Devanagari Text and Calligraphy Recognition Using ICF & ACF

V M Lomte, D. D. Doye
Research Scholar, Computer Science & Engineering, Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India
Professor, Department of Electronics & Telecommunication Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India

Abstract:
In India Multilanguage is used for communication. On an average, 600 million people use the Devanagari script for documentation. As it is a national script; hence, the information & communication technology era gives more scope to look at OCR strategies. Recognition of handwritten language text has been a demanding task in the computing field. This paper provides an adept approach for recognizing printed and handwritten documents for English and Devanagari Script. The major contributions in this paper are
1. Preparation of the datasets for handwritten Devanagari compound characters, words and calligraphy text
2. To provide an adept approach for recognizing English and Devnagari text
3. Analysis of pattern recognition methods based on Wavelet, Contour and SVM
4. Detailed discussion on results obtained for several text categories like Devanagari compound, alphanumeric, confusing, mirror characters and calligraphy.
The present work consists of wavelet transform, statistical parameters Inter correlation and autocorrelation functions, support vector machine and contour. Novelty of work is recognition of text and calligraphy using Correlation functions -ICF, ACF and template generation for new pattern. . Standard dataset HPL (Devnagari characters and numbers) and own dataset (2146 – mentioned categories) are used experimentation. The recognition accuracy achieved for printed text is 100%, handwritten text is 98.89 and devanagari compound character is 98.87. This throughput shows superior performance than existing work. Handwritten character recognition is useful solution for society for day to day activity.
Keywords: OCR – Optical Character Recognition, SVM -Support Vector Machine, ICF - Inter-correlation function, ACF -Auto-correlation function, Wavelet, Contour.
DOI: 10.24297/j.cims.2023.1.7

1. Introduction

In a Multilanguage country like India, people use Devanagri script to convert required data present in bookish hard copies into digital form and communication tool as a matrubhasha in substantial quantities. It is a national language information & communication technology era

giving more scope to contribute towards OCR-Optical Character Recognition strategies. Recognition of handwritten text always has a demanding task for the computing field. Today's highest priority is to develop an efficient and accurate OCR.

Devanagari script is chosen as our problem script as it is the world's fourth-largest spoken language widely used in many states of India. It is an ample opportunity for a researcher to address handwritten Devanagari character recognition issues, which would be help- hand towards archeologists, the historical, government sector, and most general-purpose documents recognition. Live input approach & template generation method for the unknown or new pattern.OCR is a familiar technique for the conversion of printed documents into digital documents.OCR converts Hardcopy of a paper into digital, editable form with the help of word processors. The traditional method document contains the manual process of filing and storing data in hardcopies. Hence it was time-consuming, with chances of typing errors and degrading paper quality for storing for an extended period.

2. Difficulties occurred in dev-ocr development:

Formation of a meaningful word, single characters joined with a header line called "Shiro Rekha." A single or broken line drawn on the upper side of n number of characters joined together. How to draw Shiro Rekha is dependent upon the writing habits of writers. It creates problem in isolating an individual character from the words. (Chaudhuri 1997)

Devanagari script consist of many vowel modifiers, anuswar, isolated dots, visarg & Chandra bindu are the bidy part of devnagari script which add up uncertainty to the recognition.
Formation of Composite is the process of combination of two characters:

म्र ज्ज व्य द्व

Recognition of Handwritten text is more tricky than printed because of the unique writing style and various writing tools. Applicability of Handwritten Devanagari Recognition will be in e-book generation, historical document preservation, form processing, mobile scanner recognition, etc. Handwritten text recognition is useful to recognize the handwritten prescription written by the doctor. In this many time while taking medication it is difficult to understand the text in the

prescription. Hence OCR development is also useful to recognize handwritten text and convert it in to digital format for better understanding. The Devanagari script is a combination of 47 basic characters, including consonants-33, vowels-14 and Devanagari script writing is significantly used. Pronunciation of the language reflects the script orthography.The compound character of the Devanagari text is the combination of two characters. Devnagari text recognition is a complicated process due to its confusing structure; hence, compound character and half character recognition are a big challenge for future researchers. Preprocessing, segmentation, feature extraction, classification are five fundamental stages of the image processing life cycle.

3. Literature review

Wavelet Transformation, Support Vector Machine (SVM), and Contour are effective in text recognition.

2.1 Wavelet Transformation

Chaudhuri B. B, et al., 1997[1] proposed a method for Wavelet decomposition-Single. It generates the approximation coefficients. The rate of recognition obtained for work is up-to 95%.G. G. Rajput et al., 2010[2] Worked on Recognition of the handwritten script. First, Document at 300 dpi scanned grayscale and stored. The 512 x 512 block size of an image has been considered for further processing. Binarization is done using Otsu's method. Feature extracted using discrete cosine transform & discrete wavelet transform ,K-Nearest Neighbour classifier used for classification. The database created contains 800 handwritten image blocks and each script contains one hundred. The researcher obtained recognition accuracy of 94.2% average value.Sushama Shelke et al., 2011[3] proposed a method modified Wavelet for multistage handwritten Marathi compound character recognition. The dataset contains 35,000 samples. An image preprocessed using morphological operations. Features extracted using modified wavelet features and the Multi-layer Perceptron classifier used. The accuracy obtained 96.23%. Binu P. et al., 2012[4] worked on method wavelet energy having a single hidden layer with extreme programming - feed-forward network. Dataset created by 3000 writers- 9,000 characters of 30 different classes collected from others. Accuracy for recognition obtained 90.69 %. Several wavelets classifiers evaluated on DB at level 3 & recognition rate got 67.33%.Shitala Prasad et al., 2012[5] proposed the recognition of handwritten character Multilingual data. They have first done data preprocessing using binary ratio-based normalization and filtering done by using Cascade interpolation decimation filter for noise reduction. Thresholding, Thinning and thickening did on image. Bounding boxes- Edge-based segmentation processed in

segmentation. Features extracted by Wavelet transform method and SVM and KNN classifiers used & Accuracy obtained -95.30%.Ashutosh Aggarwal et al., 2012[6] system developed using Gradient representation for extracting features. Results showed features with cross-validation method performance accuracy of 94%.Adwait Dixit et al., 2014[7] authors worked on Handwritten Devanagari character recognition. In preprocessing image was resized into pixel size of 300 * 300, morphological operations, normalization of input into 128x128 & Otsu's Segmentation performed on the image. Wavelet transform method extracted features .Artificial Neural Networks classifier used in the classification process. Resize factor was 64*64. 100 people have written 2000 characters for dataset preparation. Shirorekha plays an important role in individual letter class generation. The recognition accuracy for this work was 70%. This system can utilize further document analysis. Rajiv Kumar et al., 2014[8] proposed handwritten Devanagari digit recognition. In this research, Otsu's Method, morphological operation – erosion, is used for preprocessing. Labeling operation used for segmentation. Gradient & Wavelet transform used for feature extraction. Cascade neural network (CCN), Fitness function network(FFT) along with Neural network classifier which analyzed on dataset CPAR-2012 contains 82,609 handwritten isolated characters, numerals, 2,000 + 2,000(unconstrained & constrained pangram) text, and 4,000 digitized data collection forms. Hence the author titled this work benchmarking on the new dataset, with the recognition accuracy of 97.87%. The authors suggested that the researcher share results with other researchers to help future research and add to the dataset.Pratibha Singh et al., 2016[9] contributed a Devanagari handwritten text recognition. Elastic zoning and bounding box used for preprocessing. Segmentation processed by global and local histogram-based zone boundaries. The Gradient Features were extracted and provided for classification using Multilayer Perceptron (MLP). Dataset Numeral Dataset ISI-CVPR used, and recognition accuracy was 97.26%.Malika Ait Aider et al. [10], 2018, worked on recognizing handwritten character recognition. Wavelet transformation used feature extraction and SVM classified on the MINIST dataset with samples 12,000. Recognition accuracy obtained 98.76 %.Anirban Mukhopadhyay et al., 2020[11] proposed Indic Handwritten Script Recognition using Theory of Evidence Dempster–Shafer. Gray level co-occurrence matrix (GLCM) and Gabor wavelet transform used complementary sets of features in feature extraction for handwritten lines and words written by different 12 official scripts. Sixty feature values passed to MLP classifier. Hidden layers of 40 neurons with 0.8 learning rate as features of MLP Authors created their own database through a laboratory for handwritten documents. The size of the data set was 3600 text lines. The precision rate was GLCM – 90.58 %, and overall recognition accuracy was 92.06%. Panyam Narahari et al., 2020 [12]

Introduced method for Offline Handwritten Characters recognition 2D-FFT for English and Hindi language. Threshold 0.7, binarization preprocessing. Features extracted using 2D-FFT. The matching technique is used to match the input with the reference image. The data set contains 50 samples for both Hindi & English characters prepared by own. Performance evaluated & accuracy got 71.41%.

2.2 Support Vector Machine (SVM)

Shalini Puri et al., 2009 [13] introduced an efficient Devanagari classification for printed and handwritten documents. Input preprocessed by using Skew correction & normalization. The input information was segmented using project profiles- horizontal & vertical projection profiles. Features extracted using geometrical character structure contained black pixels with some values used as features. SL characters and SL modified characters geometric-based features extracted and used in training. SVM selected for classification. Data set contains documents prepared by two writers in all three languages, in total sixty data samples created for the dataset. 40% for testing & 60% of the dataset utilized for training & Testing. The accuracy obtained for Sanskrit printed images 98.77%, Hindi printed 100% handwritten documents accuracy of 99.23%, Marathi printed 99.86%, and for Handwritten 98.61%.Mariette Awad et al., 2015 [14], proposed a sparse kernel base method for Support Vector Machine. Result evaluated on prepared dataset 2000 sample. Accuracy obtained 94%.Sandhya Arora1 et al. 2012[15] analyzed SVM classifiers and ANN for recognition of handwritten Devanagari data. Statistical process investigated for segmentation & feature extraction using Shadow Features of character, Chain Code Histogram of Character Contour, View-based features, and Longest-run Features methods. Classification method comparison done on artificial neural network, multiple classifiers, the combination of kernel-based methods, neural classifier, Majority Voting, Support Vector Machine, and Weighted Majority voting. ISI dataset prepared training -3430 data samples and 1470 for testing and authored have prepared their dataset with data samples 2254. The average accuracy obtained was 97.91%.Jamimamul Bakas, et al., 2012[16] contributed to SVM classifiers. The authors compared different classifiers like KNN, SVM, MPL etc KNN gives good results as compared to others. Deepika Wadhwa et al. 2012[17] proposed online handwritten Hindi numerals recognition. Bezier interpolation method used for preprocessing. Features extracted using the Radius of curvature, and the image was classified using SVM. Recognition accuracy obtained 98.90%. The data set used contains 1000 samples. Authors quoted on results - direction angle and curvature features played a vital role in recognizing and increasing the recognition process. These two features give better results rather than being used

individually for character recognition.B. R. Bombade et al. 2016[18] proposed the HTDDC method for character recognition. Chain code features were extracted and Used SVM classifier for recognition. Author Worked on Recognition of Devanagari and Kannada text. Accuracy yields 94 % on the HCL dataset.Ujwal Sinh Vohra et al., 2016[19] proposed a method for Multilingual Recognition. Multilingual means text use of different languages like Devanagari and English for writing together. Binarization, skew correction, etc. used for Preprocessing. HOG feature extracted and passed to classification using SVM . Dataset prepared 50 samples per character and obtained accuracy 95.048 %. Authors mentioned accuracy challenges in paper-like broken characters containing lines, half written or incomplete, and several writing styles are few challenges in character recognition and degradation factors like Paper quality and handwriting style during inputting & preprocessing stage. While working special characters like!; : "? = % etc for processing, the author also quotes that system producing noise for some unknown characters and depicted as ¤.Martin Rajnoha et al., 2017[20] investigated a system for Offline Handwritten Text Recognition. An input was preprocessed by using thresholding. In segmentation, Rectangle used a Character window set 25x25 pixels. Features extracted by doing a binary conversion (625 features). Down sampling the input image max pooling-position invariance is used over a larger area. SVM classifier used for classification. 22,556 and 23,392 numerals of Devanagari and Bangla scripts are available in the database to evaluate the system performance. The accuracy obtained for multilingual characters is 92.86 %. Output post-processed in spell checker feature. Rajib Ghosh et al., 2019[21] worked on a combination of several classifiers –online text and non text handwritten devanagari script for the construction of centric ellipses around data samples. The data set is divided into many ellipse regions. Structural and directional feature values of ellipse regions analyzed for data available in the form text and non text got analyzed. Two classifiers Hidden Markov Model (HM) Support Vector Machine (SVM) are utilized for analysis of throughput. Dataset prepared contains a sample of images for 2000 text samples and 2000 non-text samples. The accuracy obtained for this work was 95.91%.

2.3 Contour

Naresh Kumar Garg et al. 2010,[22] produced efficient techniques for line segmentation. Hindi text considered as input data for line segmentation. Image segmented based on a header & baseline detection. Projection – 2-strip method used in segmentation process. Recognition accuracy obtained 97%.Madhav Vaidya et al., 2016[23] proposed OCR recognition based on Segmentation for the Marathi handwritten characters. Input image preprocessing using

binarization, Noise reduction, and normalization of data. Segmentation processes with n three stages as line, word and character. Segmentation accuracy yields 99%.Dhanya Sudarsan et al. 2018 [24] proposed a novel approach to recognize handwriting in Malayalam scripts. Contours were detected under segmentation and provided for the classification stage by using Convolutional Neural Network for classification. Data collected from museums for database preparation. The author quoted less overhead without compromising the accuracy of the proposed work.Pinaki Saha et al. 2020[25] have proposed a method for handwriting recognition using active contours. In preprocessing grayscales and smoothing using Gaussian algorithm filters and reshaping. Line, Word & character segmented in segmentation phase. Convolution neural network (CNN) technique extracted features also classified it. Recognition accuracy yields 95% & 79.3% for UCI and E-MNIST dataset respectively. Jyoti Pareek et al., 2020[26] chosen problem statement - Gujarati Handwritten Character Recognition. Scanning and Resizing, Noise Removal and Binarization, Skew Correction used in segmentation. Lines, and words, and then character segmentation used. The Histogram Projection Profile method got selected for the detection of Text Blob Contour. Convolution Neural Networks and MLP for classification. Data set prepared of 10,000 images from 250 different peoples for the Gujarati language. The accuracy obtained 64.87% for MLP & 97.21% accuracy yield using CNN. This review perceives that the Devnagari handwritten text recognition work has a lot of challenges. Hence researchers have more scope to work in this area. Devnagari text recognition is challenging due to its confusing structure and writing style of the writer. In Indian, the Devnagari script is very popular. The Marathi language is primarily spoken and used for documentation of Government offices in Maharashtra (India). Hence maximum attention should be given to Devanagari character recognition and focus on H – OCR technology. It isn't easy due to script structure complexity and handwriting styles, and medium.

4. Proposed Methodology

Devnagari handwritten character recognition is a promising task due to the difficulty level in its construction parameters and confusing nature. It makes it more interesting to recognize this script for history lovers and handwriting experts. Usually, Marathi textshave a combination of kana, matra, ukar & vanjan. (Boser 1992) The Marathi language is the commonly used communication tool for Government documentation in Maharashtra. So, the researcher should focus on the development of an efficient OCR system. Parameters like wide range of handwriting, medium, and style of writing, etc., take into consideration while working on OCR. (Pawar 2014)

The method introduced to recognize printed and handwritten Devanagari characters, numbers, confusing, complicated, poorly written, alphanumeric and joint characters, and printed and handwritten English characters. We have considered English alphabets, numbers, Devanagari mulaakshar, numbers, confusing, poorly written, complicated, alphanumeric, joint/compound, and mirror characters .Confusing characters are similar in shapes as described in Fig.ure 03. They have characteristics of similarities in attributes.

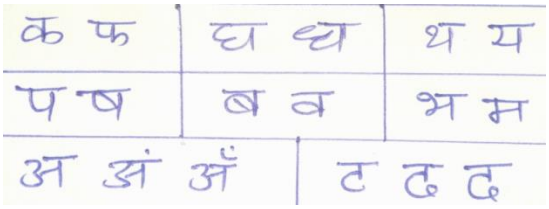


Fig. 1 a Similar shape characters

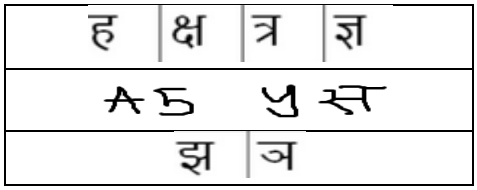


Fig .1 b Devnagari characters categories

Compound character recognition is a complex task in Devanagari script recognition. It is a combination of two different or same characters. Some of the characters possess similar shapes like म and भ so it is tough to recognize. When we move towards confusing joint characters like फ्क this, it becomes a complicated task and will increase recognition difficulty. Joint and similar nature characters frequently occur in the script, demanding developing a system for efficient and accurate recognition purposes. Recognition of such a combination of all types of characters under the Devanagari script provides the researchers an opportunity to work on complex real-time problems. Difficult characters of Devanagari handwritten character shapes are complicated in structure. So, it increases the difficulty level of recognition.

In Fig.ure 3b shows Alphanumeric character is a combination of alphanumeric characters written in numbers & text together. Some people are not good at handwriting. Their handwriting is not easy to understand because of poorly written text. It gives an excellent opportunity to the researcher to contribute to handwritten text recognition. (Bombade 2016)

Mirror Characters are available in English and Devnagari Marathi text like m & w, n & u, p & q, etc. similarly in Devanagari M ३६&९ ०r characters are

This work gives an approach for recognition of distinctive recognition methods to help to reduce the recognition process arduously. System has novel feature to recognize handwritten text in patterns. Patterns are like different color pens, sketch pens, black board chalk writing,

different, different shapes and style of writing etc. Handwritten Devnagari text in patterns is giving less as accuracy than normal text.

Table 1: System recognition parameters for images

Patten	Image type	Patten	Image type
P1		P2	
P3		P4	
P5		P6	
P7		P8	
P9		P10	
P11		P12	
P13		P14	

We have considered a total 14 patterns for recognition represented in table 1.

Fig.ure 2 elaborates system architecture designed to recognize printed and handwritten English & Devanagari texts efficiently. Online & offline used to accept the input. The offline input approach will take input as scanned paper images saved in the database. In the online or live input approach, place input written on paper in front of a webcam and take input live. The developed system is allowing inputting by both online and offline strategies. After capturing input, it is passed to the preprocessing phase, producing processed images and recognizing text by extracting features and classifying. An image is classified into an extract group and correctly identified after feature extraction. (Singh 2016)

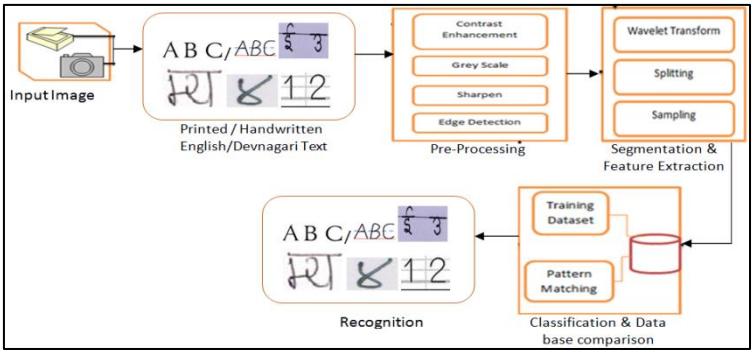


Fig. 2 Proposed HDTR System Architecture

Devanagari script is a combination of multiple characters with multiple shapes. The main base of forming a shape is a line.

$$m = ax + c \quad \text{.....(1)}$$

Intercept slop form Point slope form:

$$y - y1 = m(x - x1) \quad \text{.....(2)}$$

Vector counter r of length k defined as:

$$r = yo + y1 + y2 + y3 + \cdots + yk$$

As the scalar product of contours, r and N called such complex number:

$$n = (r, N) = \sum_{n=0}^{k-1}(Yn, Vn) \quad \text{.....(3)}$$

Equation 1 represents the line. We are forming vectors by using equation 2. Contour analysis-need to find contour; hence, equation 3 is contour identification.

Image Acquisition

An input image takes through a camera or scanner. File formats are JPEG, JPG, PNG, etc., with the maximum resolution image, increasing the output's accuracy. (Rajput 2010)We can use a stored image from the system or can capture the live image from a webcam. The handwritten characters are rest scanned and obtained in image format.

Preprocessing

Images undergo operations such as contrast enhancement, resizing, morphological operations, and binarization to reduce ambiguity and noise. This stage involves noise reduction, reducing blur effect, resizing, etc., and finally stored in the database. (Sultane 2017, 2018)Counter Analysis is used for pattern recognition, transporting the image for further processing. It detects the boundary of each character. A random complex number is used to assign each pixel of an image. Also, the addition of all complex numbers is zero. It also allows relating, accumulating, comparing, and finding the objects on exterior outline contours. This technique effectively resolves the main problem of transposition in pattern recognition, rescaling, and turning an item's image. (Kamble 2015)The extraction of character from an image is to be simple if the boundaries of the character can be found through a few methods. The boundaries of a character can be found in numerous ways relying upon numerous variables like the edges of the character, force of the content present in the information. (Yawalkar 2018)Contour is one of the best

methods to be applied on an image after detecting the edges of the character. To begin with, the primary pixel of a character should be recognized. This first pixel can be available at any point of the character. This striking of the position of the first pixel relies upon the manner in which we track the information picture to extract characters. Naturally, as the beginning of a picture lies at the top most pixels and the upper left corner of the character, when the primary pixel is distinguished, different pixels are distinguished by following the way along the edge from the first distinguished pixel. This assists with following along the edge of a character. According to this logic the edge recognition provides closed, there is consistently an opportunity that we arrive at the first recognized pixel subsequent to visiting any remaining pixels along the edge. These aides as a breaking condition for distinguishing the boundaries of the given character. When the boundaries of the character are known in the source input character, it tends to be extracted and used for additional references.

Initially take any input image convert it into binary form using thresholding and then obtain the edges using for given image character. (Singh 2015) The edges of each character object are found along with its internal information. Contours are more valuable in these kinds of recognition which are shaped with curved and closed loops and helps in discovering object boundaries present in the character. The Extracted characters then compared with template for matching. Resizing is done as per template size before matching.

The Canny Edge Detection Algorithm

To remove noise and unwanted details - smooth the image: Gaussian Filter.

$$g(i,j) = G\sigma(i,j) * f(i,j)$$

$$G\sigma = \frac{1}{\sqrt{2\pi\sigma^2}}\ Exp(-\frac{i^2+j^2}{2\sigma^2}\ )$$

Gradient $g(i,j)$ Computation: Intensity image has large magnitudes

$$K(i,j) = \sqrt{g_i^2}(i,j) + g_j^2(i,j)$$

$$L_T(i,j) = \{L(i,j) \quad if\ L(i,j) > T0 \quad Otherwise$$

$$T - all\ egde\ elements\ are\ kept\ by\ noise\ supression$$

Apply non-maximum suppression method to mark local maxima marked as edges

$$Each\ nonzero\ K_T\ (i,j) > Its\ two\ neibhourhoods\ with\ \ griadient\ k(i,j)$$

$$Keep\ L_T\ (i,j)\ is\ unchanged\ otherwise\ set\ 0$$

Apply double threshold to determine the potential edges

Edge tracking by hysteresis

Segmentation & Feature Extraction:

After preprocessing, the input passes to the segmentation phase. Segmentation separates the background and foreground parts of the image. (Vaidya 2017)We have selected Otsu's method and contour methods in our approach.

Algorithmic Steps:

1. Analysis of the Contour
2. Apply four filters on Contour Vector
3. Calculate Wavelet Convolution for Auto Correlation function

Where V = Auto Correlation function

j = 4 * i / Count // count = auto correlation count and value of i varies from 0 till Count value

4. Calculate Sum[i] = filter[j] * V

5.Calculate Auto Correlation Descriptors

AutoCorrDesc = 100 * sum[i] / count // count is auto Correlation count

6. Draw Rectangle with the help of x and y with height and width

Feature Extraction is a significant phase of the image processing life cycle. Here one of the most famous methods for feature extraction is Wavelet transform. It instigates the features such as translatability, multi-resolution and capability, etc.

The discrete wavelet transform [4] $f(x,y)$of size $M \times N$

$$W_{(j0,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)_{j0,m,n}\ (x,y)} \cdots\cdots \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(4)$$

$$W_{\varphi}^{i}(j,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\varphi_{j,m,n}^{i}\ (x,y) \ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(5)$$

Where

$$W_{j,m,n}(x,y) = 2^{j/2}\,\varphi\,(2^{j}x - m, 2^{j}y - n) \quad \text{.....(6)}$$

And

$$\mathrm{W}_{\mathrm{j}}^{\mathrm{i}}\mathrm{m,n}(x,y) = 2^{\frac{j}{2}}\varphi^{i}(2^{j}\,y - n) \quad \text{.....(7)}$$

Equations (4) to (7) represent the basic structure of the Wavelet.

We can extract some key features from an image like Geometrical and Topological, Global Transformation, Series Expansion, & Statistical Features types. (Bombade 2016)

Wavelet transform technique in equations (8) and (9) denotes Fourier Transform coefficients. Once we obtain these coefficients, then we can reconstruct, denoise, and compress an image. It represents the wavelet domain which can use to distinguish image content and noise. (Chacko 2012) (Dixit 2014)

$$X(k) = \int_{-\infty}^{\infty} x(m).e^{-k}\,dm \quad \text{.....(8)}$$

$$X(m) = \int_{-\infty}^{\infty} x(k).e^{k}\,dk \quad \text{.....(9)}$$

$$k = \sum_{m=0}^{m-1} \frac{-2d * Pi * m}{count}$$

$$X(m) = \int_{-\infty}^{\infty} x\left(\sum_{m=0}^{m-1} \frac{-2d * Pi * m}{count}\right).e^{k}\,dk \quad \text{.....(10)}$$

$$X(m) = -2d\int_{-\infty}^{\infty} \sum_{m}^{m-1} \frac{Pi * m}{count}.e^{k}\,dk \quad \text{.....(11)}$$

Above Fourier transform equation (10) & (11) decomposes an image to a complex exponential function of different image contours (Image uniform regions)

$X(k)$ = Multiple Points

Where k denotes multiple points of the contour

$\mathbb{X}(m)$ = Single Points

Where m denotes single points of the contour

Pi = All points of x and y index pointing to itself.

$$R^2 = \sqrt{a*a+b*b} \quad \text{.....(12)}$$

Here equation (12) is used to calculate the normalization function, i.e., calculating the Norma (R^2).

In Normalization equation (13), we calculated Inter Correlation Function (ICF) using Maximum Deviation (d)[43]

$$d = \sum_{i=0}^{count} \frac{P*a}{R^2} * \frac{P*b}{R^2} \quad \text{.....(13)}$$

Splitting attempts to divide an image into uniform regions:

Set ContourList = IMAGE

Repeat

Extract the first element of ContourList

If the region is uniform, then add to ContourArea

Else

split the region into four sub-regions and add these

to ContourList

Until (all regions removedfrom ContourList)

Initially, take the complete image to be the area of interest. Select the image portion where we want to work: AOI(Area of interest) and check if all pixels appear in the region, which satisfies similarity constraints. If we find AOI successfully, then the process continues until further splitting. When the areas are in one pixel, this happens in the worst case; else, it will divide the AOI into four equal sub-areas and, in each turn, consider each of the sub-areas as the AOI. Process will continue until no further splitting is required (until one area pixel). Sampling consists of digitization of coordinate values. In sampling, the image is split into several samples, each having its co-ordinate value. There are some variations in the sampled signal, which are random. Taking more samples reflects the collection more indirectly, resulting in better image quality

with less noise present.

Interactive template generation method:

The proposed method introduced a new feature of generating a new template for character and adding it to the database.

The template method is a unique feature of the system. We can generate a template for a new character, number, alphanumerical, joint character, etc., which is not available in the database. That can be added into the database and can be used further for future reference to recognize that character, number, etc. Using this feature, we can also generate new templates for any unique character. Fig.. 5 represents template generation for a new pattern of character.

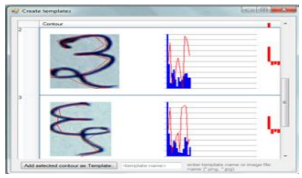


Fig. 3 Template generation

3.4 Classification:

Support Vector Machine: It is a classification method in the statistical learning theory era, which introduced great success in pattern recognition and bioinformatics applications. Here SVM plays a vital role in maximizing the distance of separating boundaries among the two or more than two classes (c1,c2, and so on). (Puri 2019)The distance between separating planes is maximized by SVM . The width of the margin between the two classes, say c1 and c2is the SVM optimization criteria. The decision boundary of the empty area specified by the nearest training pattern called Support Vectors. SVM's aim to find a hyper plane in N-dimensional space(N — the number of features) that distinctly classifies the data points. (Wadhwa 2012) (Singh 2015)

Dataset are divided into training data and testing data for classification and recognition(Vohra 2016). Based on extracted features, classier work on classification stage – support vector machine. Characters classified on matching with particular classes. Contours were detected around the character. Multiple contours of a single character get stored and used in detection for different users and handwriting styles.

5. Data Base Preparation

We have prepared our database of around 1000 character/number samples from various writers for Handwritten English & Devanagari – Marathi characters. We have also used the HPL standard dataset used for Devanagari characters and numerals- There are 20 classes of characters with 270 samples of each 111 . We have also considered a few images for English and Marathi printed paragraphs containing 300 + characters. We can make the dataset large by adding more templates into it by template generation facility in real-time.

HPL dataset is downloaded from

https://archive.ics.uci.edu/ml/datasets/Devanagari+Handwritten+Character+Dataset

Devanagari language contains various characters similar in shape, structure, curves, segments, with only a slight difference notable by a human. Examples of such characters are ख and य, म and भ, घ and ध, etc. The system is trained on such characters precisely to avoid ambiguity by the system with such characters. We are considering 16 confusing characters for recognition and 30 samples for fourteen patterns.

6. Results and discussion

Results yield 99.20, 98.89 and 98.87 for printed, handwritten and devanagari compound character images. devanagari compound characters respectively. We considered character samples from various writers, the dataset is highly scalable. If a particular character is not available in the database then a new template can be generated by using the interactive template generation method and can be added to the database for future reference. We are taking fourteen handwritten patterns of Devanagari text like sketch pen drawing, blackboard writing, letters written in a cartography, etc.The developed method gives 94.11% accuracy for recognizing confusing characters, 80% accuracy for complicated images and patterns-calligraphy 45.4%.

Fig. 4 English & Devanagari Character Recognition

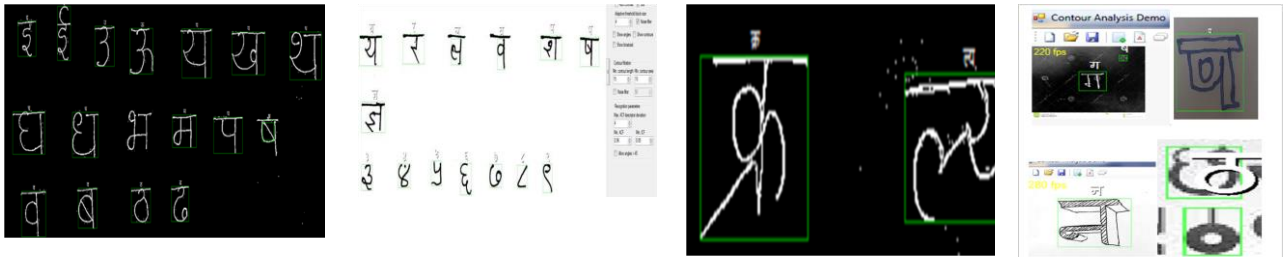


Fig 5 Devanagari Character Recognition for different categories

Figure 4 representing recognition of printed text in English & Devanagari, which indicates superior recognition in accuracy and processing speed. Handwriting varies according to the writer' s style. It makes it difficult to recognize handwritten text written in Marathi & English. Fig.ure 5 shows Recognition of Handwritten English & Devanagari Character, indicates good results.

The Proposed work is efficient to recognize English/Devanagari printed & handwritten text. It also gives opportunity to recognize confusing characters, mirror characters and few patterns. In this research 17 similar shape character samples are taken for recognition and found system was responding with impressive results. We are come with new approach for pattern recognition which successfully recognizes 14 different patterns of devnagari text like character written by chock on board with pattern, calligraphy etc. gives good recognition rate.

Table & Graph

Table 2: System recognition parameters for images

Image (P - Patterns)	Adaptive Threshold Block	Contour Filterization		Recognition Parameters		
		Length	Area	Max. ACF Descriptor Deviation	Min. ACF	Min. ICF
P1 - I1	4	15	10	4	0.96	0.85
P1 - I2	4	15	10	4	0.96	0.85
P1 - I3	4	15	10	4	0.96	0.85

P1 - I4	4	15	10	4	0.96	0.85
P2 - I5	7	16	21	4	0.38	0.45
P2 - I6	9	19	21	4	0.96	0.85
P3 - I7	17	15	15	4	0.57	0.85
P4 - I3	26	37	43	44	0.96	0.85
P5 - I9	30	15	10	4	0.96	0.85
P6 - I10	4	15	10	4	0.7	0.8
P7 - I11	4	15	10	4	0.68	0.87
P8 - I12	9	15	10	4	0.64	0.072
P9 - I13	9	15	10	4	0.64	0.72
P10 - I14	4	15	10	4	0.89	0.77
P11 - I15	4	15	10	4	0.85	0.65
P12-I16	19	27	34	27	0.96	0.85
P13-I17	10	15	10	4	0.96	0.85
P14-I18	31	15	10	4	0.96	0.85

Above table2 shows different statistical parameter used in recognition process, which represents recognition parameters using ACF & ICF Feature extraction techniques & contour filterization.

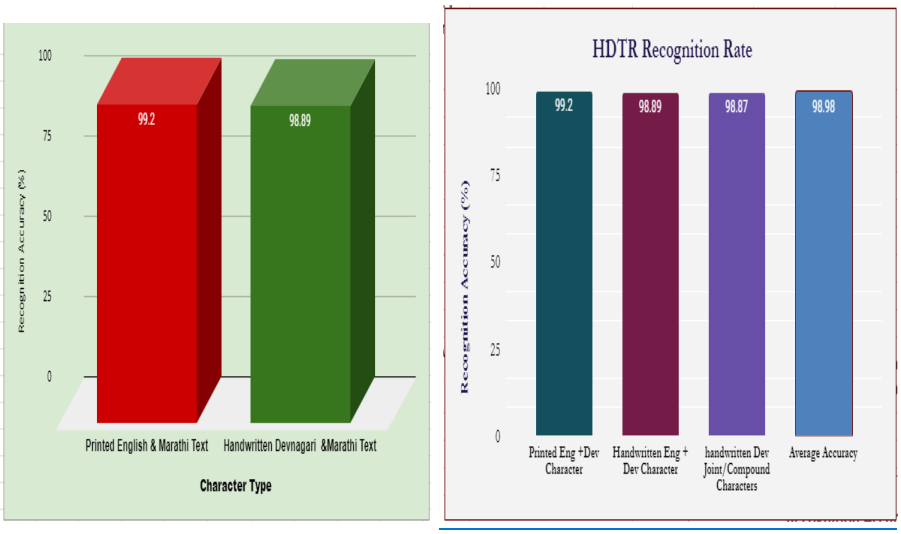
Table 3: Comparison of features considered.

Properties	Existing Methods	Proposed Work
Camera-Based real-time recognition	No	Yes
Multiple Script Recognition	Yes	Yes
Contour Analysis	Yes	Yes
Template generation and reuse	No	Yes
Customization in processing	No	Yes
Output at each stage	No	Yes

Table 3 indicates proposed method consist of properties listed above which to make system more robust.

The classification techniques using SVM classifiers with their accuracy are listed in table 3. Accuracy comparison for Marathi Compound Characters with other feature extraction methods is shown in table 4.

Outcomes for English and Devanagari printed, handwritten text recognition shown in Fig.ure 06, which represents results for Handwritten Devnagari Confusing Character, Alphanumeric, Compound characters and patterns.



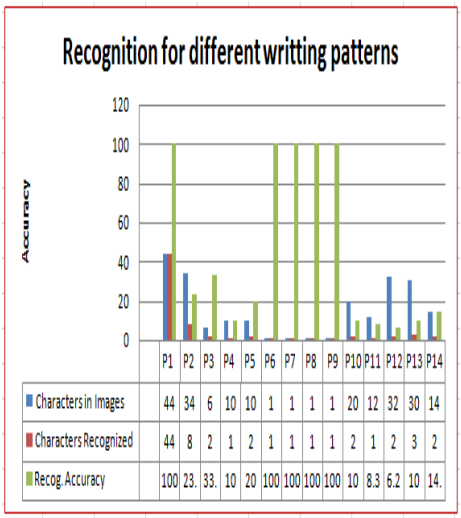
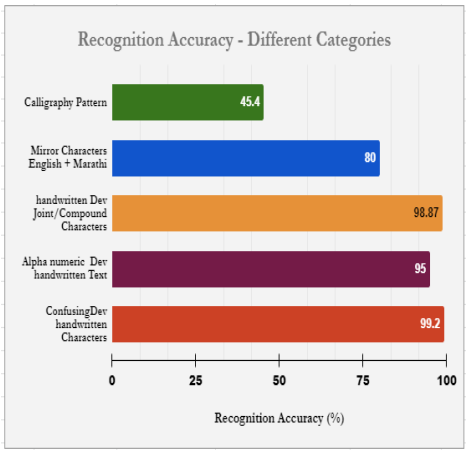
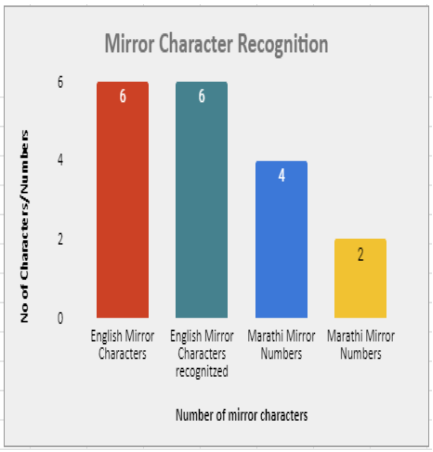


Fig. 6 Recognition Accuracy for printed & handwritten text.

Fig. 7 Recognition accuracy for different Handwriting patterns





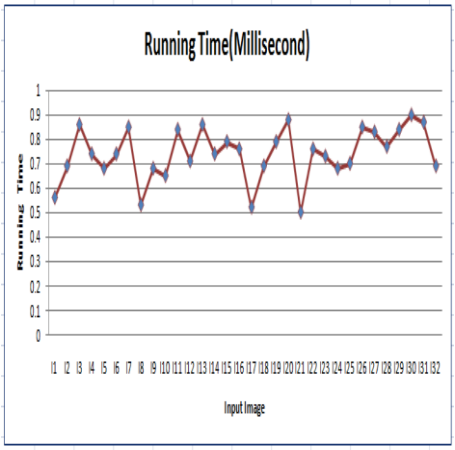


Fig. 8 Accuracy for Recognition handwritten Devanagari text for different categories. Fig. 9 System elapse Time

Fig.ure 6 shows accuracy obtained for printed & handwritten text with some categories. Fig.ure 7 represents variation in accuracy for input images, recognized images and recognition rate. Accuracy prediction for various recognition categories like mirror characters for English & Marathi as shown in Fig.ure 8. The processing time required to obtain results for various input images are shown in Fig.ure 09.

Table 4: Comparative Analysis with Existing Systems

Reference work	Existing approach	Dataset	Accuracy(%)
Chaudhari B. B. et al.,1997 [1]	Wavelet decomposition – Single, Multilayer perception, K-Nearest Neighbour	--	95
G. G. Rajput et al., 2010 [2]	D.C.T. & DWT, K-NN classifier	800 handwritten characters	94.2

Sushama Shelke et al., 2011[3]	Modified Wavelet	35000	96.23
Binu P et al., 2012[4]	Wavelet Energy, SLFN	9,000	90.69
Shitala Prasad et al.,2012 [5]	Wavelet Transform, KNN,SVM	64 Hindi characters	95.30
Ashutosh Aggarwal et al., 2012[6]	Gradient feature, SVM	7200	94
Adwait Dixit et al., 2014[7]	Wavelet transform (db1 family)	2000 characters	70
Rajiv Kumar et al., 2014[8]	Gradient & Wavelet Transform, KNN classifier	CPAR-2012	97.87
Pratibha Singh et al., 2016[9]	Gradient feature-stochastic gradient descent	ISI-CVPR	97.26
Malika Ait Aider et al, 2018 [10]	Wavelet Transform, SVM	MNIST database	98.76
Anirban Mukhopadhyay et al., 2020 [11]	Gabor Wavelet & Gray level co-occurrence matrix	3600 samples	92.06
Panyam Narahari et al, 2020 [12]	2D –FFT, Nearest Neighbourhood Classifier	1500	71
Sandhya Arora et al., 2012[15]	statistical methods, artificial neural networks, support vector machines	2254	97.91
Ujwal Sinh Vohra et al 2016 [19]	Support Vector Machine,	50 sample per characters	95.04
Martin Rajnoha et al.,2017 [20]	Support Vector Machine	22,556	92.86
Nilesh Kumar Garg et al. 2010 [22]	Contour	300	97
Pinaki Saha et al., 2020 [25]	Contour	UCI dataset E-MNIST	95 79.3
Jyoti Pareek et al., 2020 [26]	Contour	10,000 Gujrati Language	MLP -64.87 CNN - 97.21

B. R. Bombade, et al. 2016 [18]	Support Vector Machine	HPL dataset	94
Hammandu et al., 2007	Fuzzy set	HPL dataset	88.58
Arora et al.,2009	Shadow ,chain code histogram, MPL	HPL dataset	89.71
Arora et al., 2011	Chain code histogram, MPL	HPL dataset	92.67
Pal et al., 2008	Gradient curvature features, Modified QDF, Support vector machine	HPL dataset	93.84
Vidhya et al., 2013	Support vector machine ,Positional features	HPL dataset	84.52
Mukherji et al., 2009	ACD coding features, Fuzzy set	HPL dataset	92.70
Deshpande et al., 2008	Directional features, Min Edit Distance	HPL dataset	85.93
Proposed Method - HDTR	Contour +Wavelet-Co-relation Factors +SVM	HPL + Own dataset	**98.89**

We have compared our system with existing methods for Devnagari character recognition plotted in table 4.Many researchers have studied the Wavelet feature extraction technique, the accuracy of the wavelet classifier indicates well compared to other methods. Handwritten English and Devanagari text provide remarkable printed Devanagari and English text recognition. It's also giving excellent recognition results for handwritten text recognition for Devanagari and English text.

Table 5: Comparative analysis for recognition accuracy -Marathi Compound Characters

Reference work	Existing approch	Dataset	Accuracy (%)
Pratibha Sing et al., 2016[9]	Wavelet, SVM	12000	98.76
Prashant Yawalkar et al ,2018[33]	--	1050	40
Snehal Goiait et al. 2017 [37]	discrete wavelet transform	--	95

Snehal Goiait et al. 2016 [38]	Minutiae Extraction	--	90
Amol A. Kadam et al., 2019 [40]	zoning-based , support vector machine and K-NN classifier	3500	SVM -96.49 KNN- 95.67
Sushma Shelke et al. 2014[41]	Wavelet ,SVM	--	96.23
Proposed Method-HDTR	Contour +Wavelet-ICF+ACF +SVM	2146	**98.87**

Table 5 gives detailing of Devanagari compound characters recognition comparative analysis. From the above tables 4 & 5 can be concluded that our system is working very well for handwritten Devanagari confusing characters in similar shapes and another category also. It is giving superior results than pervious methods. The system is contour-based, so it is not giving good results for broken handwritten Devanagari text.

Statements and Declarations

Funding

The authors did not receive support from any organization for the submitted work.

Conflicts of interest/Competing interests

Authors have no conflict of interest.

7. Conclusion

This work introduces an efficient technique which contains Wavelet, Contour and support vector machine together. It contributes in printed & handwritten English, Devanagari character recognition. The performance evaluated on HPL dataset & created dataset on parameters-accuracy & time. Average recognition accuracy achieved 98.98 % and for handwriting patterns 45.40%. Finally, it concludes the HDTR method is efficient with template generation add on feature which improve performance in recognition rate and processing speed (millisecond) along with live and offline input approach. But, our method gives less accuracy for calligraphy text. System may show poor performance for break characters due to use of contour which will be a big challenge for future researchers.

Acknowledgment

I am pleased to express my gratitude to all who contributed their valuable help to complete this paper successfully. I am grateful to my guide Dr. D. D. Doye, for his valuable guidance and continuous support for the same. I also wish to extend my special thanks to Dr. U. V. Kulkarni, and Dr. B. R. Bombade for their expert suggestions.

I am happy to take this opportunity to thank the reviewers of this paper for putting the expertise and efforts into reviewing my research work, without which it would not be possible to maintain a high standard of peer-reviewed journals.

References

1. Chaudhuri B B, Pal U Skew angle detection of digitized Indian script documents,(1997) IEEE Trans Pattern Anal Mach Intell 19(2):182–186, https://doi.org/10.1109/34.574803.
2. Rajput, G. G., & Anita, H. B. (2010). Handwritten script recognition using DCT and wavelet features at block level. IJCA, Special issue on RTIPPR (3), 158-163.
3. Shelke S V, Dhangare (2014) A H Handwritten Character Recognition using Wavelet Transform for Feature Extraction International Journal of Multidisciplinary Educational Research ISSN: 2277-7881; IF-2.735; V:5.16; Vol 3, Issue 3(7).
4. Chacko, B. P., Vimal Krishnan, V. R., Raju, G., & Babu Anto, P. (2012). Handwritten character recognition using wavelet energy and extreme learning machine. International Journal of Machine Learning and Cybernetics, 3(2), 149-161.
5. Prasad S., Verma, G. K., Singh, B. K., & Kumar, P. (2012). Basic handwritten character recognition from multi-lingual image dataset using multi-resolution and multi-directional transform. International Journal of Wavelets, Multiresolution and Information Processing, 10(05), 1250046. https://doi.org/10.1142/S0219691312500464.
6. Aggarwal, A., & Renudhir, R. R. (2012). Recognition of Devanagari handwritten numerals using gradient features and SVM.
7. Dixit, A., Navghane, A., & Dandawate, Y. (2014, December). Handwritten Devanagari character recognition using wavelet based feature extraction and classification scheme. In 2014 Annual IEEE India Conference (INDICON) (pp. 1-4). IEEE.
8. Kumar, R., & Ravulakollu, K. K. (2014). On the performance of Devnagari handwritten character recognition. World Applied Sciences Journal, 31(6), 1012-1019,DOI: 10.5829/idosi.wasj.2014.31.06.2062

9. Singh, P., Verma, A., & Chaudhari, N. S. (2015). On the performance improvement of Devanagari handwritten character recognition. Applied Computational Intelligence and Soft Computing, 2015.https://doi.org/10.1155/2015/193868
10. Aider, M. A., Hammouche, K., & Gaceb, D. (2018). Recognition of handwritten characters based on wavelet transform and SVM classifier. Int. Arab J. Inf. Technol., 15(6), 1082-1087.
11. Mukhopadhyay, A., Singh, P. K., Sarkar, R., & Nasipuri, M. (2020). Handwritten Indic script recognition based on the Dempster–Shafer theory of evidence. Journal of Intelligent Systems, 29(1), 264-282.
12. Panyam Narahari Sastry, Syed Sameer, Mohammed Sameer Syed(April 2020) Recognition of Offline Handwritten Characters using 2D-FFT for English and Hindi Scripts, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-4.
13. Puri, S., & Singh, S. P. (2019). An efficient Devanagari character classification in printed and handwritten documents using SVM. Procedia Computer Science, 152, 111-121, https://doi.org/10.1016/j.procs.2019.05.033.
14. Awad, M., & Khanna, R. (2015). Support vector machines for classification. In Efficient learning machines (pp. 39-66). Apress, Berkeley, CA.
15. Arora, S., Bhattacharjee, D., Nasipuri, M., Malik, L., Kundu, M., & Basu, D. K. (2010). Performance comparison of SVM and ANN for handwritten devnagari character recognition. arXiv preprint arXiv:1006.5902.
16. Bakas, J., Mahalat, M. H., & Mollah, A. F. (2016). A comparative study of various classifiers for character recognition on multi-script databases. Int J Comput Appl, 155(3), 1-5.
17. Wadhwa, D., & Verma, K. (2012). Online handwriting recognition of hindi numerals using svm. International Journal of Computer Applications, 48(11)., 10.5120/7391-0250.
18. Bombade B. R., Doye D. D., Kulkarni U. V.(2016), An Efficient Method for Devanagari Character Recognition,(International Journal of Mathematics and Computation, ISSN 0974-570X(Online), Vol. 27, Issue No. 2.
19. Vohra, U. S., Dwivedi, S. P., & Mandoria, H. L. (2016). Study and Analysis of Multilingual Handwritten Characters Recognition using SVM Classifier. Oriental Journal of Computer Science and Technology, 9(2), 109-114.
20. Rajnoha, M., Burget, R., & Dutta, M. K. (2017, February). Offline handwritten text recognition using support vector machines. In 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 132-136). IEEE.

21. Ghosh, R., Shanu, S., Ranjan, S., & Kumari, K. (2019). An approach based on classifier combination for online handwritten text and non-text classification in Devanagari script. Sādhanā, 44(8), 1-8.
22. Garg, N. K., Kaur, L., & Jindal, M. K. (2010, April). A new method for line segmentation of handwritten Hindi text. In 2010 seventh international conference on information technology: new generations (pp. 392-397). IEEE.
23. Vaidya Madhav, Joshi Yashwant Joshi, Bhalerao Milind (August 2017) Segmentation Based Optical Character Recognition for Handwritten Marathi characters, International Journal in Advance Research in Science & Engineering, Vol. 06, Issue-17,.
24. Sudarsan, D., & Joseph, S. (2018, September). A novel approach for handwriting recognition in malayalam manuscripts using contour detection and convolutional neural nets. In 2018 international conference on advances in computing, communications and informatics (ICACCI) (pp. 1818-1824). IEEE.
25. Saha, P., & Jaiswal, A. (2020). Handwriting recognition using active contour. In Artificial Intelligence and Evolutionary Computations in Engineering Systems (pp. 505-514). Springer, Singapore.
26. Pareek, J., Singhania, D., Kumari, R. R., Purohit, S. (2020). Gujarati handwritten character recognition from text images. Procedia Computer Science, 171, 514-523.
27. Boser, B. E. (1992). isabelle M. Guyon, and Vladimir N. Vapnik. "A Training Algorithm for Optimal Margin Classifiers." COLT, 92.
28. Pawar, V. R., & Gaikwad, A. (2014). Multistage Recognition Approach for Offline Handwritten Marathi Script Recognition. International Journal of Signal Processing, Image Processing and Pattern Recognition, 7(1), 365-378.
29. Singh, E. B., & Rani, E. S.(June 2016) INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY PUNJABI CHARACTER RECOGNITION USING WAVELET AND CURVLET TECHNIQUES WITH SIFT ALGORITHM.
30. Sultane Chetan A., Bhalerao Milind V., Bonde Sanjiv V. (2017), Research Article Character Recognition Based On Skeletonization: A Survey", Int. J. Adv. Res. 5(6), 1503-1519, Issn: 2320-5407. 10.21474/IJAR01/4564.
31. Kumar, D., & Gupta, D. (2018). Review on optical character recognition for off-line Devanagari handwritten characters & challenges. Int J Sci Res Comput Sci Eng Inf Technol, 3(3), 1364-1367.
32. Kamble, P. M., & Hegadi, R. S. (2015). Handwritten Marathi character recognition using R-HOG Feature. Procedia Computer Science, 45, 266-274.

33. Yawalkar, P., & Kharat, M. U. (2018). Effective Thinning Algorithm for Recognition of Hand Written Devnagri Compound Characters Using Neural Network. International Journal of Applied Engineering Research, 13(12), 10539-10550
34. Singh, A., & Maring, K. A. (2015). Handwritten Devanagari character recognition using SVM and ANN. International Journal of Advanced Research in Computer and Communication Engineering, 4(8), 123-128.
35. More Vijay, Kharat M U, Gumaste S V(2018) Study of Different Features and Classification Techniques for recognition of Handwritten Devanagari Text, International Journal of Engineering & Technology,7 (4.19) 1055-1059, 10.14419/ijet.v7i4.19.28285.
36. Jinturkar, A. A., & Khanale, P. B.(2017) Problems and Technology Evolution in Handwritten Numeral Recognition of Marathi.
37. Golait, M. S. S., Malik, L., & Thomas, A. (2017). Handwritten Marathi Compound Character Recognition using Combined Discrete Wavelet Transform and Edge Map Features. IJCSIS, 15(10).
38. Malik, L. (2016). Handwritten Marathi compound character segmentation using minutiae detection algorithm. Procedia Computer Science, 87, 18-24.
39. Golait, M. S. S., Malik, L. G., & Thomas,(2017) A. Handwritten Marathi Compound Character Recognition using Structural and Statistical Features.
40. Kadam, A. A., Bhalerao, M. V., & Tanurkar, M. N. (2019). Handwritten Marathi Compound Character Recognition. International Journal of Engineering Research & Technology (IJERT), 8(07), 742-747. (IJET), Volume 08, Issue 07 (July 2019).
41. Shelke, S., & Apte, S. (2011). A multistage handwritten Marathi compound character recognition scheme using neural networks and wavelet features. International Journal of Signal Processing, Image Processing and Pattern Recognition, 4(1), 81-94.