

Character Segmentation of Handwritten Text Using Machine Learning

Dr Chetanpal Singh¹, Dr Rahul Thakkar²
Victorian Institute of Technology, Melbourne VIC 3000,

Abstract: Considering the importance of handwritten documents in everyday transaction technologies like optical character recognition will be a valuable addition to the new set of technologies. This technology allows the translation of different documents and images into editable, analyzable, and searchable information. Researchers have successfully integrated technologies like machine learning and artificial intelligence to automatically analyze printed or handwritten documents for converting them into electronic formats. At the time of recognizing a text, one needs to process the input image, extraction functionality, and classification schemes. This is the training stage of the system for acknowledging specific text. In this phase, the system is trained to find out similarities and differences between handwritten sample documents. Technology makes use of images of hand transcription and then transforms these images into a digital copy. The primary aim of this research report is to understand and explain the procedure for the development of character recognition systems. It will give further direction to the technology for identifying research documents. The systematic literature review includes evaluated, collected, and synthesized articles on this technology. Most of these research articles were published between 2019 and 2021. For this research report, various electronic databases were evaluated to understand the predefined review protocol. Previously published articles were searched with the help of keywords. After a comprehensive selection process around 20 articles were shortlisted on this. This review report aims at presenting the current state of results and methodologies of OCR and also provides a new dimension in the research process. It also helps in addressing the literature gaps.

Index Terms: Character Recognition, Image Processing, Recurrent Neural Network (RNN), CN (Convolutional Neural Network), Character Segmentation, Machine Learning

1. INTRODUCTION

Nothing has impacted the human race more than computers and phones. However several people still prefer traditional handwritten documents over digital alternatives. This method has been in use for hundreds of years and helped humans evolve into what they are today. Even after the growing demand and availability of smart technologies people still choose to take down notes in a traditional way using pen and paper. However, there are some drawbacks to this approach.

Most importantly, the hard copy of documents needs to be kept safe and stored properly. Searching anything on these documents and sharing them with another person is another challenge using physical documents. Handwriting recognition technology allows interpreting handwritten documents from sources like documents, screens of other devices, or other digital formats [1]. A handwriting recognition system can handle formatting and also performs character segmentation. It is also capable of recognizing the most plausible words. The translation of handwritten characters into digital formats is becoming more and more popular. Over time, the hard copies may get spoiled, but anything that has been stored in the system will remain accessible [2].

However digital files can also get lost if deleted or gets corrupted. Storing handwritten documents in digital format is gaining immense popularity.

Optical character recognition technology is a tool that can convert text into a machine-encoded form. At present, OCR technology is being used for the digitization of handwritten scripts [3]. It also facilitates the conversion of typewritten text into digital format. This has also simplified the retrieval of information from huge piles of documents. Organizations can now easily access historic data from digital files [4]. This technology is also being used in domains like academics, law, etc.

The performance of an OCR system primarily depends on its level of extraction functionality and categorization of patterns. Handwritten OCR is gaining huge popularity as a subfield of this technology. The offline system comprises of static mechanism form using scanned images in a n online system [5]. The information that is put into the system is not static and can vary as per the motion of the pen tip with specific projection angle, velocity, position, and locus point. Online systems have a more complicated layout with advanced features. It is capable of dissolving overlapping issues of input data available in offline systems.

After the texted document is fed into the system and an HD image is obtained, it starts by segmenting every character from the image for identifying the letters. In the next stage, the letters are recognized by the system, and the image is used for the detection of words. Machine Learning Algorithms are applied to go through these processes that are based on the data obtained in the training fees [6]. The machine sends the expected output in a word file format. The system can be easily trained for processing large sets of data that comprise different shapes and styles. Machine learning plays a crucial role in the segmentation of handwritten documents. It is also beneficial for companies to store necessary documents in a handwritten format. It becomes a lot faster in easier to complete with the help of such technology.

Scanning images of documents and converting them to searchable text can be done by a huge number of Optical Character Recognition software. However, this process works well only when the image is printed and not handwritten [7]. Therefore the focus of this paper is to create a solution for identifying English handwritten numbers and alphabets and save them.

The trend of using software capable of perceiving information from scanned documents has become big nowadays. However, issues occur when the scanned documents are of the printed form and not handwritten [5]. Hence the demand for saving the data of these forms in storage devices and then using them later has also grown. This idea of saving forms in computer storage and then searching and reading the available information is called Processing the Forms [3]. For paper forms, scanning and then saving them as images can be the simplest way out. However, line-by-line and word-by-word searching and the use of specific data become problematic in this scenario. The reason behind this is the dissimilarity between handwritten and printed characters which makes it impossible for computers to identify them. Hence, a capable OCR system is the ideal document processing software that is needed at the moment.

A. Research Objective

Setting up various kinds of forms in the system by creating software that will allow it is the central problem. The capability to perceive English handwritten numbers and alphabets concerning the form which is set up should be the main feature of the solution. It should also be able to save those as information.

To highlight weaknesses in the research process and eliminate potential issues affecting further research.

- A. Summarizing the existing research work in different languages with the help of a handwritten character recognition system.
- B. Identification of new technologies for character recognition and segmentation.

B. Research Motivation

It is not easy to read handwritten papers. Handwriting recognition technology is one of the most challenging research works in the field of pattern recognition and image processing. At present, it is one of the largest contributors towards the advancement of automated systems and helps in enhancing the interface between machine and man for different applications. Multiple research work has been conducted to focus on new methods and techniques for reducing the pre-processing duration to obtain higher accuracy [1]. Technologies like handwriting recognition are quite complex as it involves multiple phases of processing and segmentation. Besides, every people have a different style of writing. The system requires processing a large number of characters like small letters, capital letters, special symbols, and digits [8]. Incorporation of a large data set is necessary for training and OCR system. It is also known as an offline character recognition tool that can scan and recognize still images of handwritten characters.

2. LITERATURE REVIEW

One of the most promising fields in computer vision is character recognition. Humans can also identify handwritten transcription. Every type of language has a different pattern hence humans can identify the text with more accuracy. However, the machines are not capable of identifying hand transcription. It is particularly difficult to spot differences in the text by the system. At the time of processing a text, one needs to process the input image classification schema, and extraction of this system requires extensive training for recognizing and segregating the text [9]. During this phase, the tool is trained extensively to identify similarities and differences between different samples of handwritten text. The application is programmed to take the image of a handwritten transcription which is then transformed into a digital text.

Good segmentation abilities play a major role in obtaining high resolution. The rate of recognizing isolated letters is much higher in comparison to connected words and characters [10]. This is also influenced by the segmentation capabilities of the system. The method of separating image documents into smaller parts is termed segmentation. It helps in dividing the document image into easily understandable components. This phase is crucial for recognizing and classifying handwritten text documents. The segmentation process also happens to be the most complex process of the handwritten text recognition system [5].

The methodologies for handwritten text recognition systems involve several challenges for offline character recognition. Researchers are working towards exploring new techniques in this domain. The process of offline handwriting recognition comprises steps like text line segmentation, pre-processing, feature extraction, classification, and unit segmentation.

The character recognition mechanism of an image is also influenced by the composition of the image, preprocessing methodologies, selection techniques, and recognition model. This research report highlights methods of character recognition using different algorithms. It also promotes the use of deep learning for better recognition inaccuracy.

Process Flow of Character Recognition

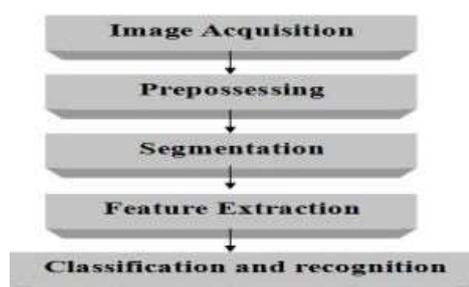


Figure 1: Proposed Flow of Process [6]

Image Acquisition

A mobile phone camera or scanner is utilized for scanning the form images. Then to eliminate the background noise and use binarization for generating image pixels within 0s and 1s, the image is provided for pre-processing.

Roll No.:	C S I 1 8 0 3 1
First Name:	S I D D H A N T
Middle Name:	M A L I K
Last Name:	R A T H O D
Date:	0 1 0 5 1 9 9 6
Age:	0 2 1
Adhar No.:	9 8 4 1 0 3 2 1 6 5 7 3
Mobile No.:	1 2 3 4 5 6 7 8 9 0
Address:	S E C T O R 1 8 . N E A R D L F N O I D A D E L H I
Pin code:	1 1 0 1 2 1
City:	D E L H I
PAN No.:	A B C 7 8 9 1 0 2 3
State:	M A H A R A S H T R A

Figure 2: Sample Form [11]

Pre-Processing

Gray-level, color, or binary document images with text and/or graphics requires Preprocessing methods.

Pre-processing steps are: Size normalization: For an image with a normal size, Bicubic interpolation is utilized.

Binarization: Threshold is used to convert a gray scale image into a binary image. This process is called Binarization. There are black elements in the foreground and white pixels in the background. 0 or 1 is the pixel value in a binary image.

Edge detection: To increase the intensity of characters' edges, a morphological gradient operator is utilized for edge detection.

Smoothing: The objects' borders are smoothed by dilation and erosion.

Opening filter- Dilation is preceded by erosion. Converting tiny objects to the background after removing them from the foreground is the main work of an opening filter.

Closing filter- An erosion process wipes out tiny holes in the foreground. It also converts tiny regions to foreground from background. An erosion process occurs after dilation in the closing filter.

Roll No.	C	S	1	1	8	0	3	1											
First Name:	S	I	D	D	H	A	N	T											
Middle Name:	M	A	L	I	K														
Last Name:	R	A	T	H	O	D													
D.O.B:	0	1	0	5	1	9	9	6											
Age:	0	2	1																
Adhar No.	9	8	4	1	0	3	2	1	6	5	7	3							
Mob No.	1	2	3	4	5	6	7	8	9	0									
Address:	S	E	C	T	O	R		.	1	8									
	N	E	A	R		D	L	F											
	N	O	I	D	A														
	D	E	L	H	I														
Pin code:	1	1	0	1	2	1													
City:	D	E	L	H	I														
PAN No.	A	B	C	7	8	9	1	0	2	3									
State:	M	A	H	A	R	A	S	H	T	R	A								

Figure 3: Pre-Processed Form [3]

Segmentation

The character segmentation issue can be managed by connected component analysis. The characters are written in not- connected, vertical histogram profile, horizontal histogram profile, and in "print fashion". The image can be divided into its subcomponents. To recognize the valid segmentation points within characters, a basic heuristic segmentation algorithm is used for scanning written text. Identifying the minima or arcs between letters is the base of this segmentation.

The explanation of the division of pre-processed images in terms of words, characters, and lines is given down.

Word Segmentation: - The gaps within words are given by Word Segmentation.

Character Recognition: Within characters, the spaces are given by them.

Line Segmentation: It is used for distinguishing text lines.

The output generated by both the algorithms is of almost the same accuracy, but we will compare the accuracy of the Support Vector Machine and neural network and train using both of these.



Figure 4: Segmented using Contour [10]

Feature Extraction

Character geometry and feature extraction based on it : - The kind of line that amalgamates to create various characters is extracted by it.

A Universe of Discourse-Smallest Matrix capable of fitting the entire character skeleton is the definition of universe of discourse.

Zoning-Windows of equal size is created by dividing the image. Then feature is performed on every window

starters- window Intersection start- the number of neighbor pixels contained in it is more than one.

Use of Gradient Features for Feature Extraction: Utilizing the Sobel operator for altering small neighbor pixel's intensity.

In chain code, decomposition of gradient image is done and the gradient vector of every pixel is achieved.

Distances and Crossings: The number of crossings done by a line segment of a contour in a distinct direction is a famous statistical characteristic. The extraction of the features of every region is done after the character frame is segregated into a set of regions in various directions.

Projections: Collection of pixel grey values on lines in various directions can be done for representing characters. Hence, from a two-dimensional image, a one-dimensional signal can be created with this representation. Representation of character image can be done using this.

Border Transition Technique (BTT)- Assumption that vertical orientation of every character takes place is the essence of Border Transition Technique. Every character is segregated into four equal quadrants. In every division, the calculation and scanning of zero-to-one transition take place. This is done in both horizontal and vertical directions.

Graph Matching Method- The utilization of a character's structural feature is what the graph matching method is all about. There are three features in this method. One pixel which contains information of position is connected to the endpoint in the initial stage. After that, more than three pixels containing feature information are connected to a branch point. This method is good for rotation or changing fonts.



Figure 5: Character Extraction [9]

Classification

The planned artificial neural network is possessed by us:

Representation and capturing of intricate input/output relationships is done by a capable data modeling tool called a neural network. The willingness to create an artificial system is the key reason behind the development of neural network technology. This artificial system is developed in such a way that it can work on a similar line with the human brain and perform "intelligent" tasks.

There are two ways by which the human brain and neural networks are similar:

1. Learning is the process through which a neural network gains information system.
2. Synaptic weights refer to interneuron connection strengths where the knowledge of a neural network is kept.

The relative positions of the fields are subjected to change in form. Therefore, the next step should be segregating the image into various blocks and then apply the neural network to every block. This is important to consider every field and train the network in them. It can be done by working on fields like name, age, date of birth, etc of an image. Considering particular fields, dividing images based on them can be problematic and hence is one of the toughest works of this project.

Support Vector Machine:

For regression analysis and classification, data analysis is done by supervised learning models with features for associated learning algorithms. These networks are called Support Vector Machines (SVM). An SVM training

algorithm creates a model for assigning new examples to categories when provided with a set of training examples where each is marked as affiliated to any one of the two categories [12]. This makes it a non-probability binary linear classifier. Points in space can be seen as examples of SVM model representation. These are mapped so that a clear and wide gap can be used to divide examples from different categories. In that same space, new examples are mapped a gain and the category based on the side of the gap is predicted.

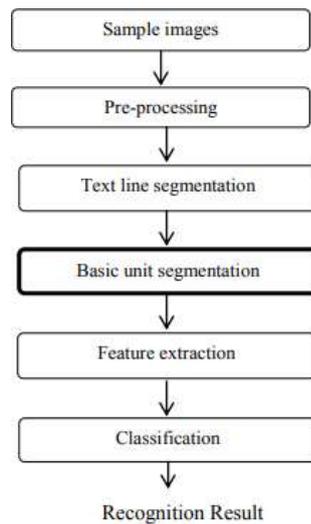


Figure 6: Offline Handwriting Text Recognition Framework [1]

As can be seen in the diagram, sample collection is the very first step of the handwriting recognition mechanism. For offline handwriting text recognition tools, the writing style of sample text does not always follow writing rules as every person has their style and characteristic of writing [1]. To make the process a little more usable the following step involves a preprocessing where denoising, binarization, and normalization technique is necessary.

Character Segmentation

One can define segmentation as an essential aspect of segmenting an image into smaller sections based on the specific condition. It is referred to as one of the most vital steps in the handwriting recognition process. Segmentation of text lines is the very first step for recognizing handwritten documents [10]. The entire text image requires text line segmentation to obtain a basic unit for each line. This step is specifically useful for understanding and evaluating of fine handwritten text. Some researchers prefer line segmentation instead of basic unit segmentation [4]. After the basic unit segmentation is complete, the structure of every text and script is analyzed. For recognizing languages like English, a word or letter serves as the basic unit. Similarly, for the identification of languages like Chinese, Chinese character is considered as the smallest unit. For the Arabic language, the smallest unit may be a connected component or word that can be easily segmented. For languages like Uyghur, the smallest possible unit can be a connected component or word derived from Arabic. Appropriate segmentation of the basic units is critical for the identification and extraction of text [13]. In this stage, the classifier selection is necessary to obtain correct results. When a document is segmented into basic units the next tasks include the development of word recognition.

For offline handwritten text recognition systems, the recognition and segmentation processes are interconnected to the basic unit [8]. Similarly, the accuracy of this system is largely dependent on the segmentation technique. Both segmentation and recognition are mutually dependent and largely influenced by the outcome.

At present, researchers are working on different models of machine learning like support vector machine (SVM), Random forest, decision tree, neural networks, and k nearest neighbor [14]. Researchers are

continuously combining these machine learning techniques along with image processing technology to enhance the accuracy of the OCR system. Currently, researchers are working to develop new methods for digitizing handwritten documents [1]. This approach is primarily based on Deep learning techniques. It has created a paradigm shift due to adopting a cluster of computing devices and GPU. It also involves Long Short-Term Memory Network (LSTM), Recurrent Neural Network (RCN), and Convolution Neural Network (CNN).

[11] preferred to use a non-linear kernel residual network to identify handwritten documents. They have suggested the use of three different network models for extending the use of the nonlinear kernel. A primarily intermediate convolution model is suitable for preprocessing images. The composite residue structure with the dropout layer facilitates parameter optimization. These reports compare different models to conclude that they offered good recognition along with data training.

To develop an Arabic handwritten character recognition system the researchers have proposed developing a mobile application with the help of Google API and cloud computing [15]. The cloud computing platform will provide a more advanced approach for handwriting recognition. I t sent the recognized text to the mobile phone for applying on the Google API.

Another report discusses the different types of segmentation classification and features. [16] use the clustering method for the segmentation of the image. The feature extraction is a comparative approach and makes use of deep learning for character recognition. It offers more accurate results for Sanskrit scripts.

The use of deep CNN is also useful for the development of more advanced forms of Meetei Mayek character recognition systems. It is found that deep learning methods are more accurate for recognizing two different characters [17].

Similarly, a convolution neural network will completely revolutionize the handwriting recognition system by offering a more accurate processing method. It will also facilitate the development of a faster CNN framework for recognizing the location of historical texts [18]. It will make use of deep learning methods to further enhance the accuracy of processing.

The research is applied to different projection techniques for the segmentation of digits and CNN [19]. For extraction of textual data from medical reports, the researchers have suggested the use of deep learning methods.

Research and evaluation network for verifying the data sets is more usable. Deep learning techniques have 97% accuracy [20]. They also suggested improvements for the CNN model for developing a more dynamic and then character recognition system.

Literature gap

The research work that was extracted for further analysis in this particular review was from various publications on the six most commonly spoken languages. It was noticed that there was a disparity between the techniques being used and the respective script as some techniques proved to be more accurate than the others [1]. For example, a technique called multilayer perceptron classifier was accurate for Bengali as well as Devanagari numerals. The result was average for the other languages in general. The type of dataset that is used for identification and the characteristics of it may be responsible for these contrasting results. The studies that are published in this particular research area suggest that there should be one technique for all languages [7]. There are quite a few datasets that are available publicly that are well-aligned. However, they do not correspond well with the scenarios that are seen in real life [18]. It is mainly concerning the thickness of characters, the style of writing and the distortion in strokes etcetera. Convolutional Neural Networks is the technique that has been most commonly used for recognizing both handwritten materials as well as characters printed in machines. The superior detection of images by CNN plays an important role for it is universally

accepted and used [19]. If there was an object that needed to be identified from an image, CNN was routinely used for it. ResNet and AlexNet are some of the important architectures that are used by CNN for visually recognizing images.

3. METHODS AND MATERIALS

The presence of inclusion as well as exclusion criteria ensures that only relevant articles are considered for the systematic review. Any article or literature of credit from journals, symposiums, workshops, and conferences has been a part of the inclusion criteria. The timeline that was included ranged from 2019 to 2021. The initial search was made based on keywords and it yielded several legitimate results which talked about the use of OCRs in different languages. The articles that were excluded from the review at this point were the ones that were not a part of handwritten OCRs but had been primarily included based on keywords. It was followed by exclusion based on duplication, unavailability of full texts and articles that did not comply with the research questions.

Both automatic and manual search strategies were employed. The automatic ones brought forward studies with a broader perspective which were eventually included in the review as well. The references of the studies that were obtained from automatic search were further screened using the manual search technique. The different standardized database was used for initiating the automatic search. These include Springer, IEEE Xplore, Scopus, Elsevier, and ISI Web of Knowledge. Although a considerable amount of literature could be gathered from other sources like newspapers, magazines, blogs etcetera, they were not included in the study as they would compromise the study quality to some extent.

4. RESULTS AND DISCUSSION

To recognize various kinds of handwritten text, a system can be developed with machine learning integrated with image processing knowledge [5]. The automated process of manual entry can get available with this implementation and this can be further enhanced with a huge amount of data sets and suitable training. The recognition of handwritten characters faced quite a few challenges. Obtaining solutions for the recognition of handwritten characters have several limitations.

- a) Speed of detection- the rate at which multiple images are processed is increased considerably. It is due to the time taken by advanced algorithms.
- b) Rate of Error- even though several algorithms mostly can be used for the identification of handwritten texts; the accuracy of these algorithms is still questionable [9].
- c) Detectors that are scalable- The data is always on the rise. Developing an algorithm that is scalable and detectable is important.

A comparative analysis of the literature has also been performed to know the outcome of the research carried out.

Comparative Analysis of Literature

Authors	Year of Publication	Approach	Results
[1]	2021	Deep learning techniques	Digitizing of handwritten documents
[4]	2018	Line Segmentation	Accurate analysis of the structure of every text and script
[5]	2019	Segmentation	For recognizing and classifying handwritten text documents
[6]	2020	Offline	Interconnection
		handwritten text	of recognition and segmentation processes

		recognition systems	
[7]	2020	Text Line Segmentation	Helps in understanding and evaluating offline handwritten text
[8]	2018	non-linear kernel residual network	good recognition along with data training
[9]	2020	Support Vector Machines (SVM)	For accurate regression analysis and classification
[10]	2019	Classifier selection	Appropriate segmentation of the basic
[11]	2020	Different models of Machine Learning	To enhance the accuracy of optical character recognition
[12]	2019	Development of mobile application with the help of Google API and cloud computing	Offering an advanced approach for handwriting recognition
[13]	2019	clustering method	more accurate results for Sanskrit scripts
[14]	2019	Deep CNN	Accurate recognition of two different characters
[15]	2020	Convolution Neural Network (CNN)	development of a faster CNN framework for increased accurate processing
[16]	2021	Segmentation of digits and CNN	Easy extraction of textual data from medical reports
[27]	2019	Deep learning techniques	97% Accuracy and improvement in CNN Model

5. CONCLUSION

The method of recognizing handwritten characters from an image that is scanned is known as deep learning. It can be used to convert any handwritten information into an electronic form. This paper concludes that a sanguine tool or method needs to be developed that can help in this conversion of characters that are handwritten to a digital format. Deep learning may be immensely useful in this endeavor as image processing takes place excessively smoothly.

The reason why handwritten character recognition has hit a roadblock is because of the large number of characters that are present in various languages. Various characters have contrasting architecture as well. The research is mainly focused on segmenting the process, extracting information from it, and using an algorithm for classification. Machine learning is a potent solution for it. Nowadays, there have been massive technological advancements as well. This ensures that handwriting recognition becomes more efficient shortly. The classifiers are not able to function optimally because a large variability is created when the characters are written by different individuals. Deep learning methods have ensured that the classification of characters has become accurate; it has also made the algorithm more complex. Thus, issues related to handwritten character recognition have still not been resolved in this context.

Work that can be done in the future- Large amounts and varieties of datasets can be taken and embedded models can be used on it to make the study even more extensive. Pen and paper will soon take a backseat as technology starts playing a vital role and most of the writing will be done using touchpads. The development of software that will detect these texts automatically and convert them to a digital format will become the need of the hour. This will simplify the searching as well as an understanding of the texts.

The common languages are used for research in OCR which is due to the easy availability of literature based on them. However, this leaves a lot of scope for future studies as well. The regional and not so commonly spoken languages can also be considered. Using endangered languages for this purpose can also play a role in cultural preservation and improvement in global synergism. The research about OCR also needs to be more commercialized. This will help in the building of OCR systems at a cost-effective rate which can be used to convert the information of valuable proportions into digital texts in no time.

6. REFERENCES

1. P. Sunitha, S. Hegde, A. V. Shreyashree, B. Prasad and Chinmayi, "Machine Learning Approach for Translating Handwritten Document to Digital Form," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 2, pp. 11-16, 2021.
2. S. Preetha, I. Afrid, K. Hebbar and S. K. Nishchay, "Machine Learning for Handwriting Recognition," International Journal of Computer (IJC), vol. 38, no. 1, pp. 93-101, 2020.
3. S. S. Mor, S. Solanki, S. Gupta, S. Dhingra, M. Jain and R. Saxena, "Handwritten Text Recognition: with Deep Learning and Android," International Journal of Engineering and Advanced Technology (IJEAT), vol. 8, no. 3S, 2019.
4. M. Kumar, S. Jindal, M. Jindal and G. Lehal, "Improved recognition results of medieval handwritten Gurmukhi manuscripts using boosting and bagging methodologies," Neural Process. Lett., vol. 50, p. 43–56, 2018.
5. A. Singh and A. S. Bist, "A Wide Scale Survey on Handwritten Character Recognition using Machine Learning," INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING, vol. 7, no. 6, pp. 124-134, 2019.
6. A. Shaikh, S. Arsalan, K. M. Altaf and S. Athar, "Handwritten Text Recognition of Document Form Using Machine Learning," International Research Journal of Engineering and Technology (IRJET), vol. 5, no. 2, pp. 2117-2120, 2018.
7. S. Aqab and M. U. Tariq, "Handwriting Recognition using Artificial Intelligence Neural Network and Image Processing," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 11, no. 7, pp. 137-146, 2020.
8. B. Rajyagor and R. Rakhliya, "Handwritten Character Recognition using Deep Learning," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 6, pp. 5815-5819, 2020.
9. S. R. Zanwar, U. B. Shinde, A. S. Narote and S. P. Narote, "A Comprehensive Survey On Soft Computing Based Optical Character Recognition Techniques," INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, vol. 8, no. 12, pp. 978- 987, 2019.
10. A. A. Ahmed Ali and M. Suresha, "Survey on Segmentation and Recognition of Handwritten Arabic Script," SN Computer Science, vol. 1, no. 192, pp. 1-31, 2020.
11. Z. Roa, C. Zeng, M. Wu, Z. Wang, N. Zhao, M. Liu and X. Wan, "Research on a handwritten character recognition algorithm based on an extended nonlinear kernel residual network," KSII Transactions on Internet and Information Systems, vol. 12, no. 1, p. 413–435, 2018.
12. M. Sharma and V. Dhaka, "Segmentation of handwritten words using structured support vector machine," Pattern Anal. Applic., vol. 23, p. 1355–1367, 2020.
13. R. Khan, A. Meyer, H. Konik and S. Bouakaz, "Saliency- based framework for facial expression recognition," Frontiers Comput. Sci., vol. 13, no. 1, p. 183–198, 2019.
14. J. MEMON, M. SAMI, R. A. KHAN and M. UDDIN, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," IEEE Access, vol. 8, pp. 142642-142668, 2020.
15. N. Shorim, T. Ghanim and A. AbdelRaouf, ". Implementing Arabic Handwritten Recognition Approach using Cloud Computing and Google APIs on a mobile application," in n Proceedings of the 14th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 2019.
16. J. Pratikshaba, "A Review Character Recognition and Digitization of Sanskrit Characters," Journal of the Gujarat Research Society, vol. 21, no. 13, p. 551–560, 2019.
17. N. D. Chingakhham, D. Das and M. D. Haobam, "A Meetei Mayek Basic Characters Recognizer Using Deep Features," Communications in Computer and Information Science, vol. 1031, pp. 306-315, 2019.
18. S. Ahlawat, A. Choudhary, A. Nayyar, S. Singh and B. Yoon, "Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN)," Sensors, vol. 20, no. 3344, pp. 1-18, 2020.
19. T. T. Zin, S. Thant, M. Z. Pwint and T. Ogino, "Handwritten Character Recognition on Android for Basic Education Using Convolutional Neural Network," Electronics, vol. 10, no. 904, pp. 1-18, 2021.
20. K. Subramani and S. Murugavalli, "Recognizing ancient characters from tamil palm leaf manuscripts using convolution based deep learning," International Journal of Recent Technology and Engineering, vol. 8, no. 3, p. 6873–6880, 2019.

INFO

Corresponding Author: *Dr Chetanpal Singh, Victorian Institute of Technology, Melbourne VIC 3000.*

How to cite this article: *Dr Chetanpal Singh and Dr Rahul Thakkar, Character Segmentation of Handwritten Text Using Machine Learning, Asian. Jour. Social. Scie. Mgmt. Tech. 2021; 3(6): 22-32.*