e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

COMPARATIVE ANALYSIS OF ML CLASSIFIERS FOR RECOGNITION OF HANDWRITTEN DEVANAGARI SCRIPT

Komal Gupta*1

*1Research Student Department Of Information Technology, B.K. Birla College Of Arts, Science And Commerce (Autonomous) Kalyan, Thane, India.

DOI: https://www.doi.org/10.56726/IRJMETS29901

ABSTRACT

Devanagari has been widely adopted across India and Nepal to write Sanskrit , Marathi , Hindi , Hindi dialects , Konkani , boro and various Nepalese languages. Millions of people all in India use this script for writing documents in Hindi and Marathi. Many of the Indian mythology is written in this script. Therefore, it has received a tons of popularity over the decades. In this paper, we have taken a dataset of handwritten Devanagari characters which consist of 61200 handwritten photographs of the characters. Along with the dataset we have also proposed the deep learning algorithms for recognition of those characters named as "multi-layer perceptron classifier", "Random Forest classifier" and " K-Nearest Neighbor classifier ". The sole purpose of this paper is to compare all the three classifiers on the basis of their accuracy for recognition of the images and to come up with the best one.

Keywords: Devanagari, Dataset, Handwritten Characters, Algorithms, Classifiers, Recognition.

I. INTRODUCTION

Recognition of handwritten characters has been an extremely popular research area for many years because of its different application potentials [1]. It is getting a high amount of attention because of its wide range of applications. India is a multi-script multi-lingual country and have 22 languages. Eleven Scripts are used to write these 22 languages and in these Devanagari is the most popular Script in India. Many approaches have been proposed by the researchers towards handwritten character recognition and many recognition system for isolated numerals/characters. The Devanagari Script has different characters such as complexity in shape, presence of modifiers, similarities between characters which makes it difficult to recognize.

The recognition system techniques associate a symbolic identity with the image of a character. These characters are then preprocessed and after preprocessing of characters features are extracted from them. Features which are extracted from the characters are then encoded to the structural characteristics of the character shape.

The handwritten character recognition is classified into two parts :-

1. Online character recognition:

In online character recognition the characters are written as soon as they are recognized. It gives better performance than offline character recognition as it does not have to locate the character and also have time information.

This process consist of three main steps i.e. Image segmentation , extraction of features and classification. On every step different algorithms and methods are used.

2. Offline character recognition:

Offline character recognition is further divided into following:

a. Printed characters recognition
b. Handwritten character recognition

The classification of handwritten character recognition is more difficult due to the shape of characters, greater variation of character symbol and quality of image. Because of great variation of characters he writing style also changes from person to person which in turn results in difficulty in recognition. This process consist of four main steps i.e. image Acquisition, Data preprocessing , feature extraction and classification.[2]

The recognition of handwritten Devanagari characters using convolutional neural networks and Dense neural networks for extracting features and various multiclass classifiers such as MLP () classifier , Random forest classifier and KNN() classifier for classifying characters. The used dataset consist of 36 unique Devanagari

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

characters with 1700 images of specific character is being used for training as well as testing the model. Predicting each character by the full image.

II. RELATED WORK

Comparative Study of Devnagari Handwritten Character Recognition using Different Feature and Classifiers:-

This paper contains a comparative study of 12 different classifiers for recognizing handwritten Devanagari characters and it has been concluded that mirror image learning gave best results among all the 12 classifiers and showed highest results which was 95.19% accuracy [3] .

Devanagari Handwritten Character Recognition using Hybrid Features Extraction and Feed Forward Neural Network Classifier (FFNN)

In this paper a 100 image dataset is evaluated for recognition of the characters. It proposed the final method and the framework using combination of different features like efficient FFNN and appropriate segmentation method. It concluded that FFNN have gave the best accuracy (i.e. 91.3%) than others [4].

The third most popular language in the world after Chinese and English is Indian national language Hindi (Devanagari script).

Some of the leading institutes in India doing research on Devanagari OCR are International Institute of Information Technology in Hyderabad, the Indian Statistical Institute in Kolkata, the Indian Institute of Science in Bangalore and the Indian Institute of Technology in New Delhi. [5]

III. METHODOLOGY

This paper uses CNN (convolutional neural network) and DNN(Deep neural network) for extracting features and some classifiers are proposed for classifying the image.

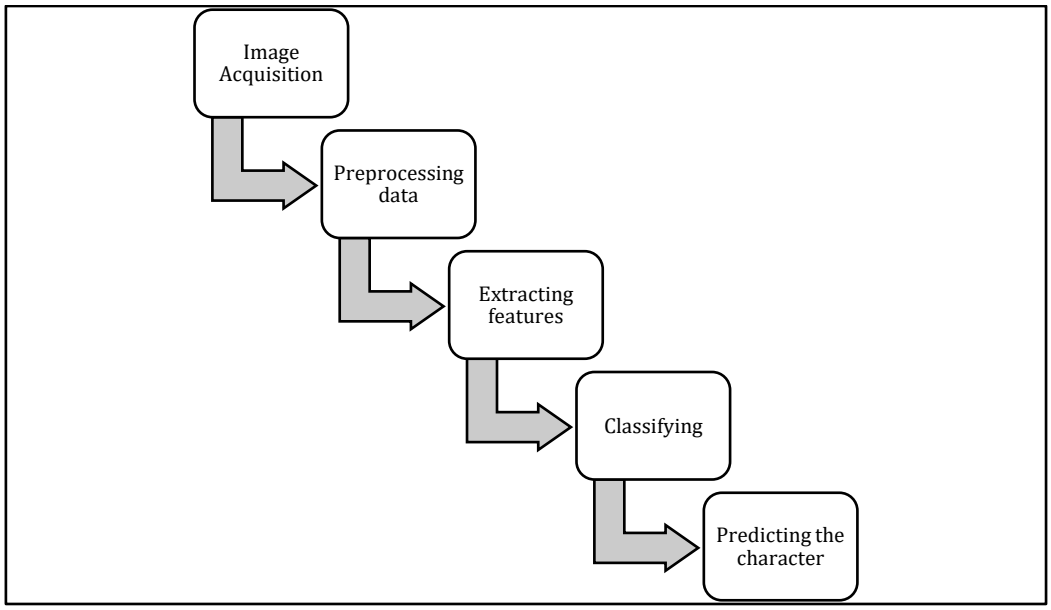

Figure 1: steps for recognition

1. Image Acquisition :-

Image Acquisition :- scanning of images

Due to a huge growth in technology now a days we have a good digital camera on our such a smaller devices such as mobile phones. The image which is clicked from the mobile phone is almost better for preprocessing of image. Some issues while capturing the image using digital camera are blurring , shadows and text-skewing. These issues can be rectified by using some image enhancement techniques.[6]

2. Preprocessing data :-

Preprocessing data :- noise removal , binarization , slant angle correction, resize

To enhance and to make it suitable preprocessing performs series of operation[7]. Preprocessing step involves Gray into binary conversion by using threshold value obtained by Otsu's method noise removal generated during document generation [8].

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

Some filters like mean filter , min-max filter, gaussian filter, etc. can be used for the removal of noise from the image. Binarization changes the colored or gray scale image to black and white image. If the document is not perfectly horizontally aligned, so it can be aligned by using slant angle correction. To improve the speed of processing and to reduce dimensions we can resize the document. But the drawback of reducing the dimensions of image is that we may remove some of the useful features too.

3. Extracting features :-

The features should be selected in such a way that it minimizes the intra-class variability and maximize the inter-class discriminability in the feature space. CNN has been the best for the feature extraction neural network. The scope has been tested and experimented by using dense neural network in the combination to CNN. Activation layer used are

Input and hidden layers :- "RELU"

Output layer :- "SIGMOID" [9]

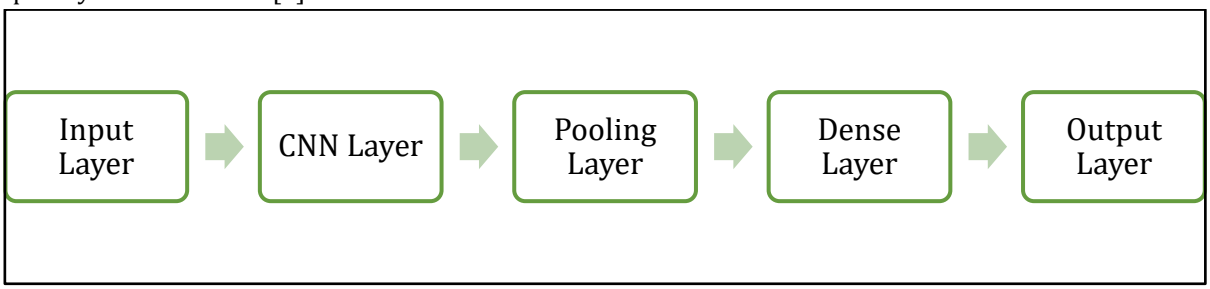

Figure 2: Architecture of neural network

```
print(cnn.summary())
```

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 30, 30, 4)         40
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 28, 28, 4)         148
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 14, 14, 4)         0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 12, 12, 4)         148
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 10, 10, 4)         148
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 5, 5, 4)           0
_________________________________________________________________
flatten_1 (Flatten)          (None, 100)               0
_________________________________________________________________
dense (Dense)                (None, 20)                2020
_________________________________________________________________
dense11 (Dense)              (None, 1024)              21504
_________________________________________________________________
dense_1 (Dense)              (None, 36)                36900
=================================================================
Total params: 60,908
Trainable params: 60,908
Non-trainable params: 0
_________________________________________________________________
None
```

Fig 3: Architecture of CNN

4. Classifying :-

The extracted features from neural network layers are then passed to different classifiers. The extracted features are taken as input to the neural network . the classifiers compare the input features with the stored pattern and finds the most matching class for input.[10]

5. Prediction :-

In this paper the model is trained on full character handwritten images of Devanagari character with high accuracy. So we can predict the output by using different classifiers.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

IV. RESULTS AND DISCUSSION

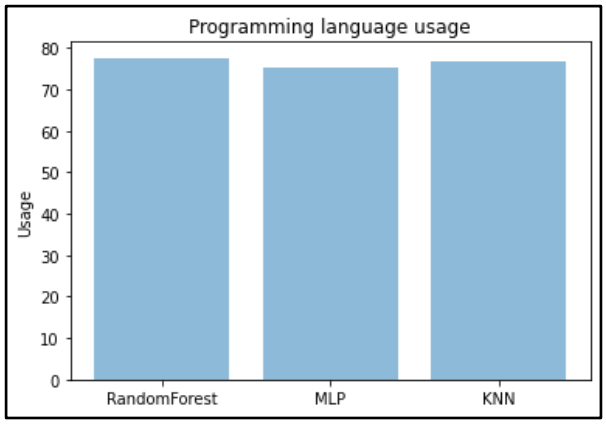

Figure 4: Bar Graph on Accuracy of classifiers

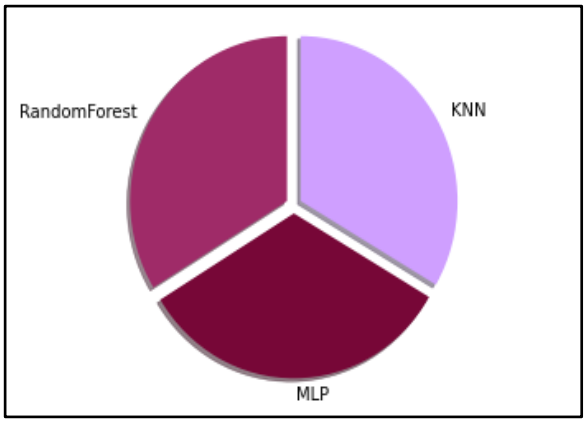

Figure 5: Pie chart on accuracy

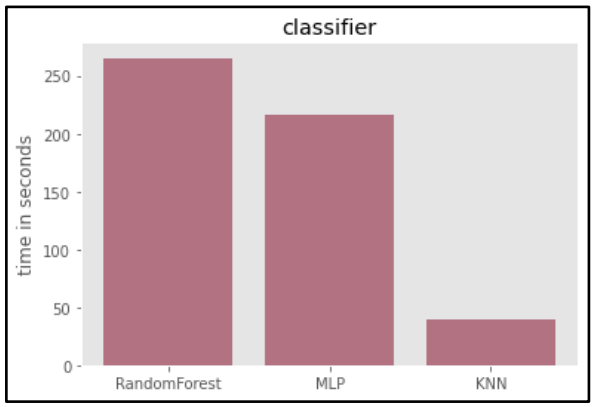

Figure 6: Accuracy based on time(in secs)

	classifier_name	accuracy
0	RandomForest	77.6
1	MLP	75.3
2	KNN	76.9

Figure 7: Accuracy

V. CONCLUSION

From the above experiment it is clearly visible that random forest classifier have shown most accuracy that the other two. If we see according to the time so it can be concluded that KNN takes lesser time to give the output so the speed is higher in KNN, whereas on the other hand random forest classifier took the maximum time to give the output. MLP classifier with 75.3% accuracy , random forest classifier with 77.6% accuracy and KNN classifier with 76.9% accuracy on the basis of these experimental results random forest has shown its best performance. And overall KNN is also good for the character recognition because it took lesser time and shown a certain good amount of accuracy.

ACKNOWLEDGEMENT

A heartfelt thankyou is given to Prof. Swapna Augustine Nikale, Department of Information Technology of B.K. Birla College of Arts, Science and Commerce (Autonomous) Kalyan, Thane.

VI. REFERENCES

[1] "PROJECT.pdf - VISVESVARAYA TECHNOLOGICAL UNIVERSITY 'Jnana Sangama', Belagavi - 590018 A Project Report on "Hindi Handwritten Characters Text | Course Hero."
https://www.coursehero.com/file/113645092/PROJECTpdf/ (accessed Sep. 12, 2022).

[2] M. Patel and S. P. Thakkar, "Handwritten Character Recognition in English: A Survey," IJARCCE, pp. 345–350, Feb. 2015, doi: 10.17148/IJARCCE.2015.4278.

[3] U. Pal, T. Wakabayashi, and F. Kimura, "Comparative Study of Devnagari Handwritten Character Recognition Using Different Feature and Classifiers," in 2009 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 1111–1115.
doi: 10.1109/ICDAR.2009.244.

[4] S. Ansari and U. Sutar, "Devanagari Handwritten Character Recognition using Hybrid Features Extraction and Feed Forward Neural Network Classifier (FFNN)," Int. J. Comput. Appl., vol. 129, no. 7, pp. 22–27, Nov. 2015, doi: 10.5120/ijca2015906859.

[5] EC Department, Gujarat Technological University, India, A. A. Malanker, and Prof. M. M Patel, "Handwritten Devanagari Script Recognition: A Survey," IOSR J. Electr. Electron. Eng., vol. 9, no. 2, pp. 80–87, 2014, doi: 10.9790/1676-09228087.

[6] K. R. Shah and D. Dattatray Badgujar, "Devnagari handwritten character recognition (DHCR) for ancient documents: A review," in 2013 IEEE CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGIES, Thuckalay, Tamil Nadu, India, Apr. 2013, pp. 656–660.
doi: 10.1109/CICT.2013.6558176.

[7] Y.-K. Chen and J.-F. Wang, "Segmentation of single- or multiple-touching handwritten numeral string using background and foreground analysis," Pattern Anal. Mach. Intell. IEEE Trans. On, vol. 22, pp. 1304–1317, Dec. 2000, doi: 10.1109/34.888715.

[8] V. L. Sahu and B. Kubde, "Offline Handwritten Character Recognition Techniques using Neural Network: A Review," 2013.

[9] D. Mangnani, J. Sukheja, H. Mohinani, and P. Kanade, "HANDWRITTEN CHARACTER RECOGNITION," vol. 9, no. 6, p. 8, 2021.

[10] Hitesh Mohapatra, "Handwritten Character Recognition (HCR) Using Neural Network," 2009,
doi: 10.13140/RG.2.2.21287.24488.

[11] Y. Zamani, Y. Souri, H. Rashidi, and S. Kasaei, "Persian handwritten digit recognition by random forest and convolutional neural networks," in 2015 9th Iranian Conference on Machine Vision and Image Processing (MVIP), Nov. 2015, pp. 37–40. doi: 10.1109/IranianMVIP.2015.7397499.

[12] M. B. Abdulrazzaq and J. N. Saeed, "A Comparison of Three Classification Algorithms for Handwritten Digit Recognition," in 2019 International Conference on Advanced Science and Engineering (ICOASE), Apr. 2019, pp. 58–63. doi: 10.1109/ICOASE.2019.8723702.

[13] "1702.00723.pdf." Accessed: Sep. 12, 2022. [Online]. Available:
https://arxiv.org/ftp/arxiv/papers/1702/1702.00723.pdf

[14] K. R. Shah and D. Dattatray Badgujar, "Devnagari handwritten character recognition (DHCR) for ancient documents: A review," in 2013 IEEE CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGIES, Thuckalay, Tamil Nadu, India, Apr. 2013, pp. 656–660. doi:
10.1109/CICT.2013.6558176.