

Handwritten Word Recognition Using MLP based Classifier: A Holistic Approach

Ankush Acharyya¹, Sandip Rakshit², Ram Sarkar³, Subhadip Basu³, Mita Nasipuri³

¹Department of Information Technology, Jadavpur University,
Kolkata-700032, India

²Army Institute of Management,
Kolkata-700027, India

³Department of Computer Science and Engineering, Jadavpur University,
Kolkata-700032, India

Abstract:

Holistic Word Recognition is one of the new modalities for handwritten word identification. The holistic paradigm in handwritten word recognition treats the word as a single, indivisible entity and attempts to recognize words from their overall shape, as opposed to recognize the individual characters comprising the word. In the present work reports a longest-run based holistic feature, that has been used to classify word images belonging to different classes, using a neural network based classifier. To evaluate the technique, a few words from the handwritten documents of the *CMATERdb1.2.1* dataset have been used. Frequently occurring English words are manually extracted from the handwritten pages and the accuracy of the technique is evaluated using a three fold cross-validation method. The best-case and average-case performances of the technique to the said data set are 89.9% and 83.24% respectively.

Keywords: *Holistic word recognition, Longest-run features, MLP classifier, Handwritten Document*

1. Introduction

Identification of handwritten words is an important step in the process of Optical Character Recognition (OCR) of handwritten document images, which is a vital perspective for various document retrieval purposes. Handwriting word recognition (HWR) technologies are devised for either online or offline purposes.

In Online HWR the trajectories of pen tip movements are recorded and analyzed to identify intended information. With the latest technological advancements it is very common to write on an ordinary paper and immediate wireless transmission of handwritten annotations to a remote server is possible[1][2]. In Online HWR the writing is done with a special pen on an electronic notepad or a tablet and where temporal information, such as the position and velocity of the pen along its trajectory, is available to the recognition algorithm. Since most algorithms for Online HWR attempt to recognize the

writing as it is being written, sometimes it is also referred to as “real-time” HWR.

On the other hand, off-line HWR deals with the recognition of handwritten words after it was written[3]. Moreover, there is little or no control in most offline scenarios of the type of medium and instrument used. The artifacts of the complex interactions between medium, instrument, and subsequent operations such as scanning and binarizations present additional challenges to algorithms for offline HWR. It is, therefore, generally regarded as much more difficult compared to its online counterpart[3], [4]. While for some applications of online HWR, a single writer’s assumption can be made and the algorithms tuned to a particular style of writing, but this assumption cannot generally be made for the offline problem[4]. Here the recognition algorithm must deal with the variety of writer’s own idiosyncrasies. The handwritten word or phrase may be constrained by the application to be in a particular style. For example, forms often request that the responses are hand printed. In general, however, handwritten words may be cursive, purely discrete, touching discrete, or a mixture of these styles (see Fig. 1).

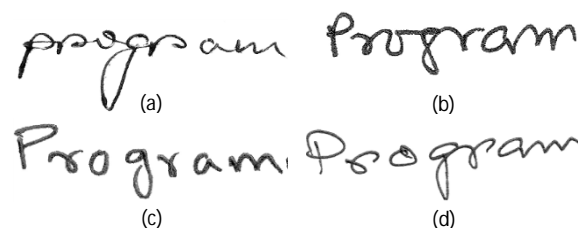


Fig. 1: Examples of (a) cursive (b) touching discrete (c) purely discrete (d) mixture of cursive and discrete words.

A handwritten word is typically scanned in from a paper document and made available in the form of a binary or grayscale image to the recognition algorithm. The handwritten text must be located, extracted, made free from the artifacts like underlines and background from the check leaf, boxes in forms, postal marks on the piece of mail etc.. Text line identification is also a requirement. Then, from the text lines, words are to be extracted they are to be used for the recognition purpose. These extracted words are used only for the process of problem solving.

The Present work focuses on the task of offline HWR. It deals with the identification of the holistic features from handwritten word . However, the discussion is pertinent to the online problem as well. There can be mainly two approaches for the word recognition purpose for any handwritten documents. 1) Analytical approach: It treats words as a collection of simpler subunits such as characters and proceeds by segmenting the word into those subunits, and then identifies the units. As an example to identify any word using this approach the letters are to be identified first then they are used for the total word recognition. 2) Holistic approach: It treats the word as a single, indivisible entity and attempts to recognize it using features of the word as a whole. So, in this approach the whole word image features are used for the recognition purposes.

In case of holistic approach, as said earlier the words are not divided into subunits, rather is concentrate on the feature values of the total images and match for those patterns[6]. Here, different words are treated as different classes. Then some pattern matching technique is used to recognize them. Given that words of large variety of writing styles holistic approaches have been used in application scenarios where the classes are few. For example, in case of postal automation system only the city names are considered[7].

Researchers have utilized many different approaches for both the segmentation [8]and recognition tasks of word recognition. Some researchers have used conventional, heuristic techniques for both character segmentation and recognition [9],[10], some have used a convex hull based [11]and recursive contour following algorithms[12], while others have used heuristic techniques for segmentation followed by ANN based methods for the character/word recognition process[13–15].Hidden Markov Model based techniques are also used widely for both offline and online hand-written document recognition[1],[2],[16].

For printed and cursive handwriting[17]some of the most successful results have been obtained with the use of techniques that possess tightly coupled segmentation and recognition components, which also uses run length

encoding[14],[18],[19].These techniques are lexicon directed and compute the best segmented for character images and sub-images (primitives) of a word so that they can be assembled and matched to represent a possible string in a lexicon. These techniques do not employ complex segmentation algorithms. As a result, the number of segmentation points found can be relatively high. By employing more powerful segmentation techniques, it is possible to reduce the number of false segmentation points found[10].This in turn can increase the speed and efficiency of the system by reducing the number of primitive combinations that need to be assembled and checked against the lexicon of words. Recently, some researchers have turned to ANNs to assist in the segmentation process[20],[15].

1.1 Motivation:

Though so many research papers are available in the literature, still handwritten word recognition is an open problem, especially in the domain of feature design and classification methodology. As discussed before, researchers have attempted to address this problem from different research perspectives. Most of such attempts are not comparable due to, 1) non availability of standard data sets and, 2) variations in research objectives.

The work presented in this paper uses a public domain handwritten document database[21]and proposes a longest-run based [22]multi-level feature descriptor for holistic recognition of isolated handwritten word images. An MLP based classifier is designed to assess the prediction of accuracy. Please note that we have worked on a multi-script database and automatic word extraction[6],[23] and script separation tasks [17][24][25][26]are not explicitly explored in the current work.

2. Methodology

In the present work, a method is developed to identify the handwritten words from a dataset namely *CMATERdb1.2.1* of handwritten pages. The document pages are written in Bangla script mixed with English words. We have used our technique in English words. So firstly, we have checked the frequency of the English words and then we have selected twelve words which have occurred in the database mostly. Sample word images from the database set are shown in Fig. 2.

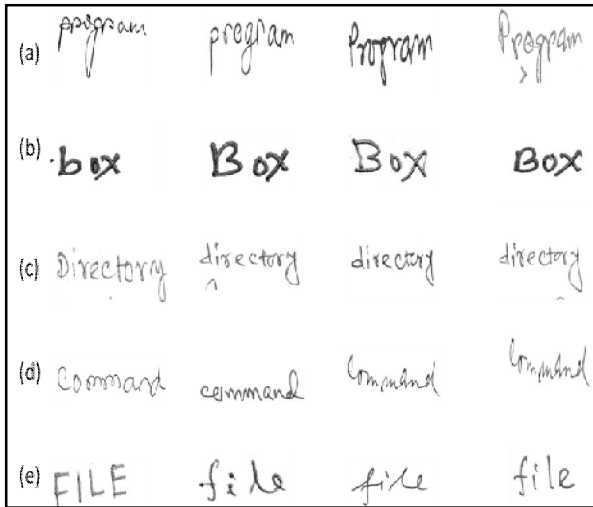


Fig. 2: Sample of few high frequent words taken from the dataset CMATERdb1.2.1 (4 such different forms of 5 words are shown)

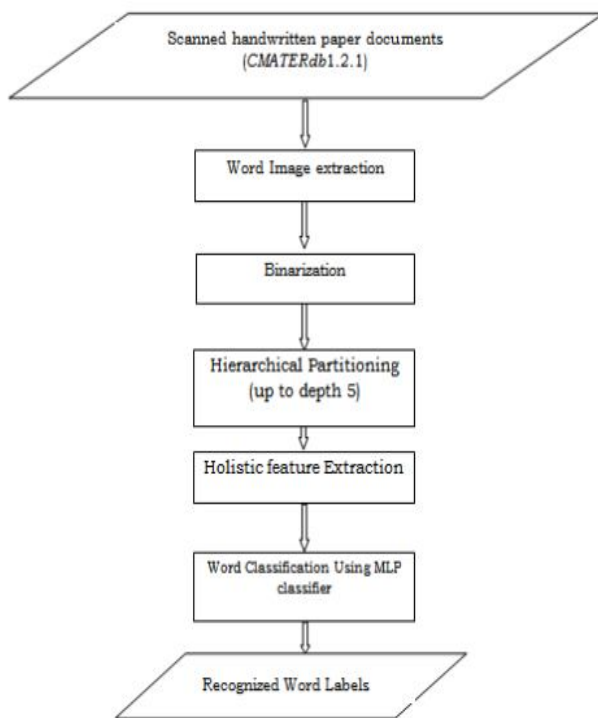


Fig. 3: A schematic flow chart of the proposed work

The word images are binarized using a simple adaptive binarization technique, where the threshold is considered as the average of the maximum and minimum intensity of the image. Then, holistic features are computed from the whole word image. Next the word images are hierarchically partitioned into vertical segments for

extraction of the local features. This hierarchical partitioning process continues up to depth-4. Based on these feature values, computed at different depths, an MLP classifier is trained to recognize the word images. The overall process is explained in the flow chart (as shown in Fig. 3).

2.1. Design of Feature Set

The grayscale word images, extracted from the database, are first converted into their respective binarized images as mentioned earlier using an adaptive thresholding technique. The adaptive threshold value is considered as the average of the maximum and minimum gray scale values of the respective images.

Design of a feature set has been always a challenging issue in the pattern recognition domain. The challenge gets incremented if the documents are handwritten. To deal with this problem, in the present work, longest run features [8] are used. These features are computed in four directions; row wise (east), column wise (north) and along the directions of two major diagonals (northeast and northwest). The row wise longest run feature is computed by considering the sum of the lengths of the longest bars that fit consecutive black pixels along each of all the rows of the region. They are calculated in a forward and backward direction of the longest runs [6],[11],[5]. In fitting a bar with a number of consecutive black pixels within a rectangular region, the bar may extend beyond the boundary of the region if the chain of black pixels is contained there.

The three other long-run features within the rectangle are computed in the same way. Each of the longest run feature values is to be normalized by dividing it with the product of the height and the width of the image. These feature values are effectively computed as the sum of running length values along the four matrices and they are divided by the total row count for that particular image for getting the feature values.

2.2 Hierarchical partitioning of the word images

To get the more discriminating information of a particular word image, a hypothetical line is considered vertically along the center of gravity (CG) of the word image, generating two sub-images or segments. Each of these segments is again partitioned in the similar way. This partitioning procedure is continued up to depth 5 (see Fig. 4 for illustration). In each level of the partitioning procedure longest run features are calculated in a similar way. More specifically, in depth-0 we compute 4 features from the original image. In depth-1, we compute 8 features from two segments. Likewise, up to depth-5, we compute

252 (=4+8+16+32+64+128) feature values are computed for classification of the particular image.

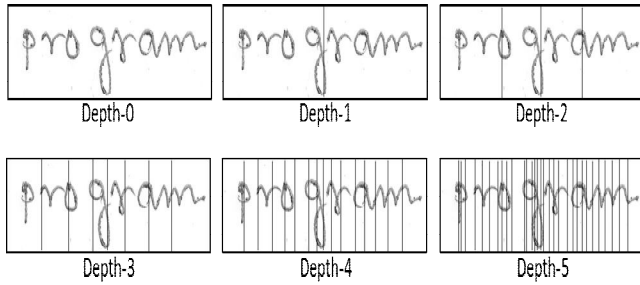


Fig. 4. A sample word image is hierarchically partitioned upto depth-5

2.3. Design of the Classifier:

The features, computed hierarchically on each word image are fed to an MLP based classifier, for identification of an image in one of the 12 categories of word images. MLP networks with back-propagation training algorithm are widely - accepted in pattern recognition applications.

The sigmoid function is used for the training of the neural network follows a parameter value [0, 1]. The total data set is divided in 2:1 ratio for training and testing purpose and the method is validated using a 3-fold cross validation technique.

3. Experimental Results and Discussion

To test and evaluate the present work the dataset *CMATERdb1.2.1* is used. The database consists of 50 handwritten document images written in Bangla script mixed with English words. We have extracted the English words from the document pages and then total 12 word classes consisting of 291 highly frequent image sets are used for the learning purpose. Fig. 5 shows the sample images of each category of word classes.

The selected words with are of varying frequency with a minimum frequency of 7 to the maximum frequency of 68. They are labeled as class levels 0 to 11. The cropped images (words) are segmented along the CG values up to depth-5, having 4 feature values in each segment resulting in total 252 feature values. The word classes are partitioned into 3 mutually exclusive sets. Then a 3-way cross validation technique is applied over the database. This way 191 data has been used for training in each fold and 100 word images are used for testing purpose.

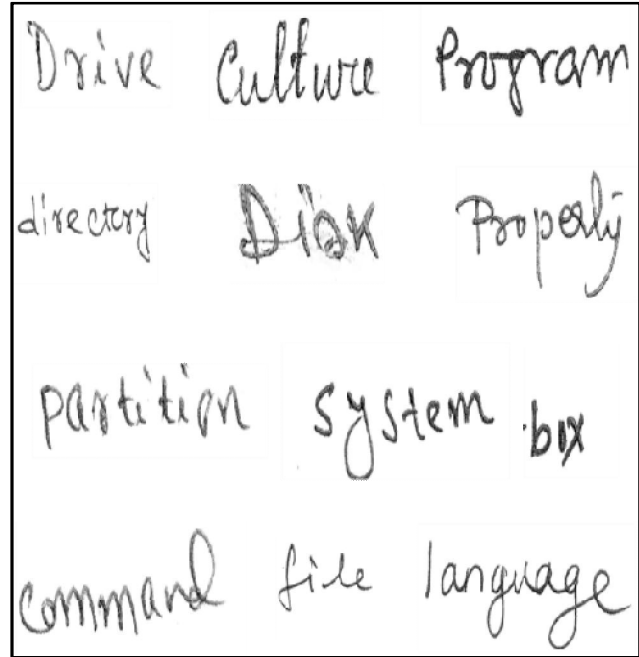


Fig. 5. Sample images of the 12 categories of word images, considered under the current work.

During training the neural network with a learning rate of 0.8, acceleration factor is 0.8 and error tolerance is 0.000001, number of iterations 25000 is used for the experiment. To decide over the network topology, the number of hidden neurons (in the single hidden layer) is varied from 120-130 in steps of 5. The MLP with the topology 252-120-12 gives the maximum success rate of 76%, 83.81% and 89.90% respectively for the 1st, 2nd and 3rd fold. The details experimental results using the 3-fold cross validation are illustrated in Table 1. The success rate of the present technique is determined by the following formula:

$$\text{Success rate} = \frac{\text{Number of successfully classified words for each class}}{\text{Total number of words in that particular class}} \times 100$$

The experiment has successfully identified the word images (as shown in Fig. 6(a-c)) into respective classes in most cases. In some cases, however the technique has failed to identify the word images (see Fig. 7(a-c)), as the words appear as a mixture of cursive and touching handwriting and the feature values mismatch due to the presence of extra features with those used for the training purposes. In Fig. 7(a), the letters are overlapping and as a result the feature values extracted get distorted. On the other hand Fig. 7(b) suffers from loss of data due to bad word extraction and in case of Fig. 7(c) the character spacing are too large, in comparison to the training set.

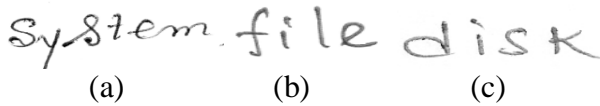


Fig. 6: Sample word images successfully identified by the present technique

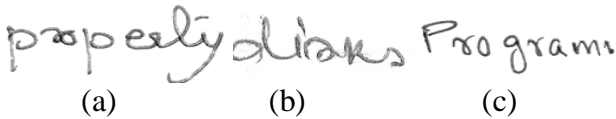


Fig. 7: Sample word images misclassified by the present technique

4. Conclusion

In this paper, we have developed a holistic word recognition technique from handwritten documents using a MLP classifier. The overall objective of the work is to recognize specific keywords from an unconstrained handwritten document. One important presumption for this work is the presence of isolated word images. As discussed before, word extraction from handwritten documents is an important research area and not explicitly considered within the scope of the present work. Therefore, we have generated the training and test samples manually from a benchmark document image database. We can say that the developed technique is user independent, as we have considered a wide intra-class variability. Most of the images contain significant noise and have a variable stroke - width. In many cases, words of a class have mixed-case writing, i.e., some images are written in lower-case, and some are in upper-case. This situation is evident in Fig. 2. Despite these challenges, the developed technique successfully recognizes the 12-class word images with an average of 83% cases. In the second fold of the cross-validation experiment, the best case performance is close to 90%. Whereas, in case of the first fold the performance drops to around 75%. One possible reason for this variability may be linked to the limitation of training samples, and also to the wide intra-class variability in some patterns. In future we may attempt to add more number of writing samples in each class and by possible consideration of intra-class clustering technique to group dissimilar writing patterns in each class of word images.

Table 1: Results observed over the 3 fold datasets

Number of output class	Number of features	Number of iterations	Folds	Number of Hidden neurons	Success Rate (%)
12	252	25000	1	120	75.00
				125	76.00
				130	76.00
			2	120	89.90
				125	89.90
				130	88.89
			3	120	82.76
				125	83.81
				130	83.81
Average success rate (%):				83.24	

This work is done for the offline handwritten documents which may be extended for the online documents as well. In this work, we have tested the technique on English words vocabulary only. But the developed technique is script independent; it can be applied to classify word images of different scripts. In a nutshell, the work presented here outlines possibilities to recognize unconstrained isolated word images of handwritten documents with reasonable accuracy, highlighting the scopes for further application in related fields of research.

Acknowledgement

Authors are thankful to the Center for Microprocessor Application for Training Education and Research (CMATER), Department of Computer Science and Engineering, Jadavpur University for providing the necessary research documents.

References

- [1] Y. Bengio, Y. LeCun, C. Nohl, and C. Burges, "Lerec: A NN/HMM hybrid for on-line handwriting recognition," *Neural Computation*, vol. 7, no. 6, pp. 1289–1303, 1995.
- [2] S. Bercu and G. Lorette, "On-line handwritten word recognition: an approach based on hidden Markov models," in *Proceeding Third Int. Workshop on Frontiers in Handwriting Recognition*, 1993, pp. 385–390.
- [3] A. Vinciarelli, "A survey on off-line cursive word recognition," *Pattern recognition*, vol. 35, no. 7, pp. 1433–1446, 2002.
- [4] R. Plamondon and S. N. Srihari, "Online and off-line handwriting recognition: a comprehensive survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 1, pp. 63–84, 2000.
- [5] S. Madhvanath and V. Govindaraju, "The role of holistic paradigms in handwritten word recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 2, pp. 149–164, 2001.

- [6] A. Khandelwal, P. Choudhury, R. Sarkar, S. Basu, M. Nasipuri, and N. Das, "Text line segmentation for unconstrained handwritten document images using neighborhood connected component analysis," *Pattern Recognition and Machine Intelligence*, pp. 369–374, 2009.
- [7] Y. Wen, Y. Lu, and P. Shi, "Handwritten Bangla numeral recognition system and its application to postal automation," *Pattern recognition*, vol. 40, no. 1, pp. 99–107, 2007.
- [8] H. Lee and B. Verma, "Binary segmentation algorithm for English cursive handwriting recognition," *Pattern Recognition*, 2011.
- [9] R. M. Bozinovic and S. N. Srihari, "Off-line cursive script word recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 1, pp. 68–83, 1989.
- [10] M. Blumenstein and B. Verma, "A new segmentation algorithm for handwritten word recognition," in *Neural Networks, 1999. IJCNN'99. International Joint Conference on*, 1999, vol. 4, pp. 2893–2898.
- [11] N. Das, S. Pramanik, R. Sarkar, S. Basu, and P. K. Saha, "Recognition of Isolated Multi-Oriented Handwritten/Printed Characters using a Novel Convex-Hull Based Alignment Technique," *International Journal of Computer Applications IJCA*, vol. 1, no. 23, pp. 40–45, 2010.
- [12] A. Bishnu and B. B. Chaudhuri, "Segmentation of Bangla handwritten text into characters by recursive contour following," in *Document Analysis and Recognition, 1999. ICDAR'99. Proceedings of the Fifth International Conference on*, 1999, pp. 402–405.
- [13] H. Yu and B. M. Wilamowski, "Efficient and reliable training of neural networks," in *Human System Interactions, 2009. HSI'09. 2nd Conference on*, 2009, pp. 109–115.
- [14] D. You and G. Kim, "An approach for locating segmentation points of handwritten digit strings using a neural network," in *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*, 2003, pp. 142–146.
- [15] A. Bivens, C. Palagiri, R. Smith, B. Szymanski, and M. Embrechts, "Network-based intrusion detection using neural networks," *Intelligent Engineering Systems through Artificial Neural Networks*, vol. 12, no. 1, pp. 579–584, 2002.
- [16] M. Liwicki and H. Bunke, "HMM-based on-line recognition of handwritten whiteboard notes," in *Tenth International Workshop on Frontiers in Handwriting Recognition*, 2006.
- [17] H. Bunke, "Recognition of cursive Roman handwriting: past, present and future," in *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*, 2003, pp. 448–459.
- [18] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern recognition*, vol. 33, no. 1, pp. 25–41, 2000.
- [19] E. Kussul and T. Baidyk, "Permutative coding technique for handwritten digit recognition system," in *Neural Networks, 2003. Proceedings of the International Joint Conference on*, 2003, vol. 3, pp. 2163–2168.
- [20] A. C. Tsoi, M. Hagenbuchner, and A. Micheli, "Building MLP networks by construction," in *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, 2000, vol. 4, pp. 549–554.
- [21] R. Sarkar, N. Das, S. Basu, M. Kundu, M. Nasipuri, and D. KUMAR BASU, "CMATERdb1: a database of unconstrained handwritten Bangla and Bangia—English mixed script document image," *International journal on document analysis and recognition*, vol. 15, no. 1, pp. 71–83, 2012.
- [22] S. Basu, N. Das, R. Sarkar, M. Kundu, M. Nasipuri, and D. K. Basu, "A hierarchical approach to recognition of handwritten Bangla characters," *Pattern Recognition*, vol. 42, no. 7, pp. 1467–1484, 2009.
- [23] S. Rakshit, S. S. Das, K. S. Sengupta, and S. Basu, "Automatic Processing of Structured Handwritten Documents: An Application for Indian Railway Reservation System," *International Journal of Computer Applications IJCA*, vol. 6, no. 11, pp. 26–30, Sep. 2010.
- [24] S. Fernández, A. Graves, and J. Schmidhuber, "Sequence labelling in structured domains with hierarchical recurrent neural networks," in *Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI, 2007*, vol. 20, p. 2007.
- [25] V. Frinken, A. Fischer, R. Manmatha, and H. Bunke, "A novel word spotting method based on recurrent neural networks," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 2, pp. 211–224, 2012.
- [26] E. Indermuhle, V. Frinken, A. Fischer, and H. Bunke, "Keyword Spotting in Online Handwritten Documents Containing Text and Non-text Using BLSTM Neural Networks," in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, 2011, pp. 73–77.

Ankush Acharyya is an M.Tech student of Jadavpur University

Sandip Rakshit has been an Assistant professor and computer coordinator at the ARMY Institute of Management (Ranked A++ Category B-School in the country) , Kolkata, India Mr. Rakshit is Pursuing Ph.D. in Computer Science Engineering from Jadavpur University He is also a member of IEEE (USA) and ACM (USA) for last five years.

Ram Sarkar received his B. Tech degree in Computer Science and Engineering from University of Calcutta, in 2003. He received his M.C.S.E and Ph. D. Degrees from Jadavpur University, in 2005 and 2012 respectively . He joined J.U. As a lecturer in 2008. His areas of current research interest are document image processing, line extraction and segmentation of handwritten text images. He is also a member of IEEE (USA)

Subhadip Basu received his B.E. Degree in Computer Science and Engineering from Kuvempu University, Karnataka, India, in 1999. He received his Ph.D. (Engg.) degree thereafter from Jadavpur University (J.U.) in 2006. He joined J.U. As a senior lecturer in 2006. His areas of current research interest are OCR of handwritten text, gesture recognition, real-time image processing. He is also a senior member of the IEEE, USA

Mita Nasipuri received her B.E.Tel.E., M.E.Tel.E., and Ph.D. (Engg.) degrees from Jadavpur University, in 1979, 1981 and 1990, respectively. Prof. Nasipuri has been a faculty member of J.U since 1987. Her current research interest includes image processing, pattern recognition, and multimedia systems. She is a senior member of the IEEE, USA, Fellow of I.E (India) and W.A.S.T., Kolkata, India.