International Journal on Advanced Science Engineering Information Technology

Vol.8 (2018) No. 3
ISSN: 2088-5334

Efficient Handwritten Digit Classification using User-defined Classification Algorithm

R. Vijaya Kumar Reddy[#], U. Ravi Babu[*]

[#] *Research Scholar, Dept. of. CSE, Acharya Nagarjuna University, Guntur, India.*
E-mail: vijayakumarr285@gmail.com

** Principal, DRK College of Engineering & Technology, Hyderabad, India.*
E-mail: uppu.ravibabu@gmail.com

Abstract— **In automatic numeral digit recognition system, feature selection is most important factor for achieving high recognition performance. To achieve this, the present paper proposed system for isolated handwritten numeral recognition using number of contours, skeleton features, Number of watersheds, and ratio between the numbers of foreground pixels in upper half part and lower half-part of the numerical digit image. Based on these features the present paper designed user-defined classification algorithm for handwritten digit recognition. To find the effectiveness of the proposed features, these features are given as an input for standard classification algorithms like k–nearest neighbor classifier, Support Vector Machines and other classification algorithms and evaluate the results. The experimental result proves that the proposed features are well suited for handwritten digit recognition for both user and standard classification algorithms. The novelty of the proposed method is size invariant.**

Keywords— **digit recognition; classifier; k–nearest neighbor; support vector machines classifier; hand-written digit.**

I. INTRODUCTION

Handwritten digit recognition plays an important role in pattern recognition and Optical Character Recognition (OCR). It has a wide range of practical applications in real life, such as zip code recognition in postal mail sorting [1], writer identification and verification, form processing, and handwritten digit recognition on bank check. Over the past decades, lots of machine learning methods have been employed for effective handwritten digit recognition , such as Linear and Non-Linear Classifier, Support Vector Machines (SVMs),Neural Networks (NNs), Boosted Stumps, CNNSVM Classifier, etc.

Despite the fact that the Handwritten Digit Recognition problem has been studied extensively for a number of years, one of its benchmark databases is still actively used by researchers till date [2, 3, 4]. The conventional applications of OCR are numerous and include diverse a reassuch as automatic bank cheque processing, immigration data processing, health data record conversion into electronic format, tax forms data conversion and many other applications. Despite being one of the earliest research and application areas of AI, the digit and character recognition remains an active research topic. The present dayS popularity and mass availability of smart phones with sophisticated camera technology enable users to capture images of handwritten notes containing digits and characters. Subsequently there is a need to extract the handwritten notes information from these images to convert them into text files. Analysis of documents and images with texts continues to be active research topics [5,6,7,8, 9]. Hence the need for developing efficient handwritten digit and character recognition algorithms and techniques are still live in today.

The 3feature extraction plays major role in numerical recognition system. Numbers of feature extraction methods are stated in the literature, like template matching, projection histogram, zoning , and various moment techniques [10] to enable for specific applications. Some methods include fuzzy features [11,12] invariant moments features [16], template and deformable Templates [13, 14] structural and statistical features [12,15] extraction.

Ravi *et.al* [16] proposed an approach for Telugu printed numeral digits recognition using number of contours, skeleton feature, water reservoir features, and ratio of length of Top line to bottom line of the image. The proposed method is applied with success to a database of 3,150 printed multi-font printed Telugu numerals. Ravi *et.al* [17] proposed another approach to offline handwritten numeral recognition based on structural and statistical features. Five skeleton features and two-geometrical features and distribution features are used for the recognition of numerals. The proposed method experimented with MNIST database. A Euclidian minimum distance criterion is used to find minimum distances and k-nearest neighbor classifier is used to classify the numerals. MNIST database is used for both training and testing the system.

Ravi *et.al* [18] proposed another approach for handwritten digit recognition system. This approach extracts digit image features based on distance measure and derives an algorithm to classify the digit images. The distance measure can be performed on the thinned image.

Kumar *et.al* [19] proposed an approach for handwritten digit recognition system. The present paper extract digit image features based on distance measure and derive an algorithm to classify the digit images. The distance measure can be performed on the Skelton image. The present approach experimented with MNIST database, CENPARMI, CEDAR and newly collected data.

Based on the critical survey of existing methods, all the existing methods are used standard classification algorithms are used for classification. The present study aimed is to classify the isolated handwritten digits using user defined classification algorithm not with standard classification algorithms. The main objective of the present study is that classify the isolated handwritten digits with a minimum number of features. To achieve this, we proposed a new approach for classification without using any standard approchas for isolated digits. It is generally based on the feature extracts such as number of contours in an image [12], Number of end points and Junctions of a thinned digit image, number of water reservoir sheds and ratio between the number of foreground pixels in upper half part and lower half-part of the numerical digit image. Totally five features are extracted for this algorthim. No such method is available to classify the digits with a minimum number of digits.

Rest of the paper is organized as follows; Section II contains Database and the preprocessing. Feature extraction process is described in Section, the proposed algorithm and Classifications method. results and discussion obtained are presented in Section III. Section IV contains the conclusion part.

II. MATERIAL AND METHOD

The proposed method is mainly consists of 4 phases steps. In the first phase, collecting the numerals data from various data bases and gathering images from various people in both telugu states. After collecting the numeral data preprocess data i.e. elimination of noise and conversion of gray scale images into binary images and also the normalization of the binary images by using the normalization techniques in the second phase. In the third phase, extracted the features from individual digit image and derive an algorithm for recognition of handwritten numerals system in the last phase. The block diagram of the proposed method is shown below figure 1.

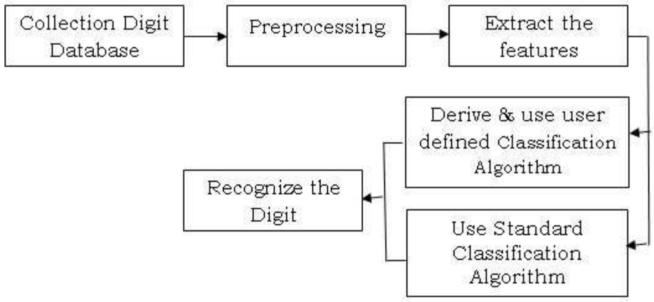

Fig 1: Block diagram proposed handwritten digit recognition system

A. Database creation and the preprocessing

1) Collection of numeral image database:

Several standard datasets of digits are found in English. Some of them are CENPARMI, CEDAR, and MNIST datasets. The CENPARMI (Centre for Pattern Recognition and Machine Intelligence) digit dataset [20] is available from CENPARMI, Concordia University. In this dataset 4,000 images (400 samples per class) are specified for training purpose and 2,000 images are used for testing purpose. These digit datasets were collected from United States Postal Service (USPS). The Center of Excellence for Document Analysis and Recognition (CEDAR) digit dataset is available from CEDAR, The State University of New York, and Buffalo. The training and test sets contain 18,468 and 2,711 digits, respectively. The number of samples in both training and test sets differ for each class. The Modified National Institute of Standards and Technology (MNIST) dataset [21] was extracted from the NIST datasets SD3 and SD7. The training and test sets are composed of both SD3 and SD7. Samples are normalized into 20×20 grayscale images with aspect ratio reserved, and the normalized images are located in 28×28 frame. The number of training and test samples is 60,000 and 10,000 respectively. The sample images of the MNIST dataset is shown in figure 2.

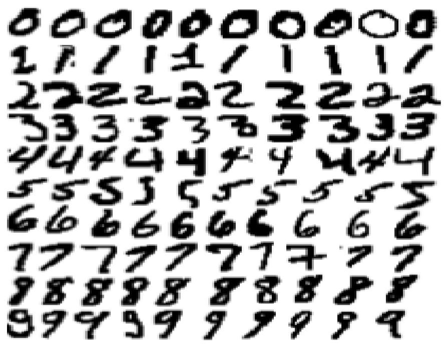

Fig 2: MNIST database

Fig. 2. Sample digit images of MNIST database The plain paper was used for data collection. Each person was instructed to write the digits (fully unconstrained) along the vertical direction. The dataset contains about 100 isolated samples each of 10 numerals written by 1,000 native writers including university graduates, high

school children, and adults. Around fifty percent data is from high school children. A flatbed scanner was used for digitization, with images in gray tone at 300 dpi. These were stored has Bit Map File (BMP) format using a standard technique for converting them into black and white images. Data was manually extracted from scanned images and normalized into 50×50 size using a standard bi-cubic approach. After processing scanned images about digits and a total of 14,4000 (100×1440) images of numerals are obtained. Dataset developed planned to be made available publicly for research purpose. Some of the sample images after extracting from the scanned image are shown in figure 3.

Fig 3. Sample Scanned document of digit images Step

2) Preprocessing:

The recognition of handwritten numerals can achieve high performance based on preprocessing stage also. The handwritten digit database images used in this approach are in gray scale images, convert images in database into binary images. Converting the gray level images into binary based on the threshold value. After converting the images into binary , images may have surplus elements one's (black) at undesirable places in the background image. For efficient classification, need to remove the surplus from undesirable places.

To remove these unwanted one's from the background, noise removal algorithm[8] is used. The present paper used own 3×3 templates are used. It is assumed that the pixel (p, q) as center pixel and neighbors of the center pixel are (p- 1, q), (p-1, q+1), (p, q+1), (p+1, q+1), (p+1, q), (p+1, q-1), (p, q-1), and (p-1, q-1), as is shown in figure 4. Noise removal in an image depends on the type of noise in that image. The noise can identified by using 3×3 templates. The template shown in figure 3 isused to identify the single pixel surplus data in digit image. In the same way 2-pixel, 3-pixel and 4-pixel noise can be identified by using the templates shown in figure 4, 5 and 6 respectively.

Once the 1-4-pixel noise (Black pixel) identified, the black pixel is replaced with white-pixel. Figure 9 shows the digit database contains some amount of noise . Figure 10 shows the resultant images after removing the noise by using the 3×3 window patterns shown figures from 5 to 8.

P_1(p-1, q-1)	P_2(p-1, q)	P_3(p-1, q+1)
P_8 (p, q-1)	P_1(p, q)	P_4 (p, q+1)
P (p+1, q-1)	P_6 (p+1, q)	P_5(p+1, q+1)

Fig 4: 8-neighbors of a 3×3 window

0	0	0
0	1	0
0	0	0

Fig 5: Isolate pixel pattern

0	0	0
0	1	1
0	0	0

0	0	1
0	1	0
0	0	0

0	1	0
0	1	0
0	0	0

1	0	0
0	1	0
0	0	0

0	0	0
1	1	0
0	0	0

0	0	0
0	1	0
1	0	0

0	0	0
0	1	0
0	1	0

0	0	0
0	1	0
0	0	1

Fig 6: Two pixel patterns for noise removal

0	0	0
0	1	1
0	0	1

0	1	0
0	1	1
0	0	0

1	1	0
0	1	0
0	0	0

1	0	0
1	1	0
0	0	0

0	1	0
1	1	0
0	0	0

0	0	0
1	1	0
1	0	0

0	0	0
1	1	0
0	1	0

0	0	0
0	1	0
0	1	1

Fig 7: Three pixel patterns for noise removal

1	1	0
1	1	0
0	0	0

0	1	1
0	1	1
0	0	0

Fig 8: Four pixel patterns for noise removal

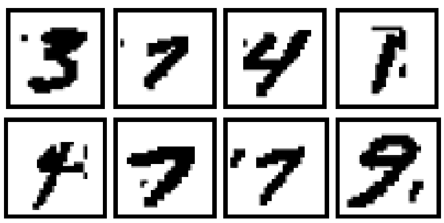

Fig 9: Before preprocessing numerical digit aimages

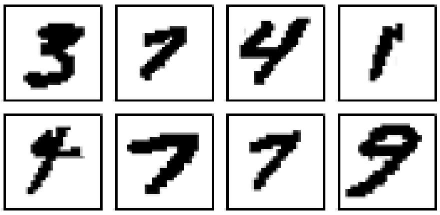

Fig 10: After preprocessing numerical digit images

B. *Feature Extraction*

1) The number of contours / aholes:

The number of digit image contours is a structural feature. The contour is a boundary of object, a population of points (pixels), and separating object from a background. The contour contains the necessary information on the object shape. In systems of computer vision, so many approaches are there to find the number of contours of a digit image . Some of the approaches are Freeman chain code approach [22], two-dimensional coding system [23], polygonal coding, and the connected component labeling algorithm [24] and etc. From them, the connected component labeling algorithm is most popular. In connected component labeling algorithm, the number of contours is equal to the number of background components (white components) minus one. For example, the digits 0, 2, 4, 6, 9 have two contours. The number of contours of the digits from 0 to 9 shown in table1.

TABLE I
NUMBER OF CONTOURS IN NUMERICAL DIGIT

Digit	0	1	2	3	4	5	6	7	8	9
Number of contours	2	1	1 or 2	1	1 or 2	1	2	1	3	2

2) Skeleton feature

Thinning is usually a pre-processing stage in character recognition where the character image is reduced to a simplified one-pixel wide skeleton. In this paper use Fast thinning algorithm which gives good skeletons for digitimages [25] for finding the Skeleton feature: Endpoint (EP) and Junctions.

Thinning: A binary digitized picture is defined by a matrix M where each pixel M(i, j) is either 1 or 0. The pattern consists of those pixels that have value 1. Each stroke in the pattern is more than one element thick. Iterative transformations are applied to matrix IT point by point according to the values of a small set of neighboring points. It is assumed that the neighbors of the point (p, q) are (p - 1, q), (p - 1, q + 1), (p, q + 1), (p + 1, q + 1), (p+ 1, q), (p + 1, q - 1), (p, q- 1), and (p - 1, q - 1), as is shown in Figure 4.

In parallel picture processing, the new value given to a point at the nth iteration dependson its own value as well as those of its eight neighbors at the $(n - 1)^{th}$ iteration, so that all picture points canbe processed simultaneously. It is assumed that a 3×3window is used, and that each element is onnectedwith its eight neighboring elements.

The algorithm requires only simple computations.

Parallel Thinning Algorithm:

It consists of two Sub iterations

- Step1: deleting the south-east boundary points and north-west corner points
- Step2: north-west boundary points and south-east corner points are deleted
- Step3: end points and pixel connectivity are preserved

The above procedure is done until there are no changes in the image.

Procedure:

In order to preserve the connectivity of the skeleton, divide each iteration into two sub iterations. In first iteration, the contour pixel P_1 is deleted from pattern if the following conditions are satisfied

(a) $2 <= B(P_1) <= 6$
(b) A(P) = 1
(c) $P_2 * P_4 * P_6 = 0$
(d) $P_4 * P_6 * P_8 = 0$

where (P_1) is the number of 01 patterns in the orderedset P_2, P_3, P_4, - • - P_8, P_9 that are the eight neighbors of P_1 (Figure 4), and B(Pi) is the number of nonzero neighbors of P_1,that is,

$$B(P_1) = P_2 + P_3 + P_4 + \cdots + P_8 + P_9.$$

If any condition is not satisfied, e.g., the values of P2,P3, P4, • • - P9 as shown in Figure 2, then A(Pi) = 2 Therefore, P_1 is not deleted from the picture.

In the second sub-iteration, only conditions (c) and (d) are changed (Figure 11) as follows:
(c') $P_2*P_4*P_8 = 0$
(d') $P_2*P_6*P_8 = 0$ and the rest remain the same.

By conditions (c) and (d) of the first sub-iteration, it will be shown that the first sub-iteration removes only the south-east boundary points and the north-west cornerpoints which do not belong to an ideal skeleton which shows in figure 12. The proof for the first sub-iteration is given, that is,the points to be deleted satisfy conditions:

1

0	0	1
1	P_1	0
1	0	0

2

Fig 11: Countinga the 01 patterns in the ordered set P2, P3, P4, ... P8, P9.

		North		
		P_2		
West	P_8	P_1	P_4	East
		P_6		
		South		

Fig 12: Points under consideration and their locations

(c) $P_2*P_4*P_6 = 0$ (1)
(d) $P_4*P_6*P_8 = 0$ (2)

The solutions to the set of equations (1) and (2) are $P_4 = 0$ or $P_6 = 0$ or ($P_2 = 0$ and Ps = 0). So the point P_1,which has been removed, might be an east or south boundary point or a north-west corner point. Similarly, it can be proved that the point P_1 deleted in the second sub-iteration might be a north-west boundary point or a south-east corner point.

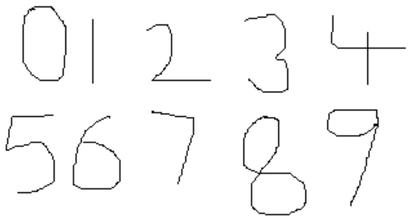

Fig 13: Thinned digit images

By using the skeleton (single pixel width) of the digit's body to extract one feature: feature point. can be easily found from the skeleton by examining the eight immediate neighbors of every black pixel p shown in figure 6. The resultant image of the thing operation of digit database in shown in figure 13.

3) End point:

End point (EP) [26] is a digit point (black pixel) with only one black neighbor . The patterns used to find endpoints of an image when the image is scan from top left corner to bottom right corner is shown in figure 14. The table2 shows the number of end points of the each numerical digit. The figure 14 shows vthe end points of each numeral digit.

TABLE II

NUMBER OF END POINTS OF IN NUMERICAL DIGIT

Digit	0	1	2	3	4	5	6	7	8	9
Number of contours	0	2	2	2 or 3	2 or 3 or 4	2	1	2 or 4	0 or 1	1

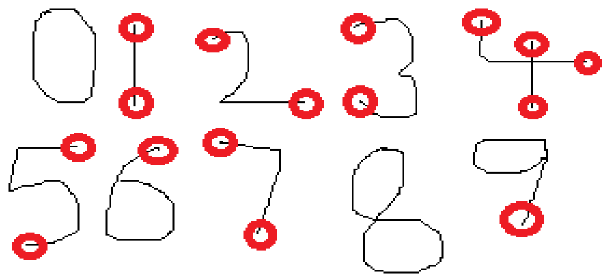

Fig 14: End point Images

4) Junction in thinned image:

Junction (JP) [25] is a digit point (black pixel) with any 4 black neighbors. Generally the digit 2, 4 and 8 forms the junctions. The present approach uses the number of junction points (NoJ) is used as feature for the digit recognition.

5) Number of Water Reservoir Area:

The water reservoir principle is as follows. If water is dispensed from a side of a component, the cavity regions of the component where water will be stored are considered as reservoir [27]. The opening regions of the components where water will be stored are considered as reservoir. Handwritten digits generate reservoirs which are used for classification. The water reservoir extraction scheme of Handwritten Digits.

Top reservoir: By top reservoirs of a digit, it means the reservoirs obtained when water is poured from top of the digit. The water reservoir is appeared when digit image is unconnected. The figure 15 shows Top reservoir area of the Digit4. If digits are unconnected then the cavitya regions are generated. Figure 16 illustrates the unconnected digits of 0, a6, 8 and a9

Fig 15:. Top reservoir area of the digit 4

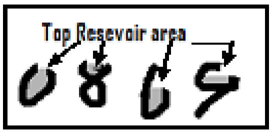

Fig 16: Unconnected digit that causes generate the cavity regionsa.

Bottom reservoir: By bottom reservoirs of a digit, it means the reservoirs obtained when water is poured from bottom of the digit. A bottom reservoir of a digit is visualized as a top reservoir when water will be poured from top after rotating the digit image by 180 degrees. Generally bottom reservoir is not obtained for any image except if the image is slant towards the bottom. The figure 17 shows that illustrations

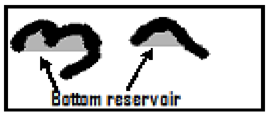

Fig 17. Bottom reservoirs of the digits 3 and 7 when they slant to abottom.

Left (right) reservoir: If water is poured from the lef (right) side of a c omponent, the cavity regions of the digit where water will be stored are considered as left (right) reservoirs. The figure 18 and 19 illustrates the left and right reservoirs.

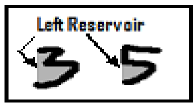

Fig 18: Left reservoirs of the digits 3 and 5

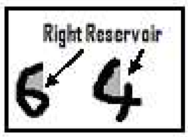

Fig 19: Right reservoirs of the digits 6 and 4

The present paper uses that the digit form how many water reservoir area as feature for the recognition of digits.

C. Proposed user defined classification algorithm for handwritten digit Recognition

Algorithm1: Recognition of Handwritten Digits
Input: Isolated numeral image
Output: Classification of the Numeral Digit 0 or 1 or 2 or 3 or 4 or 5 or 6 or 7 or 7a or a8 or 9.

Method: Structural, Statistical features aextraction.

- Step1: Pre-Process the image
- Step2: Find the Number of Contours (NoC) and classify the digit as Group1 or Group2 or digit 8.
- Step3: Find out the Skeleton of input image and extract the Skeleton feature number of end points (NoEP) in Group1 and Group2.
- Step4: Based on NoEP in digits in Group1 further divide into 3 sub-groups Group1, Group1b, Group1c, Group2 and classify the digits 0, 2, 4
- Step5: Calculate the Number of watersheds (NoW) of the Digits in Group1.
- Step6: Based on NoW values further divide into Subgroup Group1a nd classify the digits 1, 7, and 2.
- Step7: Calculate the Number of Joints of digits in Group1.
- Step8: Based on the NoJ values classify the image is either digit 3 or 5.
- Step9: Crop the Digits in Gropu1b.
- Step10: Calculate the Number of watersheds (NoW) of the Digits in Group1b.
- Step11: Based on NoW values classify the digits either 3 or 4.
- Step12: Calculate the ratio between the numbers of foreground pixels in upper half part and lower half-part (ULR) of the numerical digit images in Grop1c.
- Step13: Based on the ULR value classify the digit either 4 or 7.
- Step 14: Calculate ULR value of the digits in Group2a
- Step15: Based on the ULRvalue classify the digit either 6 or 9.

The graphical representation of the user define algorithm is shown in figure 20 nd 20b.

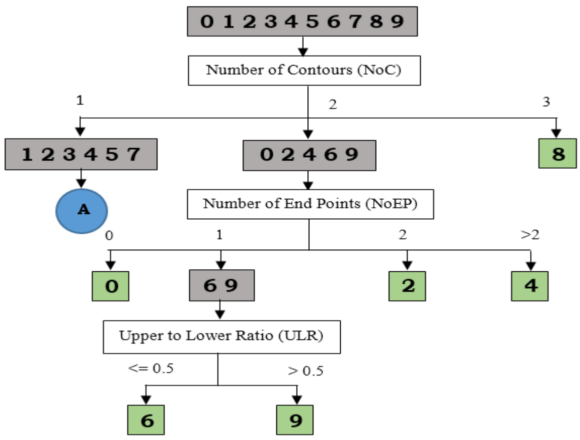

Fig 20a: Flow chart of user defined algorithm

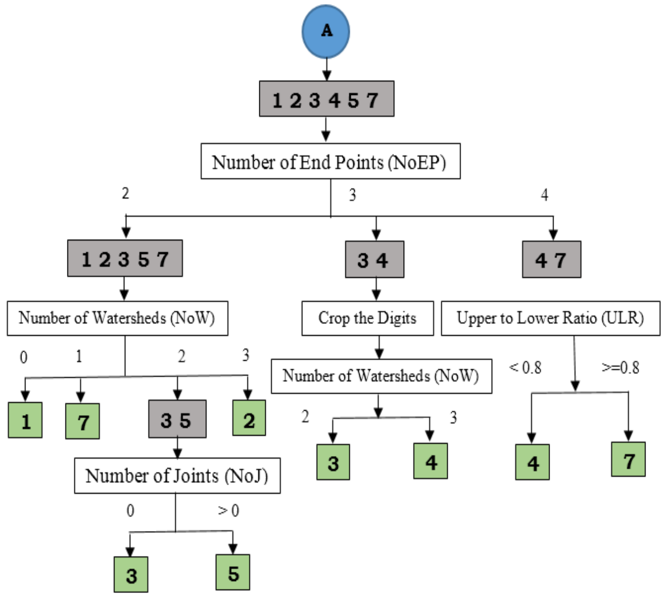

Fig 20b: Flow chart of user defined algorithm

From the above algorithm, it is identified that the proposed method needs the feature at maximum 4 for the classification of handwritten digits every digit does notneed all these 4 features. The number of features required to recognize the digits is shown in the table3.

TABLE III
FEATURE COUNT REQUIRED FOR FOR EACH DIGIT

Digit	0	1	2	3	4	5	6	7	8	9
Number of Features	2	2	2/3	3/4	3	4	3	3	1	3

III. Result and Discussion

We evaluated the performance of user-defined recognition algorthim with a standard handwritten numerals database set which are collected from MNIST database that contains of 60,000 training and 10,000 testing datasets where as from CENPARMI database with 4,000 training set and 2,000 testing set, CEDAR database with total training and testing datasets 21,179 respectively.And finally USPS database which consists of 7,291 training set and 2,007 testing test. The some of the images collected from students in Hyderabad city who are studied from first class to UG by giving the format to the students is shown in figure 3. Scan the individual digits and created own database which consists of 1,20,000 images as a training set and 24,000 images as a training set. Collectively, the dataset contains 2,12,470 digit images used for training and 38,007 images for testing. No such method has tested using with huge database.

All 3experiments are carried out on a PC machine with i3 processor 2.7GHz CPU and 3 GB RAM memory under MatLab 11.0 aplatform. From the training set images extract the feature values which are specified in section 3 and stored in Feature Vecto r(FV) and extract the features of test database and stored in Test Vector (TV).

Classifier performance depends greatly on the characteristics of the data to be classified. There is no single classifier that works best on all given problems (a phenomenon that may be explained by the no-free-lunch theorem). Various empirical tests have been performed to compare classifier performance and to find the characteristics of data that determine classifier performance. Determining a suitable classifier for a given problem is however still more an art than a science. An intriguing problem in pattern recognition yet to be solved is the relationship between the problem to be solved (data to be classifieda) and the performance of various pattern recognition algorithms (classifiers). In this paper used user-defined classification algorithm, k-Nearest-Neighbor Classifier and Support Vector Machines (SVM) are used individually and classify test database digit images.

A. User defined algorithm

By using the algorithm defined in section 3, the digits are classified. The proposed algorithm is tested with 38007 images and the individual results of the test database is listed in table 4.

TABLE IV
RESULTS OF NUMERICAL RECOGNITION FOR 38007 TEST IMAGES USING USER-DEFINED CLASSIFIER

Digit	No.of images	Correctly Classified	Not Correctly Classified	% accuracy
0	3,878	3,850	28	99.28
1	3,891	3,854	37	99.05
2	3,743	3,717	26	99.31
3	3,907	3,868	39	99.00
4	3,679	3,645	34	99.08
5	3,815	3,794	21	99.45
6	3,962	3,924	38	99.04
7	3,756	3,738	18	99.52
8	3,794	3,753	41	98.92
9	3,582	3,549	33	99.08
Average Recognition Percentage				99.17

B. K-Nearest neighbor classification (k-nn Classifier)

The proposed method uses k-nearest neighbor (k-nn) classification algorithm for classifying the test database digit images in test database using the feature vector of training database. The k-nearest neighbor algorithm (k-NN) is classification technique to classify the digits based on training features space. In K-nn object is classified to a particular class which has majority of votes. In the k-nn classification, compute the distance between feature values of the test sample and the feature vector values of every training image and the class of majority mong the k-nearest training samples is based on the Euclidian distance ameasures. The training vector is a multidimensional array. Each row in an array contains feature values and corresponding class label of the training images where as test vector contains only feature values. In classification process, for each row in test vector assign the class label based on the Euclidian distance measures and number of neighbors (k) considered. The k value is defined by the user.

The algorithm is executed with the value of k is 1, 3, aand 5. The graphical representation of the accuracy of classification in using various k values are shown in figure 10 and the overall classification results are listed out in table 5. From table 5, it is clearly evident that the optimal value of k is 3 for classification of considered numerical digits by using k-nearest neighbor classification technique. The recognition rate of the individual digits in test samples by using k-nearest neighbor classification lgorithma (with k value 3) is listed in table 6 and from that table the overall recognition rate of the test database ias 98.62%.

TABLE V
ACCURACY RATE USING DIFFERENT VALUES OF K WITH KNN CLASSIFIER

NN classifiers with different K values	Numbera of training samples	Numbera of testinga digits images	Accuracy in percentage
1	2,12,470	38,007	98.15
2	2,12,470	38,007	99.44
3	2,12,470	38,007	97.87

TABLE VI
RESULTS OF NUMERICAL RECOGNITION FOR 38007 TEST IMAGES USING K-NN CLASSIFIER

Digit	No.of images	Correctly Classified	Not Correctly Classified	% accuracy
0	3,878	3,829	49	98.74
1	3,891	3,833	58	98.51
2	3,743	3,696	47	98.74
3	3,907	3,847	60	98.46
4	3,679	3,624	55	98.51
5	3,815	3,773	42	98.90
6	3,962	3,903	59	98.51
7	3,756	3,717	39	98.96
8	3,794	3,732	62	98.37
9	3,582	3,528	54	98.49
Average Recognition Percentage				98.62

C. Support Vector Machines (SVM) based classification

In Artificial Intelligence (AI) and Machine Learning (ML), Support Vector Machine (SVM) algorithm is a powerful classification tool. It is a supervised learning model used for classification and regression. The SVM algorithm was developed by Vapnik developed the SVM algorthim for ML tasks. In this amodel, the present approach is given a set of training examples ain which each one of them is marked to be belonginga to one of the two classes.

SVM is a training algorithm for linear classification and also for non-linear classification. The solution is expressed as a linear combination of supporting patterns, which are the subset of training patterns close to the decision boundary, called the support vectors. For non-linear case, SVM mapped the data sets of input space into a higher dimensional feature space, which is linear and the large margin learning algorithm is then applied. The present paper uses SVM classifier and statistical feature sets for classification.

The extracted features of both training data set and test database and are given to a SVM machine. The recognition rate of the individual digits in test samples by using SVM algorithm is shown in table 7 and from that table the overall recognition rate of the test database is 98.49%.

TABLE VII
RESULTS OF NUMERICAL RECOGNITION FOR 38007 TEST IMAGES USING SVM CLASSIFIER

Digit	No.of images	Correctly Classified	Not Correctly Classified	% accuracy
0	3,878	3,824	54	98.61
1	3,891	3,828	63	98.38
2	3,743	3,691	52	98.61
3	3,907	3,842	65	98.34
4	3,679	3,619	60	98.37
5	3,815	3,768	47	98.77
6	3,962	3,898	64	98.38
7	3,756	3,712	44	98.83
8	3,794	3,727	67	98.23
9	3,582	3,523	59	98.35
Average Recognition Percentage				98.49

When comparing the three classifiers for the same test database and same feature set, User defined classifier shows better results it concludes that classification results not only depends on features extracted from the image but also depends on type of classification technique used. The overall recognition rate is about 98.76%. The time consumed for the testing of the test database for user defined algorithm is about 19.0025 sec, for k-nn classifier need 59.032 sec and SVM classifiers need 60.732 sec. The advantage of use defined classifier is once the algorithm is e defined no need to calculate the feature values. Any numbers of images can be tested.The confusion matrix generated by the user-defined classifier is shown in table 8.

TABLE VIII
CONFUSION MATRIX GENERATED WHEN USER DEFINE CLASSIFIER APPLIED ON TEST DATABASE

Digit	0	1	2	3	4	5	6	7	8	9
0	**3850**	**4**	**2**	**0**	**0**	**0**	**3**	**8**	**6**	**5**
1	**2**	**3854**	**9**	**0**	**0**	**2**	**3**	**12**	**3**	**6**
2	**0**	**5**	**3717**	**3**	**2**	**2**	**2**	**3**	**2**	**7**
3	**0**	**2**	**1**	**3868**	**2**	**3**	**4**	**5**	**13**	**9**
4	**0**	**0**	**3**	**4**	**3645**	**5**	**2**	**3**	**11**	**6**
5	**0**	**0**	**2**	**3**	**0**	**3794**	**9**	**0**	**0**	**7**
6	**3**	**2**	**2**	**0**	**0**	**7**	**3924**	**7**	**9**	**8**
7	**0**	**12**	**0**	**1**	**0**	**0**	**0**	**3738**	**2**	**5**
8	**1**	**6**	**4**	**15**	**0**	**3**	**5**	**0**	**3753**	**7**
9	**5**	**0**	**3**	**4**	**2**	**1**	**7**	**3**	**8**	**3549**

D. Comparison of the proposed approach with other existing approaches:

The efficiency of the proposed method is compared with other existing methods like aDeep Learning proposed by *Anuj Dutt et.al* [28], MCS HOG Features and SVM classifer by *Hamayun* [29], and Convolutional Neural Network as a Classifierproposed by *Jisha et.al* [30]. Anuj Dutt proposed Machine Learning, Deep Learning and Computer Vision algorithms. The overall efficiency of the deep learing approach is about 98.72%. *Hamayun* proposed Multiple-Cell Size (MCS) approacha was aproposed for utilizing Histogram of Oriented Gradient (HOG) features and a Support Vector Machine (SVM) based classifier for efficient classification of Handwritten Digits. The HOG based techniques sensitive to the cell size selection used in the relevant feature extraction computations. Hence a new MCS aapproach has been used to perform HOG analysis and compute the HOG features. This approach got 99.36% when applied on MNIST database. But while applying on other databases it is decreased to 97.13. Jisha proposed Convolution Neural Networks (CNNs) consist of multiple layers. It is a powerful technique for classification of visual inputs like handwritten digits nd faces recognition. The classification task is performed using aConvolution Neural Network (CNN). The actually purpose of developing the multilayer neural network is to reduce the meansquare error, between the actual output and the final output. But each subnet between the input and the hidden-layer are initialized with random weights and also trained with different feature maps. This is applied on only 1000 image but applied on large database it decreased to 96.36%. The performance evolution of the proposed method with other existing methods is listed out in table 10 and the classification graph is represented in figure 11. From table 9 and figure

21, it is clearly evident that, the proposed method exhibits a high recognition rate than the existing methods.

TABLE IX
ACCURACY OF HWD CLASSIFICATION OF DIFFERENT APPROACHES

Database	Deep learning [28]	HOG based technique [29]	Convolution Neural Networks [30]	Proposed Approach
MNIST	98.98	97.28	95.72	98.95
CEDAR	98.73	97.45	96.39	99.24
CEPARMI	98.33	96.52	97.45	99.23
Scanned images	98.85	97.28	95.88	99.29

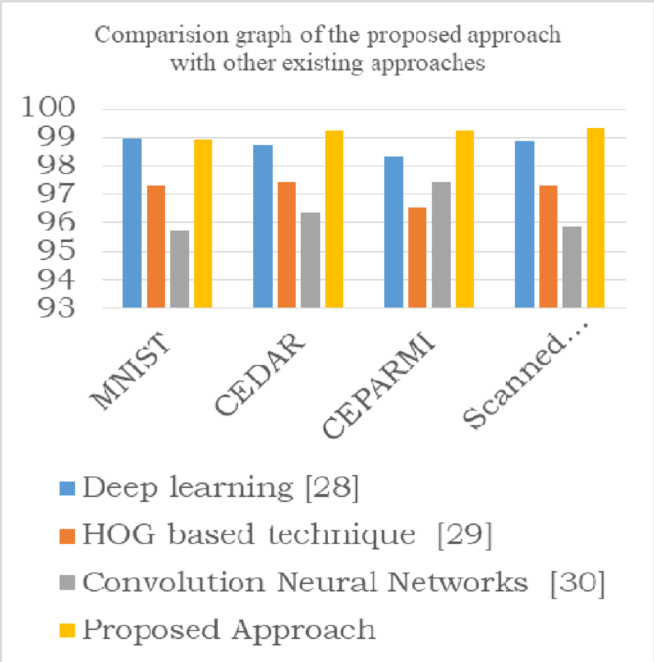

Fig 21: Graphical representation of the % recognition of the proposed method and other existing method

IV. CONCLUSIONS

The novelty of this method is that, free from size normalization, fast, accurate, independent of size and Writer style/ink independent. The present paper defined classification of isolated handwritten digits with good classification results. The proposed method was tested with large database. No such method is available in the literature to test the large test data set. The proposed approach extracts only 5 features. The time complexity of the proposed method is also very less. The proposed method shows high recognition rate when compare with the test results of standard classification algorithms.

REFERENCES

[1] Y.LeCuu, B.Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard and L.D. Jackel, Handwritten Digit Recognition with a Back-Propagation Network. AT&T Bell Laboratories, Holmdel, N.J.07733, 1990.

[2] Zhou, Z.H. and Feng, J, Deep Forest: Towards an Alternative to Deep NeuralNetworks.Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, pp. 3553-3559, 2017

[3] Nielsen, M.A, Neural Networks and Deep Learning. Determination Press, http://neuralnetworksanddeeplearning.com, 2015.

[4] Bengio, Y.I., Goodfellow, J. and Courville, A, Deep Learning. MIT Press, Cambridge, MA, http://www.deeplearningbook.org, 2016.

[5] Farulla, G.A, Murru, N. and Rossini, R. A Fuzzy Approach for Segmentation of Touching Characters, arXiv:1612.04862v1, https://arxiv.org/abs/1612.04862, pp.1-18, 2016.

[6] Chen, G., Li, Y. and Srihari, S.N, Word Recognition with Deep ConditionalRandom Fields. Proceedings of the IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, 25-28 September 2016.

[7] Liu, Z., Li, Y., Ren, F. and Yu, H. A Binary Convolutional Encoder-DecoderNetwork for Real-Time Natural Scene Text Processing. arXiv: 1612.03630v1, 2016.

[8] Jaderberg, M., Simonyan, K., Vedaldi, A. and Zisserman.A, Reading Text inthe Wild with Convolutional Neural Networks. International Journal of Comput Vision, vol 116, pp.1-20, 2016.

[9] Yin, X.C., Yin, X., Huang, K. and Hao, H.W, Robust Text Detection in NaturalScene Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 36, pp. 970-983, 2013.

[10] Oivind Trier, Anil Jain, Torfiinn Taxt, A feature extraction method for character recognition-A survey, pattern Recognition, vol 29, pp641-662, 1996.

[11] Shamic Surel, P.K.Das, Recognition of an Indian Scripts Using Multilayer Perceptions and fuzzy Features. Proc. Of 6th Int. Conf. on Document Analysis and Recognition (ICDAR), Seattle, pp 1220-1224, 2001.

[12] P.Nagabhushan, S.A.Angadi, B.S.Anami,A fuzzy statistical approach of Kannada Vowel Recognition based on Invariant Moments, Proc. Of NCDAR-2003, Mandy, Karnataka, India, pp.275-285, 2003.

[13] J.D. Tubes, A note on binary template matching. Pattern Recognition, vol 22, pp:359-365, 1989.

[14] Anil K.Jain, Douglass Zonker, Representation and Recognition of handwritten Digits using Deformable Templates, IEEE, Pattern analysis and machine intelligence, vol 19, pp:1386 – 1390,1997.

[15] L.Heutte, T.Paquest, J.V.Moreau, Y.Lecourtier, C.Oliver, A structural/ statistical feature based vector for handwritten character recognition, Pattern Recognition,vol 19, pp.629-641, 1998.

[16] U Ravi Babu, Y V V Satyanarayan nd S. Marthu Perumal , Printed Telugu Numeral Recognition based on Structural, Skeleton and Water Reservoir Features, International Journal Of Computers & Technology, Vol 10, pp.1815-1824, 2013.

[17] U Ravi Babu , U Venkateswralu and Anil Kumar, Handwritten Digit Recognition Using Structural, Statistical Features and K-nearest Neighbor Classifier. International Journal of Information Engineering and Electronic Business, vol 1, pp.62-68, 2014.

[18] U Ravi Babu and G Charles babu, Novel Technique for the Handwritten Digit Image Features Extraction for Recognition. International Journal of Video & Image Processing and Network Security IJVIPNS-IJENS Vol 15, pp: 1-9, 2015.

[19] Pullela S V V S R Kumar, Appana Naga Lakshmi, U Ravi Babu, New approaches for the Features Extraction on Handwritten Digit Image for Recognition. International Journal of Digital Content Technology and its Applications (JDCTA) Vol 10, pp: 33-44, 2016.

[20] M. D. Garris, J. L. Blue and G. T. Candela, NIST form-based handprint recognition system. National Institute of Standards and Technology, 1997.

[21] P. J. Grother, NIST special database 19 hand printed forms and characters database. National Institute of Standards and Technology, 1995.

[22] G. G. Rajput, Rajeswari Horakeri, Sidramappa Chandrakant, Printed and Handwritten Kannada NumeralRecognition Using Crack Codes and Fourier Descriptors Plate. IJCA Special Issue on RecentTrends in Image Processing and Pattern Recognition pp.53-58 2010

[23] Graham, D.N, Image transmission by two-dimensional contour coding. Proceedings of the IEEE Vol55, pp.336 – 346, 1967.

[24] U. Pal, T. Wakabayashi, N. Sharm nd F. Kimura,Handwritten Numeral Recognition of Six Popular IndianScripts. In Proc. 9th International Conference on Document Analysis and Recognition. pp. 749-753, Curitiba,Brazil, September 24-26, 2007.

[25] T. Y. Zhang and C. Y. Suen, A Fast Parallel Algorithm for Thinning Digital Patterns, Image Processing andComputer Vision, Vol 27,pp.236-239, 1984.

[26] Seong- W han Lee,Multi layer cluster neural network for totally unconstrained handwriritten numeralrecognition. Neural Networks. Vol 8, pp.409-418, 1984.

[27] U Pal and P.P.Roy, Multi-oriented and curved text lines extraction from Indian documents. IEEE Trans on system,Man and Cybernetics-Part B, Vol 34, pp.1667-1684, 2004.

[28] Anuj Dutt, AashiDutt, Handwritten Digit Recognition Using Deep Learning. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol 6, pp: 990-997,2017.

[29] Hamayun A. Khan, MCS HOG Features and SVM Based HandwrittenDigit Recognition System. Journal of Intelligent Learning Systems and Applications, vol 9, pp.21-33, 2017,.

[30] Jisha Mol L, Femy John, Handwritten Digit Recognition: ConvolutionalNeural Network as a Classifier. International Journal of Innovative Research in Computerand Communication Engineering, Vol. 5, pp. 6729-6734, 2017.