**IJFEAT**

# INTERNATIONAL JOURNAL FOR ENGINEERING APPLICATIONS AND TECHNOLOGY

## A Survey on Handwritten Character Recognition using Machine Learning

**Vaishnavi Kharate[1], Vaishnavi Tajne[2], Vaishnavi Yadao[3] , Anagha Uplenchwar[4] ,Asst. prof. Sachin Inzalkar[5]**

*IVth Year BE (CSE Department), JDIET, Yavatmal- vaishnaviytl@gmail.com*
*IVth Year BE (CSE Department), JDIET, Yavatmal- vaishnavivt7598@gmail.com*
*IVth Year BE (CSE Department), JDIET, Yavatmal- vaishnaviyadao412@gmail.com*
*IV[tth] Year BE (CSE Department), JDIET, Yavatmal- anaghauplenchwar24@gmail.com*
*Assistant Professor, JDIET, (CSE Department),  Yavatmal –sachininzalkar@gmail.com*

## Abstract

In this paper, a comparative analysis of latest techniques for character recognition is done. Our cause is to pick out the effect of machine learning inside the area of man or woman identification. Character reputation has numerous applications within the fields of banking , healthcare and different fields for searchability , storability, readability, editability, accessibility, etc. To ease up diverse processes. Traditional gadget learning techniques like a neural network, aid vector device, random forest, etc. were used as classification techniques. Now with the improvement in the field of computer hardware and efficient research in artificial intelligence field have given emergence to deep learning algorithms. Recent articles are the usage of deep gaining knowledge of for character identification. They also depict how various functions make  the performance  better in the filed of pattern recognition over time. The primary purpose of this paper is to inspire younger researchers towards this domain and consequently research and work closer to attaining novelty inside the field.

***Index Terms:*** *Machine learning, Deep Learning, Handwritten character recognition etc.*

-------------------------------------------------------------------- *** --------------------------------------------------------------------

## 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, 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 pre-processing, 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. MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Machine learning involves the process of designing a prediction algorithm based on experience. The important part is learning, and it requires data in the concerned domain after that prediction network organizes itself according to error. The current scenario has attained high complexity because the same field has attracted the attention of researchers. Various models are evolving, and some of them are as follows:

- Decision Trees
- Nearest Neighbour
- Random forest
- Artificial Neural Network
- Logistic regression
- Linear Regression
- Apriori Algorithm
- Support Vector Machine
- K-Means Clustering Algorithm
- Naive Bayes Classifier
- Neural Network

Deep Learning has attained pace due to various advancements of hardware and at the same time, algorithmic research that has been done on deep network information processing. Some of the essential algorithms of deep learning are:

- Recurrent Neural Network
- Autoencoder
- Restricted Boltzmann Machine
- Convolutional Neural Network
- Deep Belief Network
- Deep Neural Network
- Deep Extreme Learning Machine
- Localized Deep Extreme Learning Machine

## 3. CHARACTER RECOGNITION SYSYTEM

There is a range of challenges within the written character recognition system. method of the written recognition system is shown in Fig 1. There area unit 2 classes in character recognition: on-line and offline character recognition. on-line character recognition involves a digital

pen and pill. Offline recognition includes written and printed characters. written characters have loads of varieties. Segmentation and while not segmentation is concerned for written words. additional steps involve feature choice. Optimization are often accustomed speed up the method of classification. later on, there is a demand of a

classification formula for reading options. Finally, a trained model is employed for desired tasks.
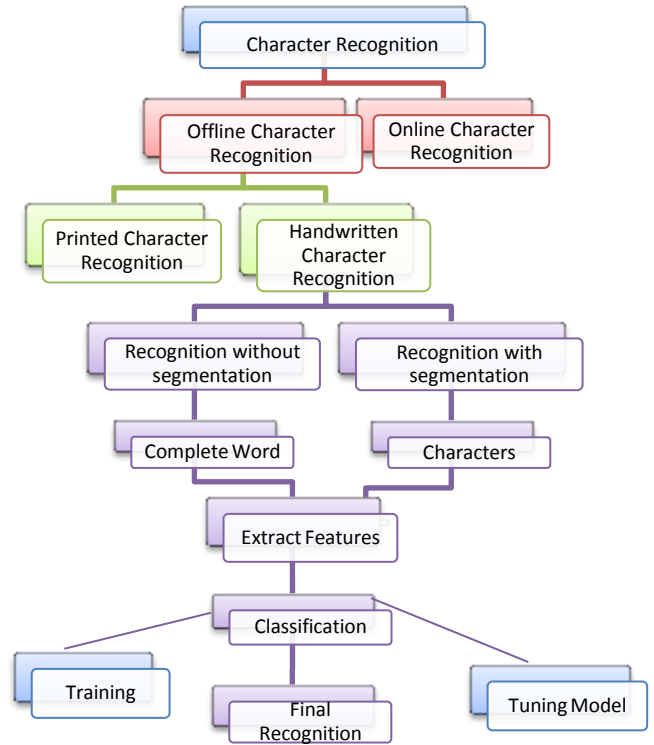


**Fig-1: General Steps of Character Recognition**

## 4. DATASET

Appropriate written character recognition system desire suitable info that contains completely different handwriting. In this , we have got thought-about four languages: Arabic, Devanagari, Chinese and Bengali. In analysis articles, authors have taken various datasets from on-line sources similarly as self-reproach data sets. IFN/ENIT contains 946 Tunisian villages/towns names &amp; communication codes achieved by 411 individuals. ICDAR twenty09 and ICDAR comprise 20,575. Arabic words from one hundred sixty five completely different writers. Semitic technology centre generated the massive knowledge set, that contains one,000 writers, 5,000 pages, 175,000 words, and approx. one million characters. Arabic info contains 154 paragraphs 194000 characters from forty eight writers. Different databases are set for scripts Bengali and Roman.

It contains the mixed text of one hundred fifty pages. The info is employed for Nagari script and Roman script. It includes mixed script of one hundred fifty pages. And alternative is for Bengali and contains 18931words. Another is employed for Nagari script and contains 15528 words .The set for Roman includes 103331 words. It include one.2 million handwritten characters, that involves Japanese, Chinese, Latin, and numeric characters. It include 2956 Chinese,71 samples collected from 4000 individuals.

The info include 783 categories, and alternative info have 1309 samples, taken from the time period state of affairs .Other holds 2 databases, info A incorporate480 writers and info B include one hundred writers. The complete info involve 3214-character category of 2965, 10 numerals, eighty two hiragana, 157 characters consisting of English, Katakana, and symbols. HCI2000 accommodate 3755 Chinese characters, written by a thousand individuals. ITRI contains 5401 Chinese character categories, and every category has 1000 samples. 4MSL comprise 4060 Chinese characters. The main conclusions of the study could also be bestowed in a very Short Conclusion Section. during this section, the author(s) ought to also concisely discuss the constraints of the analysis and Future Scope for improvement.

## 5. CHALLENGES IN AUTOMATIC WRITTEN

## DIGIT RECOGNITION

1. Challenges in written character recognition

Solutions of written character recognition have several limitations.

a.) Error Rate: - As shown within the literature , numerous algorithms are designed to unravel the issue of written character recognition, but exact detection continues to be a difficult question. conjointly interpret identical state of affairs of Bengali character .

b.) Detection Speed: - Advancement algorithms and deep networks take time in coaching therefore to develop multiple images, detection time mechanically expands.

c.) extensile Detectors: - Development of extensile detection algorithms which will sight the increasing data properly may be a burning concern of written character recognition. Poor Quality, Poor Inking, and Obsolete Fonts: - As written within the heading, these factors regulate the speed of detection accuracy. correct knowledge set and its preparation is additionally a central issue.

## 6. EXPERIMENTAL ANALYSIS

We experimented on various state of the art and other standard methods for Handwritten Digit Recognition. The performance of the methods, namely the Auto Encoders and Dense Net models, were recorded on various changing parameters. The best performing activation functions were applied to the network, including Google's new SWISH activation function and ELISH activation function. Table5: Observation for MNIST Database. Activation Function Accuracy Auto encoder Relu 0.9953599962234497 Swish 0.9956799955368042 E-swish 0.9956899953842163 Elish 0.9955199964523316 Selu 0.9951099985122681 Activation Function Accuracy Dense Net rely 0.982619 Swish 0980000 E-swish 0.982143 Elis 0.982247 sell 0.969524

| Activation Function | Accuracy Auto encoder |
|---|---|
|  |  |
| Relu | 0.9953599962234497 |
| Swish | 0.9956799955368042 |
| E-swish | 0.9956899953842163 |
| Elish | 0.9955199964523316 |
| Selu | 0.9951099985122681 |
|  |  |
| Activation Function | Accuracy Dense Net |
| Relu | 0.982619 |
| Swish | 0980000 |
| E-swish | 0.982143 |
| Elish | 0.982247 |
| Selu | 0.969524 |

**Table-1: Observation for MNIST Database**

As can be seen from the table that the best performing function on average is the E-Swish Activation Function. The results also describe the accuracy of the techniques in recognizing the Handwritten Characters. Conclusions of the study may be presented in a short Conclusion Section. In this section, the author(s) should also briefly discuss the limitations of the research and Future Scope for improvement

## 7. CONCLUSION

Handwritten character recognition is a complex problem because of a variety of character in different languages. The complex architecture of characters is another major reason that makes the handwritten character recognition task stuff. Research in this direction focuses on segmentation procedures, feature extraction procedure, and classification algorithms. Various machine learning techniques have been used for solving the same problem. Now with the advancements in hardware and the efficient algorithm has given birth to deep learning, and it is widely used for solving handwritten character recognition. In this paper, we presented a survey on handwritten character recognition. Initially, we presented a procedure of handwritten character recognition. Four languages, Devanagari, Bangla, Chinese, and Arabic, are taken for analysis. We presented a study in tabular form that reflects the various techniques. Used & accuracy attained in the handwritten character recognition task. Challenges in the concerned domain are also discussed.

The wide use of handwritten character recognition for commercial products like mobile phones, PC, etc. attracts the attention of the research community towards this problem. As stated above that deep learning is catching attention the modified version of deep learning algorithms like Discriminative Restricted Boltzmann Machines (DRBM) , Conditional restricted Boltzmann machines (CRBM) , CBIR (Content-based image retrieval), CDBNs (Convolutional deep belief network) , Separable deep encoder , Recursive Convolutional network (RCN) , Convolutional restricted Boltzmann machine (CRBM) , Dense convolutional neural network  etc. have been developed in past literature. Analysis and exploration of these algorithms, along with advance feature extraction algorithms  will be used in the future. I hope that this intuition will be helpful for those who are working in this direction.

## REFERENCES

[1]. Bhowmik, Tapan Kumar, Ujjwal Bhattacharya, and Swapan K. Parui. "Recognition of Bangla handwritten characters using an MLP classifier based on stroke features." International Conference on Neural Information Processing. Springer, Berlin, Heidelberg, 2004.

[2]. Basu, Subhadip, et al. "A hierarchical approach to recognition of handwritten Bangla characters." Pattern Recognition 42.7 (2009): 1467-1484.

[3]. Das, Nibaran, et al. "Handwritten Bangla character recognition using a soft computing paradigm embedded in two pass approach." Pattern Recognition 48.6 (2015): 2054-2071.

[4]. Pal, Umapada, et al. "Multi-oriented Bangla and Devnagari text recognition." Pattern Recognition 43.12 (2010): 4124-4136 etc.