



# HANDWRITTEN CHARACTER RECOGNITION

<sup>1</sup>Gaurav Surve, <sup>2</sup>Saurabh Sadhu, <sup>3</sup>Aditya SV, <sup>4</sup>Prof. Madhu Nashipudimath

Department of Information Technology  
Pillai College of Engineering, University of Mumbai, New Panvel, India

**Abstract:** In the current digital era, technology has become omnipresent in people's lives and some human tasks such as face recognition, handwritten recognition, voice recognition are now gadget tasks. Pattern recognition has a wide range of applications, one of which is handwriting recognition. Handwritten recognition allows a machine to detect the characters written in a user's handwriting automatically. It is concerned with extracting the majority of the necessary information from raw data. Deep learning is perhaps the most sophisticated machine learning approach widely available. Offline Handwritten Recognition, which works on scanned images, and Online Handwritten Recognition, which requires input at run time. Offline handwritten handwriting is the most tricky to recognize. Hence, the present idea is to implement a Real-Time Hand-Written Character Recognition System using a Convolutional Neural Network(CNN). The suggested system will extract texture features to detect handwritten letters, words, or text. The flexibility of establishing this intended character recognition system by integrating it into an existing system has been enhanced. This system will be divided into three parts: training, testing, and recognition. CNN is used as the training algorithm and then tested on the IAM words and lines dataset. The IAM dataset has a total of 96000 handwritten images of words and a total of 13353 handwritten images of lines. The proposed model classifies with 89 percent overall accuracy.

**Index Terms -** Handwritten Character Recognition, Convolutional Neural Network, Artificial Intelligence, IAM words, lines dataset.

## I. INTRODUCTION

The Recognition System is basically the identification of the character from an image. With the help of availability of dynamic information, the science of identifying a character from image data or any other data is performed. Based on its availability, recognition can be performed in two ways which are known as online or offline character recognition. The available data can be in formats such as alphabetic, numeric, punctuation. The images may be composed or in printed form and may have various styles and textual sizes. Offline handwritten recognition is a more prominent research problem and is an important aspect of pattern recognition. Handwritten character recognition, especially in the real-time part, might be considered a special type of human-computer interaction that adds significant value.

Handwritten digit recognition has a wide range of applications, including identifying postal codes on envelopes, processing large-scale financial statements, and processing bank form input. Despite the fact that there are ten fixed categories of such samples, the handwritten characters of different people or regions vary greatly, there are a few manual recognition errors in some cases, and in actual use, a relatively high recognition rate is required frequently, making it much harder to implement. It was often thought that people who used computers for business had to change their input style to match what the computer expected, whether they were typing or filling out forms with letters. Pattern recognition is employed in biometric traits that are used for personal identification. To produce the most accurate results, biometric data of each individual is collected, features are extracted, and matched with features accessible in a database. These findings could be used for a variety of purposes, including personal identification, criminal investigations, and other security concerns. The domains such as analysis of digital images, categorization of text data, recognition of characters is done using CNN.

**HANDWRITTEN RECOGNITION BY USING MACHINE LEARNING APPROACH :**

Deep Learning is the most powerful machine learning technique in the current generation. This paper presents the result of handwritten recognition using deep learning. Handwriting is unique to each individual. So the handwriting is different from one person to another person. Handwritten Recognition can be done in two ways. One is Online Handwritten recognition and another one is Offline Handwritten Recognition. Online Handwritten recognition system, which takes the input at run time and Offline Handwritten Recognition which works on scanned images.

**DESIGN AND SIMULATION OF HANDWRITTEN RECOGNITION SYSTEM :**

Handwritten recognition is a major problem since there are lots of difficulties in identifying various variations in each individual's handwriting. One way to identify is by extracting some features like curvature of letters, spacing between characters. This paper is to describe the recognition of handwritten characters for the ease of use of it by using the various machine learning algorithms. This paper involves the use of supervised and unsupervised machine learning algorithms to provide the betterment of result in the form of accuracy level. Algorithms such as Random forest, Logistic regression, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) algorithms are compared to provide its efficiency with the result comparison.

**DIAGONAL BASED FEATURE EXTRACTION FOR HANDWRITTEN ALPHABETS RECOGNITION SYSTEM USING NEURAL NETWORK :**

An off-line handwritten alphabetical character recognition system using a multilayer feed forward neural network is described in the paper. This system will be suitable for converting handwritten documents into structural text form and recognizing handwritten names.

**HANDWRITTEN CHARACTER RECOGNITION USING NEURAL NETWORK AND TENSORFLOW :**

In this paper the offline handwritten character recognition will be done using Convolutional neural network and Tensorflow. A method called SoftMax Regression is used for assigning the probabilities to handwritten characters being one of the several characters as it gives the values between 0 and 1 summing up to 1. The purpose is to develop the software with a very high accuracy rate and with minimal time and space complexity and also optimal.

**HANDWRITTEN DIGIT RECOGNITION USING DEEP LEARNING :**

Handwritten digit recognition has recently been of very interest among the researchers because of the evolution of various Machine Learning, Deep Learning and Computer Vision algorithms. In this report, the author compares the results of some of the most widely used Machine Learning Algorithms like SVM, KNN & RFC and with Deep Learning algorithms like multilayer CNN using Keras with Theano and Tensorflow.



## SUMMARY OF RELATED WORK

The summary of methods used in literature is given in the Table.

Summary of Literature survey

| Literature   | Method  | Advantage and Disadvantage   |
|--|---|--|
| Design and simulation of handwritten recognition system  | SVM(support vector machine) and KNN(K nearest neighbor) | <p><b>Advantage:</b> Using knn, the dataset pertaining image dataset is classified and the characters are recognized with highest accuracy.</p> <p><b>Constraint:</b> The accuracy level of character recognition would have been increased with enhanced techniques upon describing various machine learning algorithms.</p>  |
| Handwritten recognition by using a machine learning approach.  | CNN(Convolutional neural network) and MNIST dataset     | <p><b>Advantage :</b> Maximum accuracy 98.5% was obtained in MNIST dataset using the multilayer perceptron neural network technique.</p> <p><b>Constraint :</b> Accuracy of recognition would have been improved by improving pre-processing of data which is fed into deep convolutional neural network</p>   |
| Diagonal based feature extraction for handwritten alphabets recognition system using neural network. | diagonal based feature extraction                       | <p><b>Advantage:</b> The diagonal method of feature extraction yields the highest recognition accuracy of 97.8 % for 54 features and 98.5% for 69 features.</p> <p><b>Constraint:</b> This system will not be eminently suitable for several applications including postal/parcel address recognition, bank processing, document reading and conversion of any handwritten document into structural text form needed with better- quality recognition rates.</p> |

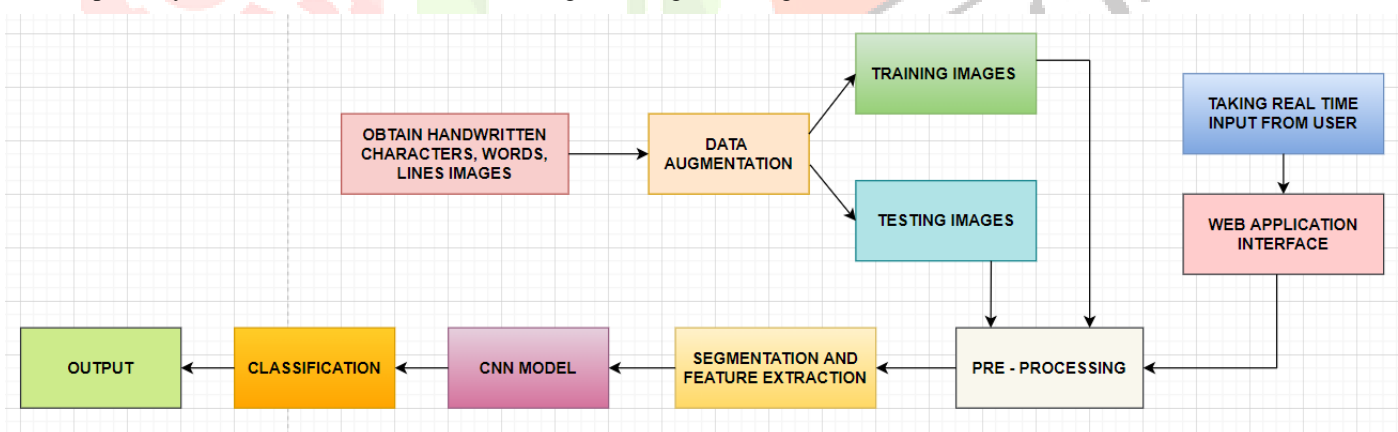
|   |  |   |
|---|--|---|
| Handwritten Character Recognition using Neural Network and Tensor Flow. | SVM, KNN, CNN                                    | <b>Advantage:</b> The feature extraction method like diagonal and direction techniques are way better in generating high accuracy results compared to many of the traditional vertical and horizontal methods. Also using a neural network with the best tried layers gives the added feature of having a higher tolerance to noise thus giving accurate results.<br><b>Constraint:</b> The bigger our training data set and better neural network design, the better accurate is the result. |
| Handwritten Digit Recognition Using Deep Learning                       | CNN Convolutional neural networks and TensorFlow | <b>Advantage:</b> Using Deep Learning techniques, a very high amount of accuracy can be achieved.<br><b>Constraint:</b> The complexity of the code and the process is a bit more as compared to normal Machine Learning algorithms.   |

### III. PROPOSED WORK

Recognition of human character is still a difficult task due to various styles and curves and also the way of writing like the inclination of the characters. So, the proposed system works to find out better results and more features. This Handwritten Character Recognition system recognizes handwritten words and lines written by the user. The proposed recognition system is described here. It consists of pre-processing, segmentation, classification and recognition, and post processing stages.

#### 3.1 SYSTEM ARCHITECTURE

The Proposed system for handwritten character recognition is given in figure.



#### 1. Image Acquisition

The input image is sent to the recognition system during the image acquisition stage. The input can be in the form of an image (JPEG, PNG, etc. ), a scanned picture, a digital camera, or any other acceptable digital input device.

#### 2. Data Augmentation

The dataset is divided into two parts where 80% is used for training and 20% is used for testing. To boost the efficiency of the model, the diversity in the data is artificially boosted using Data Augmentation. By rotating the images of the dataset to form images with different angles.

#### 3. Splitting of dataset

The dataset is splitted into two parts that are testing and training. 80% of the dataset is used for training and 20% for testing.

#### 4. Preprocessing

Pre-processing is the first step in character recognition and is crucial in determining the recognition rate. Preprocessing helps to normalise the strokes and remove any variations that could slow down the accuracy rate. The major focus of preprocessing is on

numerous distortions such as irregular text size, points lost during pen movement, jitters, left-right bend, and uneven spacing. Noise reduction, binarization, and normalizing are all part of the process.

#### 5. Segmentation

Segmentation is a technique for breaking down a large input image into individual characters. Word, line, and character segmentation are the approaches employed. It's usually done by separating single characters from a word graphic.

#### 6. Feature Extraction

Deep learning itself extracts features with deep neural networks and classifies itself. Compared to traditional algorithms its performance increases with increase in dataset.

#### 7. Classification

CNN is used to classify the image based on its features and along with the trained dataset. The Handwritten Character Recognition recognizes an image when it is given as an input, its important features are extracted and is provided as an input to the CNN classifier. It compares the featured image with the trained dataset and classifies the image with higher accuracy.

#### 8. Taking real time user input

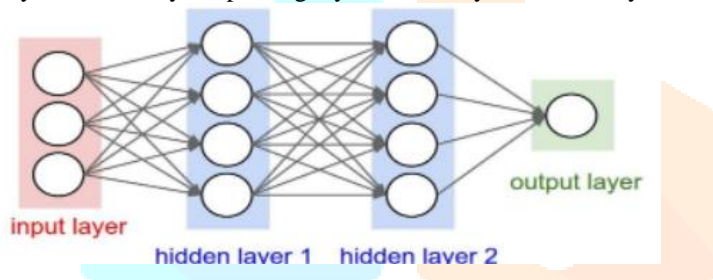
Now the model is used to recognise real time characters written by the user. Users need to write on paper and upload the scanned copy of the same to predict the text written by them.

#### 9. Output

The output specifying what was written in the input image is obtained.

### 3.2 CONVOLUTIONAL NEURAL NETWORK

An input layer, an output layer, and hidden layers are all present in CNN. The hidden layers usually consist of convolutional layers, ReLU layers, pooling layers, and fully connected layers.



- **Input** : This layer contains the image's raw pixel values.
- **Convolutional Layer**: The results of the neuron layer that is connected to the input regions are obtained by this layer. The number of filters that will be used in this layer. Each filter may be a 5x5 window that slides across the input data and outputs the pixel with the highest intensity.
- **Rectified Linear Unit Layer**: On the image data, this layer uses an element-by-element activation mechanism. Back propagation is used by a CNN. So we use the ReLU function to keep the pixels' values the same and prevent them from being modified by back propagation.
- **Pooling Layer**: This layer performs a volume down-sampling operation along the spatial dimensions (width, height).
- **Fully Connected Layer**: This layer is used to determine which score classes have the highest score according to the input character.

### IV. Requirements Analysis

#### 4.1 HARDWARE

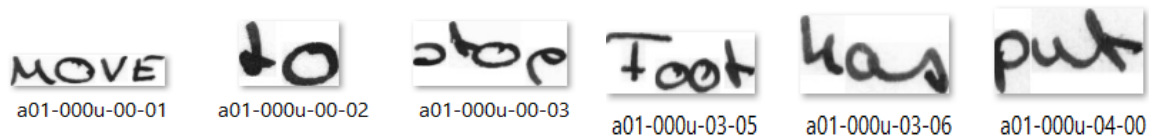
- Processor : Intel Core i5
- HDD: 1TB
- RAM: Minimum 2GB; Recommended 4GB

#### 4.2 SOFTWARE

- Operating system : Windows 10
- Dataset: IAM Dataset (Words, Lines)
- Programming Language: Python
- **Numpy** : Core package providing powerful tools to manipulate data arrays, such as our character images.
- **OpenCV** : OpenCV is a large open-source library for image processing, character recognition, and machine learning. It can scan handwritten images.
- **Autocorrect** : It is used to correct the spelling. It supports many languages.
- **Tensorflow** : Tensorflow is the core open source library to help you develop and train Machine Learning models.

### 4.3 Dataset

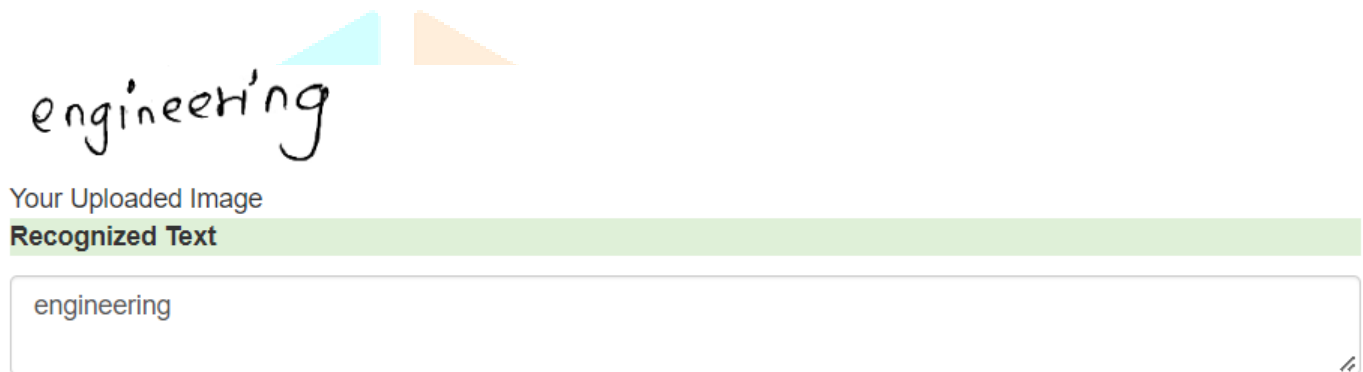
The IAM words dataset has a total 96000 handwritten images of words and its size is 800\*64.



The IAM lines dataset has 13353 images of handwritten lines and its size is 2200\*95.



## V. RESULT



## VI. CONCLUSION AND FUTURE SCOPE

Since the main aim of this paper was to develop the web application for Handwritten Character Recognition that is capable of recognizing handwritten words and sentences written by a user in real time, we can definitely say that the application that is developed is completely capable of doing said work. Using deep learning CNN algorithm, 89% of accuracy is achieved here. However, for a very high amount of accuracy or additional accuracy, reduced training and testing time, the Use of GPU is required. Using GPU's one can get much more parallelism and attain much better results.

## REFERENCES

- [1] P.Thangamariappan, Dr.J.C.Miraclin Joyce Pamila(2020). Handwritten recognition by using a machine learning approach. In 2020 International Journal of Engineering Applied Sciences and Technology. Available FTP:
- [2] D. Prabha Devi, R. Ramya, P. S. Dinesh ,C Palanisamy, G. Sathish Kumar,(2019). Design and simulation of handwritten recognition systems.
- [3] J.Pradeep , E.Srinivasan and S.Hymavathi(2011). Diagonal based feature extraction for handwritten alphabets recognition system using neural network. International Journal of Computer Science & Information Technology (IJCSIT).
- [4] Megha Agarwal, Shalika, Vinam Tomar, Priyanka Gupta (2019). Handwritten Character Recognition using Neural Network and TensorFlow. International Journal of Innovative Technology and Exploring Engineering (IJITEE).
- [5] Anuj Dutt, Aashi Dutt(july 2017). Handwritten Digit Recognition Using Deep Learning. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET).
- [6] Rohini G. Khalkar, Adarsh Singh Dikhit, Anirudh Goel, Manisha Gupta(2021).Handwritten Text Recognition using Deep Learning.International Advanced Research Journal in Science, Engineering and Technology(IARJSET).
- [7] S. Anandh Kishan, J. Clinton David, UG Scholar(2018). Handwritten Character Recognition Using CNN.International Journal of Research and Analytical Reviews (IJRAR).
- [8] Verdha Chaudhary, Mahak Bansal and Gaurav Raj(2018). Handwritten Character Recognition using Convolutional Neural Networks (CNN).International Journal of Research in Advent Technology(IJRAT).

[9] Anuj Dutt, AashiDutt(2017).Handwritten Digit Recognition Using Deep Learning.International Journal of Advanced Research in Computer Engineering & Technology (IJARCET).

[10] Megha Agarwal, Shalika, Vinam Tomar, Priyanka Gupta(2019).Handwritten Character Recognition using Neural Network and TensorFlow.International Journal of Innovative Technology and Exploring Engineering (IJITEE).

