

A Hybrid Model for Recognizing Handwritten Bangla Characters using Support Vector Machine

Shyla Afroge

Assistant Professor, Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology,
Rajshahi-6204, Bangladesh.

Boshir Ahmed

Professor, Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology,
Rajshahi-6204, Bangladesh.

ABSTRACT

Considering the real time scenario, hand written bangla recognition getting a drastic part to the research community. Though various studies have been performed for Bengali handwritten recognition, but a robust model for Bangla Handwritten classification is still in practice. Therefore a hybrid model is presented in this paper, which intent to classify Bangla handwritten characters. The Proposed model combines Zernike moments, raw binary pixels and histogram of oriented gradients features for recognizing Bangla hand written characters which is feed to the Support Vector Machine classifier. It is observed that, the proposed model outsails existing models with smaller epochs. Proposed model is trained and test with “Bangla Lekha Isolated” dataset which consists of 30000 characters where 24,000 for training dataset and 6,000 for testing .This system shows 46.98% for Zernike Moments, 66.60% for Raw Binary Pixels and 87.62% for Histogram of Oriented Gradients where overall combined features achieve an accuracy of 94.88% in recognizing characters which achieves the best accuracy rate reported till date for this dataset.

Keywords

Hand written character recognition; Histogram of oriented gradients; Zernike moments; raw binary pixel ;support vector machine; Bangla OCR

1. INTRODUCTION

The increasing trend of digitization of books and manuscripts demands an immediate solution for their easy electronic access. This will make the valuable documents to be searchable and usable by the users. A traditional solution to this problem is Optical Character Recognition (OCR) which takes scanned images of paper documents as input and converts them into digital format for computer aided data processing. OCRs have wide range of applications like library and office automations, form and bank cheque processing, institutional repository captcha recognition and so on. Two types of OCR can be used. One is online another is offline. Online handwriting recognition deals with a data stream which is coming from a transducer while the user is writing. On the other hand Offline handwriting recognition deals with the images of document pages which have been obtained by scanning the documents. For Offline recognition document pages may be printed or hand written. Recognition of text is always more complex, in case of handwritten document than the printed or typed one due to the complex nature (like cursive, discrete or mixture of both) of handwritten words. In last few decades, enormous work has been developed in this field and as a result many high quality OCR systems are commercially available in the market for certain languages

In the field of pattern recognition an important section is detection of characters such as automatic data extraction, institutional repository, captcha recognition, optical music character recognition and many more. Bangla is currently the world's seventh most used language in regard to population and second in the Indian sub-continent. So the need for digitizing Bangla characters is increasing day by day. As the world a progress, Bangla language is used in many systems for this huge population that uses Bangla as their mother tongue. To improve the recognition of Bangla characters the following system is hereby proposed.

In the past many research works have been implemented on recognizing Bangla characters however this system faced many challenges like accuracy of recognition characters, extensive calculations which slowed the operations. Most of these problems can be solved by improving the feature extraction techniques. In the past many feature extraction techniques have been used as well as different classifiers for Bangla OCR. In this paper a hybrid feature extraction technique is proposed which is followed by justification of the improvement.

There are many types of feature extraction methods which can be classified into three major groups [1]: Statistical Features: Zoning, Characteristic loci, crossing and Distances etc fall in the category of statistical features. Global Transformation and Series Expansion Features: Fourier Transforms, Gabor Transforms, Wavelets, Karhunen Loeve Expansion, Moments these are members of Global Transformation and Series Expansion Features. Geometric and Topological Features: Chain codes, Stroke Direction and Bays, Stroke are Geometric and Topological features.

In our proposed system we have used three different feature extraction techniques they are Histogram of Oriented Gradients (HOG) features, Zernike moments and raw binary pixel of the image. These three techniques are used to capture the features of an image and hence used for classifying an image accordingly. The reason for using these three techniques is that they each possess different strong points based upon which they extract features from an image and collectively they provide a much better feature than each of them individually. For classification of the data we have used Support Vector Machine (SVM) which provides a very good performance for multiclass classification [2].

During the implementation of the system a training set of 24,000 characters and testing set of 6,000 characters were used for developing the system. The dataset is collected from a repository [3].

2. PREVIOUS WORK

In Literature there is not enough works for Bangla handwritten characters [4-8]. Niaran Das et al. [4] proposed an MLP and SVM based approach for recognition of bangla basic and compound characters using longest run feature. They have not considered all 160 compound characters where only first 55 characters with 90% accuracy to identify different groups. It is also found that the classes forming each group are very close to each other in shape, and therefore, it is difficult to distinguish them. In our experiment, each of these groups is treated as a single class and they are trained and tested like that. Thus the total number of compound character classes has further been reduced from 55 to 43 classes. Then 50 basic characters are combined with 43 classes. Here, a 204 element, quad tree based shadow and longest run feature set has been calculated for each of the characters of these 93 classes. An MLP based classifier is trained and tested using this feature set.

Rahman et al.[5] and Bhowmik et al [6] used multistage approach for handwritten characters recognition and MLP is used for recognition. Feature set include matra, upper part of the characters, disjoint section of the character, vertical line and double vertical line and for the MLP classifier, the feature set is constructed from the stroke features o characters.

S. Bhowmik et al. [8] a holistic word recognition approach is followed, where a word is inseperable as a unit. Here each word sample belongs as a class or object. For extracting HOG feature initially handwritten word images are smoothed with Gaussian filter for the removal of noise. After smoothing, horizontal and vertical gradients of the input image are computed. A 3-fold cross-validation method was used to evaluate the performance of the system. In each fold, 680 word images are used for training and 340 word images are used for testing. In [9] a survey on OCR systems is provided by the authors where a long history of OCR systems is discussed. The classifiers used to recognize characters have evolved over time from Neural Network [10, 11] to Support Vector Machines (SVM) [12]. An online recognition method for hand sketched symbol is presented in [13]. Zernike moments descriptors are used to represent symbols and three different classification techniques are compared: Support Vector Machine (SVM), Minimum Mean Distance (MMD) and Neural Network (NN) and obtained 97% accuracy rate on a dataset consisting of 7,410 sketched symbols using Zernike moment features and a SVM classifier.

3. PROPOSED METHODOLOGY

3.1 Image Preprocessing

Before performing feature extraction of image the steps that are performed are part of preprocessing. The steps can be shown in figure 1. These steps include image acquisition, binarization of the image, noise removal, skew detection, segmentation and scaling.



Fig.1 Preprocessing of images

3.1.1 Binarization

After acquiring an image the first step is to make it a grayscale image from the RGB image. Since processing RGB images takes much longer time than grayscale images hence it

is useful to process grayscale image. Then binarization process is performed on the image based on a particular threshold value. Here Otsu's method [14] is used for binarizing the image. Figure shows binerization of handwritten character.

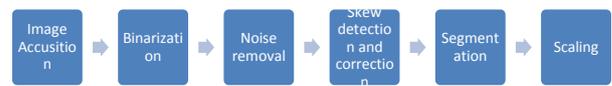


Fig.2 Binarization of an image.

3.1.2 Noise removal

Noise removal is an important section of preprocessing an image. By doing so the image provides better visibility and features. In this paper Median Filter is used upon the images. It smoothes the image surface but at the same time blurs the edges as well.

3.1.3 Skew Detection and Correction

Skew is an angle to which the image displaces when an image is acquired. This displacement causes erroneous behavior from the OCR the system. We can perform skew correction in the following way. If theta is the angle that is displaced when acquiring the image if we rotate it in the opposite direction with same value then the skew will be removed.

3.1.4 Segmentation

Segmentation is the process of creating many sub images from a particular image. There are three types of segmentation. These are line segmentation, word segmentation and character segmentation. Character segmentation can be further classified into Matra line detection, working zone detection, Baseline detection [15].

To perform line segmentation the image is scanned horizontally and is checked for a particular number of rows for which there is no interference with characters. Through line segmentation lines of a paragraph are separated and those separated lines can be further processed. Word segmentation can be performed in the similar manner but in this case the image is scanned vertically rather than horizontally. Next step is character segmentation. In Bangla language each character of a word is connected with a line called Matra. Detection and removal of Matra is similar to line segmentation. Rownak, Ahnaf Farhan, *et al* have proposed a system through which segmentation[15] can be done with quite good accuracy.

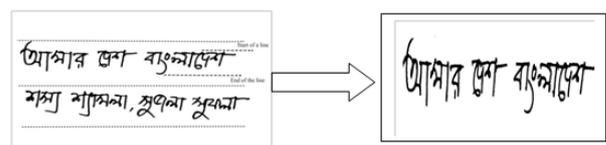


Fig. 3 Line Segmentation

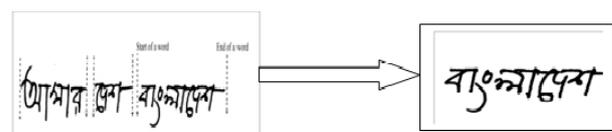


Fig.4 Word Segmentation

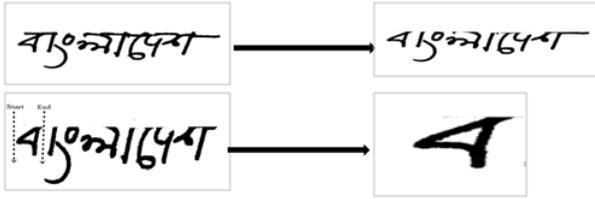


Fig.5 Character Segmentation.

3.1.5 Scaling

After all the previous steps have been done the final step in preprocessing an image is scaling. Scaling refers to the size of the image upon which feature extraction techniques will occur. As the size of the image increases more features can be extracted accurately but it the cost of computation will increase as well. So, that is keeping in mind this contradictory situation for this research the images are scaled to 32 x 32 matrices.

3.2 Feature Extraction

In feature extraction stage, each character represent as a feature vector, which becomes its identity. The major goal of feature extraction is to extract features which maximize the recognition rate with the least amount of elements and to generate similar feature set for variety of instances of the same symbol. Due to the nature of hand writing with its high degree of variability and imprecision obtaining these features, is a difficult task. Feature extraction methods analyze the input document image and select a set of features that uniquely identifies and classifies the character. Here we have used three techniques for Feature Extraction those are Histogram of Oriented Gradients, Zernike Moments, and Raw Binary Pixels.

3.2.1 Histogram of Oriented Gradients



Fig. 5 Feature Extraction using Histogram of Oriented Gradients.

Histogram of Oriented Gradients (HOG) features has become very popular in object detection in recent years specifically in human detection [16]. The process of extracting features is the shown in figure 5. HOG is used due to the fact it provides a few advantages, the shapes of local objects can be characterized very well by the distribution of gradients or in other words edge directions. If translation and rotation are smaller than orientation bin size than they do not have much impact.

The image is at first divided into small spatial regions called cells. These cells can be either rectangular or radial. Each cell has a 1-D Histogram of gradient directions [16]. The gradient orientation is quantized into 9 bins spaced from 0° to 180°. The votes are gradient magnitude. To reduce aliasing votes are interpolated bi-linearly between neighboring bin center. In order to get good performance contrast normalization is very important. The final descriptor is the then the vector of all components of normalized cells of all blocks. Generally the blocks are overlapped which seems redundant but in order to get good normalization it is important and this also provides better performance [16].

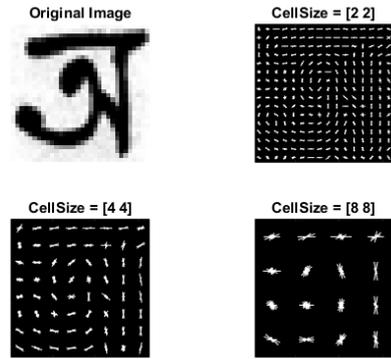


Fig.6 HOG features of Bangla character 'আ' for different cell size.

3.2.2 Zernike Moments

Moments or functions of moments are very popular feature extraction technique in pattern recognition. These are used in many application [17][18][10]. Regular Moments can be defined as

$$m = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy$$

Where $m_{(p+q)}$ is the $(p+q)^{th}$ order moment of the image defined by $f(x, y)$. For digital images the integrals are replaced by summations and thus $m_{(p+q)}$ becomes

$$m = \sum_x \sum_y x^p y^q f(x, y)$$

In [20], Zernike proposed a set of complex polynomials which created a complete orthogonal set over the interior of the unit circle, i.e., $x^2 + y^2 = 1$ [17]. Zernike Moment provides the advantage of being translation, rotation and scale invariant for recognizing an image. This complex polynomial is defined as, $V_{nm}(x, y)$. It can be denoted as

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(jm\theta)$$

Where

n Positive integer or zero, m is positive and negative integers, $|m| \leq n$ and $n - |m| = \text{even}$ ρ is Length of the vector from origin to (x, y) , θ is the Angle between the vector ρ and x axis in counterclockwise direction

$R_{nm}(\rho)$, is the Radial Polynomial. It is denoted by,

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2}-s\right)! \left(\frac{n-|m|}{2}-s\right)!} \rho^{n-2s} \text{ And } R_{n,-m}(\rho) = R_{nm}(\rho).$$

Zernike moments are projection of the image function onto this orthogonal basis function [17]. The Zernike moment of order n with repetition m for a continuous image function $f(x, y)$ is

$$A_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} f(x, y) V_{nm}^*(\rho, \theta) dx dy$$

Where $*$ is the complex conjugate operator. For a digital image the integrals are replaced with summations

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V_{nm}^*(\rho, \theta) dx dy, \quad x^2 + y^2 \leq 1$$

Below Table I list the first 36 moments of up to order 10.

Table I: List of no. of moments up to order 10

Order(n)	Zernike Moment of order n with repetition $m(A_{nm})$	No. of Moments
0	A_{00}	1
1	A_{11}	1
2	A_{20}, A_{22}	2
3	A_{31}, A_{33}	2
4	A_{40}, A_{42}, A_{44}	3
5	A_{51}, A_{53}, A_{55}	3
6	$A_{60}, A_{62}, A_{64}, A_{66}$	4
7	$A_{71}, A_{73}, A_{75}, A_{77}$	4
8	$A_{80}, A_{82}, A_{84}, A_{86}, A_{88}$	5
9	$A_{91}, A_{93}, A_{95}, A_{97}, A_{99}$	5
10	$A_{10,0}, A_{10,2}, A_{10,4}, A_{10,6}, A_{10,8}, A_{10,10}$	6

3.2.3 Raw Binary Pixels

After the binarization process an image becomes a 2D matrix with values of 1 and 0 only. A feature vector can be created from those pixel values by taking all the values of the image and making it a row vector. Now this row vector is capable of being a feature extraction technique where each column corresponds to a feature of that image.

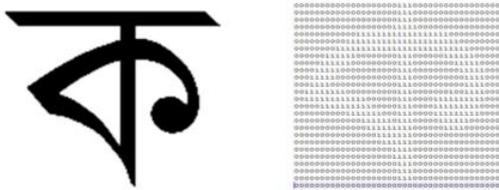


Fig.7 Raw Binary pixels of an image

4. CLASSIFICATION

In this section we will review about Support Vector Machine for classification problems [21]. Recently Support Vector Machines have been implemented successfully for pattern recognition tasks. This was mainly designed for binary classification for determining the optimal hyper-plane to maximize the margin of separation between positive and negative data set [22]. Although this can be extended to multiclass classification by using the one versus all technique. The goal of SVM is to predict the output y on the testing set based on the features provided by the testing set after being trained using the training data set. For a given a training data set $\{(x_i, y_i), 1 \leq i \leq n\}$ where n is the total no. of data points, $x_i \in \mathbb{R}^n$ is the i^{th} input pattern and $y_i \in \mathbb{R}^n$ is the i^{th} output pattern. The SVM maximizes the cost function $\frac{1}{2} W^T W$, which is subject to constraints

$$x_i \cdot w + b \geq 1 \text{ for } y_i = 1 \text{ and}$$

$$x_i \cdot w + b \leq -1 \text{ for } y_i = -1$$

Where w is the weight factor and b is the bias term. In case a separating hyper plane in this higher dimensional space does not exist, variables ξ are introduced such that

$$y_i[w^T \varphi(x_i) + b] \geq 1 - \xi_i, i = 1, \dots, n,$$

$$\xi_i \geq 0, i = 1, \dots, n$$

Where φ is a nonlinear function which maps the input space into a higher dimensional space. SVM now tries to determine w which minimizes $\frac{1}{2} W^T W + C \sum_{i=1}^n \xi_i$

A kernel is similarity function. It takes two inputs and returns how similar they are. Some popular used kernels are: Polynomial of degree d , $K(x, x_i) = (x x_i + 1)^d$. Radial Basis Function (RBF) or commonly known as Gaussian kernel: $K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2})$. Both of these two kernels are used for non-linear SVM.

5. EXPERIMENTAL RESULT AND ANALYSIS

As there is no standard dataset for Bangla OCR for the experimental purpose we have used the dataset which is collected from [4]. For each character 500 data has been used a total of 30,000 characters for training and testing Bangla characters and numerals. 24,000 characters have been used for training and 6,000 characters have been used for testing. Before collecting features and passing it to the classifiers all the preprocessing steps mentioned above have been performed for both the training data set and testing dataset. For feature extraction in Histogram of oriented gradients we have used cell size of [8X8] which makes the calculation computationally faster. In Zernike Moment the moments of an image are collected till the order of 17. Support Vector Machine was implemented here using linear kernel and one vs one coding scheme. The below Table II shows the accuracy of identifying characters using these three feature extraction technique individually and in a combined manner which has been proposed here.

Table II: Accuracy rate corresponding to different feature extraction techniques for different characters.

Input	No. of Testing data	No. of Correctly Identified Data			
		Zernike Moments	Raw Binary Pixels	Histogram of Oriented Gradients	Combined
স	100	42	88	87	91
৩	100	46	91	93	98
অ	100	68	96	99	99
আ	100	78	73	91	99
ক	100	24	65	83	99
গ	100	33	77	88	95
ঙ	100	67	38	91	92
য	100	28	42	75	90
স	100	33	79	77	95
ঝ	100	50	52	85	97
ঢ	100	57	46	92	92

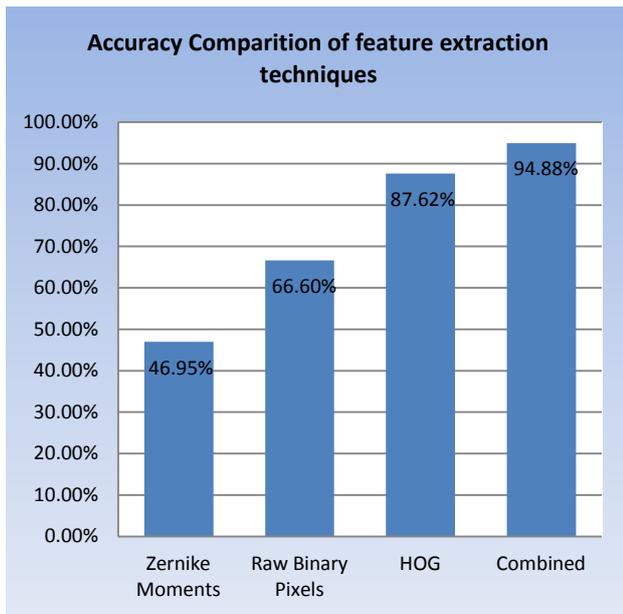


Fig.8 Accuracy rates of different feature extraction techniques

Table III: No. of correctly identified characters when using different feature extraction techniques.

	Total Testing Data	Correctly Identified Data	Accuracy
Zernike Moments	6000	2819	46.98%
Raw Binary Pixels	6000	3996	66.60%
Histogram of Oriented Gradients	6000	5257	87.62%
Combined	6000	5693	94.88%

Here Figure 8 shows a comparative accuracy rate for different feature extraction techniques based on correct recognition of characters. A detailed Table III is shown below to acknowledge the number of correctly identified data. Proposed model is trained and test with “Bangla Lekha Isolated” dataset which consists of 30000 characters where 24,000 for training dataset and 6,000 for testing .This system shows 46.98% for Zernike Moments, 66.60% for Raw Binary Pixels and 87.62% for Histogram of Oriented Gradients where overall combined features achieve an accuracy of 94.88% in recognizing characters which achieves the best accuracy rate reported till date for this dataset.

6. CONCLUSION

In this correspondence main focus is emphasized on a feature extraction technique which combines three individual feature extraction techniques i.e. Zernike Moments, Histogram of Oriented Gradients and Raw binary pixels. Using this combined technique it has been shown that it can be more reliable and it correctly recognize characters to 94.88% accuracy. Proposed model is trained and test with “Bangla Lekha Isolated” dataset which consists of 30000 characters where 24,000 for training dataset and 6,000 for testing .This system shows 46.98% for Zernike Moments, 66.60% for Raw Binary Pixels and 87.62% for Histogram of Oriented

Gradients where overall combined features achieve an accuracy of 94.88% in recognizing characters which achieves the best accuracy rate reported till date for this dataset. However, this accuracy depends largely on skew correction, line segmentation, word segmentation, removal of matra. As these are performed more efficiently the recognition rate will improve as well. In the near future these issues will be taken care of and accuracy will be increased.

7. ACKNOWLEDGMENTS

The authors are grateful to the Department of Computer Science and Engineering (CSE) of Rajshahi University of Engineering & Technology (RUET) for providing their all out support during this work. Authors are also grateful to the researcher for providing the datasets.

8. REFERENCES

- [1] G. Kumar, P. K. Bhatia, “A Detailed Review of Feature Extraction in Image Processing System,” International Conference on Advanced Computing & Communication Technologies, 2014.
- [2] Burges, Christopher JC. “A tutorial on support vector machines for pattern recognition.” *Data mining and knowledge discovery* 2.2 (1998): 121-167.
- [3] Mohammed, Nabeel; Momen, Sifat; Abedin, Anowarul; Biswas, Mithun; Islam, Rafiqul; Shom, Gautam; Shopon, Md. (2017), “BanglaLekha-Isolated”, Mendeley Data, v2.
- [4] N. Das , B. Das, R. Sarkar, S. Basu, M. Kundu, M. Nasipuri, “Handwritten BanglaBasic and Compound character recognition using MLP and SVM classifier”, *JOURNAL OF COMPUTING, VOLUME 2, ISSUE 2, FEBRUARY 2010, ISSN 2151-9617*
- [5] A.F.R. Rahman, R. Rahman, M.C. Fairhurst, “Recognition of Handwritten Bengali Characters: a Novel Multistage Approach,” *Pattern Recognition, vol. 35, p.p. 997 1006, 2002.*
- [6] T. K. Bhowmik, U.Bhattacharya and S. K. Parui, “Recognition of Bangla Handwritten Characters Using an MLP Classifier Based on Stroke Features,” in *Proc. ICONIP, Kolkata, India, p.p. 814-819, 2004.*
- [7] S.Basu, N.Das, R.Sarkar, M.Kundu, M.Nasipuri, D.K.Basu, “Handwritten Bangla Alphabet Recognition using an MLP Based Classifier,” in *Proc. Of the 2nd National Conf. on Computer Processing of Bangla, pp. 285-291, Feb-2005, Dhaka*
- [8] S. Bhowmik, M.G. Roushan, R. Sarkar, M. Nasipuri, S. Polley, S. Malakar, “Handwritten Bangla Word Recognition using HOG Descriptor” , *Fourth International Conference of Emerging Applications of Information Technology, 2014.*
- [9] G. Nagy, “At the frontiers of OCR”, *IEEE 80, pp. 1093–1100 (1992).*
- [10] X. Tong, A. D. Evans, “A statistical approach to automatic OCR error correction in context”, In: *Proceedings of WVLC, pp. 88–10 (1996).*
- [11] E. Francesconi, M. Gori, S. Marinai, G. Soda, “A serial combination of connectionist-based classifiers for OCR.” *International Journal on document analysis and recognition, 3(3), pp. 160-168 (2001).*

- [12] H. Byun, W. S. Lee, "Applications of support vector machines for pattern recognition: a survey". In: Proceedings of 1st International Workshop on Pattern Recognition with Support Vector Machines, pp. 213–236 (2002).
- [13] H. Hse and A. Richard Newton, "Sketched Symbol Recognition using Zernike Moments" Department of Electrical Engineering and Computer Sciences University of California at Berkeley, CA 94720, U.S.A.
- [14] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms." *IEEE transactions on systems, man, and cybernetics* 9.1 (1979): 62-66
- [15] Rownak, Ahnaf Farhan, et al. "An efficient way for segmentation of Bangla characters in printed document using curved scanning." *Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on*. IEEE, 2016.
- [16] Dalal, Navneet, B. Triggs. "Histograms of oriented gradients for human detection." *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE, 2005
- [17] Khotanzad, Alireza, and Y. H. Hong. "Invariant image recognition by Zernike moments." *IEEE Transactions on pattern analysis and machine intelligence* 12.5 (1990): 489-497.
- [18] Abu-Mostafa, Yaser S., and D. Psaltis. "Image normalization by complex moments." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1 (1985): 46-55.
- [19] Hu, Ming-Kuei. "Visual pattern recognition by moment invariants." *IRE transactions on information theory* 8.2 (1962): 179-187.
- [20] F. Zernike, *Physica*, vol. 1, p. 689, 1934
- [21] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20.3 (1995): 273-297.
- [22] Das, Nibaran, et al. "Handwritten Bangla basic and compound character recognition using MLP and SVM classifier." *arXiv preprint arXiv: 1002. 4040* (2010).