HANDWRITTEN CHARACTER RECOGNITION USING NEURAL NETWORKS

Dr. L. SURIYA KALA

Assistant Professor, Department of Information Technology, Saradha Gangadharan College, Pudhucherry ,India

Abstract: OCR has been an interesting topic for many years now. The main use of this OCR is to make documents which can be editable from existing paper documents (or) even image files. To develop this OCR we are need of approaches . To recognize the character image processing algorithms are needed. In this paper mainly focusing about MLOCR, dataset, Prediction accuraies in Neural Netwok.

Keywords: Optical Character Recognition, CNN, Pattern recognition, Handwriting Recognition, Document Image, MLOCR.

1. INTRODUCTION

With this approach there are numerous systems are available for the character recognition of impression Devanagari, Bengali, Tamil, China and so on. With the representation of symbols we can communicate with humans. With the development of the deep learning techniques convolution Neural Network has been applied in handwritten character recognition. According to data acquisition handwriting recognition can be divided into online and offline recognition. There are various methods to improve the accuracy. As for the classification techniques most famous approaches are convolutional neural network (CNN) based approaches. The purpose of this topic is to take any kind of characters as an input and process the given character train the Neural network algorithm to recognize the pattern to modify these characters to get as the output. .Pattern recognition is most common purpose of this neural network. Ocr technology is being implemented in many software products nowadays to enhance the user experience such a google Translate. It provides a fast and efficient way to communicate between two different cultures and languages irrespective of the entities involved. By Utilising the concept OCR we can create number or application products which will help us to expand our reach to acquire the knowledge among the different countries and cultures.

2. LITERATURE SURVEY

[1] Kolamunna K.G.T.D , Fernando N.I.R (JULY 2021) proposed on the deep learning based offline handwritten Chinese character recognition area to understand the gaps that should need to develop in future. Discussed about the convolution neural network mainly focused about some new techniques to improve the related area and the gaps should need to improve in future.
Merits:
Convolutional Neutral Network is the most popular classification technique that uses in this pattern recognition area.
Demerits:
To identify more new techniques to do the offline recognition..

[2] Hakan Koyuncu (JUNE 2021) The effects of noise on character images are displayed during character recognition related to Neural network properties. Neural networks are commonly used to recognize patterns among optical characters. Feed Forward neural networks are deployed in this study. Improved accuracy is observed with character recognition during the predicition phase of the neural networks. It was concluded that canny was better than Sobel in terms of numeric character recognition.
Merits:
• It identifying the best preprocessing methods for input images to enhance them in determining the optimum number of hidden layers in the Neural network.
Demerits:
• Training the neural network with noisy images lead to worse performance compared to other applications.

[3] Sasmita Kumari Pradhan, J,Thirupathi, N V Krishna rao, Jahnavi Aluru, Keerthana Suraparaju, Paresh Ranith T (May 2021) It is the process of extracting text from an image. The use of this OCR is to make documents which can be editable from existing paper documents or even image files. Character recognition also works on sentence detection which helps in preserving a document structure. Important area of research is multi-lingual character recognition .

Merits:
- OCR used in many practical applications in real-time applications.
- Increase the speed of the process of character recognition in the document processing.

Demerits:
- OCR recognition of text characters for language such as Arabic and even for urdu are still remains an open challenge for many.
- OCR is facing difficult problem in variety of languages

[4]. Diandian Zhang ,Yan Liu ,zhuowei wang and Depei Wang (June 2021)
Deep CNN model is used for text recognition , using a sliding window instead of manual segmentation. Discussed about the dataset and different quanties of sample data and with the latest model.

Merits
- The accuracy rate of text recognition is higher in deep CNN Model.

Demerits
- Recognition errors often occur when the Manchu source text is handwritten.

3.1 EXISTING METHODOLOGY

High pass and Laplacian filtering are then applied as test inputs to neural network to get the best prediction accuracies. Availibilty of huge image data sets but there are limitations in each approach. Some of the existing accuracy is not upto the mark they are lacking in the handwritten character recognition. Here they are comparing almost all the letters with the existing database images comparision and identify the letters.

4. PROPOSED METHODOLOGY

4. 1 The idea behind an OCR is to identify and analyze a document image by dividing the page into line elements, further sub-dividing into words, and then into characters. These characters are compared with image patterns to predict the probable characters. Recognition of characters can be done either from printed documents or from handwritten documents. The highest target of any OCR system is to pretend the reading ability of the human being, so that the machine (computer) can read, understand, edit and perform the exact work it does with the provided text .
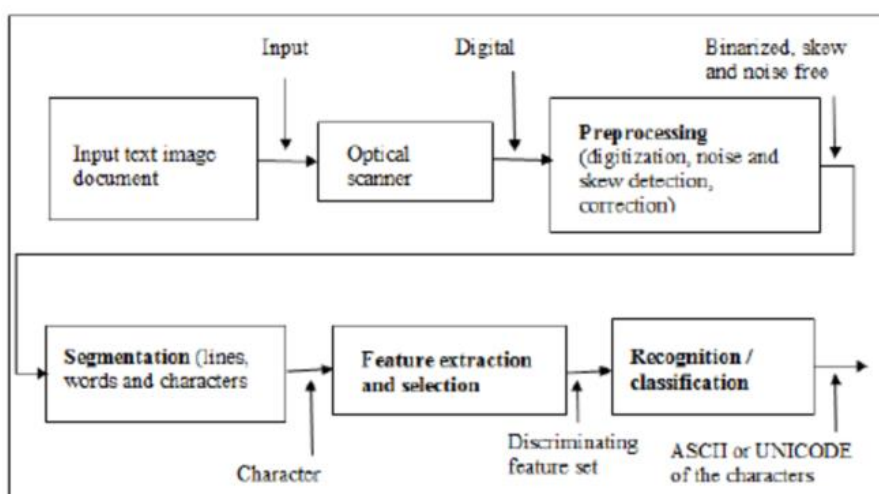
4.2. MultiLingual Indian Regional Languages
The system MLOCR is an effective way to convert the document images into an editiable text. Here the scanned image is taken and preprocessed to be converted by the OCR engine. To convert the scanned image into text by using the same concept the program is development for the most common regional languages of India. Languages used in MLOCR are hindi, Telgu, Kannada, Malayam, Tamil, English. In all these languages Images can be converted to text at a time but

only one language can be processed. Here everything is a feed forward manner this one after another each process will begin on the output of the earlier one. In every office and those who needs to prepare (or) convert a Hard copy of document into an editiable soft copy can use this MLOCR. Here we are comparing multi language that are present in india. We are not going to do any supervision here. Algorithm will itself group everything together based on the similar characteristics and it will define everything is in right or wrong. We are taking multi language in order to identify the letters neural network concepts as been used.

4.3   Neural Networks
We are facing some difficulties in the existing neural network for each and every languages the characters and the letters in each languages are different to one another. Characters and letters in each languages are different. In order to achieve the better performance or better results we can use this methods In order to achieve the best performance and better result In the present work we use the deep CNN model. Deep learning plays an important role and improves the accuracy . It Contains five types of layers input convoluting, pooling, fully connected and output. Each layer has a specific purpose like summarizing connecting (or) activating. Convolutional neural networks have popularized image classification and object detection CNN has also been applied to other areas such as neural language processing and forecasting.

4.4 In order to know the complete understanding of the each letter and unique characteristics for each languages we should use the neural networks in order to identify the same from the given input two approaches are used one is the Supervised and Unsupervised

4.4.1   Supervised
      In Supervised learning problem can be solved easily since the data is provided with the labeled dataset. In Supervised learning we train an neural network or Artifical Intelligent based machine Learning based we are providing the characteristics in order to group the input once the deep learning identifies the given input to certain character or certain letters we have to manual approve whether it is right or wrong actual processing takes and evaluated it group everything and forms and fool prove methodology to identify by the given characters.

4.4.2 Unsupervised
   Among these learning methods the Unsupervised learning method is the best approaches. Unsupervised learning refers to the use of artificial intelligence (AI) algorithms to identify patterns in data sets containing data points. In few languages we may not have the existing datasets. So have to create our own datasets. So the best approaches to find the character recognition is the unsupervised learning methods.
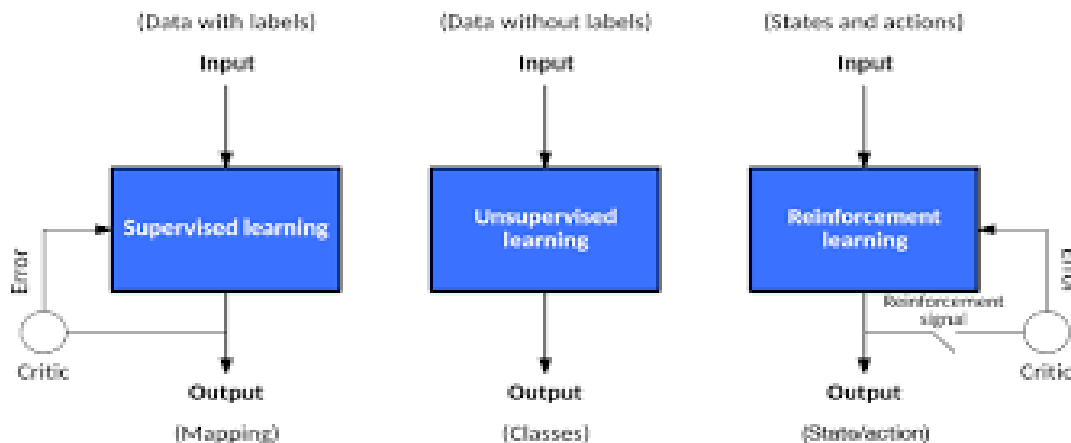

Fig 2. Supervised and UnSupervised and Reinforcement Learning

5. RECURRENT NEURAL NETWORKS (RNNS)

Here all the Inputs to a recurrent neural networks are not independent of each other, output for each element depends on the computations of its proceeding elements.
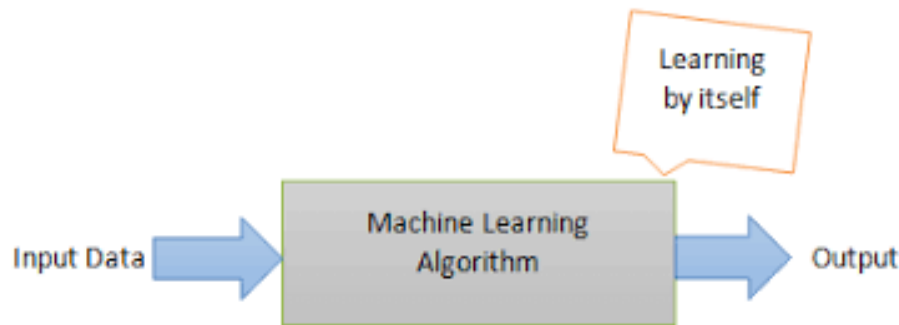
Fig 3. Supervised and Un Supervised Learning

CONCLUSION

Here the unsupervised learning is the best approaches that helps to understand the character recognition. For some languages we have thed existing datasets to identify the character recognition and for other languages the new datasets has to be created to identify the characters.

FUTURE WORK

Creating an Unsupervised neural network which will categorise and subdivide the given inputs and create data module and data set for the further classification of any given language irrespective of its origin. By utilizing both supervised and unsupervised concepts of neural network we can achieve better character identification. By combining different data sets and data models we can train a neural network to identify combined language inputs such as a document which contains more than one language can be identified and classified and translated into any language. This is the ultimate goal we need to achieve in order to communicate between two different entities irrespective of the entities involved.

REFERENCES:

1. Kolamunna K.G.T.D , Fernando N.I.R (JULY 2021) Review paper: Deep Leaning based Offline Handwritten Chinese Character Recognition.
2. Hakan Koyuncu (June 2021) A Comparative Study of handwritten Character Recognition by using Image Processing and Neural Network Techniques. Hittie Journal of Science and Engineering 2021, 8(2) 133-140 ISSN 2148-4171 DO10.17350/HJSEI9030000223.
3. Sasmita Kumari Pradhan, J,Thirupathi, N V Krishna rao, Jahnavi Aluru, Keerthana Suraparaju, Paresh Ranith T (May 2021) Manual Character Recognition with OCR Turkish Journal of Physiotherapy and Rehabilitation; 32(3) **ISSN** 2651-4451 | **e-ISSN** 2651-446X
4. Bapu Chendage1, Rajivkumar Mente2 June 2020 A Study On Digital Image Processing Techniques for Character Recognition **I** Diandian International Journal of Computer Engineering and Applications, Volume- XIV, Issue - Special Issue, June 2020, www.ijcea.com ISSN 2321-3469 NCRTCEA- 2020.
5. Muna Ahmed Awel1, Ali Imam Abidi2 Review On Optical Character Recognition International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 e-ISSN: 2395-0056.
6. Diandian Zhang ,Yan Liu ,zhuowei wang and Depei Wang (June 2021) OCR with the Deep CNN Model for Ligature Script –Based Languages
7. like Manchu, Hindawi Scientific Programming Volume 2021, Article ID 5520338, 9 pages,https://doi.org/10.1155/2021/5520338.
8. Yazan M. Alwaqfi, Mumtazimah Mohamad (2020) A Review of Arabic Optical Character Recognition Techniques and Performance.International Journal of Engineering Trends and Technology,ISSN: 2231-5381 doi: 10.14445/22315381/CATI1P208.
9. Jamshed Memon, Maira sami, Rizwan Ahmed Khan and Mueen Uddin (July 28 2020), Handwritten Optical Character Recognition(OCR): A Comprehensive SystematicLiterature Review (SLR).
10. Bhagyashree D.Upadhyay, Dr.Sachin Chaudhary, Monali Gulhane , A Review paper on Handwritten Character Recognition Using Machine Learning, International Journal of Future Generation Communication and Networking Vol 13, No 2s(2020).