Grenze International Journal of Engineering and Technology, Jan Issue

Survey on Machine Learning and Deep Learning Algorithms for Recognition of Handwritten Devanagari Characters

Mrs. Ashwini Patil[1] and Dr. Sandeep Dubey[2]
[1]Research Scholar, RKDF University, Bhopal (MP), India.
[2]Professor, RKDF University, Bhopal, (MP), India.
Email: ash82patil@gmail.com, sandeep.rkdfgroup@gmail.com

***Abstract*—The automation of an application is crucial in the modern digital environment. In the atomization process, handwritten papers must be converted into digital format. There are more than 120 spoken languages that employ the Devanagari script. Recognizing handwritten Devanagari characters is made more challenging by the characters' complicated shapes, a wide range of modifiers, and variable writing styles. Preprocessing is a crucial step in recognizing handwritten images and comprises noise reduction, segmentation, feature extraction from the image, and more. Preprocessing and recognition are done using several methods. Due to its automatic feature extraction, a deep learning algorithm like CNN is particularly effective in recognition tasks. The main goal of this paper is to conduct a study of the literature and conduct a comparative analysis of the various preprocessing and recognition strategies to help guide future research on the problem of handwritten Devanagari script recognition.**

***Index Terms*—Pre-processing, Recognition, Compound Characters, Machine, Deep Learning, Feature extraction, Segmentation, SVM, KNN, CNN.**

I. INTRODUCTION

Using more than 18 languages and 10 main scripts makes India a multilingual and multi-script nation [7]. More than 120 spoken Indo-Aryan languages, including Hindi, Nepali, Marathi, Maithili, Awadhi, Newari, and Bhojpuri, utilize the Devanagari system, the Northern Brahmic script. In the rural areas of western and central India, most communication is done in the Devanagari script. Marathi is the most widely used spoken language in Maharashtra and parts of neighboring states. The Devanagari is the script used to write Marathi. Written in the Devanagari alphabet, Hindi is the world's third most widely spoken language. The most effective method of communication is speaking each other's language. In the digital world, many applications, like Banking, Education, Agriculture, etc., use handwritten form filling and other documentation. So, recognition of handwritten scripts can be used in all such applications for automation. E.g., processing of handwritten cheques in banks and official handwritten documents in government and non-government for digitization [13]. The digitalization of handwritten documents is one of the solutions for better access and sharing of historical documents. Recognition of the Devanagari script is a critical task due to its varying structure. This paper discusses challenges in the Devanagari Script recognition and surveys various techniques used to recognize these scripts.

Grenze ID: 01.GIJET.10.1.528

II. Challenges In The Devanagari Characters

Devanagari scripts have a large number of character sets, including 12 numbers of vowels, 36 base forms of consonants, 10 numeral characters, and a few special symbols, which makes the recognition task of the Devanagari quite challenging [15].

Vowel :
अ आ इ ई उ ऊ ऋ ए ऐ ओ औ अं अः ॲ ऑ
Consonants:
क ख ग घ ङ च छ ज झ ट ठ ड ढ ण त थ द
ध न प फ ब भ म य र ल व श ष स ह ळ क्ष ज्ञ

Figure 1: Vowels and Base Consonants

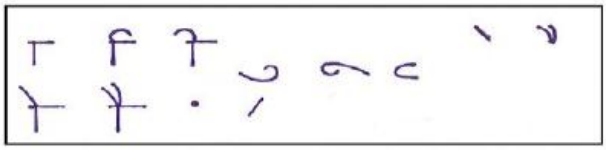

Figure 2: Modifiers based on vowel

Figure 1 shows the vowels and basic consonants of the Devanagari Script. Devanagari letters include many variants in their fundamental shapes, such as the complicated form, which has several modifiers above, below, or in line. These modifiers, in the form of vowels, change the basic shape of consonants when connected to them.

Example: If the modifiers shown in figure 2 are attached to consonant क (ka), then it changes the shape of the consonant क as follows,

क का कि की कु कू के कै को कौ कं कः कॅ काॅ

Some characters (vowels and consonants) look almost similar, making recognizing such characters more complex. Many recognition errors occur due to confusion among characters due to similar shapes. A few sample characters which look similar are shown in figure 3.

अ अं	इ ई	उ ऊ	ए ऐ
ओ औ	अ अः	क ळ	घ ध
य थ	ष ष	व ब	म भ
ढ द	ट ठ	ट ढ	द ट
प य	ड ङ		

Figure 3: Characters that look the same

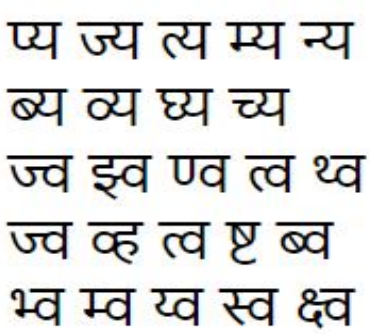

Figure 4: Sample Compound Characters

The identification process becomes more complicated when combining two or more consonants to create a distinctive compound character.
Figure 4 shows an example of some compound characters. Another obstacle to identification is the writer's range of writing styles, physical and mental health, and the pen or ink they are using while producing the document. The significant variance in the writing styles of the writer makes the handwritten script or character recognition task very challenging in natural language processing. The difficulty in handwriting character recognition increases due to confusion strokes, skewness and slanting nature, and compound characters [31].

III. Literature review of Recognition Techniques

Conversion of the handwritten script images into digital text for further editing and storing need the recognition of handwritten characters. To reduce the human effort, the task of handwritten recognition is used in various applications like postal automation [2, 9], automation of bank processes, filling of different forms, processing of forms, degradation of old documents [16] in police stations, municipal corporation offices, hospitals [38], etc. Authors are recommended to add the voice feature into the character recognition system to use the application for blind people [38].
The study on offline charter recognition first focused on reading printed characters, but it has now expanded to include handwritten characters and numerals in several Indian scripts [1, 2, 7, 15], including Devanagari [13].
Character recognition requires preprocessing, segmentation, feature extraction, classification, and matching techniques [15, 13]. These techniques can be used for Marathi character recognition as in other languages [15].
To recognize Bangla handwritten characters, Roy K. et al. [3] employed a quadratic classifier and the directional chain code histogram for feature extraction. The system proposed by the authors gave an accuracy of 90.53%, which was better than existing methods for recognizing Bangla handwritten characters. It was tested over 14879 characters obtained from different units of the population.
The handwritten Tamil characters were recognized with a two-stage recognition scheme by Bhattacharya, U. et al. [4]. The Tamil language uses 156 characters for writing. In the first stage, the authors used an unsupervised K−means clustering algorithm to form a group of charters and multilayer perceptron (MLP) classifiers that work on the chain code histogram feature in the second stage for recognizing the character class of each sample identified in the group. The overall accuracy obtained for the training set is 92.77%, and for the testing set, 89.66%.
The system that takes advantage of the directional information found in the gradient's arc tangent was proposed by Pal and U [5] and put into practice. The authors extracted the best feature using 22 mean filtering on the gray level picture, which was then normalized and segmented. The gradient image is obtained from a segmented image using the Roberts filter. The intensity of the gradient is accumulated with each of the 32 quantized directions that make up the arctangent of the gradient at first. The 392-dimensional feature vectors are obtained by Gaussian filter by down-sampling the blocks and direction [5, 7]. These features are recognized using the modified quadratic classifier. The test of the proposed system is carried out over 36172 handwritten characters collected from different writers. The accuracy of 94.24% is obtained using a 5-fold cross-validation scheme.
Hanmandlu M. et al. [6] recognized Hindi characters by fitting the modified exponential membership function to the fuzzy sets created using features including normalized distances gathered using the Box approach. The reinforcement learning method was used to train the structural parameters, improving convergence speed. Up to 90.65%, total accuracy was attained by the writers.
Support Vector Machines (SVM) and Modified Quadratic Discriminant Function (MQDF), two classifiers based on gradient and curvature data, were merged by Pal, U. et al. in their study [7]. The mechanism increases the accuracy to 95.13% on 36172 handwritten character data.
Using twelve different classifiers, including Euclidian Distance, Projection Distance, MQDF, Modified Projection Distance, Linear Discriminant Function, Subspace method, Mirror Image Learning, Compound MQDF, Compound Projection Distance (CPD), SVM, and k-Nearest Neighbor (KNN), Pal U. et al. [8] conducted a comparison study for both grey and binary images of Devanagari Out of these 12 classifiers, MIL demonstrated the best accuracy, with gradient and curvature features calculated from grey pictures having an accuracy of 94.74% and binary images having an accuracy of 94.74% and 95.09%.
Unconstrained Handwritten Devanagari Character Recognition Using Multistage Feature Extraction and Classification is done by Shelke, S., & Apte, S. [10]. As the first stage of preprocessing, binarization and morphological procedures are performed. The high-level, mid-level, and low-level attributes classify the characters. The structural characteristics are high-level and mid-level elements, such as the vertical lines in letters. The image is first cropped to determine the column with the most pixels, and the vertical projection profile is then computed. To locate the character's endpoints, hit-or-miss transformation is used to identify the mid-level

characteristics. Based on these two-stage feature extraction techniques, the characters are divided into 24 classes. The low-level characteristics are extracted from the normalized picture using the Radon transform and Euclidean distance. The neural network classifier is given these features to use. The normalized charters are also identified via template matching. A hybrid classifier then analyzes three classification results. The acceptance percentage of the suggested approach is 95.40%.

Shelke, S., & Apte [11] introduced the two-stage structural classifications approach, which classified the picture into 24 groups by extracting the image's global and local properties. Wavelet-based and modified wavelet-based methods reduce the characters for kernel creation. After template matching, several resizing factors are used to produce the kernels, and the results are then evaluated. This method allowed them to identify the 40 Marathi compound characters. Modified wavelet-based kernels outperform wavelet-based kernels in terms of recognition rate for the 16x16 and 32x32 resize factors.

A novel method for Hindi character identification was presented and put into practice by Verma, G. K. et al. [12], employing the KNN classification algorithm and the digital curvelet transform for feature extraction. Experimentation results in an accuracy of above 90%.

Ramteke, A. S., & Rane, M. E. [13] suggest that extracting the relevant feature from the image is crucial to improving classification accuracy. High recognition rates can be attained by selecting the appropriate feature extraction approach. Three categories of feature extraction exist geometric features (edges, lines), spectral features (color, tone, ratio, spectrum), and textural features (pattern, homogeneity, spatial frequency) [14].

Preprocessing is essential for recognizing handwritten characters. The adaptive threshold approach for binarization was utilized by Deshmukh A. et al. [16]. The Eigenspace approach is utilized for character recognition, extraction, and character segmentation using the center portion, and upper and lower modifiers are employed for segmentation. Various characters and modifiers are used to test the proposed approach. Reading handwritten Devanagari characters using the suggested method has shown to be highly successful.

The recognition of handwritten characters is carried out in several processes, such as preprocessing, feature extraction, and classification, using various classic networks, such as SVM, Hidden Markov Model, Fuzzy-based classification, etc. The feature extraction tasks significantly impact how well these networks recognize objects. Deep neural networks, on the other hand, do not need explicit feature extraction.

A deep convolutional neural network (CNN) automatically extracts the image's best features and classifies them [22, 23]. The Deep CNN performs the recognition job far better than conventional networks. Due to its excellent performance, the deep neural network is widely used for various tasks, including regression, classification, pattern identification, computer vision applications, etc.

According to Maitra, D. S. et al. [17], training the CNN for many characters allows for high character recognition accuracy. They trained the CNN for six-character databases, including five different 10-class numerical characters from English, Devanagari, Bangla, Telugu, and Oriya and 50 classes of fundamental characters in Bangla. The authors employed the SVM classifier. Due to CNN's auto extraction feature, the system's accuracy was found to be suitable for all six sets of characters.

Acharya S. et al. released the Devanagari Handwritten Character Dataset (DHCD), segmented from several manuscripts and includes 92 thousand images of 46 distinct classes. The depth of the network, size of the different layers, number of feature mappings in each layer, size of the kernel on each layer, choice of non-overlapping or overlapping kernel, degree of overlap, etc., all affect how effectively a CNN can detect objects. The author proposed and constructed two unique CNN models, tested them on the DHCD dataset, and adjusted the depth, width, and network parameters. They examined the impact of the dataset increment and dropout layer. The experiment's findings demonstrate that the CNNs provide very high test accuracy when implemented using the dropout layer and dataset increment technique.

Kamble utilized Rectangle Histogram Oriented Gradient (R-HOG), P. M. et al. [19] to extract features for classification using Feed Forward Artificial Neural Network (FFANN) and SVM for handwritten Marathi character recognition. They discovered that FFANN performs 97.15% accurately, which is significantly superior to SVM.

Kale P. et al. [20] proposed and implemented the recognition system using the array-based feature extraction method, and a back-propagation neural network is used as a classifier. A total of 500 characters, including 50 handwritten characters obtained from 10 participants, were used in the experiment. The system in place has been tested on both characters and sentences, providing accuracy ratings of 92% for characters and 88.25% for sentences. Compared to conventional networks, artificial neural networks are more accurate.

The transfer learning mechanism was proposed and implemented by Sonawane, P. K., and Shelke [21] for the recognition of Devanagari characters using the pertained CNN model, i.e., Alexnet. One of these networks can classify photographs into 1000 different object categories after being trained on more than a million images.

Transfer learning makes learning a new task considerably quicker and simpler by using a pre-trained model as a starting point rather than starting from scratch and training a network with weights initially assigned at random. Transfer learning is preferred for more effective training with fewer training samples. The model's training data uses about 16870 samples of 22 consonants from the Devanagari script. The model's findings are remarkable, scoring 94.49% for validation and 95.46% for test accuracy.

The DCNN was built using six distinct network architectures by Jangid, M., and Srivastava [22]. To see the first response of the DCNN, NA-1 only has 500 completely connected neurons and one convolutional-pooling layer. The next NA-2 has twice as many fully linked neurons. The goal is to evaluate the effects of improvement. Two C-P layers with varying numbers of kernels are present in NA-3 and NA-4 to further investigate the impacts of two C-P layers. Three C-P layers are present in the final NA-5 and NA-6. The authors used six optimizers to analyze these six networks' performance. They conducted several studies utilizing the ISIDCHAR and V2DMDCHAR databases provided by the Information Sharing Index (ISI), Kolkata.

With the NA-6 network design and RMSProp, an adaptive gradient approach, the highest accuracy of 96.02% was attained (optimizer). To enhance the recognition accuracy, they further trained DCNN layer-wise using NA-6 with the RMSProp adaptive gradient method and obtained the highest accuracy of 98%.

Using six distinct models, CNN-RNN, SCNN-RNN, Synth-RNN, IAM-Synth-RNN, Mixed-RNN, and Mixed-RNN-RNN-Lexicon, Dutta, K. et al. [23] suggested the hybrid architecture. They made use of the Devanagari-compatible RoyDB and LAW datasets. The RoyDB uses a customized character annotation scheme and has 16k words. Only 27k words from a narrow, specialized lexicon are used in the LAW's visuals. Because there isn't a publicly available dataset for handwritten characters in Indian scripts, authors have provided a new handwritten word dataset for Devanagari called IIIT-HW-Dev. It has 95k words gathered from 12 individuals of different ages and educational levels. Character Error Rate (CER) and Word Error Rate (WER) measures were used to evaluate the proposed model's performance using sample pictures from the IIIT-HW-Dev and Devanagari track of the RoyDB dataset. To improve word recognition, the network is altered using both fake and actual handwriting input.

Shitole, S., and Jadhav, S. [24] used Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to improve the effectiveness of the recognition approach. The three methods utilized to extract the features are chain coding, edge detection, and directional features. They employ PCA, LDA, and MLDA for data dimensionality reduction and feature extraction. As classifiers, SVM and KNN are employed. SVM generated the most accurate results for the ISI dataset: 84.3% for edge detection, 94.2% for directional features, and 86.6% for chain coding.

Thakur, A., & Kaur, A. [25] suggested a hybrid strategy of KNN combined with the neural networks for Devanagari handwritten character identification, which gave an accuracy of 94.5%. Traditional character recognition approaches are less accurate in recognizing the Devanagari scripts. The primary emphasis is on identifying each consonant and vowel separately.

Reddy, R. V. et al. [26] combined CNN with optimization methods, including Adam, RMSprop, and Deep Feed Forward Neural Network (DFFNN), to recognize handwritten Hindi characters.

When the implemented system is compared with other methodologies, such as MLP, PCA, KNN, Fuzzy model, MLP, minimal edit distance, etc., the implementation's results are satisfactory.

Deokate, S. T., & Uke, N. J. [27] gathered the sample from newspapers and other sources to create their dataset. Binarization, skew identification and correction, noise elimination, and small object removal are used to preprocess the input document. Additionally, the histogram technique is used in conjunction with the morphological operation to carry out the line, word, and character fragmentation. The authors employed KNN and CNN techniques for categorization. For KNN, they got an average of 96% precision for the test and trained dataset with k=1 to 5. For CNN, they got 100% precision for the trained and validated dataset. The proposed system is tested for Type-scripted manuscripts and handwritten manuscripts.

The system for handwritten compound characters was developed and implemented by Kadam, A. A. et al. [28] using zoning and statistical feature extraction techniques and SVM and KNN classifiers. They made use of a dataset that included 100 photos of various individuals of various ages and 35 different compound characters. Over 3500 handwritten Marathi compound character images were examined using the system, with an average SVM classifier accuracy of 96.49% and a KNN accuracy of 95.67%.

Puri, S., and Singh, S. P. [29] presented an SVM classifier for adequate recognition of handwritten and printed characters from Hindi, Sanskrit, and Marathi documents. The technique for extracting features based on geometry is employed. They found 99.54 % and 98.35% accuracy for printed and handwritten scripts.

The performance of pre-trained deep learning models, such as AlexNet, DenseNet, Vgg, and Inception, for handwritten Devanagari character recognition has been analyzed by Aneja, N., & Aneja, S. [30] in terms of accuracy attained in the first epoch, the total number of epochs required to achieve the best accuracy, the total time

taken by 15 epochs, average epoch time. They tested models with the DHCP dataset and found that Inception performed best with an accuracy of 99%, and the DenseNet model performed worst.
CNN model with Relu and Softmax activation was utilized by Bisht, M., & Gupta, R. [31]. The model's precision is evaluated using correctness metrics and the categorical cross-entropy loss function. To train the model, the scientists used the DHCD dataset. With an accuracy of 99.28%, they found that CNN, using a dropout and Stochastic Gradient Descent (SGD) optimizer, is one of the most exemplary characters and number recognition methods.
Gurav, Y. et al. [32] and Gupta, P. et al. [33] used the Deep Convolution Neural Network (DCNN) approach for Devanagari character recognition, achieving the accuracy of 99.65% and 95.6%, respectively. They found that CNN is a reliable technique for the classification of characters.
The Convolution Neural Network (CNN) was used by Chakraborty R. et al. [34] to recognize characters in the Bangla and Devanagari languages. They experimented with 10,000 Bangla primary characters and 1800 Devanagari characters, and their accuracy rates were 99.65% and 98.87%, respectively. They also intend to improve their work by using more character sets and different deep learning architectures to increase accuracy.
Kattyayan S. et al. analyze machine learning methods for Devanagari character identification [35]. They also proposed and applied the CNN model for Devanagari character recognition. The model was trained using a Kaggle dataset, with an accuracy of 98.01%.
For recognition of Modi Script, Sawant S. et al. [36] devised and implemented the CNN model with two distinct architectures. Two convolution layers are present in the initial CNN model, followed by fully linked and max-pooling layers. The second design features three fully connected layers after three convolution and max-pooling layers. Test results for the suggested model's accuracy on the Modi character were 95.44% and 95.97%, respectively. The dataset was compiled from 100 authors, representing 57 classes of characters.
Dokare I. et al. [37] used CNN to recognize Devanagari consonants, vowels, and numerals. The authors used the DHCD and Nepali Handwritten character datasets (NHCD) for training. The authors found more accuracy in the recognition task.
CNN architecture with four convolutional, two max-pooling, three dropouts, one flatten, and two dense layers was configured and implemented by Pande, S. D. et al. [38] for the recognition of Devanagari handwritten text, which extracts statistical features from isolated characters. The DHCD dataset, which contains 46 classes of characters and 2,000 unique images, was used to test the model. The implementation's results were judged superior to the recognized approaches already in use.
A 20-layer CNN model was proposed and implemented by Shinde, A., and Dandawate, Y. [39] to recognize handwritten Marathi words. They evaluated the system's performance by varying the number of layers and layer parameters, such as the number of filters and filter size. They used a dataset of 9,360 words that included 104 words with 90 images each and was compiled from people between the ages of 8 and 49 years. The segmented words are used to train the CNN model to extract features, which are then examined to produce recognition. The system, as it is now implemented, is 94.76% accurate. Deep learning may be used to recognize handwritten Devanagari characters more successfully.
Agrawal, M., Chauhan, B., & Agrawal, T [40] did a comparative study of different handwritten Devanagari character recognition techniques, including SVM, CNN, Fuzzy models, KNN, Naive Bayes, PCA, Gaussian, and others. After the study, they found that SVM and CNN perform better in recognition.

IV. Research Gap

The research gap after the literature review is given below,

- There hasn't been enough work done on recognizing compound characters in Indian handwriting, so there is a huge scope of research in this area.
- Millions of people use Devanagari, the third most popular script, for several essential languages, including Hindi, Sanskrit, Marathi, and Nepali. However, few papers have been published on script identification for offline handwriting recognition of Devanagari script [41].
- The offline handwritten character or word recognition challenge was made more difficult by the lack of publicly available handwriting datasets in Indic scripts [23].
- The available dataset should be expanded to include all vowels, consonants derived from characters, and special characters [4].
- The identification of compound characters, created by combining the principal form of one consonant with another consonant or vowel to generate another character, has not been the subject of much research [18].
- Handwriting compound character recognition has not been extensively studied in this field of research.

Multilevel compound character recognition has a lot of potentials. [5] [11].

- Different DCNN architectures can be utilized to recognize Handwritten Devanagari characters [22].

V. Review Analysis and discussion

The publications that have been reviewed are thoroughly analyzed in Table 1 based on a variety of criteria, including the language of recognition, the method used for preprocessing, segmentation, feature extraction, and recognition, the dataset used for training and testing, and the efficacy of the system.

TABLE I. ANALYSIS OF REVIEWED RESEARCH PAPERS

Paper ID	Language	Pre-processing Technique	Segmentation method	Feature Extraction Technique	Recognition Technique	Dataset Used	Recognition Accuracy (%)
[4]	Tamil Characters	Thickening, linear size normalization	_	Chain code histogram	MLP classifier	hpl-Tamil-iso-char-online-1.0, developed by HP Lab India	Train= 92.77% Test = 89.66%
[5]	Devanagari	Flatbed scanner for digitization in TIF format	Block	Arc tangent of the gradient	Quadratic classifier	36172 samples of Devanagari's characters collected from individuals	94.24 %
[6]	Devanagari-Hindi	Binarization, slant correction, and smoothing	Character segmentation	Box approach	Fuzzy Model	Database of 4750 samples collected from different writers	90.65%.
[7]	Devanagari	_	_	Gradient and curvature	SVM and MQDF	36172 samples of Devanagari basic characters collected from different Individuals.	95.13 %
[10]	Devanagari-Marathi	Removal of a header line and unwanted strokes	Character segmentation	Radon transform and Euclidean distance transform	Feed-forward back propagation neural network classifier and Template Matching	(Over 16000 samples total) 400 examples of each character from the self-generated sample were gathered from more than 40 writers.	NN (Euclidean) = 94.14% NN (Radon) = 93.91% Template Matching = 93.23 %
[11]	Devanagari-Marathi	Removal of a header line and unwanted strokes	Character segmentation	Structural features	Template matching	Own sample collected from more than 40 writers	Wavelet-based kernels – for 16x16 resize factors = 95.89% For 32x32 resize factors = 96.00% Modified wavelet-based kernels-

							for 16x16 resize factors = 96.41% For 32x32 resize factors = 97.94%
[12]	Devanagari-Hindi	Normalization, noise removal, and grayscale	Character segmentation	Digital curvelet transform	KNN	An in-house dataset containing 200 images of the character set	More than 90%.
[16]	Devanagari	Principle component analysis (PCA)	Line, word, and Character segmentation	Eigen space algorithm	Eigen space algorithm	Five different samples per character with different modifiers	_
[17]	English, Bangla, Devanagari, Oriya, Telugu numerals and Bangla's primary characters	Normalization	_	CNN	SVM	Freely available datasets -Bangla primary characters, Bangla numerals, Devanagari numerals, Oriyanumerals, Telugunumerals, English (MNIST) numerals	Bangla (Basic) 95.6%, Bangla Numeral - 98.37%, Devanagari Numeral- 98.54%, Oriya Numeral- 97.2%, Telugu Numeral- 96.5% and English Numeral- 99.10%
[18]	Devanagari	Resize, Gray scaling, suppress background pixel, padding	Character segmentation	CNN	CNN	DHCD	98.47 %
[19]	Devanagari-Marathi	Noise reduction, connect broken tiny characters, edge detection, region filling, normalization	Bounding box	Rectangle Histogram Oriented Gradient (R-HOG)	SVM, FFANN	Own Prepared dataset consists of 8000 samples each of 40 primary handwritten Marathi characters	97.15% with FFANN

[20]	Devanagari-Marathi	Noise reduction and binarization	Character segmentation with kana, Matra,Velanti separation	Array-based feature extraction	ANN	50 handwritten characters collected from 10 peoples	Character recognition: 92% and sentence recognition: 88.25%
[22]	Devanagari	_	_	Deep CNN (DCNN)	DCNN - Six different architectures of CNN (NA1 to NA6)	ISIDCHAR database and V2DMDCHAR database	96.02% with NA-6 network architecture and RMSProp
[24]	Devanagari	Noise reduction, binarization, resizing, dilation	_	Chain code, edge detection, and directional features	SVM and KNN	ISIDCHAR database provided by ISI, Kolkata	SVM: 94.2% for directional feature, 86.6 % using chain coding and 84.3 % for edge detection. KNN: 89.3% for directional feature, 75.8 % using chain coding, and 75.8 % for edge detection.
[25]	Devanagari - Hindi	Binarization, Morphological erosion, Skew Correction	Character segmentation	Fixed Center distance-based approach	A hybrid approach of KNN along with the NN	Own dataset prepared	94.5%
[27]	Devanagari	Binarization, Noise, Skew Correction, resizing, and normalization	Line, word, and character segmentation	CNN	CNN and KNN	our dataset of the typed letter with modifiers, samples from the newspapers, and other resources	KNN-average 96% of precision (k=1 to 5.) CNN-100% of precision
[28]	Devanagari-Marathi	Image resizing, grayscale conversion, Binarization, Noise reduction	_	zoning and statistical feature extraction methods	SVM, KNN	100 sets of handwritten Marathi compound character samples collected from individuals of various ages make up the author's collection.	96.49% for SVM, and 95.67% for KNN
[29]	Devanagari-Hindi, Sanskrit and Marathi	Noise removal, Skew Correction, normalization, binarization	lines, words, and character segmentation with Horizontal & Vertical Projection	character geometrical structure	SVM	A scanned copy of printed materials that were acquired from news stories, official government portal websites, and traditional shlokas	99.54% for printed and 98.35%. For handwritten

[30]	Devanagari	Resizing and normalization	Character segmentation	Transfer Learning	Used pre-trained models AlexNet, DenseNet, Vgg, and Inception	DHCD	Inception performed best with an accuracy of 99%, and the DenseNet model performed worst.
[31]	Devanagari	_	_	CNN	CNN	DHCD	99.28%
[32]	Devanagari	Gray-scaling, binarization, Bitwise NOT on image and resizing	Character segmentation	CNN)	CNN	Self-made dataset Thirty-four thousand six hundred four handwritten images of 29 consonants with no header line.	99.65%
[33]	Devanagari	_	_	CNN	CNN	MNIST Dataset is used initially, and later the image data argumentation technique is used to increase the data.	95.6 %
[34]	Bangla and Devanagari	_	_	CNN	CNN	Dataset with 10000 Bangla (having 50 classes) and 1800 Devanagari (having 36 classes) characters. IAPR tc–11 website http://www.iapr-tc11.org.	Bangle characters 99.65% Devanagari Characters: 98.87%
[35]	Devanagari	Binarization	_	CNN	CNN	46 classes having 2000 images per class with a total of 92000 images	98.01%
[36]	Modi script, which is a cursive form of Marathi	Resizing, Grayscale	Middle region Character segmentation	CNN	CNN	Dataset collected from 100 writers: 57 different classes of MODI Script characters	CNN First Architecture: 95.44% CNN Second Architecture: 95.97%
[37]	Devanagari	_	_	CNN	CNN	DHCD (Devanagari Handwritten Character Dataset) and NHCD (Devanagari (Nepali) Handwritten Character Dataset).	98% for consonants, 97.56 % for vowels, and 99% for numerical
[38]	Devanagari	Border detection, border cropping, page straightening transformation, noise reduction, skew correction, size	Character segmentation	CNN	CNN	DHCD of 46 classes of Devanagari characters, and each of these classes has two thousand different images.	99.13%

		normalization, and sharpening kernel application					
[39]	Devanagari-Marathi	Binarization, Skew Correction	Word segmentation	CNN	CNN	Own dataset with 9,360 words (104 words with 90 images each) collected form age groups of the age group of 8 to 45 years	98%

Figure 5 shows how many papers are reviewed year-wise. A total of 15 papers were reviewed and published from 2019 onward.

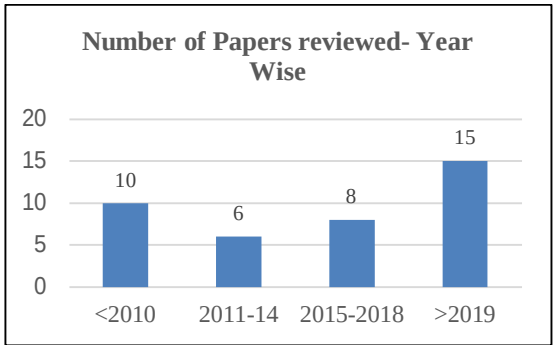

Figure 5: Year-wise analysis of reviewed papers

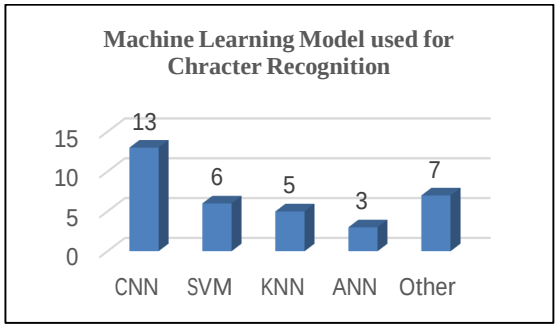

Figure 6: Machine learning model used for character recognition

Figure 6 displays the total number of papers using various algorithms. This number relates to all of the papers that were surveyed. The graph shows that a CNN, based on deep learning, is a widely popular technique for handwritten character recognition. Additionally, it is evident from the comparative table that CNN provides more accurate findings compared to other methodologies.

Compound Devanagari Handwritten Scripts can be recognized using a variety of deep learning algorithms. The goal is to use deep learning algorithms for handwritten compound character recognition and performance evaluation to select the most effective algorithm.

VI. Conclusion

Millions of people in India speak Devanagari-based languages like Marathi, Hindi, Sanskrit, Nepali, and Maithili. One of the primaries uses for image processing is character recognition. There is still much room for research into the offline handwritten recognition of the Devanagari Script, even though much work has been done on recognizing the English language and the printed form of the Devanagari script. There are many applications where it is

necessary to convert old handwritten scripts into digital forms; one of those applications is the recognition of handwritten Devanagari scripts.

Due to its more extensive complicated shape character set, which includes 12 vowels, 36 base forms of consonants, 10 number characters, several special characters, and many additional difficulties outlined in this paper, recognition of handwritten Devanagari script has become more complex.

We conducted a literature review of previous studies on Devanagari script recognition in this essay. The researchers apply several pre-processing, segmentation, feature extraction, and identification techniques that can be helpful for more studies in the field. Even yet, there is a scope for improvement in the handwritten compound Devanagari characters. The automatic feature extraction and more precise findings of the deep learning-based methodology make it the best method for handling the problems and challenges associated with recognizing handwritten Devanagari script. Therefore, research can be undertaken using a deep learning approach to recognize Devanagari's handwritten compound letters.

REFERENCES

[1] Pujari, A. K., Naidu, C. D., Rao, M. S., & Jinaga, B. C. (2004). An intelligent character recognizer for Telugu scripts using multiresolution analysis and associative memory. Image and Vision Computing, 22(14), 1221-1227.

[2] Roy, K., Vajda, S., Pal, U., Chaudhuri, B. B., & Belaïd, A. (2005, August). A system for Indian postal automation. In Eighth International Conference on Document Analysis and Recognition (ICDAR'05) (pp. 1060-1064). IEEE.

[3] Roy, K., Pal, U., & Kimura, F. (2005). Recognition of Handwritten Bangla Characters. In Proc. 2nd International Conference on Machine Intelligence (ICMI) (pp. 480-485).

[4] Bhattacharya, U., Ghosh, S. K., & Parui, S. (2007, September). A two stage recognition scheme for handwritten Tamil characters. In Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) (Vol. 1, pp. 511-515). IEEE.

[5] Pal, U., Sharma, N., Wakabayashi, T., & Kimura, F. (2007, September). Offline handwritten character recognition of Devanagari script. In Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) (Vol. 1, pp. 496-500). IEEE.

[6] Hanmandlu, M., Murthy, O. R., & Madasu, V. K. (2007, December). Fuzzy model based recognition of handwritten Hindi characters. In 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007) (pp. 454-461). IEEE.

[7] Pal, U., Chanda, S., Wakabayashi, T., & Kimura, F. (2008, August). Accuracy improvement of Devanagari character recognition combining SVM and MQDF. In Proc. 11th Int. Conf. Frontiers Handwrit. Recognit. (pp. 367-372)

[8] Pal, U., Wakabayashi, T., & Kimura, F. (2009, July). Comparative study of Devanagari handwritten character recognition using different feature and classifiers. In 2009 10th International Conference on Document Analysis and Recognition (pp. 1111-1115). IEEE.

[9] Pal, U., Roy, R. K., Roy, K., & Kimura, F. (2009, July). Indian multi-script full pin-code string recognition for postal automation. In 2009 10th International Conference on Document Analysis and Recognition (pp. 456-460). IEEE.

[10] Shelke, S., & Apte, S. (2010, November). A novel multi-feature multi-classifier scheme for unconstrained handwritten Devanagari character recognition. In 2010 12th International Conference on Frontiers in Handwriting Recognition (pp. 215-219). IEEE.

[11] Shelke, S., & Apte, S. (2011, February). A novel multistage classification and wavelet based kernel generation for handwritten Marathi compound character recognition. In 2011 International Conference on Communications and Signal Processing (pp. 193-197). IEEE.

[12] Verma, G. K., Prasad, S., & Kumar, P. (2011, March). Handwritten Hindi character recognition using curvelet transform. In International Conference on Information Systems for Indian Languages (pp. 224-227). Springer, Berlin, Heidelberg.

[13] Ramteke, A. S., & Rane, M. E. (2012). A survey on offline recognition of handwritten Devanagari script. Digital Image Processing, 4(3), 130-136.

[14] Holambe, A. N., Thool, R. C., & Jagade, S. M. (2012, January). A brief review and survey of feature extraction methods for Devnagari OCR. In 2011 Ninth International Conference on ICT and Knowledge Engineering (pp. 99-104). IEEE.

[15] Khobragade, R. N., Koli, D. N. A., &Makesar, M. S. (2013). A survey on recognition of devnagari script. International Journal of Computer Applications & Information Technology, 2(1), 22-16.

[16] Deshmukh, A., Meshram, R., Kendre, S., & Shah, K. (2014). Handwritten Devanagari Character Recognition. International Journal of Engineering Research, 3(4).

[17] Maitra, D. S., Bhattacharya, U., & Parui, S. K. (2015, August). CNN based common approach to handwritten character recognition of multiple scripts. 2015 13th International Conference on Document Analysis and Recognition (ICDAR) (pp. 1021-1025). IEEE.

[18] Acharya, S., Pant, A. K., & Gyawali, P. K. (2015, December). Deep learning based large scale handwritten Devanagari character recognition. In 2015 9th International conference on software, knowledge, information management and applications (SKIMA) (pp. 1-6). IEEE.

[19] Kamble, P. M., & Hegadi, R. S. (2015). Handwritten Marathi character recognition using R-HOG Feature. Procedia Computer Science, 45, 266-274.

[20] Kale, P., Bang, A. V., & Joshi, D. (2015). Recognition Of Handwritten Devanagari Characters Using Machine Learning Approach. International Journal of Industrial Electronics and Electrical Engineering, 3(9), 48-51
[21] Sonawane, P. K., & Shelke, S. (2018, August). Handwritten Devanagari Character Classification using Deep Learning. In 2018 International Conference on Information, Communication, Engineering and Technology (ICICET) (pp. 1-4). IEEE.
[22] Jangid, M., & Srivastava, S. (2018). Handwritten Devanagari character recognition using layer-wise training of deep convolutional neural networks and adaptive gradient methods. Journal of Imaging, 4(2), 41.
[23] Dutta, K., Krishnan, P., Mathew, M., & Jawahar, C. V. (2018, April). Offline handwriting recognition on Devanagari using a new benchmark dataset. In 2018 13th IAPR International Workshop on Document Analysis Systems (DAS) (pp. 25-30). IEEE.
[24] Shitole, S., & Jadhav, S. (2018, January). Recognition of handwritten devanagari characters using linear discriminant analysis. In 2018 2nd International Conference on Inventive Systems and Control (ICISC) (pp. 100-103). IEEE.
[25] Thakur, A., & Kaur, A. (2019). Devanagari handwritten character recognition using neural network. Int. J. Sci. Technol. Res, 8(10)
[26] Reddy, R. V. K., &Babu, U. R. (2019). Handwritten Hindi character recognition using deep learning techniques. Dept. of CSE, Acharya Nagarjuna University, Guntur, India, International Journal of Computer Sciences and Engineering, 1
[27] Deokate, S. T., & Uke, N. J. Devanagari Script Categorization by Utilizing CNN and KNN, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-5 March 2019.
[28] Kadam, A. A., Bhalerao, M. V., & Tanurkar, M. N, "Handwritten Marathi Compound Character Recognition," International Journal of Engineering Research & Technology (IJERT), 2019
[29] Puri, S., & Singh, S. P. (2019). An efficient Devanagari character classification in printed and handwritten documents using SVM. Procedia Computer Science, 152, 111-121.
[30] Aneja, N., &Aneja, S. (2019, July). Transfer learning using CNN for handwritten Devanagari character recognition. In 2019 1st International Conference on Advances in Information Technology (ICAIT) (pp. 293-296). IEEE.
[31] Bisht, M., & Gupta, R. (2020). Multiclass recognition of offline handwritten Devanagari characters using CNN. Int. J. Math. Eng. Manage. Sci., 5, 1429-1439.
[32] Gurav, Y., Bhagat, P., Jadhav, R., & Sinha, S. (2020, June). Devanagari Handwritten Character Recognition using Convolutional Neural Networks. 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (pp. 1-6). IEEE.
[33] Gupta, P., Deshmukh, S., Pandey, S., Tonge, K., Urkunde, V., &Kide, S. (2020, October). Convolutional Neural Network based Handwritten Devanagari Character Recognition. 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE) (pp. 322-326). IEEE.
[34] Chakraborty, R., Mukherjee, D., Bhattacharyya, A., Mukherjee, H., Sur, M. K., Sen, S., & Roy, K. (2020, September). Online Handwritten Bangla and Devanagari Character Recognition using CNN: A Deep Learning Concept. 2020 IEEE 1st International Conference for Convergence in Engineering (ICCE) (pp. 68-71). IEEE.
[35] Kattyayan, S., Kar, T., &Kanungo, P. (2020, November). Performance Evaluation of Learning Based Frameworks for Devanagari Character Recognition. In 2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) (pp. 1-6). IEEE.
[36] Sawant, S., Sharma, A., Suvarna, G., Tanna, T., & Kulkarni, S. (2020, April). Word Transcription of MODI Script to Devanagari Using Deep Neural Network. 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA) (pp. 18-22). IEEE.
[37] [37] Dokare, I., Gadge, S., Kharde, K., Bhere, S., &Jadhav, R. (2021, May). Recognition of Handwritten Devanagari Character using Convolutional Neural Network. In 2021 3rd International Conference on Signal Processing and Communication (ICPSC) (pp. 353-359). IEEE.
[38] Pande, S. D., Jadhav, P. P., Joshi, R., Sawant, A. D., Muddebihalkar, V., Rathod, S., & Das, S. (2022). Digitization of handwritten Devanagari text using CNN transfer learning–A better customer service support. Neuroscience Informatics, 2(3), 100016.
[39] Shinde, A., & Dandawate, Y. Convolutional Neural Network Based Handwritten Marathi Text Recognition.
[40] Agrawal, M., Chauhan, B., & Agrawal, T. (2022). Machine Learning Algorithms for Handwritten Devanagari Character Recognition: A Systematic Review.
[41] Jayadevan, R., Kolhe, S. R., Patil, P. M., & Pal, U. (2011). Offline recognition of Devanagari script: A survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41(6), 782-796.