International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 920-926 www.ijaem.net ISSN: 2395-5252

Handwritten Character Recognition Using Machine Learning Model

CH. Karthik , Dr.V.Sangeetha , CH. Satyanarayana , D. Jayanth Sri Mohan ,

G. Sai Kumar Department of CSE
Raghu Institute Of Technology

Submitted: 01-08-2021 Revised: 14-08-2021 Accepted: 17-08-2021

ABSTRACT—Handwritten character recognition is a branch of artificial intelligence that includes pattern recognition, computer vision. A computer capable of handwriting recognition can acquire and recognise characters in paper documents, photos, touch-screen devices, and other sources, and transform them into machine-encoded form. Handwriting character recognition has become a major study topic as a result of the rising use of digital technology practically in all day-to-day activities to store and pass information.

The purpose of this work is to describe the creation of a handwriting character recognition system that will be used to read students and lectures handwriting notes. Here the system is built on machine learning, when compared to other computing techniques the usage of machine learning for recognizing handwriting characters is more efficient and strong. This paper also discusses 'Handwriting Character Recognition' methodology, and implementation results of the system development.

Keywords : Artificial Intelligence, Pattern recognition, Machine Learning.

I. INTRODUCTION

Due to its practical uses in many day-to-day tasks, handwriting digits and character recognitions have become increasingly significant in today's digitized world. It can be demonstrated by the fact that several recognition systems have been created or proposed in recent years for application in various fields where high categorization efficiency is required. Computer systems are used to recognize Handwriting letters, characters and digits that can be helpful for the people to solve more complex tasks instead of being time-consuming and costly. Here we are using the concept of machine learning and it is considered as the best way to develop systems for recognizing handwritten characters and digits.

Let us consider a scenario where two persons are standing opposite to each other where the two persons are recognizing a digit that is in between them as shown in the figure below.

Here one person is recognising that the digit is six and on the other hand, the other person is thinking that it is nine. Now which person's opinion is correct, here in both of point of view their answers are correct but what is the exact answer of that digit. It is the problem occurred due to the reason that humans.

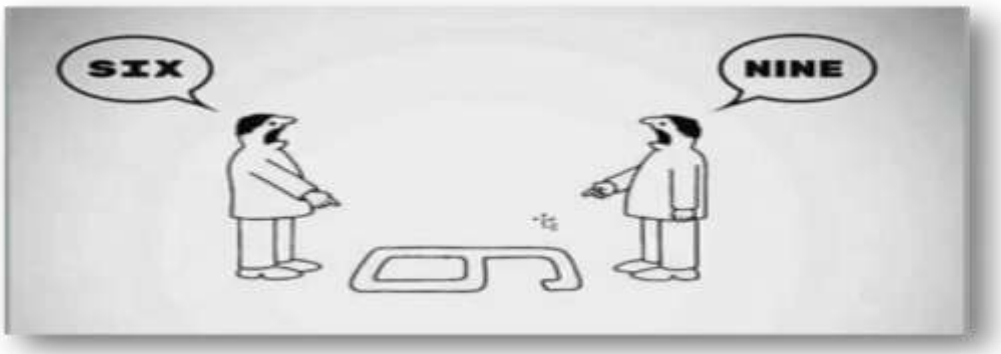

Fig. 1

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 920-926 www.ijaem.net ISSN: 2395-5252

The primary goal of this paper is to create a model that will be used to read handwriting numbers, characters, and words from the image by using the concept of image recognition. The next sections will provide an overview of the research objectives, theoretical background, architecture, methodology, implementation results, and conclusion.

II. LITERATURE REVIEW

This section discusses about the survey that has been done regarding this handwritten character recognition, there are several models have been introduced to identify and recognize the characters and digits of handwritten, though there are many drawbacks and failures in those models. As a result we are building a system that can classify and recognize the handwritten characters and digits. Sara Aqab , Muhammad Usman Tariq[1] designed handwritten recognition using artificial intelligence neural network and image processing, they trained machine learning model on the dataset with decision tree classifier and they were able to produce an accuracy of 83.4%, but they unable to recognize individual digits accurately. In [2], Boukharouba, A., & Bennia, A. they have used novel feature extraction technique for the recognition of handwritten digits, and illustrated the benefits of chain code histograms combined with transition features in digit recognition field. And they able to produce 98.46% recognition rate using 40 features, and got 98.44% when using 36 features which was slightly less than the recognition rate obtained using 46 features i.e. 98.55% and got 99.02% when 196 features were used for training. But out of a total of 20,000 digits in the testing set, 291 digits were not successfully recognized. In [3], Chandio, A. A., Leghari, M., Hakro, D., AWAN, S., & Jalbani, A. H. in 2016 had introduced a novel approach for recognizing Sindhi words in Sindhi language using Artificial Neural Network (ANN) and Self-Organizing Map (SOM) in which they have used Unsupervised learning method to train the proposed system and have collected dataset consisting of 1200 words from 60 native writes of Sindhi language. Here they have produced an accuracy rate of 83% with recognition time of 20-30 milliseconds by using 30 Sindhi language words for sampling. But here the rejection rate of recognition was also measured which was 10% to 40% due to the complexity in writing styles. In [4], Dwivedi, U., Rajput, P., Sharma, M. K., & Noida, G. in 2017 proposed Cursive Handwriting Recognition System using Feature Extraction and Artificial Neural Network where they described an offline cursive writing character recognition system in which features of each character

written in the input are extracted and then passed to the neural network datasets. Here this system efficiently recognize cursive texts and convert them into structural form. They built the neural network using 54 features and compared the recognition efficiency of proposed diagonal method of feature extraction and trained using the horizontal and vertical feature extraction methods. They had produced recognition accuracy of 97% with the diagonal method of feature extraction. In [5], El-Sawy, A., Loey, M., & El-Bakry, H. in 2017 designed Arabic Handwritten Characters Recognition Using Convolutional Neural Network where they modelled a deep learning architecture which is applied to Arabic handwritten characters. Here the CNN model had trained and tested 16800 of handwritten Arabic characters where the proposed CNN produced an average of 5.1% on misclassification error on testing data. Finally they had produced 94.9% classification rate on testing images and the miss-classification rate for training data has reached 0% on epochs 25 to 30.

There are many model have been researched on Handwritten character recognition but there are limited when it comes to models based on handwriting character and digit recognition, as a result we want to contribute our study in this field, so we have proposed a system for classification and recognition of handwritten digits and characters with machine learning model using Support Vector Machine Classifier.

III. METHODOLOGY

There are different methods and techniques which are used to ensure that the computer systems can read characters from handwriting images and documents. This paper focuses on machine learning and the Support Vector Machine.

A. Machine Learning

Machine learning technology is the inspiration from psychology and biology that focus on learning from a set of data. The central assumption of this technology is that machines can learn to perform the given tasks by learning from data[7]. A machine learning model is provided with the training data that is specific to the given problem domain and the solution to each instance of the problem. In this way, the model learns how to solve certain problems based on learning [6]. This model takes an image that has a Handwriting digit and further determines the specific digit based on the learning data.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 920-926 www.ijaem.net ISSN: 2395-5252

B. Support Vector Machine

SVM was introduced by B . E . Boser et all . in 1992 and become popular due to the success in handwritten digit recognition in 1994. It is one of the most prevailing and exciting supervised learning models with associated learning algorithms that analyze the data and recognize patterns in support vector machine. Support Vector Machine, the discriminative classifier is considered as one of the successful models that can be effective in developing handwriting recognition systems. A support vector machine is the subset of machine learning. SVM is considered as a computational algorithm that finds out a hyperplane or a line in a multidimensional space that separate classes. The separation between two or more linear classes can be achieved by the hyperplane[8,9]. This support vector machine is an approach in finding the best separation hyperplane. It is used in solving both regression and classification problems. However, it is mostly used in solving classification problems in machine learning.

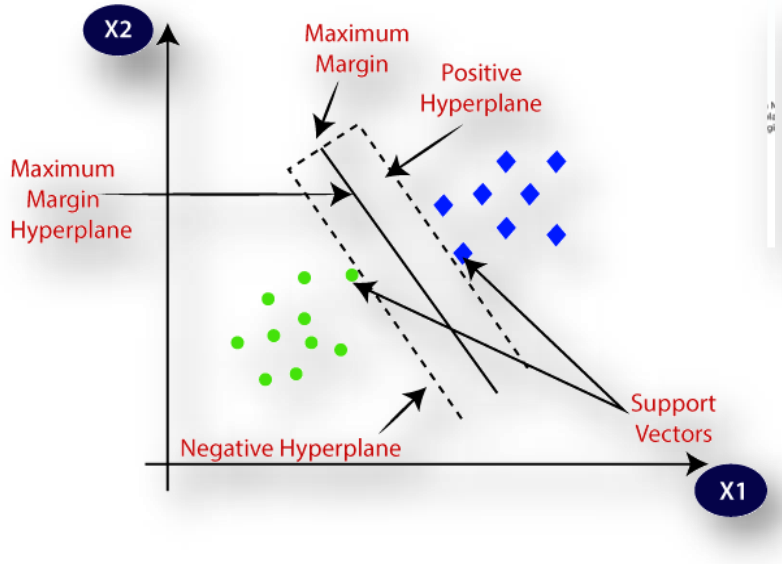

Fig.2 Support Vector Hyperplane.

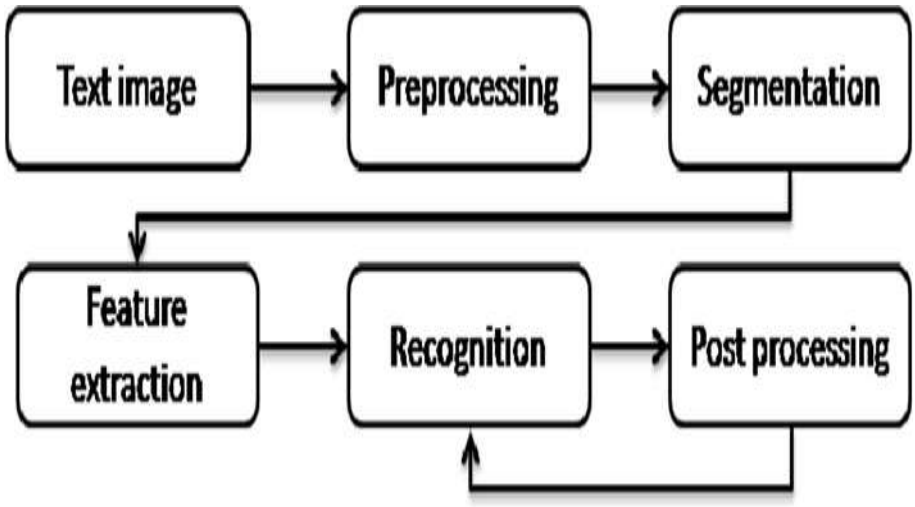

Fig.3 ARCHITECTURE OF THE PROPOSED HANDWRITTEN CHARATER RECOGNITION

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 920-926 www.ijaem.net ISSN: 2395-5252

The above figure depicts the architecture of the handwritten character recognition which include the five phases, in the next section these phases of the architecture will be discussed briefly.
The methods that are used in the handwritten character recognition are image acquisition and digitization, pre-processing, segmentation, feature extraction, and recognition.

A. Image Acquisition and Digitization

The image acquisition step involves procuring an input image or text image that contains handwriting. In this case the image should be in specific formats such as PNG or JPEG. The image is procured through a digital camera, scanner, or any other input device. On the other hand, the digitization step is used for converting the input paper into electronic format. The conversion is accomplished by scanning the original document and converting it in the form of digital image that can be saved on a computer. This digital image is required for the pre- processing phase.

B. Pre-processing

Pre-processing is the second stage of the handwritten character recognition after obtaining the digital image from the previous stage. The digitized image is pre-processed to remove noise, and then it is verified for skewing. Pre-processing is necessary for creating data that is easier to recognize using optical character recognition systems Pre-processing's major goal is to reduce the background noise, and enhance the region of interest in the image, and produce a clear difference between foreground and background.

1) **Image enhancement techniques**: To increase the image's quality and suitability by eliminating noise, enhancing contrast, blurring the image, and adding additional details. As a result, to process an image in such a way that the end result is better than the original image, as well as to provide better input for automated image processing techniques.
2) **Noise removal**: Images can be contaminated by many forms of addictive noises. As a result, noise must be removed in order to improve image quality.
3) **Binarization:** This approach is used to convert a gray scale image to black and white, effectively lowering the amount of information contained within the image from various shades of grey to a binary picture.
4) **Normalization:** This technique is used for altering the range of pixel intensity values in image processing. Its primary function is to convert an input image into a set of pixel values that are more intuitive to the human eye. The term "normalization" refers to the process of transforming photographs to a standard size.
5) **Skew correction, thinning:** This is one of the initial actions performed on scanned documents, when converting data to digital format. This method helps in obtaining a single-pixel width for quick character recognition.

C. Segmentation

Segmentation is said to be the most crucial process in character recognition techniques. Segmentation of images is usually done in testing stage. It checks for any mistake points in the incomplete image by comparing all points to the average distance between segmentation points. As a result of the technique, multiple image segments known as super pixels are created. The primary goal of segmentation is to reduce the complexity of an image's representation into something that can be easily examined. As a result, it has a beneficial effect on the script's recognition rate.

D. Feature Extraction

In this phase, The following properties are used to extract and define picture characteristics: character height, number of horizontal lines, character widths, number of circles, pixels, position of different features, and number of vertically oriented arcs, to name a few.

E. Recognition

In this phase, the characters in the image are classified and recognized using a neural network. The multiplayer perception (MLP) and Korhonen's Self-Organizing Map are the most commonly employed neural networks in optical character recognition systems.

IV. IMPLEMENTATION

This section discusses the implementation of the handwritten character recognition through the proposed system using the concept of machine learning. Here we have implemented using python programming by taking certain dataset as the input.
Problem Statement: Here we have a problem statement that is used to classify the handwritten digits. The main objective of this statement is to take the dataset of handwritten as a input and to determine what the digits or characters resided in the dataset. Here the digits in the dataset ranges from one (1) to nine (9).

Datasets:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 920-926 www.ijaem.net ISSN: 2395-5252

Here we have used Sklearn's load_digits dataset, which is a collection
of 8x8 images(64 features) of digits. Here the dataset contains a total of 1797 sample points.
The steps involved to classify the handwritten digits are:

- Firstly download the digits dataset from Sklearn's docs.
- And then pre-process the digits dataset.
- Next is to train a classifier that can categorize the handwritten digits.
- And then finally applying the model on the test set and report its accuracy.

We will be using the accuracy score to quantify the performance of the model. The accuracy will tell us that what percentage of our test data was classified correctly. From the available data, we have used 70% of the images for training the classifier and rest of the 30% is used for testing.
Given below are the figures that obtained are during the implementation of the python code :

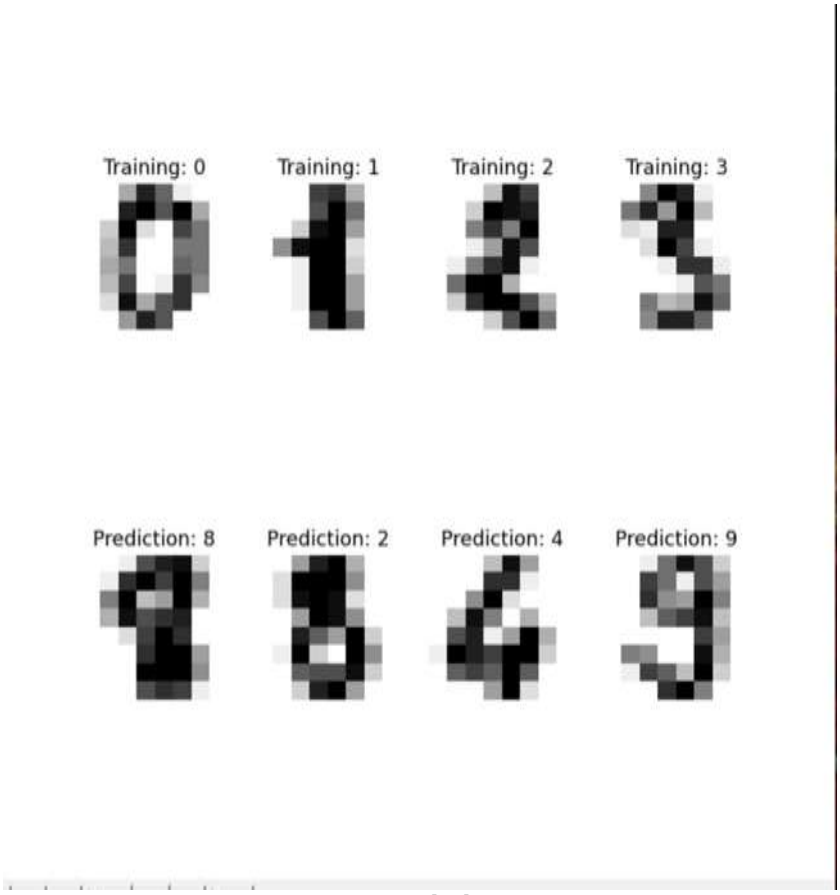

Fig.4 Sample dataset

Here from the above figure it is the sample dataset and then it visualizes the images of the handwritten digits stored in images by plotting it using matplotlib module that is imported from sklearn's library before that we have to import the numpy module. And in the output the handwritten digits are represented in the form of gray colour and the border in black colour.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 920-926 www.ijaem.net ISSN: 2395-5252

Fig. 5 Python code for character recognition

And the next step is that loading the digits dataset as a input and then importing relevant libraries and modules. Now we will prepare the data for training and reshaping the data. So finally the reduced dimension of the data will be 1797x64. In the next step it splits the data into train and test using train test split function. And finally importing SVM classifier (Proposed system) i.e. svm_classifier from sklearn library and uses predict function to make predictions of the digits, classification_report function from metrics module which gives the detailed report of every data.

V. RESULT ANALYSIS

This section describes the results observed after implementation of the python code with the proposed system. From the above implementation we can say that Support Vector Machine Classifier was the most accurate in terms of accuracy score and all other metrics in the classification report and it has the accuracy of 97%. And it also gives the best performance metrics and predictions of each digits in the dataset. By the confusion matrix we can say there are less no of non-zeroes and more no of zeroes indicates that it had predicted the accurate recognition of the characters or digits.

Classifiers	F1-Score
SVM	0.92 - 0.99
KNN	0.90 - 0.99
Stochastic Gradient Descent	0.81 - 0.97
Naïve Bayes	0.68 - 0.97
Random Forest	0.41 - 0.93
Decision Trees	0.64 - 0.91

Fig. 7 F1-Score

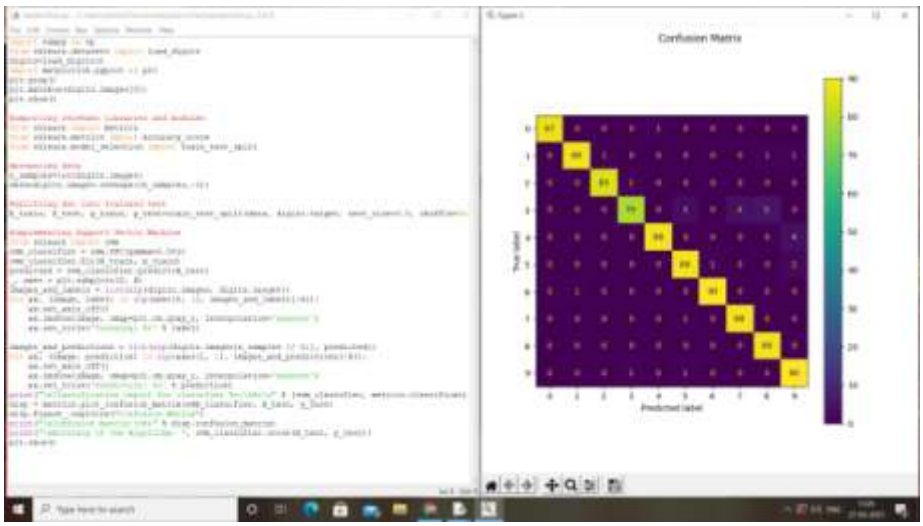

Fig. 6 Confusion Matrix

The above figure depicts the confusion matrix which is determined using plot_confusion_matrix() function.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 920-926 www.ijaem.net ISSN: 2395-5252

Classifiers	Accuracy Score
SVM	0.9688
KNN	0.9555
Stochastic Gradient Descent	0.8932
Naïve Bayes	0.8075
Random Forest	0.7532
Decision Trees	0.7352

Fig. 8 Accuracy Score

VI. CONCLUSION

The main objective of this research is to develop a system that will be useful in classification and recognition of handwriting characters and digits. So, the current system that we have proposed uses machine learning to process and classify the handwriting characters and digits. In this way, the proposed system classify and recognize different Handwriting characters and digits based on the training dataset stored in system's database. Finally, we can conclude that the final system that had proposed has satisfied the specified requirements of accuracy as well as recognition.

REFERENCES

[1] Sara Aqab, Muhammad Usman Tariq (2020). handwriting recognition using artificial intelligence neural network and Image processing. International Journal of Advanced Computer Science and Applications, Abu Dhabi School of Management, Abu Dhabi, UAE.

[2] Boukharouba, A., & Bennia, A. (2017). novel feature extraction technique for the recognition of handwritten digits. Applied Computing and Informatics, 13(1), 19-26. doi:10.1016/j.aci.2015.05.001

[3] Chandio, A. A., Leghari, M., Hakro, D., AWAN, S., & Jalbani, A. H. (2016). A novel approach for online Sindhi handwritten word recognition using neural network. sindh university research JournalSURJ (Science Series), 48(1).

[4] Dwivedi, U., Rajput, P., Sharma, M. K., & Noida, G. (2017). cursive handwriting recognition system using feature extraction and artificial neural network. Int. Res. J. Eng. Technol, 4(03), 2202-2206.

[5] El-Sawy, A., Loey, M., & El-Bakry, H. (2017). arabic handwritten characters recognition using convolutional neural network. WSEAS Transactions on Computer Research, 5, 11-19

[6] Gribaudo, M. (2013). theory and application of multi-formalism modeling. Hershey, PA: IGI Global.

[7] Lee, J. H., Shin, J., & Realff, M. J. (2018). machine learning: overview of the recent progresses and implications for the process systems engineering field. computers & chemical engineering, 114, 111-121.

[8] Abe, S. (2010). support vector machines for pattern classification. Berlin, Germany: Springer Science & Business Media.

[9] Hamid, N. A., & Sjarif N. N. A. (2017). handwritten recognition using SVM, KNN and neural network. arXiv preprint arXiv:1702.00723.