www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 13 Issue 04 April 2024, Page No. 26130-26134
ISSN: 2319-7242 DOI: 10.18535/ijecs/v13i04.4814

Performance analysis of multiple classifiers for Marathi digit recognition

A A Jinturkar, P P Agnihotri, P B Khanale

Department of Computer Science, SIES College (Autonomous), Nerul
School of Technology, SRTMUN Sub-Centre, Latur
Professor, DSM College, Parbhani

Abstract:
For many researchers over the globe; the Marathi Language Handwritten digit recognition is an area of importance.. This study has its primary application to recognize the correct Marathi handwritten numerals. Many systems have been developed for numeral recognition using soft computing paradigms such as Artificial Neural Networks, Fuzzy logic-based methods. In this paper, three classifiers were used for handwritten Marathi numeral recognition that is K- Nearest Neighbour (KNN) Classifier, Bayes classifier and SVM Classifier. In this study primary database with 250 samples were used to perform the experiment. The SVM classifier gives 98% highest recognition accuracy.

Keywords: Digit Recognition, K-means classifier, SVM Classifier, Gaussian classifier

Introduction:
Handwriting recognition is of the most fascinating area for researchers. It has been evolved and evolving continuously with the help of soft computing and other paradigms. Handwriting recognition enables computing machines to identify and recognize human writing. It can be divided as On-line and off-line recognition depending upon input source. In the former; input is obtained through writing on touch-sensitive surface while later accepts input in the form of digital images.[1]

Off-line recognition is considered as a difficult since it includes preliminary stages such as preprocessing of digital image, isolation of digits among the others. Problem becomes more complex in case of languages like Marathi due to its writing style and variations.

2. Marathi Language
Marathi language is based on Devanagari script which is also used for writing other languages such as Hindi, Kokani, Nepali with little modifications. It is the third most spoken language in India after Hindi and Bengali. It has 10 digits from 0 to 9 written as shown in fig:

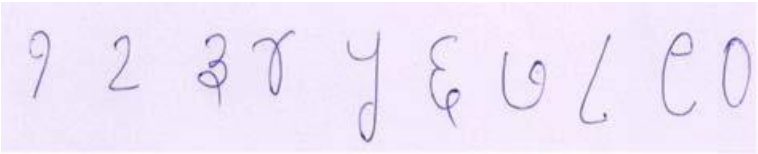

Fig. 1: Marathi Handwritten numerals

From the above image; one can observe that Marathi digit 3 and 6, 1 and 9 seems to be mirror images of one another. Moreover; people used to write digits connected to one another in many cases for e.g., writing amount on bank cheques, etc. All these issues make this problem still open for research [2]. Rest of the paper

is organized as follows: Section 3 discusses the literature review, section 4 describes proposed system and experimental set up, section 5 discusses the results and section 6 gives concluding remarks.

Following table 1 summarizes the work done by researchers and their respective results.

Table 1: Recognition accuracy of Marathi numerals of some existing researchers

Sr. No.	Author name	Technique used	Accuracy (in %)
1	M. Hanmandlu et. Al. [3]	Fuzzy Model	90.65
2	Sonika Narang [4]	MLP, NeuralNetwork (NN),CNN, SVM,Random Forest	88.95
3	Khandja D [5]	Neural network	93.4
4	Arora S [6]	Classificationdecision obtained from four Multi Layer Perceptron(MLP)	92.80
5	S. Arora [7]	Multilayer perceptron (MLP) neural networks	94.15
6	Pankaj Kale [8]	ANN Classifier	92
7	Verma G.K [9]	K-Nearest Neighbor classifier using Curvelet Transform	90

Methodology

Present study addresses the problem of classifying Marathi handwritten numeral digits into their respective classes (0-9). We have used local database which was developed as a part of research. It consists of 250 images of Marathi handwritten digits; 25 images of each digit. Dataset is then divided into Training and testing data set using k-fold cross validation technique to ensure that every sample image appears in train and test data set.

Given the dataset; the objective is to train the machine learning model using train data set and enable model to classify unknown image from test data set to classify correctly. The block diagram shown in figure 2 shows the steps involved in the development of system.

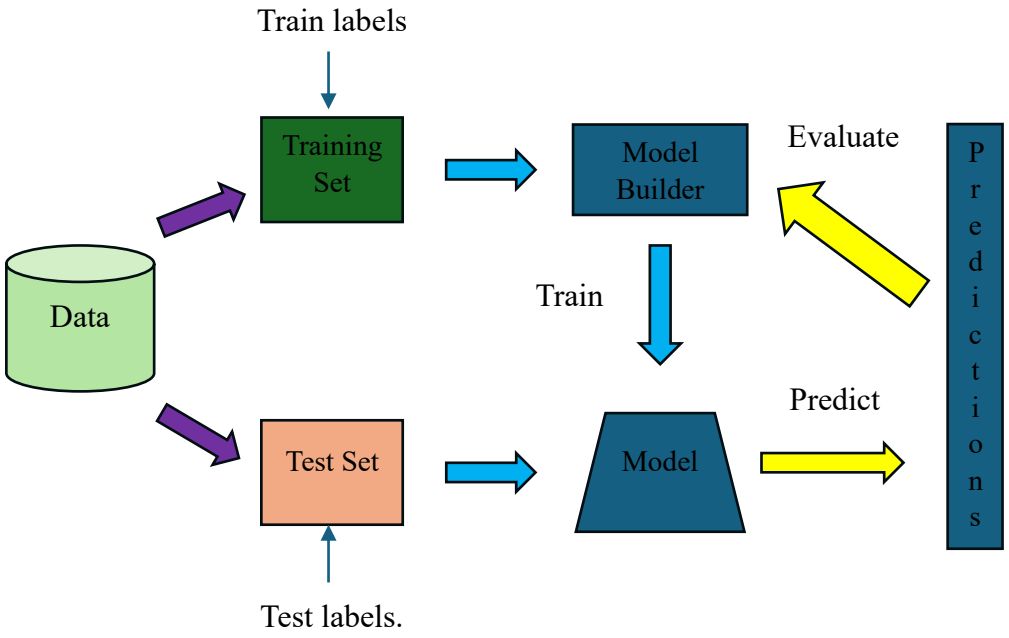

Figure 2: Proposed System

As Shown in figure 2; the primary data set is divided into training and test data sets. First select the specific classifier using Model Builder, The parameters are initialized like training algorithm, epoch and

performance function etc for selected classifier for training then perform the training operation. The Output is Trained Model which is ready for the task of recognition. Model then takes test data set and predicts the labels of images in test data set. These predictions are then evaluated for checking the accuracy of model.

Following are the highlights of proposed system:

- Three classifiers namely K-nearest neighbour, Gaussian Bayes and SVM were used.
- All three classifiers were trained using given training data set.
- These classifiers were then used for classifying sample images from test dataset.
- Predicted classes then compared with actual classes to measure the accuracy of system.

K- Nearest Neighbour (KNN) Classifier

This is the closest neighbour classifier which has straightforward approach for our problem. It constructs a classifier that accepts an image of a handwritten digit and produces a label (0–9). The k examples from the training dataset that are closest to the test image are located firstly. The label of test imagecan then be determined by simply using majority-voting. The label that most of the k closest training data points have will be the label of test image.

In order to calculate dataset's closest neighbors we can use the different distance measure techniques. For present problem; we have used Euclidean distance function. The Euclidean distance for two vectors x, y $\in R^d$ is as follows:

$$\|x - y\| = \sum_{i=1}^{d}((x_i - y_i)^2 \qquad \text{---------- Eq 1.1}$$

Following is the procedure for implementation of k-nearest neighbour classifier.

Step 1: Load train data set.
Step 2: Load Test data set.
Step 3: For each image i in test data set do
    A] Scan through each train image
    B] Compute Distance of each train image with given test image
    C] Get the k number of images which are having minimum distance.
Step 4: Evaluate the performance of classifier.

Bayes classifier

This classifier uses Gaussian generative model which is relatively faster and compact than the earlier classifier. It is one of the popular machine learning paradigm among researchers when dealing with small to medium size datasets. This classifier works on assumption that features are independent and gaussian distribution is followed in each class. Following is the procedure for implementing the above discussed classifier:

Step 1: Load train data set.
Step 2: Load Test data set.
Step 3: For each image i in train data set do
    A] Scan through each train image
    B] Compute mean and variance fo each class
Step 4: Once the training is successful it can be used to test the unknown images.
Step 5: Given an image from test data set; classifier determines the Maximum Likelihood Estimate (MLE) for image to every possible class.
Step 6: Class with maximum MLE is then assigned to that image.

SVM Classifier

SVM is one of the largely used classifier for multi-class classification problems. It uses quadratic programming techniques for separating hyper-planes by maximum margin. SVM is well suited for data with multiple variations where it becomes evident to find complex decision boundaries. They outperform many classifiers when used with appropriate feature extraction techniques. Following is the procedure for implementing the SVM classifier:

Step 1: Load train data set.
Step 2: Load Test data set.
Step 3: For each class of image; find the optimal hyperplane which separates it from the other classes in the feature space.
Step 4: Once the training is successful it can be used to test the unknown images.

Experimental Results

In order to validate the accuracy of proposed system; all three classifiers were tested on primary database. Database was organized into train and test data sets using k-fold cross validation technique with k=5. Every time four sets were used for training and remaining set was used for testing. In this way; every sample image gets chance to appear in train as well as test dataset.[10] The recognition accuracy recorded for every classifiers is given in following table 2:

Digit	Total No. of samples tested	KNN		Gaussian Bayes		SVM	
		No. of samples classified correctly	Accuracy (In %)	No. of samples classified correctly	Accuracy(In %)	No. of samples classified correctly	Accuracy(In %)
0	25	24	96	24	96	24	96
1	25	24	96	23	92	24	96
2	25	25	100	25	100	25	100
3	25	25	100	24	96	24	96
4	25	24	96	24	96	25	100
5	25	23	92	23	92	25	100
6	25	23	92	23	92	24	96
7	25	24	96	24	96	24	96
8	25	25	100	25	100	25	100
9	25	25	100	25	100	25	100
	250	242	96.8	240	96	245	98

Table No. 2 shows the result obtained by implementing the three different classifiers

Conclusion

In this study, three classifiers were used to recognize Marathi digit recognition that is K- Nearest Neighbour (KNN) Classifier, Bayes classifier and SVM Classifier. From table no. 1, it is observed SVM classifier gives highest recognition accuracy i.e.98%. It is also observed that obtained results using KNN, Gaussian Bayes and SVM classifier gives highest recognition accuracy as compared to existing research result shown in Table 1.

References

1. Mimansha Agrawal, Bhanu Chauhan, Tanisha Agrawal, “Machine Learning Algorithms for Handwritten Devanagari Character Recognition: A Systematic Review”, Journal of Science and Technology, Vol. 07, Issue 01, Jan-Feb 2022

2. A.A. Jinturkar, P.B. Khanale (2017) Problems and Technology Evolution in Handwritten Numeral Recognition of MarathiJournal of Engineering Computers & Applied Sciences(JECAS) ISSN No: 2319-5606Volume 6, No.12, December 2017
3. M. Hanmandlu, O. V. R. Murthy, and V. K. Madasu. Fuzzy Model based recognition of handwritten Hindi characters, in Proc. Int. Conf. Digital Image Comput. Tech. Appl., 2007.
4. Sonika Narang, M K Jindal and Munish Kumar. Devnagari ancient documents recognition using statistical feature extraction techniques , MS received 19 July 2018; revised 14 November 2018; accepted 4 March 2019; published online 13 May 2019.
5. Khandja D., Nain N., and Panwara S.: Hybrid feature extraction algorithm for Devanagari script, ACM Trans. Asian Low Resour. Lang. Inf. Process., 2015, 15, (1), p. 2:1– 2:10
6. Arora S., Bhattacharjee D., and Nasipuri M. et al: Combining multiple feature extraction techniques for handwritten devnagari character recognition. Third int. Conf. Industrial and Information Systems, Kharagpur, India, December 2008, pp. 1– 6
7. S. Arora, D. Bhattacharjee, M. Nasipuri, L. Malik, Classification Of Gradient Change Features Using MLP for Handwritten Character Recognition, Emerging Applications of Information Technology (EAIT), Kolkata, India, 2006
8. Pankaj Kale, Arti V. Bang, and Devashree Joshi, Recognition Of Handwritten Devanagari Characters Using Machine Learning Approach, International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982, Volume-3, Issue-9, pp. 48- 51, Sept.-2015.
9. Verma G.K., Prasad S., Kumar P. (2011) Handwritten Hindi Character Recognition Using Curvelet Transform. In: Singh C., Singh Lehal G., Sengupta J., Sharma D.V., Goyal V. (eds) Information Systems for Indian Languages. ICISIL 2011. Communications in Computer and Information Science, vol 139. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-19403-0_37
10. A.A. Jinturkar, P.P. Agnihotri, S.S. Satonkar, P.B. KhanaleMarathi Handwritten Numeral Recognition Using AdaptiveNeuro-Fuzzy System, Kala Sarovar(UGC Care Group-1Journal)ISSN:0975-4520Vol-23 No.02(VI) July-September 2020