process for human-machine interaction is mostly dependent on OCR. Bangla is the 4th most widely spoken language with approximately 200 million speakers of Bangladesh and West Bengal of Indian state [2]. But, Bangla OCR is slightly lagging behind for having multifold, modifiers, and complexity of the script. Besides, most researches were done using classifier designs which are hugely based on simple feed-forward and back-propagation neural networks and as a result, there is a lack of high performance OCR engine for Bangla [3], [4], [5], [6]. The first step of the system is pre-processing of the image before applying feature extraction methods such as projection, zoning, histograms, gabor filter, chain code, contour profiles, zernike moments etc. [7]. In this paper, Zoning method and Gabor filters have been used to extract relevant features from the images. Zoning method works with statistical distribution of pixel of several overlapping or non-overlapping regions [8]. Features generated by Gabor filters are used for their invariance to illumination, rotation, scale and translation and can be used as input to a classification or segmentation operator [9], [10]. In the final step of classification, extracted features are fed to a classifier like Artificial Neural Networks (ANN), Support Vector Machines (SVM), Back-Propagation Neural Network (BPNN), Kohonen network etc.

II. RELATED WORKS

The Bangla OCR systems started emerging in mid of 1980 and various efforts to establish a working Bangla OCR have been continued since then. Some of these efforts have emphasized on only printed but not handwritten characters. Md. Mahbub Alam and Dr. M. Abul Kashem used Freeman chain code for feature extraction and Multilayered feed forward neural network for classification [14]. Shamim Ahmed, Mohammod Abul Kashem tried to improve the accuracy of character segmentation of Bangla characters using Back Propagation Neural Network (BPNN) [15]. Md. Mojahidul Islam, Md. Imran Hossain & Md. Kislu Noman worked on Automatic Feature Extraction using XOR Operation where XOR operation was performed between the scaled image and the database image and the percentage of accuracy was calculated which showed 80% average recognition accuracy [16]. Several works have been done using digital curvelet transform and K-nearest neighbour (KNN) classifier for recognition using curvelet coefficients and the trained model values were fused later by a simple majority voting scheme to obtain the final result.

III. METHODOLOGY

The proposed system, overall works are divided into three major parts: Pre-processing, Feature Extraction, Classification of the characters which is shown below in Fig.1 diagram.

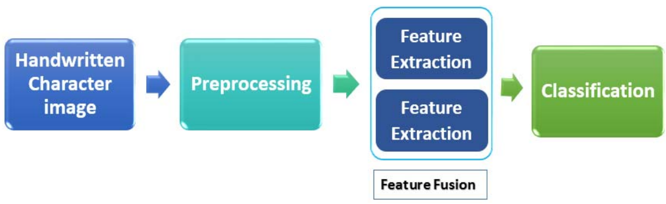

Fig. 1 Overview of the proposed system of Bangla Optical Character Recognition

A. Preprocessing

The initial stage to process the scanned image before feature extraction is preprocessing which includes binarization, noise reduction, skew detection, segmentation and normalization. These steps are discussed below.

1) Binarization: Input Character images are binarized using Otsu's method [13]. After converting the image into gray level, it is converted to binary image by turning all the pixels to zero below the threshold and one exceeding the threshold.

2) Noise Reduction: Depending on the resolution on the scanner and the success of the applied technique for thresholding, the characters may be noisy or broken. These noise causes poor recognition rates, can be eliminated by using a preprocessor or filters to smooth the digitized characters.

3) Skew Detection: During scanning document manually or mechanically, a few degrees of skew is unavoidable. Skew angle is the angle deviation between text lines and the horizontal direction. Hough transform are commonly used for skew correction. It can be done by (i) estimating the skew angle, θ and (ii) rotating the image by the θ in the opposite direction. These steps make the digital image look-alike to original image.

4) Segmentation: Segmentation process includes line segmentation, word segmentation and character segmentation. Segmentation is done by observing the histogram of horizontal or vertical scanning of image and counting foreground pixel frequency. Segmentation occurs at that point where count is zero.

5) Normalization: The size normalization or scaling is applied to obtain characters of uniform size and same dimension of feature vectors. In this paper, each character is scaled into 40x48 before proceeding to feature extraction.

B. Feature Extraction

The Feature extraction is considered as the most difficult phase of recognition. The main objective of feature extraction is to identify special characteristics of any character symbol that characterizes that individual character uniquely and eliminates unnecessary attributes from total feature space.

In this paper, two methods are used named Zoning and Gabor filter to obtain feature vectors. Following sections describe these methods theoretically.

1) Zoning: Zoning is statistical feature extraction method that extracts features from character symbol on the basis of distribution of pixels. In this method, the scaled binary images are divided into 8x8 zones of predefined sizes of 6x5 dimension to obtain 64 blocks or zones. The features are extracted by counting the number of black pixels and white pixels of each blocks to calculate pixel density where black pixels are represented by 0s and white pixels are represented by 1s. For average pixel density, d_{avg} of a zone, the total number of foreground black pixels are divided by the total number of pixels (foreground and background) of each zone [17]. Each zone contribute to the value of the overall feature vector. The average pixel density calculated by following equation for each consicutive zone helps to form the feature vectors for each character image. Equation (1) is the formula to compute d_{avg}.

$$d_{avg} = \frac{\text{Number of foreground pixels in a zone}}{\text{Total number of pixels in that zone}} \quad (1)$$

2) Gabor filters: Gabor filters are a powerful tool for feature extraction. The most advantageous properties of Gabor filters are their invariance properties to rotation, scale, and translation, robustness against photometric disturbances such as illumination of noise. The Gabor filters provide the optimal resolution in both the spatial domain and transform domain to extract local features of an image. Relations between activations for a specific spatial location are very distinctive between objects in an input character image. Unit variance and zero mean are used for normalization [18]. Gabor filters can be used as a directional feature extractor where features can be captured in various directions, which is quite similar to Human Visual System (HVS) [11],[12]. The two-dimensional Gabor filter represents a convolutional band pass linear filter whose impulse response is defined by a complex sinusoid harmonic function multiplied by a Gaussian kernel function. Equation (2) shows this expression below.

$$G_{(x,y)} = g_{(x,y)} s_{(x,y)} \quad (2)$$

Where, $g_{(x,y)}$ is a Gaussian function known as envelope and $s_{(x,y)}$ is a complex sinusoid harmonic function. Equation (3) represents the mathematical representation of the complex sinusoid, $s_{(x,y)}$.

$$s_{(x,y)} = e^{-j(2\pi(u_0 x + v_0 y) + \varphi)} \quad (3)$$

Where, (u_0, v_0) is the spatial frequency and $-$ represents the negative phase of the sinusoid. Equation (4) and (5) shows that complex sinusoid consists of real and imaginary part.

$$R(s_{(x,y)}) = \cos(2\pi(u_0 x + v_0 y) + \varphi) \quad (4)$$

$$I(s_{(x,y)}) = \sin(2\pi(u_0 x + v_0 y) + \varphi) \quad (5)$$

Equation (6) represents the 2-D Gaussian function.

$$g_{(x,y)} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} \quad (6)$$

Finally, Equation (7) shows the 2-D Gabor filter below.

$$G_{x,y,\theta,f,\sigma_x,\sigma_y} = e^{-\frac{1}{2}\left(\frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2}\right)} X\left(\cos 2\pi f x' + j \sin 2\pi f x'\right) \quad (7)$$

Where, σ_x and σ_y are the standard deviations of the Gaussian envelop along x and y direction respectively, f is the frequency and θ is the orientation. Again x' and y' can be represented by-

$$x' = x\cos\theta + y\sin\theta$$
$$y' = -x\sin\theta + y\cos\theta$$

C. Classification

Support Vector Machine (SVM) is a well-developed supervised machine learning algorithm which can be used for both classification and regression challenges. The standard SVM is a binary classifier which takes the set of input data, classifies them in one of the two distinct classes and plot each data item in n dimensional space (where n is number of features) where the value of each feature represents the value of a particular coordinate. It was determined by Cortes and Vapnik [12] in about 1946. The ability to transform data to a high dimensional space using a hyperplane to separate the data items by maximizing distance margin is considered to be basic power of SVM [12]. However, the optimization process for learning SVM with different parameters of hyperplane has great influence over recognition accuracy rate. For classification, SVM is trained by training dataset to obtain a hyperplane based on the features to predict an output over testing dataset using previous knowledge of training phase. Equation (8) can be used to represent the hyperplane.

$$w^0 x + b = 0 \tag{8}$$

Where, w^0 is the normal to the hyperplane and b is the bias from the origin to the hyperplane. The points that lie on the hyperplane must satisfy the above equation. For a given training set that contains pairs of (X_i, Y_i) , where, feature vectors can be depicted as i = 1,.......,n and the class label are y = ±1 must satisfy the following conditions.

$$w^0 x + b > 0 \qquad \text{if} \quad Y_i = 1$$
$$w^0 x + b < 0 \qquad \text{if} \quad Y_i = -1$$

The margin is maximized by the optimization process to increase the distance. Equation (9) that represents the decision boundary function for input pattern x with binary classifier is as follows:

$$f(x) = \sin(w^0 x + b) \tag{9}$$

The distance between the hyperplane margin $w^0 x + b = 1$ and $w^0 x + b = -1$ can be represented by $2/\|w\|$. Therefore, the optimization process that generates support vectors can be presented by the equation below:

$$\frac{1}{2}\|w\|^2 - (w^0 x + b) \tag{10}$$

In machine learning, kernel methods are a class of algorithms for pattern analysis that takes two inputs and spits out the similarity between them. Three kernels are used for support vector machines. Although, in this paper, linear kernel has been used only.

TABLE I
KERNELS OF SUPPORT VECTOR MACHINE

Linear Kernel	$K(x, y) = X^T Y_i$
Polynomial Kernel	$K(x, y) = (X^T Y_i + 1)^d$
RBF Kernel	$K(x, x_i) = \exp^{-y\|x-y\|^2}$

IV. EXPERIMENTAL RESULTS

A. Dataset Description:

Because of the lack of the standardization of any Bangla dataset, existing dataset has been used [19]. For total experimental purpose, including training and testing about 35000 character images have been used for 50 Bangla characters. Among these, for each character 500 data, a total of 25000 data has been used for training and for each character 200 data, a total of 10000 data has been used for testing using conventional validation of 70-30 ratio to evaluate the recognition accuracy.

B. Results of Feature Extraction Methods:

After applying two different feature extraction methods (Zoning and Gabor filters) described above, two different feature vectors can be obtained. Zoning gives a feature vector of dimension 1x64. Sample result for a basic bangla character 'ক' has been given here applying zoning method. Fig.2 (b) shows the pixel value for zone-8 of dimension 6x5. Table II shows the matrix format of corresponding feature vector for the used character.

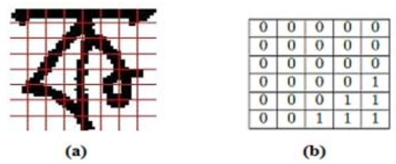

Fig. 2. (a)8x8 zoning of BW image , (b)binary value of pixels of zone-8.

TABLE II
FEATURE MATRIX BY ZONING OF CHARACTER 'ক'

0.266	0.333	0.333	0.266	0.266	0.333	0.233	0.200
0.866	1	0.966	0.266	0.400	1	0.833	1
1	1	0.433	0.600	0.233	0.433	1	1
1	0.700	0.500	0.666	0.866	0.400	0.833	1
0.933	0.300	0.966	0.566	0.933	0.300	0.200	0.900
0.800	0.366	0.366	0.466	0.866	0.300	0.433	1
1	1	0.900	0.166	0.800	0.7667	0.900	1
1	1	1	0.733	0.700	1	1	1

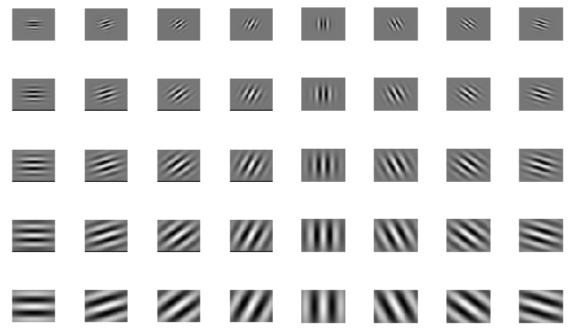

Fig.3: Real part of the Gabor filter

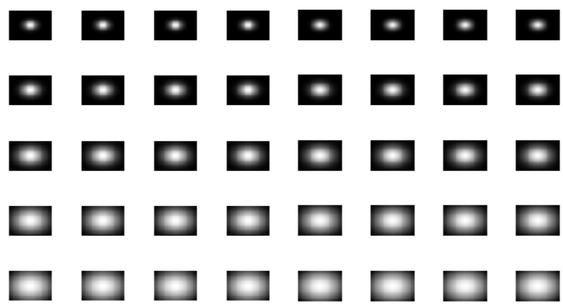

Fig.4: Magnitude of Gabor filter

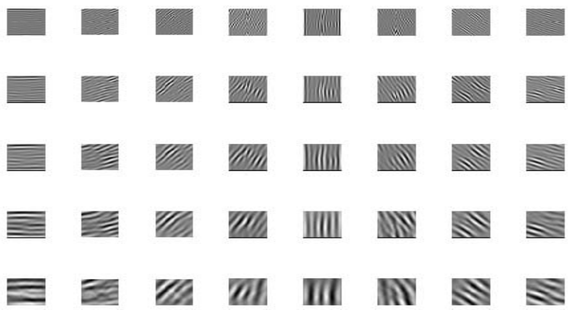

Fig. 5: Real part of the filtered character image

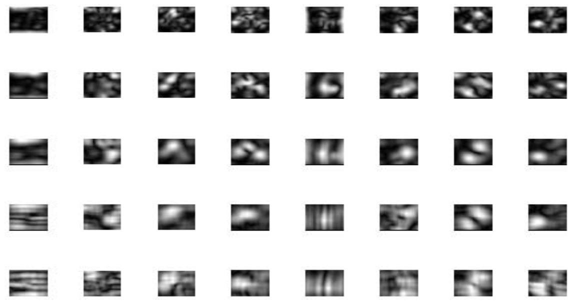

Fig. 6: Magnitude of filtered character image

Gabor produces a feature vector of dimension 1x2560 for each character image. Forty Gabor filters with a set of 5 radial frequencies and 8 successive orientations has been used here as shown in figures. Fig.3 and Fig.4 show those filters which are convolved with the sample input image resulting in a Gabor space. Fig.5 and Fig.6 are the resulting response for convolving filters with character image 'ক'. These filtered real part and magnitude provide the feature vector. Later, these two feature vectors are concatenated to form a fusion that produces a larger feature vector of dimension 1x2624 for each character image which is then passed to further steps for classification.

C. *Recognition Accuracy*

As the title depicts, Both Zoning and Gabor feature vectors are fed to SVM classifier using linear kernel and OneVsAll (OVA) coding scheme to check individual class recognition accuracy rate. Then these two feature vectors are fused together as described in previous section by concatenation producing another stronger set of feature vector which gives higher recognition accuracy as expected. This method is followed for both training and testing data. Table III shows the number of correctly recognized data and Table IV shows the recognition accuracy of different characters using following formula.

$$\text{Recognition accuracy} = \frac{\text{total no. of character image - no. of misclassified caharacter image}}{\text{total no. of character image}} * 100\%$$

Here,
Number of misclassified characters =
False Acceptance + False Rejection.
CR = Correctly Recognized
FR = False Rejection
FA = False Acceptance
ACC = Accuracy Percentage.

TABLE III
NUMBER OF RECOGNIZED CHARACTERS BY DISCUSSED FEATURE EXTRACTION METHODS.

Input Character	Number of Recognized Data by different approach (500 Training image & 200 Testing image per character)								
	Zoning			Gabor			Fusion of both		
	CR	FA	FR	CR	FA	FR	CR	FA	FR
অ	106	60	94	162	18	38	168	12	32
ঈ	134	28	66	174	14	26	180	8	20
ক	152	38	48	178	13	28	184	6	16
ঙ	134	28	66	176	23	26	186	13	14
ঞ	158	25	42	154	10	46	162	2	38
ঠ	170	103	30	190	39	10	198	31	2
ণ	168	67	32	192	14	8	192	11	8
র	154	80	46	188	19	12	194	13	6
স	104	126	43	190	48	10	194	40	6
ঢ়	174	134	26	192	5	8	194	3	6
◌ং	194	25	6	190	12	10	196	7	4

The accuracy rate of a character recognition is affected by misclassification by false acceptance and false rejection. For example, ‘ত’ is falsely accepted as ‘ভ’, again ‘স’ is falsely rejected by its own class and falsely accepted as ‘ম’. ‘অ’ and ‘আ’ are also victim of this problem. ‘ব’ and ‘র’ also lead to misclassification because of slight difference of single dot between them. This problem is mainly severe in case of handwritten characters because of pattern variation of handwriting among people. Table IV shows the percentage of accuracy for above table.

TABLE IV
RECOGNITION ACCURACY BY DISCUSSED FEATURE EXTRACTION METHODS.

Input charac-ters	Trai ning data	Testi -ng data	Zoning		Gabor		Fusion of both	
			CR	ACC (%)	CR	ACC (%)	CR	ACC (%)
অ	500	200	106	23	162	72	168	78
ঈ	500	200	134	53	174	80	180	86
ক	500	200	152	57	178	79.5	184	89
ঙ	500	200	134	53	176	75.5	186	86.5
ঞ	500	200	158	66.5	154	72	162	80
ঠ	500	200	170	33.5	190	75.5	198	83.5
ণ	500	200	168	50.5	192	89	192	90.5
র	500	200	154	37	188	84.5	194	90.5
স	500	200	104	15.5	190	71	194	77
ঢ়	500	200	174	20	192	93.5	194	95.5
◌ং	500	200	194	84.5	190	89	196	94.5

The following graphical representation in Fig. 7 shows the overall accuracy for 50 characters using 10000 testing data. These accuracy vary slightly for multiple runs and averaging them gives an approximation of final accuracy.

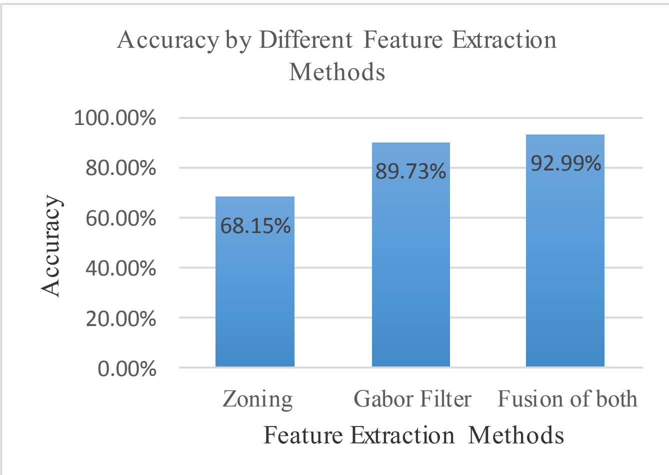

Fig. 7. Recognition accuracy by different feature extraction methods.

V. CONCLUSION

In this paper, the problem of Bangla optical character recognition has been addressed. Feature fusion technique has been addressed using Zoning and Gabor filter. Support Vector Machines are utilized for classification and recognition. From above works, it can be assumed that feature fusion based technique gives higher accuracy (92.99%) during recognition than individual ones (68.15% for Zoning, 89.73% for Gabor filter). Accuracy mostly depends on the preprocessing stage and feature extraction methods. By increasing efficiency of these phases, it might offer greater accuracy. However, some important directions for the future work may be to extend this work by incorporating modifiers along with handwritten Bangla characters, fusion of some more significant features using different fusion algorithms and try to implement other classification approaches on used dataset to evaluate comparative performance with proposed system.

REFERENCES

[1] Mori, Shunji, Ching Y. Suen and Kazuhiko Yamamoto. "Historical review of OCR research and development." Proceedings of the IEEE 80.7 (1992): 1029-1058.80.7 (1992): 1029-1058.

[2] Comrie, Bernard, ed. The world's major languages. Routledge, 2009.

[3] M.A. Hasan, P.P. Paul, M.W. Islam, S. Biswas, S. Ahmed, “Bangla digit recognition using Eigen digits”.

[4] Chowdhury, Ahmed Asif, et al. "Optical Character Recognition of Bangla Characters using neural network: A better approach." 2nd ICEE. 2002.

[5] Hasan, Md Rafiul, Mohammad Azizul Haque, and Syeda Umme Farhana Malik. "Bangla optical character recognition system." Proc. 2nd International Conference on Computer and Information Technology, ICCIT. 1999.

[6] Shatil, Adnan Md. Bangla optical character recognition using Kohonen network. Diss. BRAC University, 2006.

[7] Pradeep, J., E. Srinivasan, and S. Himavathi. "Diagonal based feature extraction for handwritten character recognition system using neural network." Electronics Computer Technology (ICECT), 2011 3rd International Conference on. Vol. 4. IEEE, 2011.

[8] LineEikvil, “Optical Character Recognition”, citeseer.ist.psu.edu/142042.html (1993).

[9] Kamarainen, J-K., Ville Kyrki, and Heikki Kalviainen. "Invariance properties of Gabor filter-based features-overview and applications." IEEE Transactions on image processing 15.5 (2006): 1088-1099.

[10] Grigorescu, Simona E., Nicolai Petkov, and Peter Kruizinga. "Comparison of texture features based on Gabor filters." IEEE Transactions on Image processing 11.10 (2002): 1160-1167.

[11] Al-Jamimi, Hamdi A., and Sabri A. Mahmoud. "Arabic character recognition using gabor filters." Innovations and advances in computer sciences and engineering (2010): 113-118.

[12] Chaudhari, Shailesh, and Ravi M. Gulati. "Script Identification Using Gabor Feature and SVM Classifier." Procedia Computer Science 79 (2016): 85-92.

[13] Otsu N (1979) A threshold selection method from grey-level histograms. IEEE Trans Syst Man Cybern 9:377–393.

[14] Alam, Md Mahbub, and Dr M. Abul Kashem. "A complete Bangla OCR system for printed characters." JCIT 1.01 (2010): 30-35.

[15] Ahmed, Shamim, and Mohammod Abul Kashem. "Enhancing the Character Segmentation Accuracy of Bangla OCR using BPNN." International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.

[16] Islam, Md Mojahidul, Md Imran Hossain, and Md Kislu Noman. "Bangla character recognition system is developed by using automatic feature extraction and XOR operation." Global Journal of Computer Science and Technology (2013).

[17] Siddharth, Kartar Singh, et al. "Handwritten Gurmukhi Character Recognition Using Statistical and Background Directional Distribution." International Journal on Computer Science and Engineering (IJCSE) 3.06 (2011): 2332-2345.

[18] Haghighat, Mohammad, Saman Zonouz, and Mohamed Abdel-Mottaleb. "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification." Expert Systems with Applications 42.21 (2015): 7905-7916.

[19] Mohammed, Nabeel; Momen, Sifat; Abedin, Anowarul; Biswas, Mithun; Islam, Rafiqul; Shom, Gautam; Shopon, Md. (2017), "BanglaLekha-Isolated", Mendeley Data, v2.