IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

Review for Handwritten Devanagari Character Recognition using ML Algorithms

Manisha Bhatnagar
Assistant Professor
ISBA Institute of Professional Studies, Indore, M.P., India

Abstract: *Devanagari Character Recognition is a system in which handwritten Image is recognized and converted into a digital form. Devanagari handwritten character recognition system is based on Deep learning technique, which manages the recognition of Devanagari script particularly Hindi. This recognition system mainly has five stages i.e. Pre-processing, Segmentation, Feature Extraction, Prediction and Post processing. This paper has analyzed the approach for recognition of handwritten Devanagari characters. There are various approaches to solve this. Some of the methods along with their accuracy and techniques used are discussed here. Depending upon the dataset and accuracies of each character the techniques differs.*

Keywords: *Devanagari Script; Optical Character Recognition; Segmentation; Convolutional Neural Network (CNN); Support Vector Machine; Image Processing.*

I. INTRODUCTION

Handwriting recognition has been one of the most enchanting and demanding research areas in today's digitalized world, which has evolved through the combination of artificial intelligence and machine learning. It contributes exceptionally to the advancements of the interface between humans and machines. Handwritten Character Recognition is basically ability of a system to identify human handwritten input. In general, it is classified into two types an on-line and an off-line handwriting recognition system. The handwriting can be from many sources, such as images, paper documents, or other devices; this is considered to be as offline system. Non-Indian languages, such as English, Chinese, German, Japanese, Korean, etc. are already grown-up as compared to Indian scripts. Although, Indic scripts have some more challenges in handwriting recognition than Latin, Chinese and Japanese because of the presence of variations in the order of strokes or symbols, half consonant, etc.

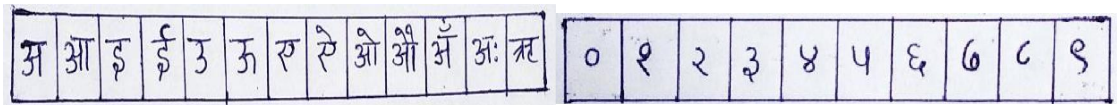
II. SOME CONCEPTS OF CHARACTER RECOGNITION

2.1 Devanagari Characters

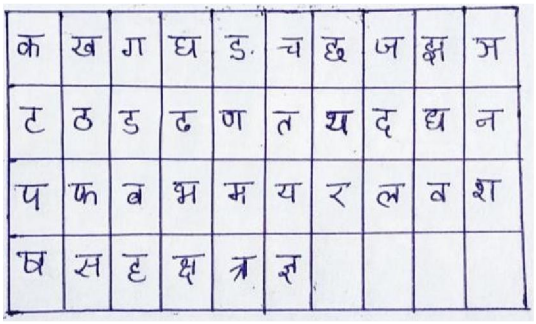
Devanagari the ancient Brahmi script, used in the Indian subcontinent and is the national font of India. Hindi is written in Devanagari script, which is also used to write Marathi, Konkani, Nepali, possibly with little modifications. Devanagari composed of 10 numeral characters (०, १, २, ३, ४, ५, ६, ७, ८, ९) and primary characters including 13 vowels and 36 consonants and is the fourth most widely adopted writing system in the world. Also, Devanagari Characters have some characters with similar structures.

For Example, the 'ङ' and 'ड' have only difference is dot. Similarly, 'ब' and 'व' are different in only the tiled line inside circle. Also, the character 'प','म ', and 'य' appear almost to be similar. The character '२' and 'र' appears almost same. Some characters like 'क्ष', 'त्र', 'ज्ञ' are derived from other previous characters like 'ग', 'य' etc. The Fig. 1 shows Hindi script consists of 13 vowels, 36 consonants, and 10 numerals.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819



(a) (b)



(c)

Fig 1. (a) Vowels of Devanagari Script (b) Numerals of Devanagari Script (c) Consonants of Devanagari Script

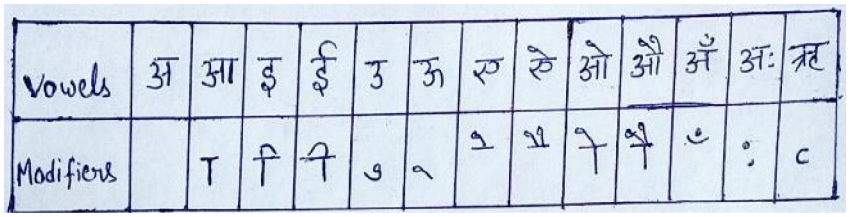


Fig 2. Vowels with Corresponding Modifiers of Devanagari Script

Vowels' modifiers have a crucial role in Hindi script. Fig.2 depicts with corresponding modifier in Devanagari script. Hindi language words can be classified as a combination of three components: a middle component (core), a top component and a bottom component. Top and bottom components consist of only swar modifiers and diacritic signs and the core component contains all, the characters, punctuation and special symbols. The top and core components are divided by a "shirorekha" (top line). A Purnaviram (full stop) is used to mark the end of a sentence or phrase which is depicted by a vertical line.

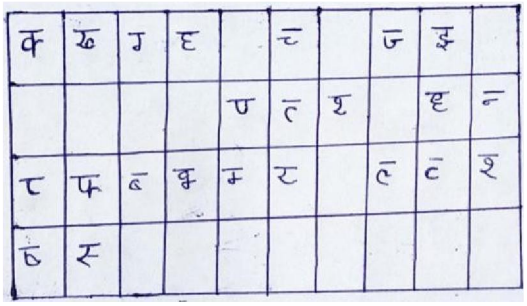


Fig 3. Half Form of Consonants of Devanagari Script

In Hindi, when two vyanjans are combined, a straight bar with a "vyanjan" can appear as a half-form. The half-form of "vyanjan" is the left side of the original "vyanjan" with a straight bar. Figure 3 shows half the "vyanjan" type of Fig.1c. The vacant positions in this figure indicate that the corresponding consonant has no half form.

2.2 Character Recognition

Character Recognition is the process of extracting digitized text from images of scanned documents. Character Recognition systems have already matured in various languages, but they still have scope in other Indian languages like Devanagari, Bengali, Marathi,etc. Character recognition could be done through various techniques, such as

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

using quadratic classifier [34], Curvelet Transform [35], Transfer Learning [36], Linear Discriminant Analysis [37], and many more.

2.3 Optical Character Recognition

Hindi OCR is a model which is basically used to recognize handwritten Hindi characters. Till today, the models developed for Indian languages like Hindi, have not shown quite good accuracy due to the complexity of the languages. The text or characters of Hindi are little difficult to segment and evaluate mainly due to their complex structures.

2.4 Working Principle of OCR

In real life applications, OCR software accepts the document image as input and produces a written text file as output. So, we are required not only to develop the recognition process, but also to develop the preprocessing and post-processing parts. The flow control of the Hindi OCR processes is shown in Figure 4. In the design of the proposed system the following steps are followed:

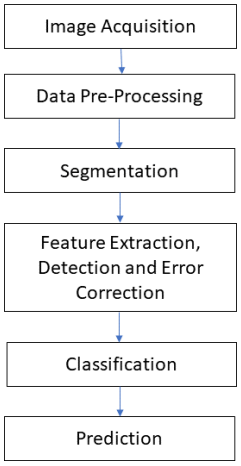


Fig 4. Flow control for Hindi OCR System

A. Image Acquisition

Image acquisition part plays the main role in OCR problem. No matter how accurate model is on testing, the real- world image will never be same. In real world image, there will be lots of noise, blur, and many other quality degradations of image.

B. Data Pre-processing

In this stage the image is converted into grayscale, and a NumPy array is prepared to store the image pixels. After this the intention is to find foreground and background colors. Removing some noise and doing threshold makes it easier for image to recognize text, and find foreground color. Here combination of threshold of Otsu and Binary is used. It is a way to create a binary image from grayscale or full-color image. This is mainly done in order to separate “object” or foreground pixels from background pixels to aid in image processing.

There are various processes in preprocessing of data such as:

1. Binarization
2. Noise Elimination
3. Skew Correction
4. Size Normalization
5. Thinning.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

D. Segmentation

Here breaking of the entire image into small fragments is done. Segmentation [38, 39] process is done here on unique way by defining the possible rows for top most part of character and the possible percentage of space between characters. Each segment individually is further passed to a prediction process.

E. Feature Extraction, Detection and Error Correction

In Feature Extraction most relevant information from the raw data is extracted and is used. The features should be selected in such a way that it reduces the intra class variability and increase the feature space. CNN has the best feature extraction Neural Network and scope. First network is FFNN (Feed Forward Neural Network) and it is the simple among the NN class.

Some feature extraction methods are:

1. Fourier Transforms
2. Wavelets
3. Moments
4. Zoning
5. Crossings and Distances
6. Projections
7. Coding
8. Graphs and Trees

F. Classification

Each segment is passed to prediction process. Before doing actual prediction, the shape of Segment must be resized as that of the neural network input. Thus, each segment is converted into 30 by 30 sized image and in addition, we added a 1-pixel borders around it with background color. Then our segments will be of 32 by 32 shape which is the input shape for our model. Then it is to the neural network. If the segment has high prediction then is assumed that the character should be shown. Prediction can be wrong also depending upon the image quality due to the false segmentation.

Character Recognition techniques can be classified as:

1. K-nearest neighbors [24,35]
2. Support Vector Machine [11,14,15]
3. Convolution Neural Networks [1,32]
4. Hybrid Network [18]
5. Artificial Neural Network [28]

These methods are mainly used but more methods could be taken under consideration.

G. Prediction

This is the overall collection of previous processes. The actual recognition is seen after there is bordered around the found character and its corresponding label as well. The final detection takes some time and gives accurate prediction. If poor image is given then it gives false prediction. So, high resolution image is recommended.

2.5 Types of Classifiers

A. Convolutional Neural Networks (CNN)

Convolutional neural networks are mainly deep neural nets primarily used to classify images, clustering them by similarity, and object recognition. CNN performs OCR to digitize text and make natural language processing possible on any sort of document like hand-written document. The efficiency of convolutional nets (ConvNets or CNN) in image recognition is one of the main reasons why people are confident in using in-depth learning.

Various layers of CNNs are:

1. Convolutional layer is the layer where convolution operation occurs that is same as image processing. A filter of same row and column or square size matrix is taken and multiplied across the window that fits filter. The element-wise product is done and then summation is done. The concept of stride is generally used, as how

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

much pixel shifts after doing one convolution. Here more the number of filters, more accuracy can be achieved but computational complexity increases.

2. Max-Pooling layer take some pixels from previous layers. Pool size is defined and then that pool size is used on input pixels. The pool matrix is moved over entire input and max value within the overlapped input is taken.
3. Dropout layer is mainly used to avoid overfitting. This layer randomly cuts the unnecessary connection between two neurons of different layers.
4. Flatten layer is one where multiple sized input is converted into 1Dvector
5. Dense layer is used to do classification after doing whole convolution process

2.6 Architecture of CNN

The main architecture of this model involves the use of Convolutional Neural Network (CNN) layers for image processing and feature extraction and detection and fully connected layers are added for recognition purpose. Fig 5 shows the general architecture of the proposed CNN model.

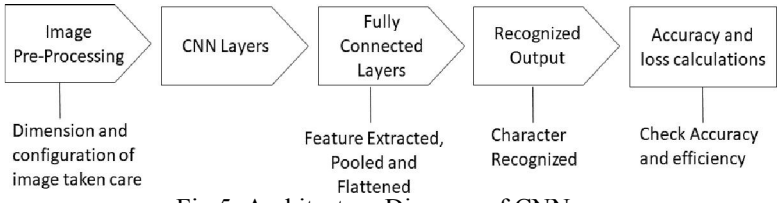


Fig 5. Architecture Diagram of CNN

A. Support Vector Machine (SVM)

Support vector machine (SVM) is considered one of the most popular and powerful supervised tool of machine learning, which can be used for both classification as well as regression. Basically, SVM takes a set of input data and gives predictions, for each given input. Given a set of training points, each marked as one of two phases, the SVM training algorithm creates a model that provides new points in one phase or another. More precisely, SVM constructs one or many hyper planes in a high or infinite dimensional space, basically used for classification, regression or other tasks. In Fig. 6 [30], a good separation is obtained by a hyper flight at a greater distance to the nearest point of any training stages because the limit is usually greater when reducing the general error of that separation.

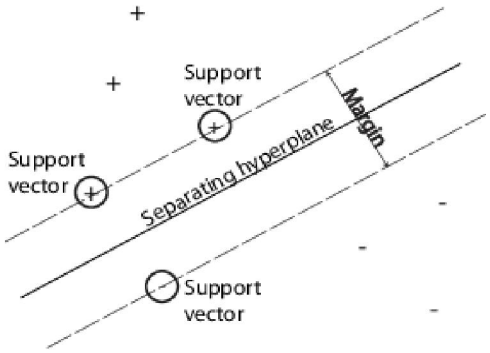


Fig 6. SVM Algorithm

B. Deep Learning

Deep learning [1, 27, 40] is a branch of machine learning in artificial intelligence (AI) that has networks capable of learning unsupervised from data that is unlabeled. Other name for Deep learning is deep neural network. Convolutional Neural Networks (CNN) along with various other Neural Networks such as ANN, RNN and other hybrid models are used on Deep Learning techniques.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

III. LITERATURE REVIEW

Acharya [1] propose a deep learning architecture mainly deep CNN for recognition of Devanagari Handwritten characters. The main focus was on the use of Dropout and dataset increment approach to improve test accuracy and hence, able to increase test accuracy by nearly 1 percent. Lakshmi [2] incorporates an effective method for recognition of isolated handwritten Devanagari numerals, in which edge directions histograms and splines along with PCA for enabling recognition accuracies is proposed. In [3], Tamil handwritten characters are recognized using Support vector machine (SVM) in which data is collected from various A4 sized documents and then preprocessed to enhance the quality of the image, finally achieved an accuracy up to 82.04%. Alternatively discussed by Hanmhunga et al. [4] s based on modified descriptive membership functions embedded in unambiguous sets found in features with standard distances obtained using the Box method. The colorful classification of Hindi characters is done using structural features such as the position of the bar, the alignment of the letter parts, and the sides open on which side etc. They tested on a database of 4750 samples and the total recognition rate was found to be 90.65%. Jawahar [5] describes the character recognition for Hindi and Telugu text. He basically used Principal Component Analysis followed by support vector classification for his bilingual OCR. The experiment was done on approximately 200000 characters, which gave an overall accuracy up to 96.7%.

The accuracy of Stroke Recognition is important for the performance of character Recognition [6]. Sekhar used SVM for the stroke recognition engine for Devanagari and Telugu scripts and applied Normalization, smoothing and interpolation for preprocessing of stroke data. He had talked about various feature extraction and stroke classification techniques such as Single recognition engine approach, Multiple Engines for Recognition and Stroke Recognition using HMMs for Devanagari Script and Telugu Script. Chaudhuri [7] experimentally did optical character recognition systems for different languages using soft computing techniques. Bansala and Sinha [8] presented a complete method for segmentation for Devanagari printed text. They have used a set of filters those are robust and used a two-level partitioning scheme and search algorithm. Various segmentation techniques were explained in his analysis. An attempt to recognize Telugu script using KNN and a compositional approach using connected components and Fringe distance is reported in [13]. Pankaj kale et. al.[28] proposed an ANN based recognition system for handwritten Marathi characters and experiments were applied on 50 handwritten characters from 10 different peoples. The data was preprocessed, and features were extracted. The accuracy obtained is 92%.

Sinha and Mahabala [29] attempt to recognize Devanagari automatically according to syntactic pattern analysis system. They choose 26 symbols and extract the structural information from the characters in terms of primitives and their relationships. However, their study is limited by the sample size, no attempt has been made to make it a commercial product and it couldn't achieve any quantitative recognition rate. Najwa et.al.[32] proposed CNN technique on self-prepared Arabic dataset called Hijja Dataset and Arabic Handwritten Character Dataset (AHCD) and achieved an accuracy of 97% and 88% on the AHCD dataset and the Hijja dataset, respectively. The Arabic language was chosen because very few studies is being done for Arabic language. Bharat et.al.[33] described the challenges in recognizing online handwriting in Indic scripts and provide an overview of the state of the art for isolated character and word recognition. Also, it shows various resources such as tools and data sets, currently available in the Indic script visual study online.

Sharma et.al.[34] proposed a Quadratic classifier-based system used for the recognition of unconstrained off-line Devanagari handwritten characters. Feature vector had dimension of 64, and the features are obtained based on the directional chain codes of the contour of the character. Encouraging results are obtained with this experiment. Gyanendra et.al.[35] experimentally proved a technique, Curvelet transform also known as the curved singularities of images. It is very useful for feature extraction to character images. Devanagari script characters have a lot of curve. Firstly, segmentation of image is done then by applying curvelet transform curvelet features are obtained after the calculation of the statistics of thick and thin images. K-Nearest Neighbor classifier is used for training the system. 200 in house images of character set is used and the model achieved an accuracy of 90%. Aneja et.al.[36] used Transfer learning technique for the recognition of handwritten Devanagari alphabets using pre trained model for Deep Convolution Neural Network. As a fixed feature extractor AlexNet, DenseNet, Vgg, and Inception ConvNet are implemented. 15 epochs for each of AlexNet, DenseNet, Vgg, V3 etc has been implemented. From the results it concluded that Inception V3 performs better for accuracy and able to achieve 99% with average epoch time 16.3

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

whereas Alexnet performs faster with per epoch with 2.2 minute per epoch and 98% accuracy.

Shitole et.al.[37] comared the performance of recognition system using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). For the feature extraction three methods are used chain coding, edge detection using gradient features and direction feature techniques. These are further reduced by using LDA. Classification of characters is done by using SVM classifier and concluded that LDA has given better result than PCA. Bisht et.al.[41] proposed a novel technique for offline handwritten modified character recognition. Two models were used, A single CNN architecture and in the second model, he used double-CNN architecture for the recognition. Dataset of Hindi consonants and Matras with acceptable accuracies was considered. Based on the results, it was concluded that duplicate CNN formats provide better results than CNN single formats, as it uses a smaller number of output classes compared to existing modes of modified characters in Devanagari text. The results also inferred that double CNN proves to be better as compared to traditional feature extraction (like histogram of oriented gradients) and classification methods (like SVM). Gurav et.al.[42] presented a system which works on a set of 29 consonants and one modifier. With the help of Consecutive convolutional layers extracting higher-level features becomes easy. Here, character-wise segmentation is done instead modifiers-wise segmentation, which is a standard approach.

IV. ANALYSIS AND DISCUSSION

Handwriting recognition has become one of the most intriguing research areas in today's world, which has evolved into a combination of artificial intelligence and machine learning, which is discussed in this survey. It contributes specifically to the development of a visual connection between humans and machines. The handwritten recognition is used in everywhere for example documents, reading postal addresses, bank check amounts, and forms, etc. In this survey, we chose a deep learning-based Handwritten Devanagari Character Recognition concept. There are approximately 42 papers selected interrelated to the survey. The effectiveness of this survey is analyzed and compared using various parameters and algorithms. In this section, the different kinds of algorithms, methods, based on the deep learning in Character Recognition concept papers are discussed effectively.

4.1 Comprehensive Study

Below table 1 shows the comprehensive study of different techniques used for handwritten recognition.

Table 1 A comparative study of various papers.

Ref. no.	Paper Name	Preprocessing Techniques	Classifier Used	Data Size	Accuracy (In Percentage)
[1]	Deep Learning Based Large-Scale Handwritten Devanagari Character Recognition	Resizing, Gray scaling, padding	CNN	92 thousand images	98.47
[2]	Handwritten Devnagari numerals-recognition with higher accuracy	-	PCA	9,800	94.25
[3]	A Novel SVM - based Handwritten Tamil character recognition system	Operations are performed on the digitized image to enhance the quality of the image.	Support vector machine (SVM) method	Data samples are collected from different writers on A4 sized documents and then scanned.	82.04
[4]	Fuzzy Model Based Recognition of Handwritten Hindi Characters	-	Fuzzy Model Based	4750 samples	90.65
[5]	A Bilingual OCR for Hindi-Telugu Documents and its Applications	Scanned document is Filtered and Binarized.	Based on Principal Component Analysis followed by support vector classification	200000 characters	96.7

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

[6]	Online Handwritten Character Recognition of Devanagari and Telugu Characters using Support Vector	Normalization, Smoothing, Interpolation	SVM, HMM	Devanagari Script: The number of examples used for training is 21780 and that for testing is 2420. The number of classes considered for Devanagari script is 91. Telugu Script: There are 253 classes and a total of 33726 samples for training and 4091 samples for testing.	97.27
[7]	Optical Character Recognition Systems for Different Languages with Soft Computing	Binarization, Noise Removal, Skew Detection	Rough Fuzzy Multilayer Perceptron (RFMLP)	HP Labs India Indic Handwriting dataset	99.8
[8]	A Complete OCR for printed Hindi Text in Devnagari Script	Thresholding method used for Binarization	Tree Classifiers	The OCR is tested from various newspapers and magzines.	93
[9]	Feature extraction based on moment invariants for handwriting recognition	-	Gaussian Distribution	2,000	92
[10]	Devnagari Ancient Documents recognition using statitical feature extraction techniques	Auto correction, Binarization, Local / Global Thresholding	MLP, Neural Network (NN), CNN, SVM, Random Forest	6152 pre-segmented samples of Devanagari ancient documents	88.95
[11]	Character Segmentation in Text line via CNN	Training data using weakly labelled data via P-N learning	SVM	-	99.6
[12]	Visualizing and Understanding Customized Convolutional Neural Network for Recognition of Handwritten Marathi Numerals	Splitting, Resizing	Customised CNN(CCNN)	80000 samples. Out of 80000 samples, written in Marathi, 70000 samples are used for training and 10000, for testing.	94.93
[13]	An OCR system for Telugu	-	KNN and fringe distance template matching	Estimated to be of the order of 10,000	92
[14]	An efficient Devanagari character classification in printed and handwritten documents using SVM	Noise Removal, Skew Detection, Normalization, Gray scaling, Binarization	SVM	Total 60 documents are considered, where 60% documents are used in training and 40% were used during testing.	99.54 for printed images and 98.35 for handwritten images

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

[15]	Performance comparison of features on Devanagari handprinted dataset	-	SVM	More than 25000 handwritten Devanagari characters	94.1
[16]	Recognition of Unconstraint-On-Line Devanagari Characters	-	The combination of multiple classifiers that focus on either local online properties, or global off-lineproperties.	Experiment with 20 different writers with each writer writing 5 samples of each character in a totally unconstrained way	86.5
[17]	Hybrid feature extraction algorithm for Devanagari script	Skeletonization, a combination of structural features of the character like number of endpoints,loops, and intersection points is calculated, quadratic curve-fitting model	Neural network	Dongre database (4000 samples)	93.4
[18]	Comparative study of devnagari handwritten character recognition using different feature and classifiers	Gray scaling, Binarization	PD (Projection Distance), SVM, MQDF (Modified Quadratic Discriminant Function), MIL(Mirror Image Learning), ED(Euclidean Distance),NN, K-NN etc.	ISI Kolkata database	95.19
[19]	A multi-scale deep quad tree-based feature extraction method for the recognition of isolated handwritten characters of popular Indic scripts	-	softmax classifier, deep learning, multi-scale convolutional neural network (MMCNN)	HPL offline character database	95.18
[20]	Combining multiple feature extraction techniques for handwritten devnagari character recognition	Conversion of Handwritten Character to Bitmapped Binary Images, Scaling of the binary character images	Classification decision obtained from four Multi Layer Perceptron (MLP)	4900 samples	92.80

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

[21]	Classification Of Gradient Change Features Using MLP for Handwritten Character Recognition	Binarization	Multilayer perceptron (MLP) neural networks	300 samples of handwritten characters	99.10 and 94.15 recognition rates on training and test sets respectively
[22]	Hindi handwritten character recognition using multiple classifiers	Histogram based global binarising algorithm, Removal of isolated dot near header line	Regular expressions of strokes	5000 samples	82.0
[23]	A system for off-line Arabic handwritten word recognition based on Bayesian approach	Base Line estimation	Variants of Bayes classifier	IFN/ENIT Tunisian city names dataset ground- truth [31]	90.02
[24]	Handwritten Hindi character recognition using curvelet transform	Normalization, noise removal and gray scale conversion.	KNN	In-house dataset containing 200 images of character set (each image contains all Hindi characters)	90
[25]	Handwritten Devanagari Character Recognition using Convolutional Neural Network	Binarization, noise removal	Convolutional Neural Network	Devanagari characters and Devanagari numerals.	91.23% for Devanagari characters and 100% for Devanagari numerals.
[26]	Devanagari Handwritten Character Recognition using fine-tuned Deep Convolutional Neural Network on trivial dataset	XnConvert batch processing, data augmentation and some regularization techniques like Dropout and Batch Normalization.	Deep Convolutional Neural Ne twork (DCNN)	5800 isolated images of 58 unique character classes: 12 vowels, 36 consonants and 10 numerals.	94.84% testing accuracy with training loss of 0.18
[27]	Deep Learning Approach for Devanagari Script Recognition	Remove noise, converted to binary image, resized to fixed size of 30×40 and then convert to gray scale image using mask operation, it blurs the edges of the images.	Deep Learning Approach	Large set of handwritten numerical, character, vowel modifiers and compound characters.	91.81

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

[28]	Recognition of handwritten Devanagari characters using machine learning approach.	Binarization, Noise Elimination	ANN Classifier	Fifty handwritten characters from 10 people resulting 500 characters are used.	Recognition of 50 individual handwritten characters is 92% and for handwritten sentences, accuracy obtained is 88.25%.
[29]	Machine recognition of Devanāgarī script	Digitized, cleaned, thinned, segmented (to extract composite characters),and labeled	-	A small set of data consisting of Resembling characters and upper and lower signs	90
[32]	Arabic handwriting recognition system using convolutional neural network	Binarization, Noise Elimination, Tilted scanned papers were manually rotated.	CNN	Two dataset used. 1. Hijja Dataset- a new dataset of Arabic letters written exclusively by children aged 7–12. The dataset contains 47,434 characters written by 591 participants. 2.Arabic Handwritten Character Dataset (AHCD) dataset	97% and 88% on the AHCD dataset and the Hijja dataset, respectively
[34]	Recognition of Off- Line Handwritten Devnagari Characters Using Quadratic Classifier	Digitized, cleaned	Quadratic classifier-based system	11270 samples of Devanagari characters (vowels as well as consonants) from different individuals and digitalized them.	For Devanagari characters 80.36% and for numerals 98.86%
[35]	Handwritten Hindi Character Recognition Using Curvelet Transform.	Segmented first then curvelet features are obtained by calculating statistics of thick and thin images by applying curvelet transform.	K-Nearest Neighbor classifier using Curvelet Transform	200 in house images of character set (each image contains all Hindi characters)	90

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

[36]	Transfer Learning using CNN for Handwritten Devanagari Character Recognition	Finetuning	Pre trained model for Deep Convolution Neural Network	Handwritten Devanagari characters has 46 classes with 2000 images of each class. We partitioned the dataset of 92,000 images into a training set of 78,200 images (0.85) and a testing set of 13,800 (0.15).	99
[37]	Recognition of handwritten devanagari characters using linear discriminant analysis	Filtering of the scanned image, Grey scale conversion, Dilation	SVM classifier	47 handwritten Devanagari characters are used of ISI dataset. Image size is 270x270 pixels each. Total 4700 images are used for training (100 images per character).	70.58
[41]	Offline handwritten Devanagari modified character recognition using convolutional neural network	Resizing of images and conversion of samples from colour to greyscale.	Double CNN	Hindi consonants and Matras dataset	90.99
[42]	Devanagari Handwritten Character Recognition using Convolutional Neural Networks	Grey Scale conversion, Edge detection, Noise Removal	Deep Convolution Neural network	Self-made 34604 handwritten images used for Devanagari script with no header line (Shirorekha) over them	99.65

4.2 Comparative Analysis

In this survey, approximately 80 papers are collected and 42 papers were taken for technical analysis. Each of the papers is referred to as the category of deep learning. The papers are selected from different kinds of journals such as Elsevier, Springer, IEEE, Conference, and others. The pie chart representation of papers selected from different kinds of journals is described in Fig. 7. From this, 7% of papers are selected from Elsevier, 21% of papers are selected from springer, 48% of papers are selected from IEEE, and 7% deep learning method papers are selected from the ACM journals. Finally, the highest number of papers are chosen from the IEEE journal, and the papers are collected from the character recognition and deep learning domain

The character recognition papers which focus on deep learning are selected from the year 2000 to 2021. In this survey, we used approximately 42 papers for technical review. There are ten papers are chosen before the year 2010, four papers from 2010-2015, seven papers from 2015-2018. The highest number of papers are selected in between 2019-2021 that is approximately twelve papers are taken. The number of papers chosen from year is noted in Fig 8.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

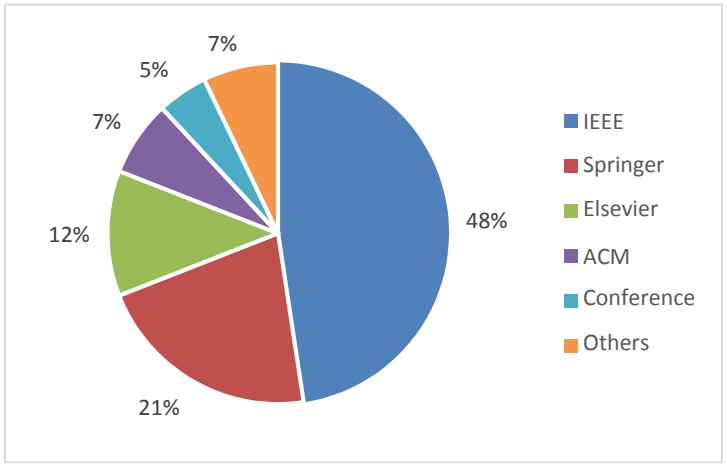


Fig 7. Representation of Character Recognition papers chosen from different journals.

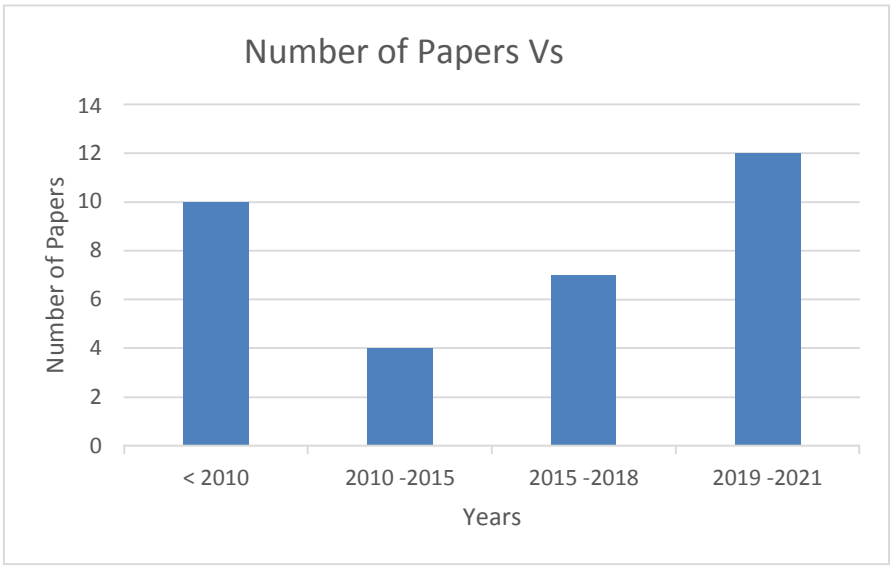


Fig 8. Analysis of Character Recognition papers published every year.

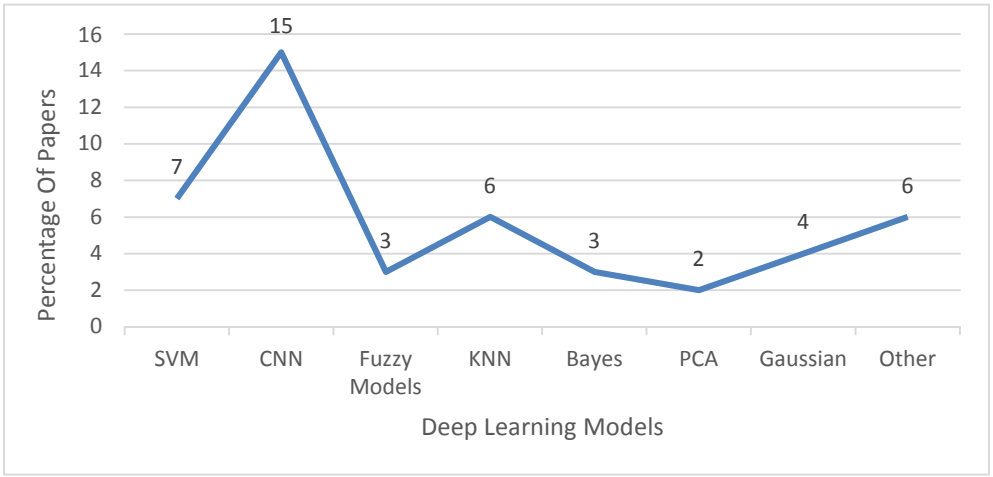


Fig 9. Analysis of character recognition papers based on deep learning methods.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

In this paper, the Devanagari character recognition based on deep learning methods is categorized into eight classes such as Convolutional Neural Network (CNN), Support Vector Machine (SVM), Fuzzy Model, K-Nearest Neighbor (KNN), Bayes Theorem, Principal Component Analysis (PCA), Gaussian Distance, Others. We have selected 7% of papers from SVM, 15% of papers from CNN, 3% of papers based on Fuzzy models, 6% papers from KNN learning, 3% of papers based on Bayes learning, 2% of papers based on PCA learning and remaining papers are based on the concept of common deep learning. Several papers belong to the recurrent neural network domain. The number of papers published based on the deep learning types is represented in Fig 9.

V. CONCLUSION

Character Recognition is one of the most common applications in image processing. Due to complexities of Indian languages, it has been recognized as one of the challenging researches is in the field of computer vision and pattern recognition. But still, a lot of research is being done on large datasets of these languages to handle the complexities and other issues.

This paper carries out a study on handwritten character recognition using deep learning with the help of various similar kinds of papers. This paper also represents a survey of preprocessing techniques, various classifiers used and recognition techniques for handwritten Devanagari character recognition. Deep learning techniques are commonly performed for character recognition due to high tolerance and less errors. This survey paper helps researches and developers to understand various techniques and the way they are implemented for recognition. The image is scanned firstly, and then data is preprocessed. The preprocessing involves various techniques such as Binarization, removal of noise and Normalization. After that, features are extracted from the preprocessed data so that the relevant data is further used to train the model. Various classification techniques are applied and the best approach is considered based on accuracy. Various models which uses multilayer perceptron compares input image with the trained set to get high accuracy. This paper has focused on various approaches effective algorithm. The study concludes that SVM classifier as well as CNN classifier both provides better results with an accuracy of 99.6% and 98.47% respectively. The study points out that the work in Devanagari scripts still in progress, so further a lot can be done in this domain.

VI. FUTURE WORK

In future, the system has a great scope of research in the area of Devanagari word as well as sentence recognition. Also, a large and complex dataset can be considered. Recognition of a full handwritten document through an OCR can be done. Still a hybrid model could be proved effective and optimized from which highly accurate results can be obtained. The system could be modified with other orthogonal moment features set Also, more work could be done on half characters as well as missing part of Hindi character recognition. The model can also predict errors of the input. Various other Indian scripts and languages can also be considered to make a generic system.

REFERENCES

[1]. Shailesh Acharya, Ashok Kumar Pant, Prashnna Kumar Gyawali. Deep Learning Based Large Scale Handwritten Devanagari Character Recognition, 2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA).

[2]. C. V. Lakshmi, R. Jain, and C. Patvardhan. Handwritten Devnagari numerals recognition with higher accuracy, in Proc. Int. onfut.Intell. Multimedia Appl., 2007.

[3]. Shanthi N and Duraiswami K. A Novel SVM - based Handwritten Tamil character recognition system, Springer Pattern Analysis & Applications,Vol-13, No. 2, 173-180,2010.

[4]. M. Hanmandlu, O. V. R. Murthy, and V. K. Madasu. Fuzzy Model based recognition of handwritten Hindi characters, in Proc. Int. Conf. Digital Image Comput. Tech. Appl., 2007.

[5]. C. V. Jawahar, M. N. S. S. K. Pavan Kumar, S. S. Ravi Kiran. A Bilingual OCR for HindiTelugu Documents and its Applications, Proc. of the 11th ICPR, vol. II, pp. 200-203, 1992.

[6]. C. Chandra Sekhar. Online Handwritten Character Recognition of Devanagari and Telugu Characters using Support Vector Machines, Tenth International Workshop on Frontiers in Handwriting Recognition, Universit éde Rennes 1, Oct 2006, La Baule (France). inria00104402.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, June 2020
Impact Factor: 4.819

[7]. A. Chaudhuri. Optical Character Recognition Systems for Different Languages with Soft Computing, Studies in Fuzziness and Soft Computing 352, DOI 10.1007/978-3-319-50252-6_8
[8]. Veena Bansala and R M K Sinha. A Complete OCR for printed Hindi Text in Devnagari Script ,IEEE 2002.
[9]. R. J. Ramteke and S. C. Mehrotra. Feature extraction based on moment invariants for handwriting recognition, in Proc. IEEE Conf. Cybern. Intell. Syst., 2006
[10]. Sonika Narang, M K Jindal and Munish Kumar. Devnagari ancient documents recognition using statistical feature extraction techniques , MS received 19 July 2018; revised 14 November 2018; accepted 4 March 2019; published online 13 May 2019.
[11]. Xiaohe Li, Xingming Zhang, Bin Yang, Siyu Xia. Character Segmentation in Text Line via Convolutional Neural Network, 2017 4th International Conference on Systems and Informatics (ICSAI).
[12]. D. T. Mane , U. V. Kulkarni. Visualizing and Understanding Customized Convolutional Neural Network for Recognition of Handwritten Marathi Numerals , IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.
[13]. A. Negi, C. Bhagvathi, and B. Krishna. An OCR system for telugu. In Proc. of ICDAR, 2001.
[14]. Shalini Puria, Satya Prakash Singh, An efficient Devanagari character classification in printed and handwritten documents using SVM, International Conference on Pervasive Computing Advances and Applications – PerCAA 2019.
[15]. S. Kumar, Performance comparison of features on Devanagari handprinted dataset, Int. J. Recent Trends, vol. 1, 2009.
[16]. S. D. Connel, R.M.K. Sinha and Anil K. Jain, Recognition of Unconstraint On-Line Devanagari Characters, Proceedings of the International Conference on Pattern Recognition, 2000.
[17]. Khandja D., Nain N., and Panwara S.: Hybrid feature extraction algorithm for Devanagari script, ACM Trans. Asian Low Resour. Lang. Inf. Process., 2015, 15, (1), p. 2:1– 2:10
[18]. Pal U., Wakabayashi T., and Kimura F.: Comparative study of devnagari handwritten character recognition using different feature and classifiers. Int. Conf. Document Analysis and Recognition, Barcelona, Spain, July 2009, pp. 1111– 1115
[19]. Sarkhel R., Das N., and Das A. et al: A multi-scale deep quad tree based feature extraction method for the recognition of isolated handwritten characters of popular Indic scripts, Pattern Recognit., 2017, 71, pp. 78–93
[20]. Arora S., Bhattacharjee D., and Nasipuri M. et al: Combining multiple feature extraction techniques for handwritten devnagari character recognition. Third int. Conf. Industrial and Information Systems, Kharagpur, India, December 2008, pp. 1– 6
[21]. S. Arora, D. Bhattacharjee, M. Nasipuri, L. Malik, Classification Of Gradient Change Features Using MLP for Handwritten Character Recognition, Emerging Applications of Information Technology (EAIT), Kolkata, India, 2006
[22]. Yadav M., and Purwar R.: Hindi handwritten character recognition using multiple classifiers. 7th Int. Conf. Cloud Computing, Data Science & Engineering – Confluence, Noida, India, Jan 2017, pp. 149– 154
[23]. Khémiri, A., Echi, A. K., Belaïd, A., et al.: A system for off-line Arabic handwritten word recognition based on Bayesian approach. Int. Conf. Frontiers in Handwriting Recognition (ICFHR), Shenzhen, China, October 2016, pp. 560–565
[24]. Verma G. K., Prasad S., and Kumar P.: Handwritten Hindi character recognition using curvelet transform. Int. Conf. Information Systems for Indian Language, Patiala, India, 2011, pp. 224– 227
[25]. A. Mohite and S. Shelke, Handwritten Devanagari Character Recognition using Convolutional Neural Network, 2018 4th International Conference for Convergence in Technology (I2CT), Mangalore, India, 2018, pp. 1-4, doi: 10.1109/I2CT42659.2018.9057991.
[26]. Deore, S.P., Pravin, A. Devanagari Handwritten Character Recognition using fine-tuned Deep Convolutional Neural Network on trivial dataset. Sādhanā 45, 243 (2020). https://doi.org/10.1007/s12046-020-01484-1
[27]. S. Prabhanjan and R. Dinesh, Deep learning approach for devanagari script recognition, Int. J. Image Graph. 17(3) (2017) 1750016. https://doi.org/10.1142/S0219467817500164

[28]. Pankaj Kale, Arti V. Bang, and Devashree Joshi, Recognition Of Handwritten Devanagari Characters Using Machine Learning Approach, International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982, Volume-3, Issue-9, pp. 48- 51, Sept.-2015.

[29]. R.M.K. Sinha and H.N. Mahabala, Machine recognition of Devanāgarī script, IEEE Transactions on Systems, Man and Cybernetics, 9(8), 1979, pp. 435-441.

[30]. Support Vector Machines for Binary Classification, Available at: https://www.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html

[31]. Pechwitz, M., Maddouri, S. S., Märgner, V., et al.: 'IFN/ENIT – database of handwritten Arabic words'. Proc. 7th Colloque Int. Francophone sur l'Ecrit et le Document, Hammamet, Tunisia, October 2002, pp. 129–136

[32]. Altwaijry, N., Al-Turaiki, I. Arabic handwriting recognition system using convolutional neural network. Neural Comput & Applic (2020). https://doi.org/10.1007/s00521-020-05070-8

[33]. Bharath A and Madhvanath S 2009 Online handwriting recognition for Indic scripts. In: Guide to OCR for Indic scripts, pp. 209–234 [34]. Sharma N., Pal U., Kimura F., Pal S. (2006) Recognition of Off-Line Handwritten Devnagari Characters Using Quadratic Classifier. In: Kalra P.K., Peleg S. (eds) Computer Vision, Graphics and Image Processing. Lecture Notes in Computer Science, vol 4338. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11949619_72

[34]. Verma G.K., Prasad S., Kumar P. (2011) Handwritten Hindi Character Recognition Using Curvelet Transform. In: Singh C., Singh Lehal G., Sengupta J., Sharma D.V., Goyal V. (eds) Information Systems for Indian Languages. ICISIL 2011. Communications in Computer and Information Science, vol 139. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-19403-0_37

[35]. N. Aneja and S. Aneja, "Transfer Learning using CNN for Handwritten Devanagari Character Recognition," 2019 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, 2019, pp. 293-296, doi: 10.1109/ICAIT47043.2019.8987286.

[36]. S. Shitole and S. Jadhav, "Recognition of handwritten devanagari characters using linear discriminant analysis," 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, 2018, pp. 100-103, doi: 10.1109/ICISC.2018.8398991.

[37]. Veena Bansal, R.M.K. Sinha, Segmentation of touching and fused Devanagari characters, Pattern Recognition, Volume 35, Issue 4, 2002, Pages 875-893,ISSN 0031-3203, https://doi.org/10.1016/S0031-3203(01)00081-4

[38]. Garg N.K., Kaur L., Jindal M.K. (2011) The Segmentation of Half Characters in Handwritten Hindi Text. In: Singh C., Singh Lehal G., Sengupta J., Sharma D.V., Goyal V. (eds) Information Systems for Indian Languages. ICISIL 2011. Communications in Computer and Information Science, vol 139. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-19403-0_8

[39]. B. Dessai and A. Patil, "A Deep Learning Approach for Optical Character Recognition of Handwritten Devanagari Script," 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, 2019, pp. 1160-1165, doi: 10.1109/ICICICT46008.2019.8993342.

[40]. Bisht, M., Gupta, R. Offline handwritten Devanagari modified character recognition using convolutional neural network. Sādhanā 46, 20 (2021). https://doi.org/10.1007/s12046-020-01532-w

[41]. Y. Gurav, P. Bhagat, R. Jadhav and S. Sinha, "Devanagari Handwritten Character Recognition using Convolutional Neural Networks," 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Istanbul, Turkey, 2020, pp. 1-6, doi: 10.1109/ICECCE49384.2020.9179193.