



Feature extraction and classification techniques for handwritten Devanagari text recognition: a survey

Sukhjinder Singh¹ · Naresh Kumar Garg² · Munish Kumar³

Received: 19 August 2020 / Revised: 12 April 2022 / Accepted: 30 May 2022 /
Published online: 9 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The character recognition system is a vital area in the field of pattern recognition. One interesting, complex, and challenging task is handwritten character recognition because of various writing styles of individuals. The accuracy of such systems highly depends upon the extraction and selection of features. Many researchers proposed a variety of feature extraction and classification methods for various scripts including Devanagari. In view of that, this article presents a broad study of feature extraction and classification methods considered so far for online and offline Handwritten Character Recognition (HCR) for Devanagari script, which is essential in Optical Character Recognition (OCR) research. This article presents techniques used by authors, the dataset used, the accuracy achieved by the methods of the work already available for the OCR research. This article is depicting the latest studies, research gaps, challenges and future perspectives for the researchers working in the Devanagari text recognition domain. Moreover, methods developed for feature extraction and classification in the area of Devanagari character recognition are presented in a systematic way as an assistance for future researchers. It has been gathered that traditional feature extraction and classifications methods are being replaced with deep learning methods to achieve higher recognition accuracy in this area.

Keywords Devanagari script · Handwritten character recognition · Feature extraction · Classification and deep learning

✉ Munish Kumar
munishcse@gmail.com

¹ Research Scholar, Department of Electronics & Communication Engineering, GZS Campus College of Engineering & Technology, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

² Department of Computer Science & Engineering, GZS Campus College of Engineering & Technology, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

³ Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

1 Introduction

Character recognition is an active research zone in the ground of pattern recognition. It automatically converts physical text information (numerals, letters, and symbols) into a corresponding digital format which is machine-readable [59]. There are two categories of character recognition, namely, online or offline. In online character recognition, one writes on an electronic surface such as an electronic tablet with a special pen or digitizer. Characters are particularly captured as a sequence of strokes, speed, and pen up/down data. In these systems, characters are recognized at real-time as soon as it is written [81]. Offline character recognition is the process of translating offline handwritten characters into a format that is understood by machines. Offline character recognition captures the information from a paper document by means of optically or magnetically scanning and hence can be further classified as optical character recognition and magnetic character recognition [57]. Offline character recognition becomes more challenging due to the shape of characters, a great variety of character symbols, document quality, and non-availability of stroke information [81]. Therefore, offline character recognition is a more challenging task than its online counterpart. The classification of character recognition has been depicted in Fig. 1.

Offline character recognition is significantly different from online character recognition as given in Table 1.

Recognition may be carried out for both printed and handwritten characters. The major issues of handwritten character recognition include a huge variety in the composition styles such as shape, speed of composing, and thickness of characters. At present, the printed character recognition framework yields more recognition accuracy as compared with handwritten character recognition frameworks. Hence handwritten character recognition has still constrained capacities. Further, handwritten character recognition may be carried out with or without using a segmentation approach. Accurate, robust, and reliable handwritten character

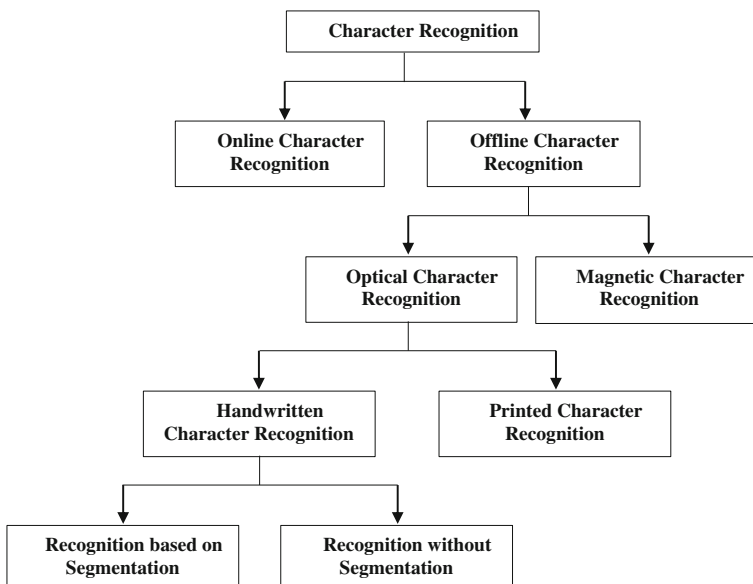


Fig. 1 Classifications of character recognition

Table 1 Online versus offline character recognition

Parameters	Online character recognition	Offline character recognition
Accuracy	Sufficiently higher	Lower
Availability of strokes	Yes	No
Raw data requirement	Samples per second	Dots per inch
Recognition speed	Sufficiently higher	Lower
Writing media	Digital pen on an electronic surface	Paper document

recognition by a computer system would be greatly useful for automatic number plate recognition, cheque reading, postcode recognition, signature verification, and as a reading aid for the blind, etc.

The remainder part of the paper is systematized as follows: Section 2 outlines the history, applications, and challenges for Devanagari Handwritten Character Recognition (HCR). Section 3 explores the overview of the Devanagari script. Section 4 presents the motivation for the readers and scholars working in the relevant domain. HCR methodology is presented in Section 5. In Section 6, the literature survey has been presented on feature extraction and classification methods considered for Devanagari HCR along with their comparative study. Research gaps are given in Section 7. Challenges for the present work are presented in Section 8. In Section 9, a few recommendations on future directions of Devanagari character recognition have been discussed. Finally, conclusions are presented in Section 10.

2 Background

To identify the significance of Optical Character Recognition (OCR) methods, in general, it is essential to present background information about the underlying problems, applications, and technical challenges. Methodically, OCR is a sub-component of pattern recognition and it gave the encouragement for developing pattern recognition and image analysis field. A brief history of machine recognition of scripts has been presented in Table 2.

2.1 Applications

Nowadays, there is a vast demand for techniques that attempt automatic optical character or script identification or recognition. Some types of techniques that cover various needs of different areas of such applications have been listed in this subsection [22, 104, 118].

- Airline ticket readers
- Automatic license plate recognition
- Bill processing systems
- Cheque reading
- Data classification through learning process
- Editing old documents
- Employee code reading/verification
- Forensic document analysis
- Form processing
- Handwritten notes reading

Table 2 History of machine recognition of scripts

Year(s) with Caption	Highlights
1870–1940 <i>Initial version of OCR</i>	<ul style="list-style-type: none"> • C. R. Carey and Nipkow invented retina scanner and sequential scanner respectively • Later in the development of devices to aid blinds • Punched cards are used to perform data entry.
1940–1950 <i>Modern version of OCR</i>	<ul style="list-style-type: none"> • OCR machines became commercially available in 1950
1950–1960 <i>1st OCR machine appears</i>	<ul style="list-style-type: none"> • Only chose text styles and shapes of characters could be read e.g. IBM 1418 • Limited number of fonts • Substantially more fit
1960–1965 <i>1st Generation OCR</i>	<ul style="list-style-type: none"> • Able to recognize regular machine-printed characters and additionally handwritten • Low cost and high performance e.g. IBM 1287
1965–1975 <i>2nd Generation OCR</i>	<ul style="list-style-type: none"> • Toshiba and Hitachi developed the first automatic letter sorting machine for postal code numbers and the first OCR machine for high performance with low cost respectively • American standard OCR character set was defined called as OCR-A. • European fonts were also designed called OCR-B, which had more natural fonts than the American standard
1975–1985 <i>3rd Generation OCR</i>	<ul style="list-style-type: none"> • Dramatic advances in hardware technology • Work was done on handwritten characters of poor print quality characters • More sophisticated OCR-machines started to appear in the market • Typing was a special niche for OCR
1985–1995 <i>4th Generation OCR</i>	<ul style="list-style-type: none"> • The sale increased considerably • Fit for scanning and recognizing characters from complex documents intermixed with texts, tables, and mathematical symbols
1995–2010 <i>Sophisticated and robust OCR</i>	<ul style="list-style-type: none"> • More complex OCRs were accessible for various scripts • Research had focused to achieve a fully automated information flow • Improving accuracy and efficiency in document processing • Preventing unnecessary errors • Image processing and pattern recognition techniques were efficiently combined with Artificial Intelligence (AI)
2010 onwards <i>OCR Today</i>	<ul style="list-style-type: none"> • OCR can be used in internet-connected mobile device applications that extract text captured using the device's camera • Use of deep learning approaches

- Human-robot interaction
- Library archival.
- Meaning translation
- Passport number reading/verification
- Postcode/pin-code recognition for postal automation
- Reading aid for blind and visually impaired users
- Recognition of ancient documents
- Sign board translation
- Signature verification
- Writer verification

Some major challenges have to be put out and handled in order to accomplish effective automation. These challenges have been discussed in the following sub-sections. Various challenges are identified and presented in this work, which will lead interest to the readers for character recognition.

2.2 Challenges for Devanagari HCR

High quality or high-resolution images (with some basic structural properties such as high differentiating text and background) are desired for achieving higher accuracy of the OCR system as it is directly dependent upon the quality of the input image. In order to achieve successful automation in OCR techniques, numerous errors have to be overcome as these often affect the quality of images dramatically [15, 52, 81, 118] and are clarified as below:

Aspect ratio Text may be short while other texts may be much longer such as traffic signs and video captions respectively. To detect text, a search procedure with respect to location, scale, and length of the text is taken into the account, which introduces high computational complexity.

Blurring and degradation Character sharpness is required for achieving greater accuracy of character recognition and character segmentation. Uneven focus results due to either a small point of view changes or catching a moving item. It results in blurring and degradation in input images which further reduces the accuracy of an OCR system [71].

Character complexity Moreover, handwritten Devanagari characters are more complex due to their structure and shape. They include a large character set with more curves, loops, and other details in the characters.

Complex background Working over a complex background may also be a much greater challenge for the OCR system than working with normal backgrounds.

Different shapes and size of characters Segmentation and classification become a challenging task for handwritten character recognition due to the different shapes and size of handwritten characters.

Existence of uneven illumination Capturing images in natural environments may result in uneven lighting and shadows. It may further cause less accurate detection, segmentation, and recognition due to degradation of the desired characteristics of the image that introduces a challenge.

Lack of standard test database Unfortunately, little standard handwritten character database of Devanagari script is available publicly as a benchmark for experimentation so that the effectiveness of recognition accuracies of various techniques can be compared on a common platform.

Larger character set is due to modifiers In the Devanagari script, there are upper and lower modifiers due to which two successive lines may overlap with each other. It may result in poor segmentation and hence lower recognition rate.

Low resolution Recognizing text captured in a photograph or scene text with low resolution remains an unsolved problem in OCR systems until the captured image is preprocessed with suitable preprocessing methods.

Noisy background Generally, it can be seen that noise gets added to the document/image during the scanning phase. Later, it becomes challenging to remove such background noise while performing digitization or binarization.

Physical and mental state of the writer Developing a framework for character recognition also poses a challenge to researchers due to the physical and mental state of the writer, writing instrument, pen width, ink color, and many other such factors.

Poor quality of documents These types of documents usually consist of holes, spots, noise, broken strokes, etc. which may result in the process of line segmentation very challenging.

Scene complexity Numerous man-made objects such as buildings, painting; appears in a natural environment having similar structural properties and appear as text. It imposes challenges in text recognition in the processed image making it difficult for OCR systems to distinguish text from non-text.

Similar-shaped characters Another challenge for character recognition is to recognize similar shaped characters or symbols. In Devanagari script, there exists many character pairs such as क-फ and व-ध that are quite similar in shape.

Skewness For optical character recognition systems, the skew correction has remained a challenge [19, 65] and various researchers have proposed easier and effective processes to correct the skewness of images such as an OJ method [80] that is suitable for any degree of rotation. Poor results may be observed if a skewed image is inputted directly into the OCR system without applying any suitable preprocessing method.

Speed of writing Characters can be represented as the trajectory drawn by the pen (up/down) on a writing medium. The nature of characters such as overlapping and touching also depends on the speed of writing which sometimes becomes a challenging problem during character recognition.

Variations of text layout or fonts Characters in cursive or italic style and script fonts of characters may cause difficulty in segmentation due to their overlapping with each other [111]. It will be difficult to recognize the characters when the class number is large i.e., has large within-class variations and from many pattern sub-spaces.

Various styles of human writing Every human being has his/her own and different style of writing which may cause difficulty to recognize the characters. Character size, shape, orientation, etc. varies from person to person.

Warping For OCR systems, warping or elastic deformation of the images could be another challenge where content or characters with varying geometry have to be recognized. Such a situation may arise when an image is captured using handheld cameras. Ulges et al. [112] and Meshesha and Jawahar [71] have found potential for rectification of warped document images called de-warping.

These factors may affect and create problems with incorrect character recognition by a computer system. Thus, there is a need for the method(s) which overcomes these challenges or

factors, so that the character recognition framework produces machine-readable correct digital form, automatically. These challenges much are considered during the designing and implementation of the recognition character system to make it more effective.

3 Overview of the Devanagari script

Devanagari belongs to the Brahmic family of scripts of India, Nepal, Tibet, and the South Asian subcontinent [2]. It is adopted by more than 500 million people and is being used for writing numerous languages viz. Hindi, Sanskrit, Marathi, Nepali including similar other languages of the South Asian subcontinent [23, 55]. The Devanagari script consists of 13 vowels, 34 consonants, and 14 modifiers of vowels as depicted in Fig. 2.

Moreover, apart from the above, it has compound or composite characters which may be formed by combining two or more basic characters. Compound characters and modifiers can be attached adjacent to each other, on the top side, or the bottom side of the basic character [21]. A vowel followed by a consonant may take a modified shape, depending on whether the vowel is placed to the left, right, top, or bottom of the consonant and are known as modifiers or matras. There is no idea of lower and upper characters and characters, including text and digits are written from left to right. Devanagari script has its own specified composition rules for combining vowels, consonants, and modifiers [13, 46]. An additional feature of Devanagari is the existence of a horizontal line on top of characters called a header line or shirorekha [30, 114]. Two or more characters are joined to form a word by joining header lines of individual characters. A word written in Devanagari script may be divided into strips viz. top, core, and bottom. This header line divides the top and core strips whereas the virtual baseline divides the core and bottom strips. Knowledge of scripting is important in a sense, if a person knows the script of a language, then he/she can easily read the words pertaining to that script on the basis of his/her mental dictionary. Figure 2 represents consonants and their corresponding half forms, vowels, and modifiers of the Devanagari script. Three strips of a word in the Devanagari script are depicted in the following Fig. 3.

4 Motivations for the readers

Research and development of HCR are growing throughout the world in various languages. Many researchers are currently working on the challenges of such systems to make the HCR

क ख ग घ ङ च छ ज झ ञ	अ आ इ ई	। ि ि
क् ख् ग् घ् ङ् च् छ् ज् झ् ञ्	उ ऊ ऋ ए	ँ ँ
ट ठ ड ढ ण त थ द ध न	ऐ ओ औ अं	ँ ँ
ट् ठ् ड् ढ् ण् त् थ् द् ध् न्	अः	ँ ँ
प फ ब भ म य र ल व श		
प् फ् ब् भ् म् य् र् ल् व् श्		
ष स ह ळ		
ष् स्र ह्र ळ्र		
	Vowels	Modifiers
Consonants and their corresponding Half Forms		

Fig. 2 Devanagari script (a) Consonants and their corresponding half forms (b) Vowels (c) Modifiers



Fig. 3 Three strips of a word in Devanagari script

system more accurate, robust, and reliable. In the present scenario, sufficient research work is available for the recognition of printed text written in non-Indic scripts such as Roman, Chinese, and Japanese. Some research work can also be traced for printed text written in Indic scripts such as Devanagari, Bangla, and Gurumukhi. For example, recently a lot of work has been presented for the Gurmukhi script [24, 25, 50, 62, 63]. The work is still going on and not matured yet to achieve higher recognition accuracy for recognition of documents written in Devanagari script within the optimal time duration. So, identifying a potential need, large application areas, and exciting challenges involved in this field is the key motivation for future researchers working in this area.

In India, many people use the Devanagari script for documentation and there has been a significant improvement in research related to Devanagari HCR systems. Although, researchers suggested different methods for online and offline HCR systems, yet they share a lot of common problems and solutions. Since offline HCR is more complex and hence requires more research compared to online and machine-printed recognition. Various library functions have been developed in MATLAB and OpenCV as a preprocessing phase for the recognition of characters. This paper attempts to address advancements in Devanagari HCR systems, especially their feature extraction and classification methods till 2022. Devanagari HCR has been the subject of intensive study in the last few decades, yet it is still an open subject to achieve the final frontier. Moreover, the study also reveals that there is a great need for efforts towards the progress of multilingual resources.

5 Handwritten character recognition approaches

In general, handwritten character recognition approaches can be broadly divided into two categories: firstly, the traditional approach that uses traditional methods of feature extraction and classification, and secondly, the deep learning approach as depicted in Fig. 4.

5.1 Image acquisitions or digitization

A handwritten paper-based document is scanned to produce a bitmap image or electronic form called digitization. It yields the digital image that can be fed to the pre-processing phase.

5.2 Pre-processing

It is a preliminary phase that aims to minimize the degradation of the acquired image and produces a normalized bitmap image. Pre-processing involves a number of steps like binarization, skeletonization, dilation of images, detection of edges, noise removal, image

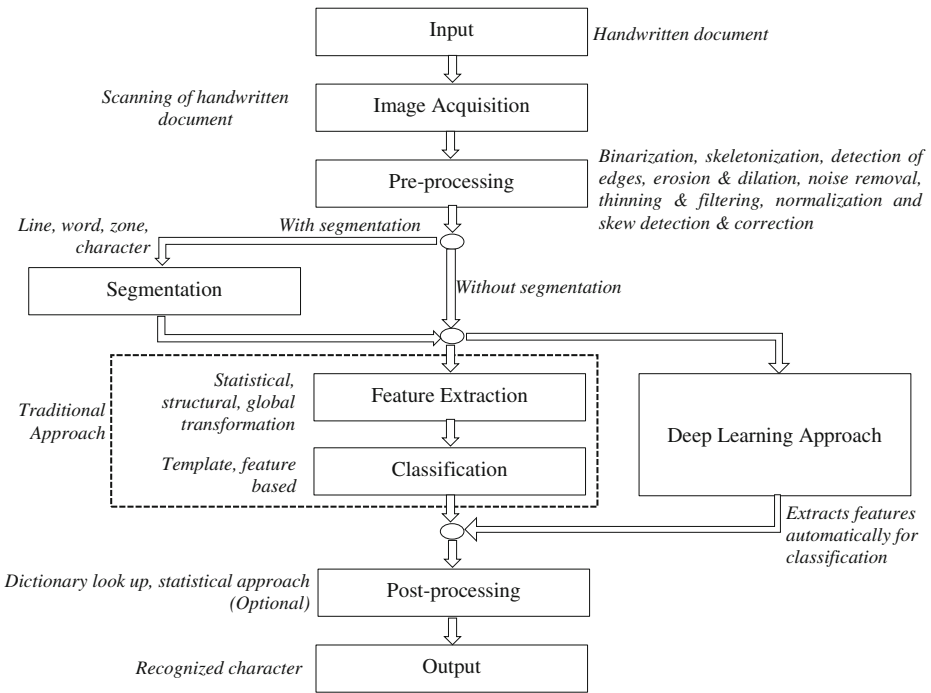


Fig. 4 Steps for handwritten character recognition

enhancement techniques for contrast stretching, thinning and filling, normalization, skew detection and correction [6, 34, 47].

Binarization It is a process that converts a grayscale image into a binary image (contains only two levels i.e., 0 and 1). Basically, binarization is used for separating foreground pixels from background pixels in an image using the required level of thresholding.

Skeletonization Foreground regions in a binary image are reduced to a skeletal remnant which is called skeletonization. It preserves the extent as well as connectivity of the original region and removes most of the original foreground pixels. Generally, it is applied to decrease the line width of the text from several pixels to a single pixel.

Detection of edges It involves the detection of edges or the selection of the outline of an object in the digitized image. Various edge detection operators such as Sobel, Canny, and first and second derivative methods can be applied for the detection of edges.

Erosion and dilation of images After locating the edges, dilation and erosion operations are used to increase or decrease objects in size to produce the pre-processed image suitable for segmentation. Erosion removes or erodes away the pixels on the image edges and results in a smaller object, whereas dilation produces a larger object by adding pixels around the image edges.

Noise removal It is carried out to remove those unwanted bits called noise that does not play a substantial role in a document for better processing. The morphological operation, filtering

such as median, Gaussian, mean, min-max and wiener, and noise modeling may be applied to remove noise from images.

Thinning and filling Thinning results improved visibility and structural information of characters in a scanned image by reducing the width while filling eliminates the gap, small breaks, and holes in digitized characters. Thinning removes selected foreground pixels in digitized characters and always extracts the features related to shape information characters [41].

Normalization Normalization is usually applied to improve the accuracy of OCR systems and it results in uniform character size, rotation, and slant by reducing the shape variation in a scanned document. It gives a remarkable reduction in data size without requiring any of the structure information of the image [66].

Skew detection and correction Skewness means the tilt or misalignment of the bit-mapped image of the scanned document. It may also happen by humans while writing a document. Skew detection and correction techniques are used to make such types of documents or images in correct alignment. These techniques include analysis of projection profile, Hough transforms, clustering, connected component, and correlation between line techniques.

5.3 Segmentation

In HCR, segmentation plays a significant role and it is used to break the scanned document into paragraphs, lines, words, and characters [93, 109]. Segmentation of handwritten characters is a challenging task due to a variety of writing styles [77]. The accuracy of HCR systems highly depends upon the detection of the best segmentation points for paragraphs, lines, words, and characters [34]. Segmentation is divided into the following parts:

Line segmentation It is a complex task and the initial stage of the segmentation phase. Researchers developed various techniques for line segmentation and are broadly divided into four groups based on projection profile, Hough-transform, smearing techniques, and thinning operations.

Word segmentation Word segmentation is used for dividing the handwritten text into words. Most of the existing techniques use a vertical projection profile for this purpose. The white space and pitch method are also used by various researchers to divide a handwritten text line into words.

Zone segmentation A line of Devanagari text can be divided into 3 horizontal zones, namely, upper or top, middle and lower or bottom. The upper zone and the middle zone are always partitioned by the header line which is named as shirorekha. The upper or top zone signifies the region or area above the headline, while the middle zone signifies the region or the area just below the headline and above the lower or bottom zone. The lower or bottom zone is the lowest part which contains some vowel components as part of vowel modifiers. In Hindi words, upper modifiers and lower modifiers have not always been necessary.

Character segmentation Character segmentation splits a text region into multiple regions of single characters. Vertical projection profile analysis was an early method for character

segmentation. Character segmentation involves extracting the individual characters without including some components of adjoining characters, even though these characters are not touching. Recognition-free and recognition-based segmentation are the methods used for character segmentation. It becomes more complex when characters are touching.

5.4 Feature extraction

Feature extraction plays a major role and an important phase in pattern recognition. The features represent precise information extracted from segmented characters (symbols or words) that distinguishes a particular character from the other characters. The recognition accuracy of an HCR system also depends upon the selection of the feature extraction techniques. Feature extraction can be carried out in a number of ways, however essential is to extract those features that can distinct dissimilar patterns or character classes that exist [113]. Features can be classified into the following major categories:

Statistical features Characteristics of the distribution of pixel values in the bitmap image are captured as statistical features. These features can be calculated from the statistical distribution of points viz. moments, zoning, histograms, or projection.

Structural features Structural features depict a pattern in terms of its topology and geometry by giving it local and global properties. These features are mainly based on geometrical properties of a symbol or character viz. loops, directions of strokes, intersections of strokes, and endpoints.

Global transformation-based features Global transformation techniques such as Fourier transform, discrete cosine transform, wavelet transform, Hough transform, and moments; have the ability to transform the pixel depiction into a corresponding denser form. These techniques generally signify the signal in terms of a linear combination of sequences of simpler well-defined functions. The sequence expansion gives a compact encoding by the coefficient of the linear combination.

Template matching based features It matches patterns pixels by pixels to identify them. Generally, character recognition in this approach does not need preprocessing such as thinning and pruning [16]. But these approaches are more sensitive to font and size variation of characters. These features are used to recognize compound characters and not suitable for noisy background documents.

Many researchers tried to amalgamate the above features, together, in order to achieve better feature extraction.

5.5 Classification

Classification or recognition is a decision-making phase that uses the features extracted in the earlier phase for deciding the class membership in the pattern recognition system [10]. It compares the input feature with the stored pattern and results in the best suitable matching class for input. It can be performed normally either using a template or feature-based methods.

Template-based method It involves the direct comparison between an unknown input pattern and an ideal pattern [11, 101]. The amount of correlation between these two patterns is considered for classification or recognition.

Feature-based method It extracts features from the input pattern and uses these features for classification or recognition models viz. like Artificial Neural Networks [14]; Kompalli et al. [52]; Pal et al., [82]; Jawahar et al., [45] Hidden Markov Models [100], Support Vector Machines [14, 35, 86] Modified Quadratic Discriminant Function [82, 97].

5.6 Deep learning approach

These approaches are proving prominent results and, in some cases, superior to human experts in the previous years. To improve the existing results, researchers are re-experimenting the existing problems by applying deep learning approaches. Researchers have been introduced different architectures of deep learning in recent year's viz. deep convolutional neural networks, deep belief networks, and recurrent neural networks. Nowadays, researchers are extensively using machine learning approaches for character recognition. The deep learning approaches are basically composed of multiple hidden layers, and each hidden layer consists of multiple neurons, which compute the suitable weight for the deep network.

5.7 Post-processing

This phase is not compulsory, sometimes is used to improve the accuracy of the HCR system. The accuracy of the HCR system can be increased if the output is constrained by a list of words that are permitted to occur in a document. It serves to further refine the results of the classification. Dictionary lookup and statistical approach are commonly used post-processing techniques for error correction [90].

6 Related work

There has been always a great need for research in the area of HCR for Indian languages, even though there are many challenges and a lack of a commercial market [105]. Research towards Indian HCR has gained much attention in recent years, even though basic research on Devanagari recognition was reported in 1977 [95] based on a structural approach. Various feature extraction methods for Devanagari HCR have been proposed in the past few decades and are briefly outlined in the following sub-sections.

6.1 Feature extraction methods

In this section, the feature extraction methods reported by various researchers towards this particular area are presented. Arica and Yarman-Vural [11], calculated both statistical and structural features while Bajaj et al., [17], considered density, moment, and descriptive component features for handwritten Devanagari numeral/character recognition. Elnagar and Harous [33], recognized handwritten Hindi numerals using end, branch, and cross point features based on strokes and cavity information about these features. Features based on

Zernike moments and zoning were extracted by Kaur in 2004 for the recognition of the Devanagari script. Kompalli et al. [52] and Kompalli et al. [53], extracted gradient, structural, and concavity (GSC) features for recognition of machine-printed and multi-font Devanagari text. Ramteke and Mehrotra [92] have extracted features based on moment invariants whereas Sharma et al. [97] used directional chain code information of the contour points of the characters as features for recognition. [42, 43] have proposed a box approach for the recognition of handwritten numerals that involve a spatial division of the numeral images into boxes. Moreover, Pal et al. [82] used chain code and gradient-based features for the recognition of Devanagari numerals. The information gained from the arctangent of the gradient and Gaussian filter was used as a feature for HCR in Pal et al. [83]. More and Rege [72], recognized handwritten Devanagari numerals using simple Geometric and Zernike moments. In recognition of handwritten Devanagari word, Shaw et al. [100] used histogram of chain-code directions in the image-strips as a feature vector taking an image-strips scanned from left to right by a sliding window. Kumar [56], carried out the comparative analysis of various feature extraction methods viz. Kirsch directional edges, distance transforms chain code, gradient, and directional distance distribution of the Devanagari handwritten dataset. Moreover, a new feature by quantizing gradient direction into four directional levels where each gradient map is alienated into 4×4 regions had been proposed in this paper. Bhattacharya and Chaudhuri [20] have extracted high-level features based on contour representations of all the four frequency components, i.e., high–high, high–low, low–high, and low–low of the wavelet-filtered image was considered for handwritten numeral recognition. To get better results of two similar-shaped handwritten characters, Wakabayashi et al. [115] discussed a feature extraction technique based on the Fisher ratio (F-ratio). Basu et al. [18], carried recognition or classification of handwritten digits using a Quad-Tree-based Longest Run (QTLR) feature. Rajput and Mali [91], have used Fourier Descriptors (FD) as features for recognition of handwritten numerals. Handwritten Devanagari compound characters had been recognized by Arora et al. [14] by calculating shadow and CH features. Aggarwal et al. [3], used the gradient representation for feature extraction. 7200-character samples used were normalized to 90×90 -pixel sizes. Experimental results using Support Vector Machines (SVM) exhibit high performance with a cross-validation accuracy of 94%. Pratap and Arya [90] have presented a general idea about the Devanagari character recognition system. An efficient character recognition system using Linear Discriminant Analysis (LDA) followed by a Bayesian discriminator function based on the Mahalonob is distance is proposed by Pourmohammad et al. [88]. In this paper, affine transformations were applied to the training samples in the first step to making the scheme robust against scaling and rotation distortion. A unique recognition system that uses wavelet features for classification and recognition was proposed by Dixit et al. [30]. The proposed system gives maximum accuracy of 70% over 2000 samples with 20 letters. Singh and Maring [106], used statistical and structured based feature extraction techniques viz. chain code, zone-based centroid, background directional distribution, and distance profile features for Devanagari HCR. They carried out experiments on more than 20,000 samples with varying image sizes: 30×30 , 40×40 , and 50×50 and achieved 97.61% overall accuracy with SVM. Statistical techniques that can be used for extracting features for handwritten character recognition were described by Ajmire et al. [5]. Tanuja et al. [110] proposed a system for handwritten Hindi character recognition using canny edge detection, distance transformation, and neural networks with backpropagation algorithms and achieved an accuracy of 95.0%.

Ansari and Sutar [9] have proposed an effective method for the recognition of isolated Marathi handwritten words for the Devanagari script. Gradient, distance transforms, regional and geometric features were computed and used as features of the images representing handwritten words. An overall recognition rate of 94.57%, was achieved for Feed-Forward Neural Network (FFNN) classifiers. The main recognition errors were observed due to abnormal writing and ambiguity among similar shaped words. The challenges involved in Indian postal system automation with a case study had been discussed and also, it throws light on the existing research literature support available for doing the postal automation [116]. A new kind of masking technique was used to extract the features using the Fisher discrimination function from ISIDCHAR (standard Devanagari database) and SVM classifier [67]. Authors improved significantly, recognition rate up to 96.58% in similar character recognition. An approach to feature extraction was proposed for handwritten Marathi characters (a version of Devanagari) using connected pixel-based features like area, perimeter, eccentricity, orientation, and Euler number [48]. They recorded the comparative accuracy of proposed methods and concluded that modified SVM gives high accuracy as compared to the KNN classifier.

Kumar et al. [60] recognized 3D handwritten Devanagari (3750 samples) words using BLSTM-NN classifier and achieved recognition performance of 50%, 68.10% 58.40%, and 63.80% respectively on features namely raw, convex, curvature, writing direction for Devanagari word samples. The authors achieved maximum accuracy using 3D curvature features for the Devanagari script. To overcome the shape similarity problem, Bhattacharya et al. [21] proposed a Sub-stroke-wise Relative Feature (SRF) for the recognition of online Devanagari cursive words. Authors achieved word recognition accuracy of 88.09% on 29,900 words as a dataset. Many researchers proposed several classification methods that utilize features extracted for the character or script identification, as described below. Kumar and Jindal [58] considered various features viz. zoning, diagonal, horizontal peak extent based, intersection and open-end point based along with different classifiers viz. k-NN, Linear-SVM, and MLP for the recognition of multi-lingual characters (English, Hindi, and Punjabi). Authors achieved 92.18%, 84.67%, and 86.79% recognition accuracy for character recognition of English, Hindi, and Punjabi respectively. Narang et al. [76] used statistical features (intersection points, open endpoints, centroid, horizontal peak extent, and vertical peak extent features) and classifiers (CNN, NN, Multilayer Perceptron, RBF-SVM, and random forest techniques) for the recognition of Devanagari ancient manuscripts by considering 6152 samples as database. The authors achieved 88.95% recognition accuracy using a combination of various features and classifiers.

Kumar et al. [64] explored hybrid features for the recognition of offline handwritten characters of Gurumukhi script. They analyzed the performance of their system by combining various features and classifiers along with AdaBoost approach. Authors obtained maximum accuracy of 96.3% on the corpus of 14,000 characters. Abuzaraida et al. [1] developed a system for the recognition of handwritten Arabic words based on structural features. Authors explored KNN classifier and obtained 99.10% of accuracy on the corpus of 2500 words. Kaur and Kumar [51] explored various feature selection approaches for the recognition of handwritten words. Authors achieved 87.42% of recognition accuracy on the corpus of 40,000 handwritten words (Gurumukhi) using Chi-Squared Attribute (CSA) based feature section and Random Forest (RF) classification.

6.2 Classification methods

The character recognition system has another important decision-making step called classification. In this, features are used to decide the class membership of various characters for their

recognition. Under this section, the classification methods adopted by various researchers towards this particular area are presented. Connell et al. [23] have achieved 86.5% recognition accuracy with no rejects, by combining multiple classifiers that focus on either local on-line property or global off-line properties of unconstrained Devanagari characters. Kaur [49] has taken the feature vector as the input of feedforward backpropagation NN to classify handwritten Devanagari characters. A quadratic classifier-based method is proposed by Sharma et al. [97]. Pal et al. [83] proposed a modified quadratic classifier for HCR. Arora et al. [12] presented a two-stage classification method for Devanagari HCR. Structural properties, namely shirorekha and spine in a character is extracted in the first stage, whereas intersection features are exploited in the second stage, which is further given to Feed-Forward Neural Network (FFNN) for classification. Hanmandlu et al. [43] have extracted different features like depending upon the availability of the vertical bar, Devanagari characters are classified into three classes viz. end-bar, middle-bar, and characters without any bar based. This coarse classification is performed prior to recognition and handwritten characters are recognized using a modified exponential membership function fitted to the fuzzy sets resulting from the features of the characters. The authors improved the speed of the learning process using a reuse policy. Deshpande et al. [27], introduced the role of Regular Expressions (RE) in Devanagari HCR. In this paper, the authors used chain-code features for translating handwritten characters into the encoded string. Two classifiers, namely, Support Vector Machines (SVM) and Modified Quadratic Discriminant Function (MQDF) were combined to achieve higher accuracy for character recognition [84]. Shaw et al. [99] proposed a segmentation-based method for handwritten Devanagari word recognition. Authors, segmented word images into pseudo-characters on the basis of the header line, and these pseudo-characters were further recognized using HMM. To recognize a handwritten word, continuous density HMM is also presented by Shaw et al. [100].

A dynamic programming-based method is proposed by Pal et al. [85] for recognition of pin code string. An Elastic Matching (EM) method is proposed on the basis of an Eigen Deformation (ED) for Devanagari HCR [72]. This method consists of two phases, namely, training (for ED estimation) and recognition phases. Pal et al. [86] carried out a comparative study of Devanagari HCR using various classifiers, namely, Compound MQDF (CMQDF), compound PD (CPD), Euclidean Distance (ED), k -NN, Linear Discriminant Function (LDF), Mirror Image Learning (MIL), Modified PD (MPD), MQDF, nearest neighbor, PD, Sub-space Method (SM) and SVM. The authors concluded that MIL classifier gives best results, whereas ED provides lowest results among above classifiers. A divide-and-conquer method is implemented for Devanagari HCR [4]. Hanmandlu et al. [44], classified, top modifiers of Devanagari script either as one touching-point or two touching-point modifiers. Further, classification was done by examining the core strip of the word. Devanagari non-compound handwritten characters are classified using two MLPs and a Minimum Edit Distance (MED) method by Arora et al. [14]. In the first phase, two MLPs are used to classify distinctly shaped characters and in a second phase, similarly shaped characters are classified using a MED method. Shelke and Apte [101] recognized Devanagari text using multistage feature extraction and classification methods. Structural features are extracted as an initial step, whereas Radon and Euclidean distance transforms have been carried out as the final step of feature extraction. These features are applied to two separate feed forward backpropagation neural networks. The hybrid classifier at the final stage takes the input from the two neural network classifiers and template matching classifier and results the final output based on a maximum voting rule. This method improves recognition accuracy over individual classifiers considerably as the proposed method

achieved a recognition rate of 95.40%. Kubatur et al. [55] have achieved a recognition rate of up to 97.2% using a neural network-based framework for HCR. Kale et al. [47], achieved an overall recognition rate of 98.25% and 98.36% for basic and compound characters respectively, considering a Legendre moment as feature descriptor and Artificial Neural Network (ANN) as a classifier. A novel part-based method is proposed in for recognizing the Devanagari characters by identifying its 40 basic classes [74].

A situation of a very large class recognition problem has been put aside by training models to classify an instance of one of these classes in any given test sample. The proposed approach gives a competitive performance than the results obtained by state-of-the-art features and classifiers for the DS1W2K dataset. Gradient Local Auto-Correlation (GLAC) algorithm is explored for HCR of Devanagari by taking two databases viz. ISIDCHAR and V2DMDCHAR [66]. In this paper, the best results obtained, using the SVM classifier on ISIDCHAR and V2DMDCHAR are 93.21% and 95.21%, respectively. Dongre and Mankar [31] used a multilayer perceptron neural network (MLP-NN) as a classifier on structural and geometric extracted features for recognition of Devanagari numerals and characters. They gained 93.17% and 82.7% recognition accuracies for numerals (using 40 hidden neurons) and characters (60 hidden neurons), respectively. Structural and directional features are extracted individually in each local zone for online HCR of Bengali and Devanagari scripts [37]. Furthermore, these features are given to SVM classifier after concatenation. Authors attained recognition accuracy of 87.48% and 84.10% for Bengali and Devanagari scripts using 4900 and 5000 test samples, respectively. Further, two zone-based feature extraction methods viz. Zone wise Structural and Directional (ZSD) and Zone Wise Slopes of Dominant Points (ZSDP) had been presented for online HCR of Bengali and Devanagari scripts [38]. Furthermore, these features are given to the SVM classifier for stroke recognition, and characters are recognized according to stroke combinations of characters with training data. It is observed that recognition performances for Bengali (9800 test data) and Devanagari (10,000 test data) scripts are 87.48% and 85.10% with ZSD, respectively, whereas, with ZSDP, the accuracies are 92.48% and 90.63%, respectively.

Pagare and Verma [81], implemented a dynamic model based on a Hopfield neural network for auto-associative recognition of Devanagari characters and numerals. Shelke and Apte [102], presented a novel approach based on multi-stage classification for the recognition of unconstrained handwritten Devanagari characters. This classification includes two steps; the first step is based on a fuzzy inference system and the second step is based on structural parameters. The recognition accuracy obtained by this method was 96.95%. Shelke and Apte [103] have presented techniques for optimizing recognition accuracy at various stages, namely, pre-classification, feature extraction, and recognition. Firstly, various structural features were used to classify characters into different classes as pre-classification. After that, features were extracted using optimized feature extraction methods, and finally, a neural network was used for recognition. Performances were analyzed by implementing different neural networks in this paper. Kumar et al. [61] recognized 3D handwritten Latin (2000 samples) and Devanagari (3750 samples) words using multiple Bidirectional Long-Short Term Memory Neural Network (BLSTM-NN) classifiers and Recognizer Output Voting Error Reduction (ROVER) framework. Authors achieved accuracies of 72.25% (for Latin) and 71.86% (for Devanagari) for their lexicon free approach. Mahesh and Sumit [68] proposed a handwritten Devanagari characters recognition system based on Deep Convolutional Neural Networks (DCNN) and adaptive gradient methods. Authors achieved the maximum recognition accuracy of 96.02% and 97.30% on ISIDCHAR database (36,172 characters), 96.45% and 97.65% on

V2DMDCHAR database (20,305 characters) and 96.53% and 98.00% on combining both databases (ISIDCHAR+ V2DMDCHAR = 56,477 characters) using DCNN and Layer-wise DCNN respectively with NA-6 and RMSProp optimizer. Further, the authors (2018b) proposed a method for the recognition of handwritten Devanagari characters using gradient-based features and an SVM classifier. They achieved 96.58% recognition accuracy on ISIDCHAR (36,172 characters) dataset in their work. Gupta and Bag [39], in their work, achieved character recognition accuracies of 95.10%, 95.57%, 96.09%, and 94.71% for the Hindi language (3000 words as a database) using random forest, SVM, MLP, and CNN classifiers, respectively. Narang et al. [75] presented a paper to recognize the Devanagari ancient manuscripts using AdaBoost and Bagging techniques. Authors achieved 90.70% (using DCT zigzag features and RBF-SVM classifier) and 91.70% (using adaptive boosting with RBF-SVM) recognition accuracy (maximum) for a database of 5484 samples. Narang et al. [78] carried out the recognition of Devanagari ancient characters (5484 samples) by considering SIFT, Gabor filter-based feature, and SVM-based Classifier. The authors achieved 91.39% recognition accuracy based on the tenfold cross-validation technique and poly-SVM classifier. Devi et al. [28] explored various machine learning algorithms (both supervised and unsupervised) namely Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) for recognition of handwritten characters. Authors concluded that KNN based system results the better recognition accuracy of 98%. Singh et al. [107] explored stroke classification based on RNN classifier for the recognition of Gurmukhi words using a corpus of 52,570 words. Authors achieved maximum of 98.67% accuracy on their corpus of collected words.

6.3 Deep learning based methods

In statistics and machine learning [96], classification algorithms such as Naive Bayes Classifier, Nearest Neighbour, Logistic Regression, Decision Trees, Random Forest, Neural Network, and KNN Classification basically analyze the training database so that classification of testing/target database can be done. Deore and Pravin [26] developed a dataset of a total of 5800 isolated images of 58 unique character classes: 12 vowels, 36 consonants, and 10 numerals. The authors implemented a two-stage VGG16 deep learning model for Devanagari handwritten character recognition. Their models gained 94.84% (First Model) and 96.55% (Second Model) testing accuracy with training loss of 0.18 and 0.12, respectively. Ghosh [36] extracted structural and directional features from publically available signature samples. They explored a deep learning network namely Recurrent Neural Network (RNN). They used two models namely Long-Short Term Memory (LSTM) and Bidirectional Long-Short Term Memory (BLSTM) for the recognition and verification of signatures in offline mode. Authors concluded that their proposed system based on RNN for signature verification performed better as compared with Convolutional Neural Network (CNN) and other state-of-the-art methods in terms of accuracy.

Narang et al. [79] used Convolutional Neural Network (CNN) for the recognition of various ancient manuscripts written in Devanagari script. They explored a deep learning model for the feature extraction and obtained 93.73% of recognition accuracy on the corpus of 5484 characters. Alrobah and Albahli [7] developed a system for the recognition of handwritten Arabic characters based on Conventional Neural Network (CNN) as feature extractor. Authors combined two classifiers namely SVM and eXtreme Gradient Boosting (XGBoost) to improve the recognition accuracy. They achieved 96.3% of recognition rate on the Arabic dataset

named as Hijaa. Singh et al. [108] developed a system for handwritten word recognition written in Gurumukhi script based on deep learning approaches. They adopted word-based approach of class labeling i.e. holistic approach so as to obtain satisfactory recognition results. Authors achieved 97% of the recognition accuracy for their dataset of Gurmukhi words.

Mushtaq et al. [73] developed a CNN architecture so as to recognize handwritten Urdu characters. Authors obtained 98.82% of recognition accuracy for their corpus of Urdu characters (74, 285 training and 21, 223 testing samples). Korichi et al. [54] developed a system for the recognition of Arabic handwriting based on Generic Feature-Independent Pyramid Multilevel Model (GFIPML). Authors used AHDB dataset for the performance evaluation of their system and achieved better results. Alrobah and Albahli [8] presented a comprehensive survey for the recognition of Arabic text using various deep learning approaches. Authors have identified some problems, issues and challenges in the recognition of Arabic text. Dey et al. [29] presented a system for the Recognition of Odia characters based on RNN and CNN. Authors have achieved 86.56% of recognition accuracy on their corpus of characters with 112 classes. Elkhayati et al. [32] developed an approach so as to segment Arabic words using Convolutional Neural Network (CNN) and Mathematical Morphology Operations (MMO) for recognition purposes. In their work, authors proposed directed CNN and achieved better results as compared with basic CNN.

Gupta and Bag [40] has compared the performance of segmentation-based and segmentation-free approaches for the recognition of Devanagari conjunct characters based on CNN and transfer learning. Authors used CNN-RNN hybrid architecture so as to minimize the intricacy of classification. They achieved 94.56% (analytic approach), 99.30% (CNN-based holistic approach) and 94.65% (CNN-RNN-based holistic approach) of recognition accuracies for various approaches adopted by them. Prashanth et al. [89] developed a corpus of 38,750 images of Devanagari numerals for recognition purpose. Authors explored various CNN architectures namely CNN, Modified Lenet CNN (MLCNN) and Alexnet CNN (ACNN) so as to recognize handwritten Devanagari numerals. Authors have achieved significant recognition results. Sachdeva and Mittal [94] developed a system for recognition of handwritten Devanagari compound characters using ResNet model of Convolutional Neural Network (CNN). They explored their in-house corpus for the experimental work and achieved good recognition results. Sharma et al. [98] developed a system for the recognition of Gurumukhi city name based on CNN model. Authors have obtained 99.13% of recognition accuracy on the corpus of 4000 words (city names) by exploring Adam optimizer with CNN model.

6.4 Comparative study

To provide a basic understanding and valuable assistance to the newer researchers in this field, brief summary of hand character recognition of Devanagari script in terms of various parameters are presented in Table 3.

Moreover, in Table 4, a comparative study of recognition results for Devanagari handwritten character recognition in terms of accuracy (%) have been presented using the same features extraction methods with dataset and classification methods considered.

The evaluation of the effectiveness of various methods has not been attempted here as the experiments are not carried out on the same standard dataset/benchmark. However, this study reveals that work done on the handwritten character recognition systems with good accuracy rates in Devanagari scripts is limited and presents a future direction.

Table 3 Brief summary of handwritten character recognition of Devanagari script

Authors	Feature	Classifier	Data Set (Size)	Accuracy (%)
Arora et al. [13]	Combined	Multi-Layer Perceptrons (MLP)	1, 500	89.58
Arora et al. [15]	Longest Run Shadow and Chain Code	MLP and Combinational	4900	90.74
Arora et al. [12]	Structural	Feedforward Neural Network (FFNN)	50, 000	89.12
Bhattacharya et al. [21]	SRF	HMM	29,900	88.09
Deshpande et al. [27]	Chain Code	Regular Expressions (RE) & MED	5, 000	82.00
Dixit et al. [30]	Wavelet	Backpropagation Neural Network	2000	70.00
Dongre and Mankar [31]	Geometric and Structural	MLP	4300	82.70
Ghosh and Roy [37]	Structural and Directional	SVM	5000	84.10
Ghosh and Roy [38]	Structural & Directional (ZSD) and Zone wise Slopes of Dominant Points (ZSDP)	SVM	10,000	85.10 (ZSD); 90.63 (ZSDP)
Gupta and Bag [39]	Shadow and cumulative stretch feature	Random forest; SVM; MLP	3000	95.10; 95.57; 96.09
Hammandlu et al. [43]	Vector Distance	Fuzzy	4750	90.65
Kale et al. [47]	Legendre Moment	Feed Forward Neural Network	27,000	98.25
Kubatur et al. [55]	Discrete Cosine Transform	Artificial Neural Network	2760	97.20
Kumar [56]	Gradient	Support Vector Machines (SVM)	25, 000	94.10
Kumar et al. [61]	Multiple BLSTM-NN	Raw, Convex, Curvature and Writing Direction	3750	79.46 (Lexicon based); 71.86 (ROVER combination)
Mahesh and Sumit [66]	GLAC	SVM	36,172; 20,305	93.21; 95.21
Mahesh and Sumit [67]	Masking	SVM	36,172	96.58
Mahesh and Sumit [68]	DCNN finds the best features automatically	DCNN and Layer-wise DCNN	36,172; 20,305; 56,477	96.02 and 97.30; 96.45 and 97.65; 96.53 and 98.00 respectively
Mahesh and Sumit [69]	Gradient based	SVM	36,172	96.58
Mane and Ragha [70]	Eigen Deformation	Elastic Matching	3, 600	94.91
Narang et al. [75]	DCT zigzag	RBF-SVM	5484	90.70
Narang et al. [76]	intersection points, open endpoints, centroid, horizontal peak extent and vertical peak extent features	CNN, NN, Multilayer Perceptron, RBF-SVM and random forest techniques	6152	88.95
Narang et al. [78]	SIFT and Gabor filter	Poly-SVM	5484	91.39
Pal et al. [83]	Gradient and Gaussian filter	Quadratic	36, 172	94.24
Pal et al. [84]	Gradient	SVM & Modified Quadratic Discriminant Function (MQDF)	36, 172	95.13

Table 3 (continued)

Authors	Feature	Classifier	Data Set (Size)	Accuracy (%)
Pal et al. [86]	Gradient	Mirror Image Learning (MIL)	36, 172	95.19
Pant et al.[87]	Geometric and Statistical	Radial Basis Function (RBF)	7380	80.25
Sharma <i>et al.</i> [97]	Chain Code	Quadratic	11,270	80.36
Shelke and Apte [101]	Multistage viz. Structural, Random Transform and Euclidean Distance	Multistage viz. Neural Network and Template Matching	16,000	95.40
Shelke and Apte [102]	Pixel Density	Multistage viz. Fuzzy Inference System and Structural Parameters	40,000	96.95

Table 4 Brief summary of handwritten character recognition of Devanagari script (feature wise)

Features	Authors	Classifier	Data Set (Size)	Accuracy (%)
Structural features	Arora et al. [12]	Feedforward Neural Network (FFNN)	50, 000	89.12
	Ghosh and Roy [38]	SVM	10,000	90.63
Statistical features	Shelke and Apte [102]	Multistage viz. Fuzzy Inference System and Structural Parameters	40,000	96.95
	Deshpande et al. [27]	Regular Expressions (RE) & MED	5, 000	82.00
	Hanmandlu et al. [43]	Fuzzy	4750	90.65
	Kale et al. [47]	Feed Forward Neural Network	27,000	98.25
	Mane and Raghya [70]	Elastic Matching	3, 600	94.91
Gradient features	Sharma et al. [97]	Quadratic	11,270	80.36
	Kumar [56]	SVM and MLP	25, 000	94.10; 91.90
	Mahesh and Sumit [66]	SVM	36,172; 20,305	93.21; 95.21
	Mahesh and Sumit [69]	SVM	36,172	96.58
	Pal et al. [83]	Quadratic	36, 172	94.24
	Pal et al. [84]	SVM & Modified Quadratic Discriminant Function (MQDF)	36, 172	95.13
	Pal et al. [86]	Mirror Image Learning (MIL)	36, 172	95.19
Global Transform features	Dixit et al. [30]	Backpropagation Neural Network	2000	70.00
	Kubatur et al. [55]	Artificial Neural Network	2760	97.20
Structural and Statistical features	Arora et al. [13]	Multi-Layer Perceptrons (MLP)	1, 500	89.58
	Arora et al. [15]	MLP and Combinational	4900	90.74
	Dongre and Mankar [31]	MLP	4300	82.70
	Ghosh and Roy [37]	SVM	5000	84.10
	Ghosh and Roy [38]	SVM	10,000	85.10
	Pant et al. [87]	Radial Basis Function (RBF)	7380	80.25
	Gupta and Bag [39]	Random forest; SVM; MLP	3000	95.10; 95.57; 96.09
Deep features	Kumar et al. [61]	Raw, Convex, Curvature and Writing Direction	3750	79.46 (Lexicon based); 71.86 (ROVER combination)
	Shelke and Apte [101]	Multistage viz. Neural Network and Template Matching	16,000	95.40
	Narang et al. [79]	CNN	5484	93.73
	Prashanth et al. [89]	a) CNN, b) Modified Lenet CNN (MLCNN) and c) Alexnet CNN (ACNN).	38,750	a) 96 b) 99 c) 99
	Gupta and Bag [40]	CNN-RNN-based	20,000	99.30

7 Research gaps

The following are some research gaps for future research directions in the field of optical Devanagari character recognition:

- a. Recognition of handwritten mathematical expression is still a challenging area in the field of character recognition.
- b. Character segmentation may result in additional problems due to overlapping, touching, and broken characters.
- c. It is desired and challengeable to detect best segmentation points for lines, words, and characters in isolation, to avoid incorrect segmentation as well as incorrect recognition. Segmentation of a handwritten text is a challenging task due to the variety of writing styles of individuals.
- d. Another challenging task is due to the non-uniform background which may cause poor recognition results.
- e. Recognition of historical documents is also a challenging problem due to the low quality of documents, availability of non-standard alphabets and unknown fonts, etc.
- f. Shape similarity of various characters written in Devanagari script such as क-फ, ख-स, घ-ध, थ-य, व-व, भ-म, and प-ष. The recognition of similar shaped characters is one of the main reasons for misclassification. It is challenging for researchers to identify the small difference (called the critical region) among similar shaped characters that are used by human beings to discriminate them.
- g. To make the document more attractive, people sometimes use artistic text (nonlinear) such as circular, triangular, curve, arc-form. Existing character recognition systems are unable to recognize artistic text. Hence there is a need to develop conversion models to translate artistic text into simpler linear text so that further character recognition can be carried out successfully in such situations.
- h. Typically, as the size of class space increases, it becomes more challenging to design a classifier and find out adequate samples of all possible classes for training.
- i. There is no general approach that is suitable for all kinds of documents such as degraded or historical and various environments.

8 Deep learning based approach and research challenges

Deep learning-based approaches may be applied to various fields of pattern recognition including character recognition [26]. This shall help to solve many complex tasks/steps such as feature extraction and classification of character recognition due to its powerful potential by adjusting the structure and parameters of various deep learning models. Although deep learning-based approaches have great potential to replace other conventional approaches, however, there are still some research challenges [117]:

- a. It is a challenging job to determine/decide the number of network layers in deep learning-based models and hence, a further number of neurons.
- b. There is a need for a larger dataset/database as the accuracy depends upon the training samples.

- c. There are various parameters of the networks of deep learning-based models and hence, determining/decide the optimal parameter is also a research challenge.
- d. To develop efficient deep learning-based models by reducing/minimizing various parameters viz. memory space, computational calculations, and bandwidth requirement are challenging jobs.

Nowadays, developing a character recognition framework using a deep learning approach is still worth exploring.

9 Suggestions on future directions

In the handwritten character recognition field, numerous directions are possible for future research as existed algorithms used for segmentation, feature extraction, and classification can be extended further for improving recognition accuracy of character recognition systems. The following are some suggestions on future research directions in handwritten character recognition:

- a. Development of appropriate and effective preprocessing techniques: Recognition accuracy can be improved by developing appropriate and effective preprocessing techniques such as detection and correction of degradation/wrapping, orientation, tilting in text. Further, a suitable technique can be developed for translating the artistic text into linear text so that the accuracy of character recognition systems can be improved.
- b. Preserve the shape of characters: After binarization or normalization, the characters may change their shape and significant information can be lost. Hence, there is a need to preserve the shape of the character.
- c. Refinement of segmented characters: Segmented characters may be refined in order to achieve better accuracy rates.
- d. Adding some more features: The performance of HCR systems can be improved by adding some more features to the existing features.
- e. Exploring combination of various classifiers: Researchers can combine the merits of one classifier (say Convolutional Neural Network) with another (say Recurrent Neural Network) to handle poor recognition accuracy due to various factors such as complex background and blur/noisy/poor quality documents.
- f. Use the different optimizers: Researchers may use the various optimizers with a deep learning approach where the deep convolution neural network can be trained with various optimizers to improve their recognition rates.
- g. Use of multiple classifier architecture: Character recognition results may be improved by combining decisions of different individual classifiers. Based upon the results produced by the individual classifier, the combination can be done according to their architecture such as cascading, parallel and hierarchical.

10 Conclusion

Hindi is the most widely spoken language in India, which is based on the Devanagari script. Devanagari is one of the working scripts for the Hindi language in government offices in India apart from English. In view of that, research on the Devanagari script is focused on in this

article so as to serve as a guide and update for readers working in the area of handwritten character recognition. This paper presents a widespread survey on feature extraction and classification methods considered so far for online and offline HCR for Devanagari script, which is essential in OCR research as presented in Tables 3 and 4. However, it is very hard to judge the success of the results of HCR systems in Devanagari script in terms of accuracy (%) as having different constraints, the data set (size), and sample space. There exists no assessment tool to test the performance of individual stages or phases as well as the overall performance of HCR systems. It has also been gathered that there exists always a trade-off between data acquisition quality and complexity of the methods, which limits the accuracy (%). In the past few years, a lot of efforts have been made by various researchers for HCR with Devanagari script; some major improvements have been achieved; however, machines cannot recognize human writing with the same fluency as humans. Moreover, available methods suffer from the lack of characterizing the handwriting generation and the perceptual process of reading, which comprises many complex phenomena.

There is a lack of a standard database on various Indic scripts for experimental work. Devanagari is one of these scripts. In this article, also various challenges are identified which will give a direction for future researchers. Even researchers are exploring the combination of multiple features to achieve good recognition accuracy. At present, there is no complete character recognition system available in India for Devanagari. There is a need to extract the features which will characterize the shape of handwritten characters in an efficient manner as the information lies in its shape and not color, texture, or edge. For text recognition, the traditional machine learning based methods mainly focus on feature extraction however the deep learning based methods mainly focus on the use of deep neural networks for effective learning. Moreover, researchers can adopt a deep learning approach for more general solutions that extract features automatically. Recognition of handwritten compound characters is still in the initial stage and the problem needs to be tackled. Future research will not be directly concerned with character recognition, but also words, phrases, and even complete document recognition.

Declarations

Conflict of interest We have no conflicts of interest in this work.

References

1. Abuzaraida MA, Elmehrek M, Elsomadi E (2021) Online handwriting Arabic recognition system using k-nearest neighbors classifier and DCT features. *International Journal of Electrical and Computer Engineering* 11(4):3584–3592
2. Acharya S, Pant AK, Gyawali PK (2015) Deep learning based large scale handwritten Devanagari character recognition. *Proceedings of the 9th international conference on software, knowledge, information management and applications (SKIMA)*, 1–6. <https://doi.org/10.1109/SKIMA.2015.7400041>
3. Aggarwal A, Rani R, Dhir R (2012) Handwritten Devanagari character recognition using gradient features. *International Journal of Advanced Research in Computer Science and Software Engineering* 2(5):85–90
4. Agrawal P, Hanmandlu M, Lall B (2009) Coarse classification of handwritten Hindi characters. *International Journal of Advanced Science and Technology* 10:43–54
5. Ajmire SP, Bagade KG, Ramteke PL (2015) An analytical study of handwritten Devanagari character recognition. *International Journal of Advanced Research in Computer Science and Software Engineering* 5(10):787–790
6. Alginahi Y (2010) Preprocessing techniques in character recognition. In: Minoru M (ed). *Character recognition*, 1–21. <https://doi.org/10.5772/9776>

7. Alrobah N, Albahli S (2021) A hybrid deep model for recognizing Arabic handwritten characters. *IEEE Access* 9:87058–87069
8. Alrobah N, Albahli S (2022) Arabic handwritten recognition using deep learning: a survey. *Arab J Sci Eng*:1–21
9. Ansari S, Sutar U (2015) Optimized and efficient feature extraction method for Devanagari handwritten character recognition. *Proceedings of the international conference on information process (ICIP)*, 11–15. <https://doi.org/10.1109/INFOP.2015.7489342>
10. Ansari S, Bhavani S, Sutar U (2016) Devanagari handwritten word recognition using efficient and fast feed forward neural network classifier. *Int J Adv Res* 4(10):2034–2043
11. Arica N, Yarman-Vural FT (2001) An overview of character recognition focused on off-line handwriting. *IEEE Trans Syst Man Cybern Part C Appl Rev* 31(2):216–233
12. Arora S, Bhattacharjee D, Nasipuri M, Malik L (2007) A two stage classification approach for handwritten Devanagari characters. *Proceedings of the international conference on computation intelligence and multimedia applications*, 399–403. <https://doi.org/10.1109/ICCIMA.2007.254>
13. Arora S, Bhattacharjee D, Nasipuri M, Basu DK, Kundu M, Malik L (2009) Study of different features on handwritten Devanagari character. *Proceedings of the 2nd international conference on emerging trends in engineering and technology*, 929–933
14. Arora S, Bhattacharjee D, Nasipuri M, Basu DK, Kundu M (2010) Recognition of non-compound handwritten Devanagari characters using a combination of MLP and minimum edit distance. *International Journal of Computer Science Security* 4(1):1–14. <https://doi.org/10.48550/arXiv.1006.5908>
15. Arora S, Bhattacharjee D, Nasipuri M, Kundu M, Basu DK (2010b) Performance comparison of SVM and ANN for handwritten Devanagari character recognition. *International Journal of Computer Science Issues* 7(3):1–10
16. Bag S, Harit G (2013) A survey on optical character recognition for Bangla and Devanagari scripts. *Indian Academy of Sciences* 38(1):133–168
17. Bajaj R, Dey L, Chaudhury S (2002) Devanagari numeral recognition by combining decision of multiple connectionist classifiers. *SADHANA* 27(1):59–72
18. Basu S, Das N, Sarkar R, Kundu M, Nasipuri M, Basu DK (2010) A novel framework for automatic sorting of postal documents with multi-script address blocks. *Pattern Recogn* 43(10):3507–3521
19. Bathla AK, Gupta SK, Jindal MK (2019) Character segmentation and skew correction for handwritten Devanagari scripts: a friends technique. *Asian Journal of Engineering and Applied Technology* 8(1):50–54
20. Bhattacharya B, Chaudhuri BB (2009) Handwritten numeral databases of Indian scripts and multistage recognition of mixed numerals. *IEEE Trans Pattern Anal Mach Intell* 31(3):444–457
21. Bhattacharya N, Roy PP, Pal U (2018) Sub-stroke-wise relative feature for online Indic handwriting recognition. *ACM Trans Asian Low-Resource Lang Inf Process* 18(2):1–16
22. Chaudhuri A, Mandaviya K, Badelia P, Ghosh SK (2017) Optical character recognition systems. *Optical Character Recognition Systems for Different Languages with Soft Computing, Studies in Fuzziness and Soft Computing* 352:9–41
23. Connell SD, Sinha RMK, Jain AK (2000) Recognition of unconstrained on-line Devanagari characters. *Proceedings of the 15th international conference on pattern recognition*, 368–371. <https://doi.org/10.1109/ICPR.2000.906089>
24. Dargan S, Kumar M (2019) Writer identification system for Indic and non-Indic scripts: state-of-the-art survey. *Archives of Computational Methods in Engineering* 26(4):1283–1311
25. Dargan S, Kumar M, Garg A, Thakur K (2020) Writer identification system for pre-segmented offline handwritten Devanagari characters using k-NN and SVM. *Soft Comput* 24:10111–10122
26. Deore SP, Pravin A (2020) Devanagari handwritten character recognition using fine-tuned deep convolutional neural network on trivial dataset. *Sadhana* 45(243):1–13
27. Deshpande PS, Malik SL, Arora TS (2008) Fine classification and recognition of handwritten Devnagari characters with regular expressions and minimum edit distance method. *J Comput* 3(5):11–17
28. Devi DP, Ramya R, Dinesh P, Palanisamy C, Kumar GS (2021) Design and simulation of handwritten recognition system. *Materials Today: Proceedings* 45(2):626–629
29. Dey R, Balabantaray RC, Mohanty S (2022) Offline Odia handwritten character recognition with a focus on compound characters. *Multimed Tools Appl* 81:10469–10495
30. Dixit A, Navghane A, Dandawate Y (2014) Handwritten Devanagari character recognition using wavelet-based feature extraction and classification scheme. *Proceedings of the annual IEEE India conference (INDICON)*, 1–4. <https://doi.org/10.1109/INDICON.2014.7030525>
31. Dongre VJ, Mankar VH (2015) Devanagari offline handwritten numeral and character recognition using multiple features and neural network classifier. *Proceedings of the 2nd international conference on computing for sustainable global development*, 425–431. <https://ieeexplore.ieee.org/abstract/document/7100286>

32. Elkhayati M, Elkettani Y, Mourchid M (2022) Segmentation of handwritten Arabic graphemes using a directed convolutional neural network and mathematical morphology operations. *Pattern Recogn* 122: 108288
33. Elnagar A, Harous S (2003) Recognition of handwritten Hindu numerals using structural descriptors. *Journal of Experimental & Theoretical Artificial Intelligence* 15(3):299–214
34. Farkya S, Surampudi G (2015) Hindi speech synthesis by concatenation of recognized and written Devnagri script using support vector machines classifier. *International Journal of Computer and Information Technology* 9(1):491–495
35. Gaur A, Yadav S (2015) Handwritten Hindi character recognition using K means clustering and SVM. *Proceedings of the IEEE 4th International Symposium on emerging trends and Technologies in Libraries and Information Services*, 65–70. <https://doi.org/10.1109/ETTLIS.2015.7048173>
36. Ghosh R (2021) A recurrent neural network based deep learning model for offline signature verification and recognition system. *Expert Syst Appl* 168:114249
37. Ghosh R, Roy PP (2015a) A novel feature extraction approach for online Bengali and Devanagari character recognition. *Proceedings of the 2nd international conference on signal processing and integrated networks (SPIN)*, 483–488. <https://doi.org/10.1109/SPIN.2015.7095313>
38. Ghosh R, Roy PP (2015b) Study of two zone-based features for online Bengali and Devanagari character recognition. *Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR)*, 401–405. <https://doi.org/10.1109/ICDAR.2015.7333792>
39. Gupta D, Bag S (2019) Handwritten multilingual word segmentation using polygonal approximation of digital curves for Indian languages. *Multimed Tools Appl* 78:1–26. <https://doi.org/10.1007/s11042-019-7286-0>
40. Gupta D, Bag S (2022) Holistic versus segmentation-based recognition of handwritten Devanagari conjunct characters: a CNN-based experimental study. *Neural Comput & Applic*, 1–17
41. Hanmandlu M, Murthy OVR (2007) Fuzzy model based recognition of handwritten numerals. *Pattern Recogn* 40(6):1840–1854
42. Hanmandlu M, Grover J, Madasu VK, Vasikarla S (2007a) Input fuzzy modeling for the recognition of handwritten Hindi numerals. *Proceedings of the 4th international conference on information technology*, 208–213
43. Hanmandlu M, Murthy OVR, Madasu VK (2007b) Fuzzy model based recognition of handwritten Hindi characters. *Proceedings of the 9th biennial conference of the Australian pattern recognition society on digital image computing techniques and applications*, 454–461. <https://doi.org/10.1109/DICTA.2007.4426832>
44. Hanmandlu M, Agrawal P, Lall B (2009) Segmentation of handwritten Hindi text: a structural approach. *International Journal of Computer Processing of Languages* 2(1):1–20. <https://doi.org/10.1142/S1793840609001993>
45. Jawahar CV, PMNSSK K, SSR K (2003) Bilingual OCR for Hindi-Telugu documents and its applications. *Proceedings of the 7th International Conference on Document Analysis and Recognition* 1:408–412
46. Jayadevan R, Kolhe SR, Patil PM, 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
47. Kale KV, Chavan SV, Kazi MM, Rode YS (2013) Handwritten Devanagari compound character recognition using Legendre moment: an artificial neural network approach. *Proceedings of the international symposium on computational and business intelligence*, 274–278. <https://doi.org/10.1109/ISCBI.2013.62>
48. Kamble PM, Hegadi RS (2017) Comparative study of handwritten Marathi characters recognition based on KNN and SVM classifier. In: Santosh, K., Hangarge, M., Bevilacqua, V., Negi, A. (eds) *Recent trends in image processing and pattern recognition*. RTIP2R 2016. *Communications in computer and information science*, vol 709. Springer, Singapore. https://doi.org/10.1007/978-981-10-4859-3_9
49. Kaur S (2004) Recognition of handwritten Devanagari script using features based on Zernike moments, zoning and neural network classifier. Ph.D. Thesis, Punjabi University, Patiala, India
50. Kaur H, Kumar M (2018) A comprehensive survey on word recognition for non-Indic and Indic scripts. *Pattern Anal Applic* 21(4):897–929. <https://doi.org/10.1007/s10044-018-0731-2>
51. Kaur H, Kumar M (2021) Performance evaluation of various feature selection techniques for offline handwritten Gurumukhi place name recognition. *Proceedings of data driven approach towards disruptive technologies*, 559–571
52. Kompalli S, Nayak S, Setlur S, Govindaraju V (2005) Challenges in OCR of Devanagari documents. *Proceedings of the 8th International Conference on Document Analysis and Recognition* 1:327–331
53. Kompalli S, Setlur S, Govindaraju V (2006) Design and comparison of segmentation driven and recognition driven Devanagari OCR. *Proceedings of the 2nd international conference on document image analysis for libraries*, 1–7. <https://doi.org/10.1109/DIAL.2006.12>

54. Korichi A, Slatnia S, Aiadi O, Khaldi B (2022) A generic feature-independent pyramid multilevel model for Arabic handwriting recognition. *Multimed Tools Appl* 81:1–21
55. Kubatur S, Sid-Ahmed M, Ahmadi M (2012) A neural network approach to online Devanagari handwritten character recognition. *Proceedings of the international conference on high performance computing and simulation (HPCS)*, 209–214. <https://doi.org/10.1109/HPCSim.2012.6266913>
56. Kumar S (2009) Performance comparison of features on Devanagari hand-printed dataset. *International Journal of Recent Trends in Engineering* 1(2):33–37
57. Kumar S (2016) A study for handwritten Devanagari word recognition. *Proceedings of the international conference on communication and signal processing*, 1009–1014
58. Kumar M, Jindal SR (2019) A study on recognition of pre-segmented handwritten multi-lingual characters. *Arch Comput Meth Eng*. <https://doi.org/10.1007/s11831-019-09332-0>
59. Kumar M, Jindal MK, RK RKS (2011) Review on OCR for handwritten Indian scripts character recognition. *Advances in digital image processing and information technology (DPPR)*. First International Conference on Digital Image Processing and Pattern Recognition (CCIS) 205:268–276
60. Kumar M, Jindal MK, Sharma RK, Jindal SR (2018) Character and numeral recognition for non-Indic and Indic scripts: a survey. *Artif Intell Rev* 52:2235–2261. <https://doi.org/10.1007/s10462-017-9607-x>
61. Kumar P, Saini RK, Roy PP, Pal U (2018b) A lexicon-free approach for 3D handwriting recognition using classifier combination. *Pattern Recogn Lett* 103:1–7
62. Kumar M, Jindal SR, Jindal MK, Lehal GS (2019) Improved recognition results of medieval handwritten Gurmukhi manuscripts using boosting and bagging methodologies. *Neural Process Lett* 50(1):43–56
63. Kumar M, Jindal MK, Sharma RK, Jindal SRGS (2020) Performance evaluation of classifiers for the recognition of offline handwritten Gurmukhi characters and numerals: a study. *Artif Intell Rev* 53:2075–2097
64. Kumar M, Jindal MK, Sharma RK, Jindal SR, Singh H (2021) Improved recognition results of offline handwritten Gurmukhi characters using hybrid features and adaptive boosting. *Soft Comput* 25(17):11589–11601. <https://doi.org/10.1007/s00500-021-06060-1>
65. Liang J, DeMenthon D, Doermann D (2008) Geometric rectification of camera-captured document images. *IEEE Trans Pattern Anal Mach Intell* 30(4):591–605
66. Mahesh J, Sumit S (2014) Gradient local auto-correlation for handwritten Devanagari character recognition. *Proceedings of the international conference on high performance computing and applications (ICHPCA)*, 1–5. <https://doi.org/10.1109/ICHPCA.2014.7045339>
67. Mahesh J, Sumit S (2016) Similar handwritten Devanagari character recognition by critical region estimation. *Proceedings of the IEEE International Conference on Advances in Computing, Communication and Informatics (ICACCI)*, pp 1936–1939
68. Mahesh J, Sumit S (2018a) Handwritten Devanagari character recognition using layer-wise training of deep convolutional neural networks and adaptive gradient methods. *Journal of Imaging* 41(4):1–14
69. Mahesh J, Sumit S (2018b) Handwritten Devanagari similar character recognition by fisher linear discriminant and pairwise classification. *International Journal of Image and Graphics* 18(4):1–18
70. Mane V, Ragha L (2009) Handwritten character recognition using elastic matching and PCA. *Proceedings of the International Conference on Advances in Computing, Communication and Control*, pp 410–415
71. Meshesha M, Jawahar CV (2008) Matching word images for content based retrieval from printed document images. *Int J Doc Anal Recognit* 11(1):29–38
72. More VN, Rege PP (2008) Devanagari handwritten numeral identification based on Zernike moments. *Proceedings of the IEEE region 10 conference on TENCON*, 1–6. <https://doi.org/10.1109/TENCON.2008.4766863>
73. Mushtaq F, Misgar MM, Kumar M, Khurana SS (2021) UrduDeepNet: offline handwritten Urdu character recognition using deep neural network. *Neural Comput & Applic* 33(22):15229–15252
74. Narang V, Roy S, Murthy OVR, Hanmandlu M (2013) Devanagari character recognition in scene images. *Proceedings of the 12th International conference on document analysis and recognition*, 902–906
75. Narang SR, Jindal MK, Kumar M (2019a) Devanagari ancient character recognition using DCT features with adaptive boosting and bootstrap aggregating. *Soft Comput* 23:13603–13614
76. Narang SR, Jindal MK, Kumar M (2019b) Devanagari ancient documents recognition using statistical feature extraction techniques. *SADHANA* 44(141):1–8
77. Narang SR, Jindal MK, Kumar M (2019c) Drop flow method: an iterative algorithm for complete segmentation of Devanagari ancient manuscripts. *Multimed Tools Appl* 78:23255–23280
78. Narang SR, Jindal MK, Ahuja S, Kumar M (2020) On the recognition of Devanagari ancient handwritten characters using SIFT and Gabor features. *Soft Comput* 24(22):1–11
79. Narang SR, Kumar M, Jindal MK (2021) DeepNetDevanagari: a deep learning model for Devanagari ancient character recognition. *Multimed Tools Appl* 80(13):20671–20686. <https://doi.org/10.1007/s11042-021-10775-6>

80. Obaida MA, Roy TK, Horaira MA, Hossain MJ (2011) Skew correction function of OCR: stroke-whitespace based algorithmic approach. *International Journal of Computer Applications* 28(8):7–12
81. Pagare G, Verma K (2015) Associative memory model for distorted on-line Devanagari character recognition. *Proceedings of the 5th International conference on advances in computing and communications*, 46–49. <https://doi.org/10.1109/ICACC.2015.42>
82. Pal U, Sharma N, Wakabayashi T, Kimura F (2007a) Handwritten numeral recognition of six popular Indian scripts. *Proceedings of the 9th International conference on document analysis and recognition (ICDAR)*, 749–753. <https://doi.org/10.1109/ICDAR.2007.4377015>
83. Pal U, Sharma N, Wakabayashi T, Kimura F (2007b) Off-line handwritten character recognition of Devnagari script. *Proceedings of the 9th international conference on document analysis and recognition (ICDAR)*, 496–500. <https://doi.org/10.1109/ICDAR.2007.4378759>
84. Pal U, Chanda S, Wakabayashi T, Kimura F (2008) Accuracy improvement of Devnagari character recognition combining SVM and MQDF. *Proceedings of the 11th International conference on Frontiers on handwriting recognition*, 367–372
85. Pal U, Roy RK, Roy K, Kimura F (2009a) Indian multi-script full pin-code string recognition for postal automation. *Proceedings of the 10th International conference on document analysis and recognition*, 456–460. <https://doi.org/10.1109/ICDAR.2009.171>
86. Pal U, Wakabayashi T, Kimura F (2009b) Comparative study of Devanagari handwritten character recognition using different features and classifiers. *Proceedings of the 10th International conference on document analysis and recognition*, 1111–1115. <https://doi.org/10.1109/ICDAR.2009.244>
87. Pant AK, Pandey SP, Joshi SR (2012) Offline Nepali handwritten character recognition using MLP and RBF neural networks. *Proceedings of the 3rd Asian Himalayas international conference on internet (AH-ICI)*, 1–5. <https://doi.org/10.1109/AHICI.2012.6408440>
88. Pourmohammad S, Soosahabi R, Maida AS (2013) An efficient character recognition scheme based on k-means clustering. *Proceedings of the 5th International conference on modeling, simulation and applied optimization (ICMSAO)*: 1–6. <https://doi.org/10.1109/ICMSAO.2013.6552640>
89. Prashanth DS, Mehta R, Raman K, Bhaskar V (2022) Handwritten Devanagari character recognition using modified lenet and alexnet convolution neural networks. *Wirel Pers Commun* 122(1):349–378
90. Pratap N, Arya S (2012) A review of Devnagari character recognition from past to future. *International Journal of Computer Science and Telecommunications* 3(6):77–82
91. Rajput GG, Mali SM (2010) Fourier descriptor based improved Marathi handwritten numeral recognition. *International Journal of Computer Applications* 3(4):9–13
92. Ramteke RJ, Mehrotra SC (2006) Feature extraction based on moment invariants for handwriting recognition. *Proceedings of the IEEE conference on cybernetics and intelligent systems*, 1–6
93. Ramteke AS, Rane ME (2012) Offline handwritten Devanagari script segmentation. *Int J Sci Technol Res* 1(4):142–145
94. Sachdeva J, Mittal S (2022) Handwritten offline Devanagari compound character recognition using CNN. In: Gupta, D., Polkowski, Z., Khanna, A., Bhattacharyya, S., Castillo, O. (eds) *Proceedings of data analytics and management. Lecture notes on data engineering and communications technologies*, vol 90. Springer, Singapore. https://doi.org/10.1007/978-981-16-6289-8_18
95. Sethi IK, Chatterjee B (1977) Machine recognition of hand printed Devanagari numerals. *Pattern Recogn* 9(2):69–76
96. Sethi R, Kaushik I (2020) Hand written digit recognition using machine learning. *9th IEEE international conference on communication systems and network technologies*, 49–54. <https://doi.org/10.1109/CSNT48778.2020.9115746>
97. Sharma N, Pal U, Kimura F, Pal S (2006) Recognition of off-line handwritten Devnagari characters using quadratic classifier. *Proceedings of the 5th Indian conference on computer vision, graphics and image processing*, 805–816. https://doi.org/10.1007/11949619_72
98. Sharma S, Gupta S, Gupta D, Juneja S, Singal G, Dhiman G, Kautish S (2022) Recognition of Gurmukhi handwritten city names using deep learning and cloud computing. *Sci Program* 2022:1–16. <https://doi.org/10.1155/2022/5945117>
99. Shaw B, Parui SK, and Shridhar M (2008a) A segmentation based approach to offline handwritten Devanagari word recognition. *Proceedings of the international conference on information technology*, 256–257. <https://doi.org/10.1109/ICIT.2008.32>
100. Shaw B, Parui SK, Shridhar M (2008b) Off-line handwritten Devanagari word recognition: a holistic approach based on directional chain code feature and HMM. *Proceedings of the international conference on information technology*, 203–208. <https://doi.org/10.1109/ICIT.2008.33>
101. Shelke S, Apte S (2010) A novel multi-feature multi-classifier scheme for unconstrained handwritten Devanagari character recognition. *Proceedings of the 12th international conference on Frontiers in handwriting recognition*, 215–219. <https://doi.org/10.1109/ICFHR.2010.41>

102. Shelke S, Apte S (2015) A fuzzy based classification scheme for unconstrained handwritten Devanagari character recognition. Proceedings of the International Conference on Communication, Information & Computing Technology (ICCICT), pp 1–6
103. Shelke S, Apte S (2016) Performance optimization and comparative analysis of neural networks for handwritten Devanagari character recognition. Proceedings of the international conference on signal and information processing (IConSIP), 1–5. <https://doi.org/10.1109/ICONSIP.2016.7857482>
104. Singh S, Garg NK (2019) Optical character recognition techniques for postal automation - a review. IEEE proceedings of the international conference on intelligent computing and control system, 1385–1389
105. Singh S, Garg NK (2020) Review of optical Devanagari character recognition techniques. In: *Satapathy S, Bhateja V, Janakiramaiah B, Chen YW (eds) intelligent system design*. Advances in Intelligent Systems and Computing 1171:97–106
106. Singh A, Maring KA (2015) Handwritten Devanagari character recognition using SVM and ANN. International Journal of Advanced Research in Computer and Communication Engineering 4(8):123–127. <https://doi.org/10.17148/IJARCC.2015.4825>
107. Singh H, Sharma RK, Singh V, Kumar M (2021a) Recognition of online handwritten Gurmukhi characters using recurrent neural network classifier. Soft Comput 25(8):6329–6338
108. Singh S, Sharma A, Chauhan VK (2021b) Online handwritten Gurmukhi word recognition using fine-tuned deep convolutional neural network on offline features. Machine Learning with Applications 5: 100037. <https://doi.org/10.1016/j.mlwa.2021.100037>
109. Smith R, Antonova D, Lee DS (2009) Adapting the tesseract open source OCR engine for multilingual OCR. Proceedings of the international workshop on multilingual OCR, 1–8. <https://doi.org/10.1145/1577802.1577804>
110. Tanuja K, Usha KV, Sushma TM (2015) Handwritten Hindi character recognition system using edge detection and neural network. International Journal of Advanced Technology and Engineering Exploration 2(6):71–75
111. Thakral B, Kumar M (2014) Devanagari handwritten text segmentation for overlapping and conjunct characters - a proficient technique. 3rd International Conference on reliability, Infocom technologies and optimization (ICRITO) (trends and future directions), 1–4. <https://doi.org/10.1109/ICRITO.2014.7014746>
112. Ulges A, Lampert CH, Breuel TM (2005) Document image Dewarping using robust estimation of curled text lines. Proceedings of the 8th International conference on document analysis and recognition (ICDAR) 2:1001–1005
113. Vamvakas G, Gatos B, Perantonis SJ (2009) A novel feature extraction and classification methodology for the recognition of historical documents. Proceedings of the 10th International conference on document analysis and recognition, 491–495. <https://doi.org/10.1109/ICDAR.2009.223>
114. Verma VK, Tiwari PK (2015) Removal of obstacles in Devanagari script for efficient optical character recognition. Proceedings of the international conference on computational intelligence and communication networks, 433–436. <https://doi.org/10.1109/CICN.2015.90>
115. Wakabayashi T, Pal U, Kimura F, Miyake Y (2009) F-ratio based weighted feature extraction for similar shape character recognition. Proceedings of the 10th international conference on document analysis and recognition (ICDAR), 196–200. <https://doi.org/10.1109/ICDAR.2009.197>
116. Wanchoo AS, Yadav P, Anuse A (2016) A survey on Devanagari character recognition for Indian postal system automation. Int J Appl Eng Res 11(6):4529–4536
117. Weng Y, Xia C (2020) A new deep learning-based handwritten character recognition system on mobile computing devices. Mobile Networks and Applications 25:402–411
118. Ye Q, Doermann D (2015) Text detection and recognition in imagery: a survey. IEEE Trans Pattern Anal Mach Intell 37(7):1480–1500

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.