Classification of Devnagari Numerals using Multiple Classifier

Shashi Kant Shukla[1], Akhilesh Pandey[2]
1 M.Tech Scholar, Information Communication, Suresh Gyan Vihar University, Rajasthan, India
2 Assistant Professor, Software Engineering, Suresh Gyan Vihar University, Rajasthan, India

Abstract: This paper presents a multiple classifier scheme for off-line hand written Devnagri numbers classification. The main purpose of this research is to find out best recognition result using multiple classifiers. This proposed technique uses simple profile and contour base triangular area representation technique for finding feature extraction and multiple classifier schemes on KNN, LDA, and KNN new neural network for classification. The performance of this technique has been tested with 36000 handwritten numerals randomly selected from CPAR datasets out of which 22000 datasets has been used for training sets and 14000 datasets has been used for test sets and we found the different result by different classifier.

Keywords: Multiple classifier schemes, Multiple Feature extraction.

1. INTRODUCTION

Recognition of Devnagari handwritten Numbers one of the biggest problem in present scenario. Devnagari numbers are not recognized efficiently and truthfully by electronic device. Many researchers and algorithm have been proposed for recognizing of numbers. For recognizing of numbers, many processes have to be performed but no single technique or algorithm can perform that recognition and give more accurate result. In the today scenario, it is more important recognizing of Devnagari numbers because in India Hindi is mother language. This system helps human being to solve their more complex problem in very easy way. In active area of research hand written numbers is a problem of the recognition. Because it is very important requirement of office automation. It is provide to effective and practical recognization of numbers. At the time of writing of a person depends on their moods and writing styles its does not lend .I helping of recognition process and all structure ,statistcaland topological information about the number in all the sort has been observed in the recognition process. In the Hand printed Hindi numbers limited variation and shape and size are consider and main attentive focus on the recognition. 35 years passed away, that researchers had been working on Hand written recognition .From last years, numbers of compares those were there in the research on the hand written recognition are gaining continually.

Nowadays, public has become attentive towards the handwritten recognition technology. aim of the creating handwritten recognition system with the rating of 100% is still not achieved as humans beings is not possible to recognized every test of any writer without any confusion most of the people cannot read their own hand writing in the very effectively . It is the responsibility of the writer to write the text in the readable format.

Optical character recognition is the miniatures form of optical number recognition. This method helps a machine to mechanically indicate numbers complete an optical mechanism. In this manner human beings identify many objects .Eyes are the aca optical machine where as the brain looks the input activities the suitability to understand there signals opposed in each person's confers to many factors .The state that face by the technologist of advancing OCR system has become easier to recognize to their variables . Firstly if a person is said to read a page of unfamiliar language he or she would be not able to identify the words but in that page only if numerical statements are gives , the person will easily will easily explain because the figures that are present are used all over the world . This make clear that OCR systems acknowledge numbers only , where as in relation some know about the whole alphanumeric number range . Secondly, numerical and alphabetical signs size has equality between them.

Thirdly, difference assistance in identifying numbers. It become hand to study the printed work which looks pitch dark backdrop either is mentioned above the words or graphics. It once more ,scheduling a system to clarify single applicable data and neglect the remaining becomes a difficult task for engineers .The documents which are present in the form of paper can be read by a human being easily but it is impossible for a computer which is a machine to learn this report straight. OCR system is flourishing to change these report into computer process able shape. Reshaping searched pictures of machine. Producing again on handwritten work, numbers alphabet and sign in a computer process able form like ASCII is the process of OCR. OCR is inspired a lot by a wish to polish man and machine broadcasting .Some products in the present time are present for identifying numbers. Items which are used to act handwritten are not on scale, even many advance have been put on .Actually, not long ago big amount of interest in putting not real neural network architecture to get rid of this trouble. Either pattern matching on statistical approaches for feature extraction was made in neural network research through the last time One of the aim of the profession a not real intelligence and machine studying it to make computers to full fill the tasks those are ordinary to humans, in agreement with the lengthy aims of survey and rival human intelligence or possibility be unexpectedly awareness. Huge development has taken place in matter both active read rates and accuracy. Keystroke data entry is much not able that data Entry via OCR. Type writer data entered in a computer wages 2450 words per minutes and can be easily analysed by desktop OCR scanners.Uniform machine print character size, shape, and any font spacing. In contrast handwriting is uneven; they can by different authors, and at different times, even by the same writer to write in many different styles

and sizes [3]. Many algorithms or methodology for hand written number recognition and each of these have their own merits and demerits[1].

Some work has been done language Devnagri, the third popular in the world. The most important is the choice of different styles and shapes of a good handwriting recognition feature vector solution. For feature extraction, we use a representative Triangle region to achieve speeds and in different digital image recognition acceptable accuracy proposed a simple outline and contour.

Printed Devanagri character recognition is attempted using Kohenen neural network and other types of neural networks in 2001[11]. Sethi and Chatterjee have described Devanagri numerals recognition using the structural approach 1976[4]. The back-propagation neural network is used in [11] for the recognition of handwritten characters. In that feature extraction is done using three different approaches, namely, ring, sector and hybrid.

Figure 1: CPAR Database image samples of Devnagari Numbers

Hanmandlu and Murthy [6] proposed a fuzzy model based recognition of handwritten Devnagri numerals and they obtained 92.67% accuracy. Bajaj et al [7] employed three different kinds of features namely, density features, moment features and descriptive component features for classification of Devnagari Numerals. They proposed multi-classifier connectionist architecture for increasing the recognition reliability and they obtained 89.6% accuracy. Bhattacharaya et al [8] proposed a multi-layer preceptron, neural network based classification approach for the recognition of Devenagari handwritten numerals and obtained 91.28% recognition accuracy. C. Vasantha Laxmi, Ritu Jain, C. Patvardhan [9] proposed a method incorporates in novel way ideas regarding edge directions histogram and splines along with PCA for enabling recognition that give 94.25% accurate results.

In this proposed work, we convert the image data in the database CPAR 48 × 48 resolution, such as a different form of the same resolution level to render a large number of different physical structures in the image of the number. This feature extraction scheme seems to be very effective in handwriting recognition. The decision has been naturalized in several categories combine usually leads to better classification accuracy because of the different categories represent the input data, we believe that the KNN, KNN new and LDA classifier training pattern classification different aspects of the purpose. The technologies have been tested in the CPAR database and get the best recognition accuracy.

In the demesne of number recognition, substantial improvement in recognition performance has been reported in a number of occasions [14,15] by considering the combination method. There are diversity of method [16,17] implementing the combination of classifiers method. For this strategy we use the multiple classifier schemes [27].

2. PROCESS OVERVIEW

Here the main allocated problem to the recognize devnagari numbers and it is main ability of the human being to recognize the handwritten numbers with do little effort. For the recognizing of Devnagari numbers the following objective to produce of main system:-

• As for the each and every input pattern all the input belong in a specific class for reorganization.

•.For develops the database algorithm to recognize Handwritten Devnagari numbers.

• To Use the classifier LDA, KNNand KNN New find the best recognition result.

The number recognition system is usually validated by running them on test on data set, on which the system has been trained. For these tests to be conclusive, the validation sets should include a fairly large number of samples to reflect the variety of writing styles that are found in real-life applications. In this work we have followed the following steps.

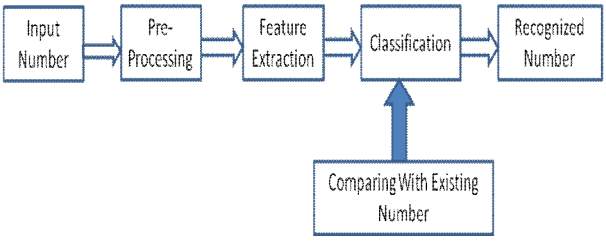


Figure 2: Block Diagram of HDCR

a) Collection of Data From different writer
b) Binarization of Image
c) Noise Reduction of Image
d) Feature Extraction of scanned Image
e) Classifying
f) Multiple classifier Scheme
g) Recognition

2.1 Database Collection

For verification purposes, we need a standard database. For handwritten Devnagari images, we have use CPAR digital database. It contains a number 36000 samples. We collected these samples from different people in different writing style collection and also use the number of different colored pen. The database also contains some samples cannot even be recognized by humans. The database is divided into two disjoint sets, one for training and the other for testing.

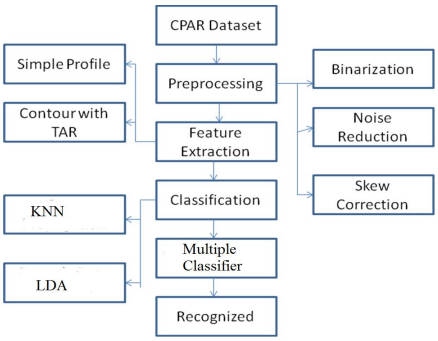


Figure 3. Block diagram of system implementation

For the training purpose we use the 22000 samples and for testing 14000 sample. If the training set contains a large number of samples with varied styles, the feature set computed from them will be able to reflect these variations in writing styles.

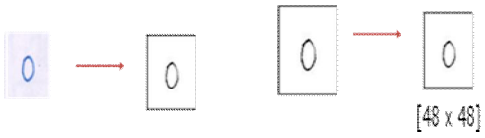
2.2 Preprocessing

a. Original Image b. Thinning Image

Figure 4. Result of pre-processing

2.2.1 Binarization

For removing of additional pixels in the digital images we cropped the images. Then, the RGB image is converted to gray scale images. When you create a database of several pre-treatment is used. Including but not limited to, the following?

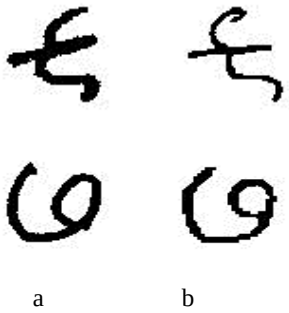


a. Binarization b. Resize of image

Figure 5. Pre-processing of the image

2.2.2 Noise Removal

The main purpose is to remove any noise removing unnecessary bit mode, this output is not in any sense. Normalized shape helps reduce the size of the extracted information data such as thinning and largely Skelton formations characters. The skeleton, we find the thin image.



a b

Figure 6. Noise Reduction a) Original Image

b) Skelton Image

2.3 Feature extraction

This step helps in based on their characteristics: the number of feature extraction classification refers to the extraction of features from each digital image, which will become its identity and help improve the recognition rate. This is a daunting task because of the nature of the handwritten changes from person to person depending on the mental state & height to get these features.

For feature extraction methods we have used

- Profile- simple profile
- Contour based Triangular Area Representation (TAR)

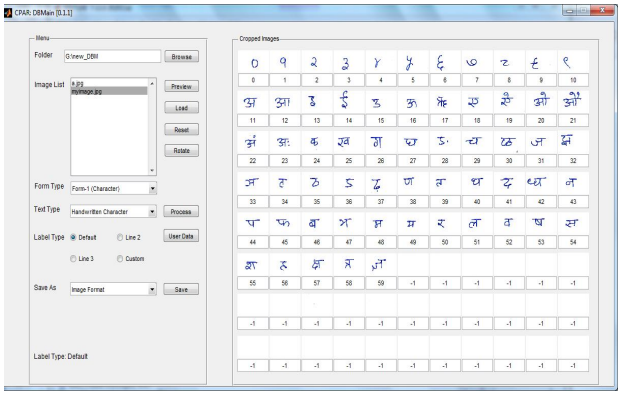


Figure 7 : CPAR Database Creation

2.3.1 Simple Profile

The profile counts the number of pixels (distance) between the bounding box of the image and the edge of the character or numeral image. Fast, less memory requirement & proved to be efficient. Profile describes the external shape of numeral and allows distinguishing between a great numbers of letters.

After training apply the classify test on the 14000 data set sample. Here we use three classifier KNN, KNN new and LDA.

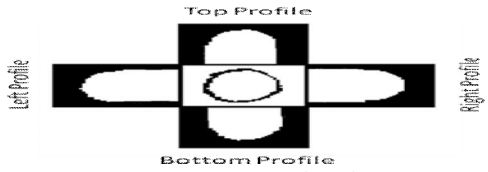


Figure 8. All Profile of number 0

2.3.2 Contour Based using TAR

2.4 Classification

Classification of human activities in the decision-making process is very important. A classification problems arose when an object needs to be allocated based on some observations about the object's attributes to a predefined group or class to. We train the neural network to identify the number of training data set samples 36000. The right to re-array network training to minimize the selected performance indicators, namely the use of LDA, KNN, New KNN algorithm.

Determine whether the training sufficiently effective strategy is to use the validation set. As more training provided to verify the recognition error decreases monotonically to a minimum value, but then began to increase, the error continues to decrease even though the training. To get better network performance, validation of the training error is minimized when the termination.

In the present work, we think before responding to network-based multi-resolution contour and contour classification purposes tar and voting scheme on a set of KNN and LDA.

The technology has been tested in several handwritten devnagri and achieved recognition accuracy and speed the most advanced technology comparable countries. The reason to get high-speed recognition rate in the proposed approach is an outline and contour feature extraction tool natural fit digital computer, its base function is defined only multiplication and addition operations - no derivatives or integration.

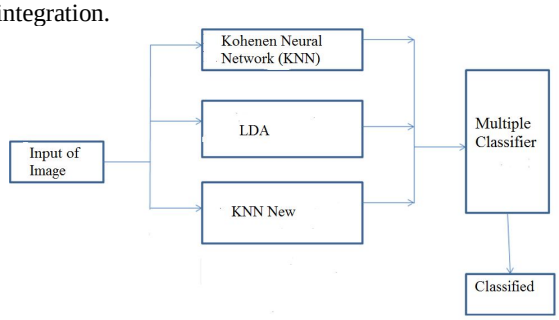


Figure 9. Block Diagram of Classification and multiple classifiers

In the present work, we think before responding to network-based multi-resolution characteristics of a simple silhouette and outline the purpose of the use of the classification scheme tar and vote on a set of LDA and KNN. The technology has been tested in several handwritten devnagri Devnagri and achieved recognition accuracy and speed comparable to state of the art technology. The reason to get high-speed recognition rate in the proposed approach is an outline and contour feature extraction tool natural fit digital computer, its base function is defined only multiplication and addition operations - no derivatives or integrals.

3 RESULTS

CPAR database includes 36,000 digital isolated handwritten Devnagri evenly distributed across all classes. This database includes age, gender, education, local, who also wrote the paper's state of mind may change in the. Therefore, there is a change of the number of people at different times of the writing style. Handwritten Sanskrit number set shooting. Follow these steps to obtain a digital image input handwritten Devnagri best accuracy from CPAR database. First, the system is trained by using different sets of data or sample to. Then the system is used for a given set of test samples, and the accuracy is measured. The data set is divided into two parts. The first part is a training system; the second is used for testing purposes. For each figure, the feature is calculated and stored for training the network. Here is a table that shows the results obtained from the program. Variance is small, but it is there.

Three input layer, network layer, twenty hidden layer and an output layer to take. If the increase in the number of neurons in the hidden layer, it is required to issue a memory allocation occurs. Further, if the error tolerance is high, the desired result cannot be obtained, the error tolerance value is changed, i.e., for example, can be obtained with high accuracy. Network error tolerance values also need to learn a few more cycles than when the margin of error is smaller than the case of high-value, which the network less learning cycles, so learning is not very fine.

	0	1	2	3	4	5	6	7	8	9	10		%
0	1087	14	0	2	3	0	2	12	2	1	1	1124	96.71
1	1	1076	5	2	0	1	0	8	1	0	28	1122	95.90
2	1	20	1012	61	1	8	1	1	6	4	7	1122	90.20
3	0	8	94	1003	0	9	0	1	1	2	5	1123	89.31
4	12	4	0	2	1067	21	5	2	6	7	2	1128	94.59
5	3	9	19	6	42	1045	3	0	0	3	0	1130	92.48
6	2	2	2	5	5	7	1030	35	3	20	12	1123	91.72
7	35	2	1	1	3	0	1	1090	0	0	0	1133	96.20
8	2	0	5	1	3	0	7	2	1100	4	5	1129	97.43
9	1	2	0	0	6	3	23	1	6	1054	2	1098	95.99
10	1	55	2	2	2	0	1	1	0	1	703	768	91.54
									Total Samples			12000	93.82

Table 1. Result of Kohenen Neural Network Classifier

	0	1	2	3	4	5	6	7	8	9	10		%
0	1087	6	1	0	0	0	2	17	4	4	1	1122	96.88
1	1	1024	2	4	1	4	0	14	0	0	20	1070	95.7
2	2	6	1026	55	2	9	7	1	11	18	4	1141	89.92
3	2	1	63	1014	6	15	6	7	3	6	3	1126	90.05
4	0	4	4	11	1042	27	10	4	4	11	9	1126	92.54
5	0	5	10	24	50	1038	7	4	0	6	2	1146	90.58
6	0	2	10	5	11	8	1005	20	13	30	20	1124	89.41
7	10	4	3	2	4	2	7	1104	2	3	0	1141	96.76
8	3	3	8	5	1	3	7	0	1095	9	4	1138	96.22
9	0	0	14	7	5	3	20	1	2	1043	3	1098	94.99
10	3	34	4	3	3	1	13	2	6	11	688	768	89.58
									Total Samples			12000	92.97

Table 2: Result of LDA Classifier

In this article, we use the TAR recognize handwritten numbers as simple contours and shape of the multi-classification scheme based on. In this approach, we consider the former back-propagation and cascade fed because in 48 × 48 resolution level of classification. Similar types using multi-classifier to improve recognition accuracy without a significant increase in the calculation of this strategy in the function.

	0	1	2	3	4	5	6	7	8	9	10		%
0	1147	15	0	2	4	0	2	10	3	0	1	1184	96.88
1	2	1138	6	1	0	1	0	10	1	0	21	1180	96.44
2	1	20	1061	61	1	8	1	1	6	4	7	1171	90.61
3	0	8	97	1032	0	9	1	1	1	1	4	1154	89.43
4	12	4	0	2	1116	22	5	3	4	4	1	1173	95.14
5	4	9	20	6	44	1065	3	0	0	2	0	1153	92.37
6	4	2	2	5	5	7	1067	33	3	10	13	1151	92.7
7	37	2	1	1	4	0	1	1144	0	0	0	1190	96.13
8	2	0	5	1	3	0	7	4	1154	1	6	1183	97.55
9	7	17	20	4	13	7	122	10	25	2137	20	2382	89.71
10	1	89	3	2	5	0	1	1	0	0	977	1079	90.55
									Total Samples			14000	93.41

Table 3: Result of Kohenen Neural Network New

4 CONCLUSION

Offline handwritten Devnagri number recognition is a difficult problem, not only because of the variation in the amount of human handwriting is large, but because of the overlap, and add a few. Recognition depends largely on the nature of the data to confirm.

REFERENCES:

1. U. Pal, B.B. Chaudhri, *"Indian Script character recognition: a survey Pattern Recognition"*, The Journal of the Pattern Recognition Society, Available at ElsevierComputerScience.com, vol. 37, pp 1887-1899, Jun. 2004.
2. J. J. Hull, S.N. Srihari, E. Cahen, C. L. Kuan, P.Cullen and P. Palumbo, *"A black-board approach to handwritten ZIP code recognition"* in Proc. United States Postal Service Advance Technology Conf. 1988.
3. R. Plamondon and S. N. Srihari, *"Online and Offline handwritten character recognition: A comprehensive survey"*, IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 22, pp 62-84, 2000.
4. I. K. Sethi and B. Chatterjee, *"Machine Recgnition of constrained Hand Printed Devengari"*, Pattern Recognition, vol 9, pp 69-75, 1977.
5. I. K. Sethi and B. Chatterjee, *"Machine Recgnition of handprinted Devengari numerals"*, J. Inst. Electron. Telecommun. Eng. 22(1976), pp 532-535.
6. M. Hanmandlu, O. V. Ramana Murthy, *"Fuzzy model based recognition of handwritten numerals"*, Pattern Recognition, vol. 40, pp 1840-1854, 2007.
7. Reena Bajaj, Lipika Dey, and S. Chaudhury, *" Devnagari numerals recognition by combining decision of multiple connectionist classifiers"*, Sadhana, col.27, part I pp-59-72, 2002.
8. U. Bhattacharya, B. B Chaudhri, R. Ghosh and M. Ghosh, *"On Recognition of Handwritten Devengari numerals"*,In proc. Of workshop on learning Algorithms for Pattern recognition, Sydeny, pp 1-7 2005.
9. R.M.K Sinha, H.N. Mahabala, *"Machine recognition of Devnagari script"*, IEEE Trans. Syst Man Cybern 9, pp. 435-441, 1979.
10. I.K.Sethi, B.Chatterjee, *"Machine recognition of handprinted Devnagari numerals"*, J Inst, Electron. Telecommun. Eng 22, pp. 532-535, 1976.
11. R. Bajaj, L. Dey, S. Chaudhury, *"Devnagari numeral recognition by combining decision of multiple connectionist classifiers"*,Sadhana 27-1, pp. 59-72, 2002.
12. V. Bansal, R.M.K. Sinha, *"A Devanagari OCR and a brief review of OCR research for Indian scripts."* Proceedings of STRANS01, IIT, Kanpur, India, 2001
13. S. Khedekar, V. Ramanaprasad, S. Setlur, V. Govindaraju, *"Text-image sepration in devanagari documents"*, Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR'03), Edinburgh, Scotland, 3-6 August, 2003, pp 1265-1269.
14. B.B. Chaudhuri, U. Pal *"An OCR system to read two Indian language scripts: Bangla and Devanagari (Devnagri")*, in: Proceedings of fourth IEEE International Conference on Document Analysis and Recognition (ICDAR'97), Ulm. Germany, August 18-20, 1997, pp. 1011-1015.
15. S. Antanani, L. Agnihotri, *"Gujarati character recognition"*, in: Proceeding of fifth IEEE International conference on document Analysis and Recognition (ICDAR'99), Banglore, India, 20-22 September, 1999, pp. 418-421.
16. M. Hanmandlu, K.R Murali Mohan, H.Kumar, *"Neural based handwritten character recognition"*, in Proceeding of fifth IEEE International Conference, pp 241-244, 1999.
17. R.M.K. Sinha, H.N. Mahabala, *"Machine recognition of Devanagari script"*, IEEE Trans. Syst. Man Cybern 9(8) (1979), pp 435-441.
18. J.C Sant, S,K, Mulick, *"Handwritten Devanagari script recognition using CTNNSE algorithm"*, International Conference on Application of Information Technology in south Asia language, AKSHARA'94, NewDelhi, India, pp. 25-26.
19. A. Elnagar, S. Harous, *"Recognition of Handwritten Devnagri Numerals using structural descriptors"*, J. Exp. Thor. Artif. Intell. 15(3), 2003, pp 299-314.
20. C.L. Liu, K. Nakashima, H. Sako, H.Fujisawa, *"Handwritten digit recognition: benchmarking of state-of-the-art techniques"*, Pattern Recognition 36(10), 2003, pp 2271-2285.
21. K.Y. Rajput and Sangeeta Mishra, *"Recognition and Editing of Devnagari Handwriting Using Neural Network"*, Proceedings of SPIT-IEEE Colloquium and International Conference, Mumbai, India, Vol. 1,66.
22. Anil k. Jain, Jianchang Mao, K.M. Mohiuddin, *"Artifical Neural Network: A Tutorial"*, IEEE 0018-9162/96, March 1996.
23. Y. Lu and M. Shridhar, *"Character Segmentation in Handwritten Words – An Overview", Pattern Recognition"*,Vol. 29, 1996, pp. 77-96.
24. R.M.K. Sinha and V. Bansal, *"On Devanagari document processing," Proc. Int. Conf. on Systems, Man and Cybernetics"* , Vancouver, BC, pp. 1621-1626,Oct. 1995.
25. R.M.K.Sinha, *"Rule based contextual post-processing for Devanagari text recognition"* , Pattern Recognition, 20(5), pp. 475-485, 1987.
26. S. N. Srihari, *"Recognition of Handwritten and Machineprinted Text for Postal Address Interpretation"*, Pattern Recognition Letters, 14, 1993, pp. 291-302.
27. U. Bhattacharya, B. B. Chaudhuri, *" A multiple classifier scheme for multiresolution recognition of handprinted numerals"*, Proceeding of seventh International Conference on Document Analysis and Recognition, 2003.

AUTHORS PROFILE:

Authors

Shashi Kant Shukla is a student of M.Tech (Dual Degree) (Computer Science + Information Communicatin) at Suresh Gyan Vihar University, Jaipur. His working on a Pattern Recognition.He as some Knowledge MATLAB and C language.

Akhilesh Pandey is Asst. Professor in Suresh Gyan Vihar University, Jaipur. Presently he is pursuing P.hd. He did his M. Tech in 2012 from Sharda University, Gr. Noida and MCA from IGNOU in 2002 and after that he worked as a faculty member in different engineering college. He guide number of student in PG and UG student. His area of Interest is Pattern Recognition and neural network.