Handwritten Offline Devanagari Compound Character Recognition Using Machine Learning

Juhee Sachdeva and Sonu Mittal

Jaipur National University, Jaipur, India

Abstract

Character Recognition is the most challenging research topic due to its diverse applicable, environment. In the field of pattern recognition, recognition of handwriting is the technique of recognizing handwritten words and characters from the images captured. It is a task of pattern recognition that can be applicable for banking automation systems, postal automation, and various other fields. Devanagari script is a complex script with a huge and complex set of characters including consonants and vowels. Consonants and vowels are joined in various ways to form compound characters. Handwritten Devanagari compound characters have large shape variations which make the task of recognition more complex. In the present article, a model for the recognition of offline Devanagari compound characters using various Machine Learning techniques are discussed. The proposed system preprocesses 5000 handwritten compound character images into 28*28 Pixel images. Feature sets of compound characters are obtained by applying the Edge Histogram feature extraction technique. These feature sets are then applied to various classifiers SVM, SMO, MLP, and Simple Logistic for recognition. We have achieved recognition accuracy of 99.88% with SVM, 99.72% with SMO model, 99.04% with SimpleLogistic model, and 97.7% with the MLP model.

Keywords

Handwritten Character Recognition, Devanagari Compound characters, EDGE HISTOGRAM, SVM, MLP, Confusion Matrix

1. Introduction

Throughout our learning, we human beings develop awareness of language learning, and as we mature, we learn ample document reading ability that can be computer printed or handwritten. It is difficult for computers to decipher this skill that is so easily carried out by humans. To replicate the human learning by computers, several researchers are trying to analyze powerful and successful techniques. Therefore, handwriting identification of characters is one of the promising aspects of pattern recognition science. Artificial Intelligence (AI) is a specialized field that aims to emulate human intellect by computers. An important aspect of AI is to train a computer or software so that it can see, translate, and read the document. This can be accomplished using Optical Character Recognition (OCR) system. OCR system helps machine to imitate human operations like reading and writing text. In OCR technology, handwritten text are converted into computer understandable format. The primary advantage of digitization process is that it is possible to conveniently archive, scan and locate the text. Handwritten forms, Questionnaires, survey forms, banking forms can be digitized using Handwritten OCR system.

Many experiments have been carried out for scripts like English and Chinese, but due to various complexities in character structure and word formation in the Devanagari script, it is very challenging to develop an OCR system for Devanagari script. OCR system of Devanagari script is a complicated process as it has huge character set with large number of vowels and modifiers. In Devanagari word formation is also very complex as characters are joined in various forms known as compound characters, also presence of modifiers also limit the Recognition accuracy. For the proposed work, we aim at developing a systematic approach for the

ACI'21: Workshop on Advances in Computational Intelligence at ISIC 2021, February 25-27, 2021, Delhi, India

EMAIL:juhee.sachdeva@gmail.com (J. Sachdeva); dr.sonumittal@jnujaipur.ac.in(S. Mittal)

recognition of offline handwritten Devanagari Compound characters using various Machine Learning algorithms. For classification problems Machine Learning algorithms are most widely used (Mohanty et al 2021; Jain et al 2021). Machine Learning algorithms are most often used for regression problems. Decision tree (J48), Naïve Bayes, SVM and MLP are few most commonly used Machine Learning algorithm for classification purpose.

2. Literature Review

A database of 27000 Marathi characters is developed and Moment feature extraction techniques are used by Karbhari V. et al. (2013) for handwritten compound Marathi script, they have used classifier MLP and KNN for their research work, accuracy rate of 98.78% is achieved with MLP classifier and accuracy of 95.65% is achieved by KNN classifier 95.65%. They developed a database that contains 9600 basic Marathi and 9000 Compound Marathi characters reported by Karbhari V. et al.(2014), 3000 split characters are also used by them. They have reported accuracy rate of 95.82% using SVM classifier and 95.82% accuracy rate with KNN classifier. 100% recognition rate is reported by Malik and Deshpande (2009) by using regular expressions for printed and handwritten Marathi characters. Shelke et al. (2011) created a dataset of 35000 Handwritten with 70 sample classes that has 30 split and 40 compound characters, they have proposed Marathi script recognition using wavelet features, and reported 94.22% accuracy rate with wavelet features and 96.23% accuracy rate using Modified wavelet features. Ajmire et al. (2015) used SVM and seventh central moment feature extraction technique for Marathi compound characters recognition system. They obtained recognition rate of 93.87%. Kibria et al.(2020) developed a Bangla compound character classifier Model with SVM. They use 3 features namely the Longest Run Feature, Diagonal feature, and Histogram of oriented gradient feature. They show promising results for their research work. Roy et al.(2018) used Deep Learning algorithm for the Bangla script and achieved an accuracy of 90.33%.

3. Devanagari Script and Compound Characters

The third widely used script around the globe is Devanagari after English and Chinese. Devanagari script has total of 44 basic alphabets which comprises of 11 'swaras' and 33 'vyanjanas'. Vowels are also called 'swaras' and consonants are known as 'vyanjanas' in Devanagari script. Swaras can be used as basic character or diacritical marks ('Matra') that are attached to character either above or below, before, or after. Devanagari script also has compound characters that are formed by combining two characters, in certain manner that the first character is converted into its half form is joined with full form of second character. The structure of these words are complicated also known as 'Jodakshre'. Compound Characters can be formed by joining two characters, where both characters can be consonants or a combination of one vowel and one consonant shown in Fig 1. These Jodakshre are often created by attaching the two characters next to each other or attaching one over the other.

म+न= म्न	ल+ल= ल्ल	ब+द= ब्द	त+य= त्य	च+य= च्य	ल+प= ल्प
क+क =क्क	ध+य= ध्य	स+क= स्क	ज+य= ज्य	च+च= च्च	क+ल= क्ल

Fig 1: Combination of Half Consonant and Consonants

It is possible to further divide the Devanagari alphabets into four classes depending on the appearance of a vertical line as seen in Fig 2.

Alphabet Group	Alphabet
Alphabet having a vertical line attached at the right side	ख,घ,च,ज,झ,ञ,त,थ,ध,न, म,ल,स,य,व,ब,भ,ष
Alphabet not attached with vertical line on the right side	ग,ण,श
Alphabet with no vertical line	ड,द,ह,ढ,र,ट,इ,ळ
Alphabet with a vertical line at the middle	क, फ

Fig 2: Classification of Devanagari Characters

4. Proposed Methodology

The proposed model uses the methodology shown in Fig 3. Different Phases of character recognition are as follows:

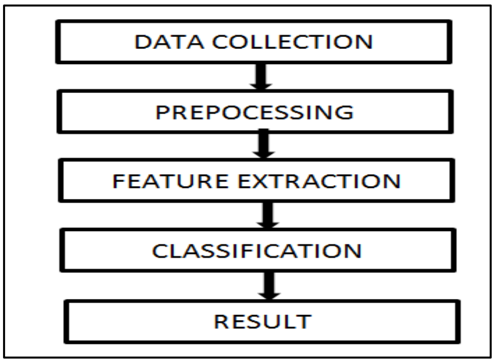


Fig 3: Offline Character Recognition System

4.1 Database Designing

To develop a system for offline handwritten Devanagari compound character classification using Machine Learning a dataset having a large no. of instances is required. The database should have all combinations and variations of compound characters. Some Devanagari databases CEDAR, MNIST, and CENPARMI are available but they don't have a compound character dataset. As no relevant dataset for handwritten Devanagari compound characters is present so we created our dataset for this work. We have used the 50 most commonly used compound characters classes for Database development as shown in Fig 4. These compound characters are selected based on the most frequent occurrence in Devanagari words.

The database is created on A4 size sheets written by different writers of different age groups with different writing styles. All the samples of characters had to be taken in boxes. The sample A4 sheet of handwritten Devanagari compound character from different users is shown in Fig 5. The same procedure is been used for taking all samples from writers, scanning these sheets and then cropping them in fixed dimensions, and finally saving these images in respective class folders in WEKA. The Dataset has 5000 having 50 classes of compound characters written by 100 writers. The Samples collected are scanned by cannon scanner at 500dpi resolution, each instance cropped in 28 pixel width by 28 pixel height and then saved in JPEG file. Sample handwritten compound characters are shown in Fig.5.

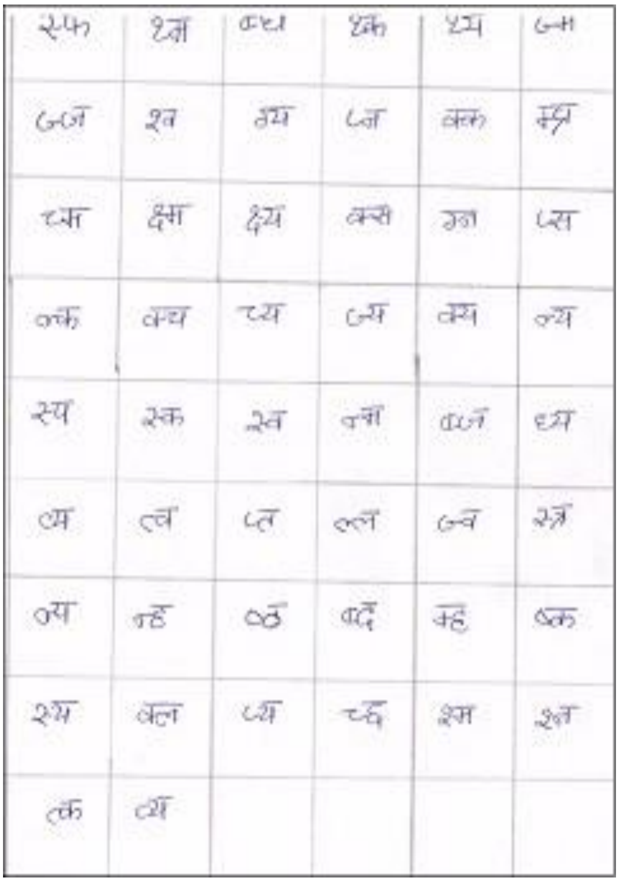


Fig 4: 50 class of Devanagari compound characters

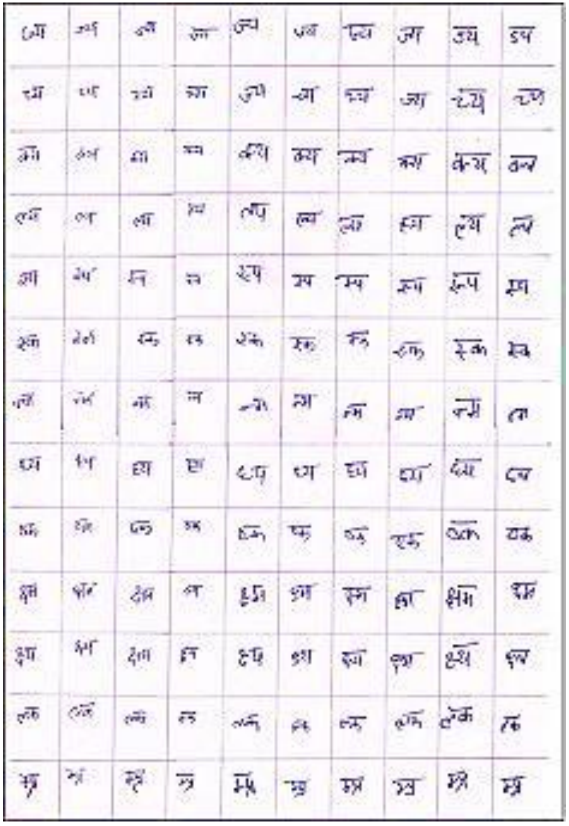


Fig 5: Sample A4 sheet for handwritten compound characters

4.2 Preprocessing

Pre-processing stage of character recognition refers to several operations that are applied on the input images at the initial stage to remove and eliminate to obtain good quality images for further processing. The primary goal of pre-processing is to eliminate all forms of image noise from images and to increase the accuracy of character image. All work has been done in WEKA

- Binarization: Binarization is the process of translation of gray scale image into a Black and White format i.e. 0 or 1 form.
- Noise Elimination: At any point, such as image acquisition, transmitting, or processing, noise may takes place. Noise degrades the quality of the image. For the removal of these noise in images, numerous filtration and thresholding are present.
- Size Normalization: Normalization stores all images in a database of uniform size. Size is normalized without modifying the picture pattern.
- Thinning: To eliminate the chosen foreground pixels from pictures, thinning is applied. Image thinning generates a skeletal structure without losing structural features of characters.

4.3 Feature Extraction Technique

This phase is a crucial step of the Recognition process it extracts various unique features from input images and stores these features in form of a feature extraction vector. These features are unique and non-redundant that convey some unique information about the input image. The accuracy of any Recognition system mostly depends on the accuracy rate of the feature extraction stage.

4.3.1 Edge Histogram Descriptor

The Edge histogram is the widely used extraction technique that calculates the global characteristic of an input character. Edge histogram Normalization leads to scale invariance as the conversion and movement of the image is invariant. Utilizing the following characteristics, the Edge histogram is quite effective in retrieval of images including indexing of character images. The intensity and positional accuracy of the image's intensity are defined by an edge histogram in the image domain. In each local region, the EHD divides 5 kinds of edges called a sub-image. The image domain is divided into 4 by 4 blocks, where no block is overlapping. For each block also known as sub-image , Edge histogram Descriptor is then generated. Edges are classified into five groups of edges: first one is vertical edge, second is horizontal edge, third is diagonal 45° edge, fourth is diagonal 135° edge, and last one is non-directional. So, for any sub-image the occurrence of any of these edges among five types describes the histogram. Therefore as a consequence, every local histogram includes five bins as shown in Fig 6. Every bin corresponds with one of five types of edges. Up to 80 bins are used (won c. et al., 2002) for a number of 16 sub-images. Edge histogram descriptor, was devised by (Park D K. et al., 1997). It incorporates five direction edge features. (Yoon et al., 2001) reported a feature extraction method which applied MPEG-7 edge histogram.

For computing EHD following steps are followed

Step 1: Segment each character image in a block of 4 by 4 subset of images.

Step 2: As seen in Fig 6, Partition each subset of image block into a group of pixels bin.

Step 3: There are five types of edges in MPEG-7. These are 0° for horizontal, 90° for vertical, 45°, 135°, and non-directional edges. Every pixel value is divided into four sub-pixel blocks, Average pixel value is then calculated. The pixel block is recognized as non-direction if the maximum value is less than a given threshold. Table 1 represents the filter coefficient of Edges.

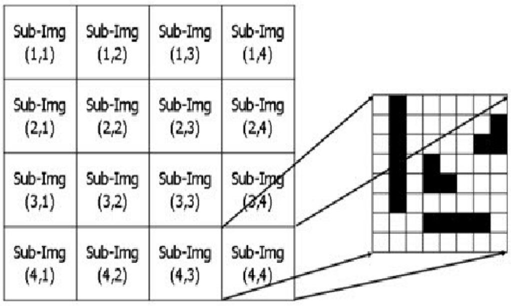


Fig 6: Partition of the image with image block (Su Jung Yoon et al., 2001)

Table 1 EHD Edge Model

Edges	Measure Value
0°	$\delta_{0°}$(i,j)=$\left|\Sigma_{k=0}^{3} a_k(i,j) f_h(k)\right|$
45°	$\delta_{45°}$(i,j)=$\left|\Sigma_{k=0}^{3} a_k(i,j) f_{45}(k)\right|$
90°	$\delta_{90°}$(i,j)=$\left|\Sigma_{k=0}^{3} a_k(i,j) f_v(k)\right|$
135°	$\delta_{135°}$(i,j)=$\left|\Sigma_{k=0}^{3} a_k(i,j) f_{135}(k)\right|$
ND	δ_{nd}(i,j)=$\left|\Sigma_{k=0}^{3} a_k(i,j) f_{nd}(k)\right|$

- **Applying Edge Histogram Descriptor in WEKA**

The Edge Histogram Descriptor has been applied for Devanagari Compound character images in WEKA 3.8 that have EHD feature extraction as Edge Histogram Filter under Preprocess tab. Edge Histogram generated 80 features corresponding to each character class ranging from MPEG-7 Edge Histogram 0 to MPEG-7 Edge Histogram 79. Table 2 represents all Histograms generated using WEKA. Mean value and Standard deviation value calculated for each features are also mentioned in Table 2.

4.4 Classification

Classification is the most important aspect for every pattern recognition scheme. Once the characteristics of each character image is extracted in which all unique characteristics are retrieved and stored in a feature vector. This feature vector work as input for the classification module and labeling to input feature vectors is done in this step. All the characters with similar features are grouped in one class and are considered a member of that class.

4.4.1 SUPPORT VECTOR MACHINE

SVM are Machine Learning techniques that follows supervised learning methods. SVM performs best for Classification and Regression issues developed at AT&T Bell laboratories. SVM gives better results when implemented for pattern recognition task. For high dimensional domain where number of classes to be recognized is higher and number of instances is also large, SVM is successfully implemented.

4.4.2 MLP

Multi-layer Perceptron follows a set of guidance rules for learning. It is a feed-forward layered network of artificial neurons. In a feedforward network information flows in one direction only, which starts through the input node and moves through hidden node followed by the output node. The process by which MLP learns is known as the Backpropagation algorithm.

4.4.3 SIMPLE LOGISTIC

Simple logistic is the WEKA implementation of linear model. Linear model of Regression method follows supervised rules, for which continuous output is calculated. Instead of seeking to identify them to classes, it is used to estimate values within a continuous spectrum. In WEKA simple logistic classifier is available under functions **Simple Logistic.**

Table 2

Statistical Data Generated Using EHD for Devanagari Compound Characters

Features	Min. value	Max value	Mean	Standard Deviation
MPEG-7 Edge Histogram0	0	5	0.659	1.19
MPEG-7 Edge Histogram1	0	6	1.149	1.504
MPEG-7 Edge Histogram2	0	7	0.801	1.55
MPEG-7 Edge Histogram3	0	7	0.568	1.303
MPEG-7 Edge Histogram4	0	7	3.66	3.013
MPEG-7 Edge Histogram5	0	6	1.163	1.535
MPEG-7 Edge Histogram6	0	7	2.462	2.081
MPEG-7 Edge Histogram7	0	7	1.569	2.276
MPEG-7 Edge Histogram8	0	7	1.287	2.054
MPEG-7 Edge Histogram9	0	7	5.528	1.911
MPEG-7 Edge Histogram10	0	6	1.684	1.589
MPEG-7 Edge Histogram11	0	7	3.812	1.671
MPEG-7 Edge Histogram12	0	7	2.052	2.03
……	……	……	……	……
……	……	……	……	……
……	……	……	……	……
MPEG-7 Edge Histogram78	0	7	0.373	1.332
MPEG-7 Edge Histogram79	0	7	1.744	1.97

4.4.4 SMO

SMO technique of Machine Learning Technology applied for computing Quadratic programming(QP) problem which occurs while training phase of Support Vector Machine. SMO algorithm divides large QP problems into smaller set equal QP problem. To obtain optimization, small QP problems are solved analytically. In WEKA, the Sequential Minimal Optimization is available as function **SMO.**

5. Performance Evaluation Strategies

5.1 Confusion Matrix

For a Machine learning problem, the outcome of any model used for classification purpose is evaluated using confusion matrix. It is also Termed an error matrix. All the rows represents predicted class where as all column values defines actual class, In confusion matrix all diagonal values determine the True Positive values that denotes the class is correctly classified. This matrix enables the efficiency to be viewed for a classifier Model. The term confusion derives from the fact that its matrix format makes it easier to visualize how the multiple classes are confused by the model and one class is classified or mislabeled as another class. Confusion Matrix is shown in Table 3.

Table 3

CONFUSION MATRIX

ACTUAL CLASS			
PREDICTED CLASS		**POSITIVE**	**NEGATIVE**
	POSITIVE	TP	FP
	NEGATIVE	FN	TN

Definitions of the terms:

Class1: Positive

Class2: Negative

- TP: Class is correctly classified.
- FN: Class Positive is misclassified as Negative class.
- TN: Negative class is misclassified as negative.
- FP: Negative class is misclassified as positive class.

5.1 Classification Accuracy

Classification Accuracy is calculated as follows:

$$\text{Accuracy}=\frac{No.of\ Correct\ Predictions}{Total\ No.of\ Samples} \quad (1)$$

5.2 Recall

Recall in confusion matrix is calculated as total no. of positive class values divided by total positive class correctly labeled. If a Recall value is high then class is correctly classified.

$$\text{Recall}=\frac{TP}{TP+FN} \quad (2)$$

5.3 Precision

$$\text{Precision}=\frac{TP}{TP+FP} \quad (3)$$

Precision is calculated by dividing the total number of correctly labeled positive class obtained by total no. of class labeled or predicted. High Precision denotes if a sample is labeled as positive is definitely positive.

If almost all positive instance are correctly classified but there are many falsely labeled (FP) positive instance in this case we get High recall and low Precision value. But if many of positive instances are missed but those we labeled or predicted are actually positive then we get Low recall value and high precision value.

6. Applying Classifiers in WEKA

We applied four different classifiers SVM, MLP, SimpleLogistic, SMO for the labeling of handwritten Devanagari compound characters. The tests were carried out on the 5000 instances of handwritten Devanagari characters. A 10-fold validation approach is used. In this method, 10 test subsets were created. For each class used in training phase, it computes the recognition rate. This method divides whole dataset in to groups of 10 equal groups or sets. In each run previous 9 sets are used for training and the remaining 10th set is used for testing. Finally, an average accuracy over 10 runs is obtained. A confusion matrix is generated for all classes.

The experiments were performed with an Edge Histogram filter on all 50 classes and 5000 image samples. The highest average recognition rate 99.88% achieved for a combination of Edge Histogram using SVM classifier shown in Table 5. Edge Histogram used with Simple Logistic classifier achieved 99.04% accuracy rate and Edge Histogram with SMO achieved an accuracy rate of 99.72%. Table 4 shows Error Report for each classifier. The detailed output accuracy of each model is displayed in Table 6. Fig 8 shows the screenshot of the confusion matrix for the SVM model.

7. Result and Discussion

This paper presented a technique for Devanagari compound characters using Edge Histogram Descriptor(EHD) with various classifiers like SVM, MLP, SMO, and Simple Logistic. As no standardized dataset is prepared till date for compound characters thus data is collected from people of all age groups. A database of 5000 samples is created with 50 distinct compound characters collected from 100 different users. The dataset characters are first preprocessed to remove all sorts of noise. A new feature extraction technique i.e. Edge Histogram technique is used for extracting a feature set of handwritten Devanagari compound characters. For this system, four different classifiers are used in combination with the EHD technique. These classifiers SVM, MLP, SMO, and SimpleLogistic when applied with the EHD technique prove to be powerful tools for the recognition system. Feature extraction uses multiple features which gives better recognition accuracy. The SVM classifier proposed in this paper gives the highest recognition rate of 99.88% shown in Fig 7. Recognition rate with SMO, SimpleLogistic and MLP are 99.72%,99.04% and 97.7%. The results of few compound characters achieved with different models are shown in Table7. Fig 8 shows screenshot of confusion matrix obtained from SVM classifier Model.

Table 8 indicates the accuracy rate of the proposed work relative to that of the other prior work published.

Table 4 Error Report for different Classifiers

Statistic	SVM	MLP	Simple Logistic	SMO
Kappa Statistic	0.9988	0.9827	0.9902	0.9971
Mean Absolute Error	0	0.0013	0.0004	0.0384
Root Mean Squared Error	0.0069	0.0215	0.0187	0.1376
Relative Absolute Error	0.1224 %	3.3803 %	1.004%	97.959 5%
Root Relative Squared Error	4.9487 %	15.324 6%	13.3256 %	98.279 3%

Table 5 Accuracy Rate for different Classifiers

Feature Extraction Used	Classifier Used	Accuracy(%)
Edge Histogram	SVM	99.88
	MLP	97.7
	SMO	99.72
	Simple Logistic	99.04

Table 6 Performance Accuracy Report for Edge Histogram with Classifiers

Statistic	SVM	MLP	Simple Logistic	SMO
Accuracy	99.88%	97.7%	99.04%	99.72 %
Error Rate	0.12%	1.7%	0.4667%	0.28%
TP Rate	0.999	0.983	0.995	0.997
FP Rate	0.000	0.000	0.000	0.000
Precision	0.999	0.983	0.995	0.997
Recall	0.999	0.983	0.995	0.997
F-Measure	0.999	0.983	0.995	0.997

Table 7 Recognition Rate in % of Devanagari Compound characters using Different Classifiers

Compound Character	SVM	MLP	SMO	Simple Logistic
ब्ध	99	98	100	100
त्त	100	99	100	92
ल्ल	99	97	98	99
ल्प	100	99	96	100
प्य	99	99	100	99
न्ह	100	99	99	98

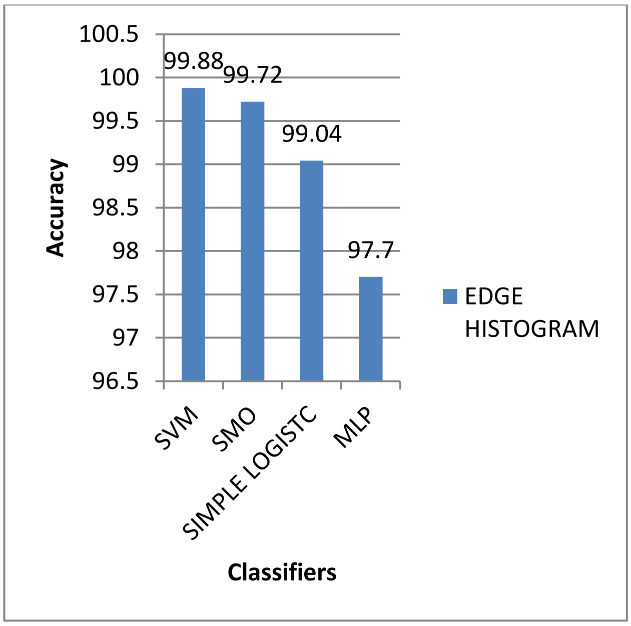
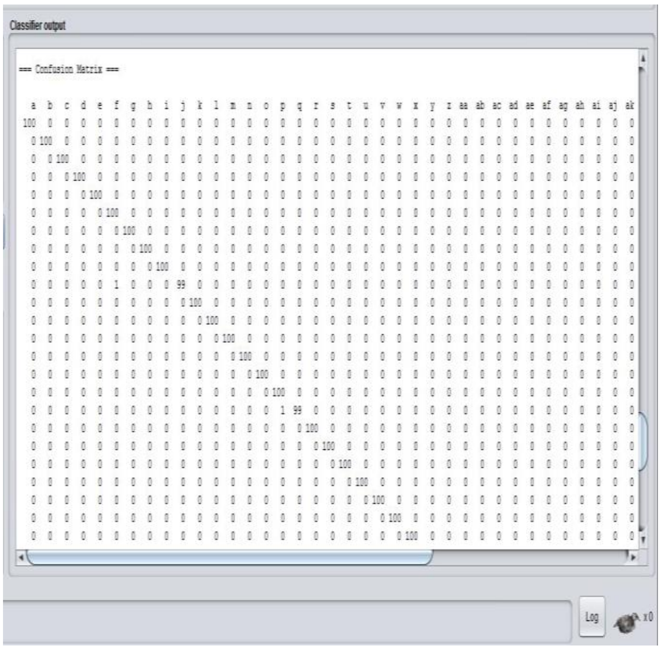


Fig 7: Accuracy Rate of Different Classifiers with Edge Histogram



```
=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   j   k   l   m   n   o   p   q   r   s   t   u   v   w   x   y   z  aa  ab  ac  ad  ae  af  ag  ah  ai  aj  ak
 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   1   0   0   0  99   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1  99   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 100   0   0   0   0   0   0   0   0   0   0   0   0   0
```

Fig 8: Screenshot of a Confusion matrix for Model EdgeHistogram and SVM Model

Table 8 Accuracy Comparisons of Current approach with other approaches in previous research

Author	Script	Feature Extraction	Classifier used	Result (%)
Kale karbhari et al.(2013)	Devanagari	Zernike Moment	SVM& KNN	SVM-98.37 KNN-95.82
Kadam A. et al.(2019)	Marathi	Zoning Feature	SVM & KNN	SVM-96.49 KNN-95.67
Hasan et al.(2019)	Bangla		Deep CNN	98.50
Proposed method	Devanagari	Edge Histogram	SVM, SMO, Simple Logistic and MLP	SVM-99.88, SMO-99.72, Simple Logistic-99.04 MLP-97.7

8. Conclusion

The present article proposes a model for offline handwritten Devanagari compound character recognition. A dataset with 5000 instances of 20 class of compound characters is created where samples are collected from persons of various age groups. This dataset is used by various models of classification for handwritten Devanagari compound characters. A matrix of unique and complex features of characters is created using the Edge Histogram technique. This feature vector is then supplied to four different classifiers SVM, SMO, MLP, and SimpleLogistic for further recognition of compound characters. We obtained 99.88% accuracy for SVM, 99.72% for SMO, 99.04% for SimpleLogistic and 97.7% accuracy for MLP model. In the future, we can apply various other models to achieve a higher accuracy rate. We would like to increase the size of the database and include compound characters with modifiers for further recognition.

9. References:

[1] Singh S., Garg N.K. (2021) 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, vol 1171. Springer, Singapore. https://doi.org/10.1007/978-981-15-5400-1_11

[2] Gonzalez, R. C. (2002). Richard E. woods. Digital image processing, 2, 550-570.

[3] Chaudhuri A., Mandaviya K., Badelia P., Ghosh S.K. (2017) Optical Character Recognition Systems for Hindi Language. In: Optical Character Recognition Systems for Different Languages with Soft Computing. Studies in Fuzziness and Soft Computing, vol 352.Springer,Cham. https://doi.org/10.1007/978-3-319-50252-6_8

[4] H.Bunke P.S.P Wang-"Handbook of character recognition and document image analysis." (1997).

[5] Verma, K., & Sharma, R. K. (2017). Recognition of online handwritten Gurmukhi characters based on zone and stroke identification. Sādhanā, 42(5), 701-712. https://doi.org/10.1007/s12046-017-0632-x

[6] Mukherjee H., Majumder C., Dhar A., Sen S., Obaidullah S.M., Roy K. (2021) A Deep Learning Approach with Line Drawing for Isolated Online Bangla Character Recognition. In: Giri D., Buyya R., Ponnusamy S., De D., Adamatzky A., Abawajy J.H. (eds) Proceedings of the Sixth International Conference on Mathematics and Computing. Advances in Intelligent Systems and Computing, vol 1262. Springer, Singapore. https://doi.org/10.1007/978-981-15-8061-1_16

[7] Vinotheni C., Lakshmana Pandian S., Lakshmi G. (2021) Modified Convolutional Neural

Network of Tamil Character Recognition. In: Tripathy A., Sarkar M., Sahoo J., Li KC., Chinara S. (eds) Advances in Distributed Computing and Machine Learning. Lecture Notes in Networks and Systems, vol 127. Springer,Singapore. https://doi.org/10.1007/978-981-15-4218-3_46

[8] Jain A.A., Arolkar H.A. (2021) A Study of Gujarati Character Recognition. In: Purohit S., Singh Jat D., Poonia R., Kumar S., Hiranwal S. (eds) Proceedings of International Conference on Communication and Computational Technologies. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-15-5077-5_21

[9] Nixon, M., & Aguado, A. (2019). Feature extraction and image processing for computer vision. Academic press.

[10] N. Singh, "An Efficient Approach for Handwritten Devanagari Character Recognition based on Artificial Neural Network," 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2018, pp. 894-897, doi: 10.1109/SPIN.2018.8474282.

[11] Deore, S.P., Pravin, A. (2019). Histogram of oriented gradients based off-line handwritten Devanagari characters recognition using SVM, K-NN and NN classifiers. Revue d'Intelligence Artificielle, Vol. 33, No. 6, pp. 441-446. https://doi.org/10.18280/ria.330606

[12] Shalini Puri, Satya Prakash Singh, An efficient Devanagari character classification in printed and handwritten documents using SVM, Procedia Computer Science, Volume 152,2019, Pages 111-121, ISSN 1877-0509 .

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

[14] M. A. Ansari and M. Dixit, "An enhanced CBIR using HSV quantization, discrete wavelet transform and edge histogram descriptor," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 2017, pp. 1136-1141,doi:10.1109/CCAA.2017.8229967.

[15] Singh S., Garg N.K. (2021) 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, vol 1171. Springer, Singapore. https://doi.org/10.1007/978-981-15-5400-1_11

[16] Paul J., Sarkar A., Das N., Roy K. (2021) HOG and LBP Based Writer Verification. In: Bhattacharjee D., Kole D.K., Dey N., Basu S., Plewczynski D. (eds) Proceedings of International Conference on Frontiers in Computing and Systems. Advances in Intelligent Systems and Computing, vol 1255. Springer,Singapore. https://doi.org/10.1007/978-981-15-7834-2_1

[17] Feng, Q.; Hao, Q.; Chen, Y.; Yi, Y.; Wei, Y.; Dai, J. Hybrid Histogram Descriptor: A Fusion Feature Representation for Image Retrieval. Sensors 2018, 18, 1943. https://doi.org/10.3390/s18061943

[18] N. B. Muppalaneni, "Handwritten Telugu Compound Character Prediction using Convolutional Neural Network," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-4, doi: 10.1109/ic-ETITE47903.2020.349.

[19] Prasad, K., Nigam, D. C., Lakhotiya, A., & Umre, D. (2013). Character recognition using matlab's neural network toolbox. International Journal of u-and e-Service, Science and Technology, 6(1), 13-20.

[20] Pramanik, R., Bag, S. Handwritten Bangla city name word recognition using CNN-based transfer learning and FCN. Neural Comput & Applic (2021).

[21] Khanderao M.S., Ruikar S. (2020) Character Segmentation and Recognition of Indian Devanagari Script. In: Fong S., Dey N., Joshi A. (eds) ICT Analysis and Applications. Lecture Notes in Networks and Systems, vol 93.Springer,Singapore.

[22] S. L. Chandure and V. Inamdar, "Performance analysis of handwritten Devnagari and MODI Character Recognition system," 2016 International Conference on Computing, Analytics and Security Trends (CAST), Pune, 2016,pp.513-516,

[23] Yegnanarayana, B. Artificial neural networks. PHI Learning Pvt. Ltd., 2009.

[24] K. V. Kale, P. D. Deshmukh, S. V. Chavan, M. M. Kazi and Y. S. Rode, "Zernike moment feature extraction for handwritten Devanagari

compound character recognition," 2013 Science and Information Conference, London, UK, 2013, pp. 459-466.

[25] Patel, H. Gujarati Ocr: Compound Character Recognition Using Zernike Moment Feature Extractor.

[26] 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, JPRR Vol 6, No 2 (2011); doi:10.13176/11.300.

[27] Sarika Jain, Ekansh Tiwari, Prasanjit Sardar (Feb 2021), "Soccer Result Prediction Using Deep Learning and Neural Networks", In: J. Hemath et al. (eds.) Intelligent Data Communication Technologies and Internet of Things. Lecture Notes in Data Engineering and Communications Technologies, vol 57, pp. 697-707. Springer Singapore. ISBN: 978-981-15-9508-0.

[28] Sarika Jain, Raushan Kumar Sharma, Vaibhav Aggarwal, Chandan Kumar (Feb 2021), "Human Disease Diagnosis Using Machine Learning", In: J. Hemath et al. (eds.) Intelligent Data Communication Technologies and Internet of Things. Lecture Notes in Data Engineering and Communications Technologies, vol 57, pp.689-696. Springer Singapore.

[29] Narang, S.R., Jindal, M.K., Ahuja, S. et al. On the recognition of Devanagari ancient handwritten characters using SIFT and Gabor features. Soft Comput 24, 17279–17289 (2020).

[30] Pramanik, Rahul; Bag, Soumen: 'Segmentation-based recognition system for handwritten Bangla and Devanagari words using conventional classification and transfer learning', IET Image Processing, 2020, 14, (5), p. 959-972, DOI: 10.1049/iet-ipr.2019.0208

[31] K. V. Kale, P. D. Deshmukh, S. V. Chavan, M. M. Kazi and Y. S. Rode, "Zernike moment feature extraction for handwritten Devanagari compound character recognition," 2013 Science and Information Conference, London, UK, 2013, pp. 459-466.

[32] Ajmire, P. E., Dharaskar, R. V., & Thakare, V. M. (2015). Handwritten Devanagari (Marathi) compound character recognition using seventh central moment. International Journal of Innovative Research in Computer and Communication Engineering, 3(6).

[33] Won, C.S., Park, D.K. and Park, S.-J. (2002), Efficient Use of MPEG-7 Edge Histogram Descriptor. ETRI Journal, 24: 23-30. https://doi.org/10.4218/etrij.02.0102.0103

[34] Su Jung Yoon, Dong Kwon Park, Soo-Jun Park and Chee Sun Won, "Image retrieval using a novel relevance feedback for edge histogram descriptor of MPEG-7," ICCE. International Conference on Consumer Electronics (IEEE Cat. No.01CH37182), Los Angeles, CA, USA, 2001, pp. 354-355.

[35] Saikat Roy, Nibaran Das, Mahantapas Kundu, Mita Nasipuri,Handwritten isolated Bangla compound character recognition: A new benchmark using a novel deep learning approach,Pattern Recognition Letters,Volume 90,2017,Pages15-21,ISSN0167-8655.

[36] Jain, Leena & Agrawal, Prateek. (2017). English to Sanskrit Transliteration: an effective approach to design Natural Language Translation Tool.

[37] Rahul Pramanik, Soumen Bag, Shape decomposition-based handwritten compound character recognition for Bangla OCR, Journal of Visual Communication and Image Representation,Volume 50, 2018,Pages 123-134,ISSN1047-3203,

[38] M. R. Kibria, A. Ahmed, Z. Firdawsi and M. A. Yousuf, "Bangla Compound Character Recognition using Support Vector Machine (SVM) on Advanced Feature Sets," 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 2020, pp. 965-968.

[39] Jameel, Mohd, Mirza Shuja, and Sonu Mittal. "Improved Handwritten Offline Urdu Characters Recognition System Using Machine Learning Technique.

[40] Khobragade R.N., Koli N.A., Lanjewar V.T. (2020) Challenges in Recognition of Online and Off-line Compound Handwritten Characters: A Review. In: Zhang YD., Mandal J., So-In C., Thakur N. (eds) Smart Trends in Computing and Communications. Smart Innovation, Systems and Technologies, vol 165.Springer,Singapore.

[41] Mohanty, S., Chatterjee, J., Jain, S., Elngar, A., & Gupta, P. (2020). *Recommender System with Machine Learning and Artificial Intelligence*. Wiley-Scrivener.