International Journal of Recent Advances in Multidisciplinary Topics
Volume 2, Issue 6, June 2021
https://www.ijramt.com | ISSN (Online): 2582-7839

Offline Handwritten Character Recognition Using Machine Learning

Ruchika SanjayAglawe[1*], Namrata Nimje[2], Kiran Shillar[3], Pritika Shahu[4], Sakshi Dokarimare[5], N. I. Jagtap[6]

[1,2,3,4,5]*Student, Department of Computer Science and Engineering, Govindrao Wanjari College of Engineering and Technology, Nagpur, India*

[6]*Professor, Department of Computer Science and Engineering, Govindrao Wanjari College of Engineering and Technology, Nagpur, India*

***Abstract*: In this paper, a comparative analysis of recent techniques for character recognition is done. Our purpose is to identify the impact of machine learning in the domain of character identification. Character recognition has a lot of applications in the fields of banking, healthcare and other fields for searchability, storability, readability, editability, accessibility, etc. to ease up various processes. Traditional machine learning techniques like a neural network, support vector machine, random forest, etc. have been used as classification techniques. Now with the advancement in the field of computer hardware and efficient research in artificial intelligence field have given emergence to deep learning algorithms. Recent articles are using deep learning for character identification. They also depict how various functions improve the performance in the field of pattern recognition over time. The primary purpose of this paper is to encourage young researchers towards this domain and thus learn and work towards achieving novelty in the field of machine learning.**

***Keywords*: handwritten character recognition, machine learning, feature extraction, deep learning.**

1. Introduction

Handwritten recognition is a typical task because there exists a variety of writing ways. Due to the same situation, the computer program does not find good accuracy for the handwritten character recognition task. Literature focuses on English, Bangla, Marathi, Devanagari, Oriya, Chinese, and Latin and Arabic languages. Machine learning and deep learning algorithms have been widely used in past literature. At the same time, feature extraction is very crucial. Graph-based features, histograms, mathematical transforms, moment-based features are some popular techniques used for this task. Some necessary steps involved in handwritten character recognition are preprocessing, segmentation, representation, training, identification, and post-processing. As far as practical applications are concerned, a variety of mobile apps and web applications are providing character recognition features to their customers again end user wants better services that can technically be defined in terms of accuracy. Significance and challenges in character recognition are, and our purpose is to explore the solutions available in the past and explore the new possibilities to find out the resolution of the concerned problem. As discussed in the literature, one of the best ways to find the solution lies in the emerging domain of machine learning and deep learning algorithms. With this motivation, we are surveying handwritten character recognition using machine learning techniques. The contribution of this study contains a comparative analysis of various machine learning and deep learning techniques for handwritten character recognition based on various factors like dataset and technique used. The organization of the paper is as follows: Section 2 gives a complete explanation of conventional and recent techniques in machine learning and deep learning field. Section 3 involves a comparative analysis of various techniques for different languages. Section 4 contains conclusion and future work. The section below describes the techniques used for past literature.

2. Problem Statement

- The handwritten digits are not always of the same size, width, orientation and justified to margins
- They differ from writing of person to person.
- One of the biggest issues is that we used variants of the MNIST (digits) and EMNIST (alphabet characters) datasets to train our handwriting recognition model.
- We introduce the concepts and algorithms of deep learning and machine learning by these projects.

3. Objective

To avoid the problem in recognition of character we are developing model of Offline handwritten character recognition using machine learning, automated and effective work of recognizing characters and numbers1) Converting the handwritten document into digital form. 2) It can be done at any time after the document has been written3) Reading postal addresses, bank check amounts, emails and forms.

*Corresponding author: aglaweruchi@gmail.com

4. Steps for Character Recognition

A. Image Acquisition

In Image acquisition, the recognition system acquires a scanned image as an input image. The image should have a specific format such as JPEG, JPG, etc. This image is acquired through a scanner, digital camera or any other suitable digital input device.

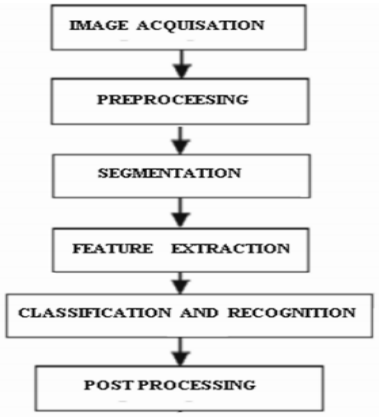

Fig. 1. Character recognition chart

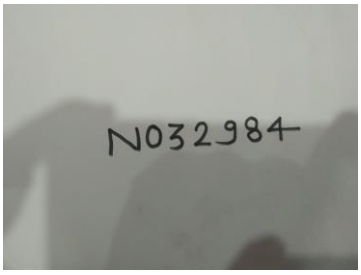

Fig. 2. Input image in JPEG format

B. Pre-Processing

The pre-processing is a series of operations performed on scanned input image. It essentially enhances the image rendering it suitable for segmentation. The role of preprocessing is to segment the interesting pattern from the background. Generally, noise filtering, smoothing and normalization should be done in this step. The preprocessing also defines a compact representation of the pattern. Binarization process converts a gray scale image into a binary image.

Fig. 3. Gray Scale Image

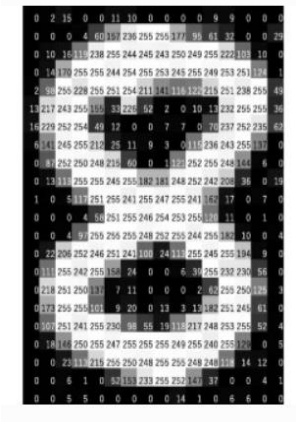

Fig. 4. Matrix form image

C. Segmentation

In the segmentation stage, an image of sequence of characters is decomposed into sub-images of individual character. In the proposed system, the pre-processed input image is segmented into isolated characters and the box can be appeared on each of them which letter can be recognized each individual character is uniformly resized into matrix.

D. Feature Extraction

In this stage, the features of the characters that are crucial for classifying them at recognition stage are extracted. This is an important stage as its effective functioning improves the recognition rate and reduces the misclassification. Diagonal feature extraction scheme for recognizing off-line handwritten characters is proposed in this work. Every character image is divided into equal zones, each of size 28*28 matrix and then further priced for recognition.

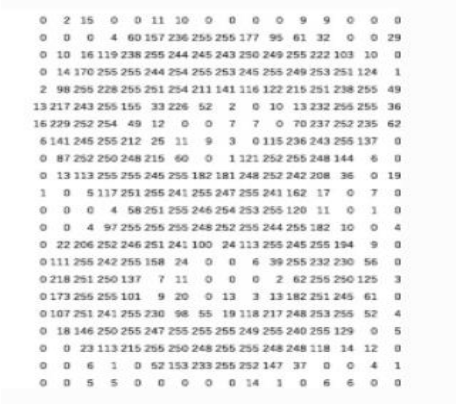

Fig. 5. 28*28 matrix of image

E. Classification and Recognition

The classification stage is the decision-making part of the recognition system. A feed forward back propagation neural network is used in this work for classifying and recognizing the handwritten characters. The matrix derived from the resized character in the segmentation stage form the input to the classifier. The neural classifier consists of two hidden layers besides an input layer and an output layer.

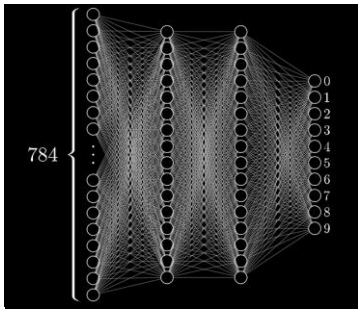

Fig. 6. Neural network image of input

F. Post- processing

Post-processing stage is the final stage of the proposed recognition system. It prints the corresponding recognized characters in the structured text form by concluding the input image using recognition index of the test samples.

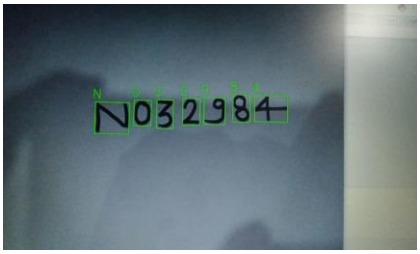

Fig. 7. Output of Image

G. Some of the outputs and their detection percentage

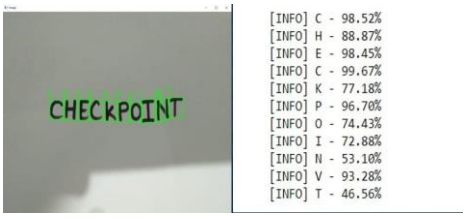

Fig. 8. Output of JPG image with percentage.

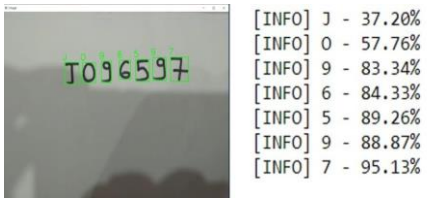

Fig. 9. Output of JPEG image with percentage.

Fig. 10. Output of PNG image with percentage.

5. Conclusion

This paper has practiced machine learning techniques including use of Tensor Flow to obtain the appropriate digit recognition. Our handwriting recognition system utilized basic computer vision and image processing algorithms (edge detection, contours, and contour filtering) to segment characters from an input image. From there, we passed each individual character through our trained handwriting recognition model to recognize each character. Our handwriting recognition model performed well, but there were some cases where results could have been improved (ideally with more training data that is representative of the handwriting we want to recognize) — the higher quality the training data, the more accurate we can make our handwriting recognition model. The error rate thus obtained is of 1.25 and training accuracy is 98% and test accuracy 97% demonstrating significant and promising performance. Thus, by practicing this we have achieved success in properly identifying the digits drawn at different angles and properly displaying the correct digit at a single turn. Hence the system would be able to recognize the introduced digit according to the formations made and according to the values in the dataset.

Acknowledgement

I would like to thank my guide Prof. Nitin Jagtap, Assistant Professor in Department of Computer Science and Engineering at GWCET (Nagpur), Maharashtra, India for motivating me to do research work on the topic Review on Offline Handwritten Character Recognition using Deep Learning Neural Network and Machine Learning. I would also like to thank my team for their continuous support, cooperation, and guidance throughout my B.E. work. Moreover, I would also like to thank my Professors who were always there at the need of the hour and provided with all the help and facilities, which I required, for my work.

References

[1] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J. et al. TensorFlow: A System for Large Scale Machine Learning. In 12thUSENIX Symposium on Operating Systems Design and Implementation (OSDI'16), vol. 16, pp. 265-283, 2016.

[2] https://www.pyimagesearch.com/2020/08/24/ocr-handwriting-recognition-with-opencv-keras-and-tensorflow/

[3] Handwritten digit recognition using state-of-the-art techniques, Cheng-Lin Liu Central Res. Lab., Hitachi Ltd., Tokyo, Japan K. Nakashima Central Res. Lab., Hitachi Ltd., Tokyo, Japan H. Sako Central Res. Lab., Hitachi Ltd., Tokyo, Japan H. Fujisawa Res. Lab., Hitachi Ltd., Tokyo, Japan, IEEE published.

[4] Offline Handwritten Digits Recognition Using Machine learning, Shengfeng Chen, Rabia Almamlook, Yuwen Gu, Proceedings on the International Conference on Industrial Engineering and Operations management Washington DC, USA, Sep pp. 27-29, 2018.

[5] Digit Recognition using TensorFlow Tool, K. V. N. Rajesh, K.V.N. Ramesh, M. Hymavathi, K. Syam Sundar Reddy, K. S. S i-manager's Journal on Pattern Recognition, vol. 4, no. 3, pp. 27-31, 2017.