



# Writer identification system for pre-segmented offline handwritten Devanagari characters using $k$ -NN and SVM

Shaveta Dargan<sup>1</sup> · Munish Kumar<sup>1</sup> · Anupam Garg<sup>2</sup> · Kutub Thakur<sup>3</sup>

Published online: 14 November 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

A biometric identification system based on single and multiple modalities has been an evolving concept for solving criminal issues, security and privacy maintenance and for checking the authentication of an individual. The writer identification system is a type of biometric identification in which handwriting of an individual is taken as a biometric identifier. It is a system in which the writer can be identified based on his handwritten text. These systems employ machine learning and pattern recognition algorithms for the generation of a framework. In this paper, the authors have presented a novel system for the writer identification based upon the pre-segmented characters of Devanagari script and also presenting comprehensive state-of-the-art work. The experiment is performed on the corpus consisting of five copies of each character of Devanagari script written by 100 different writers, selected randomly at the public places and consisting of total 24,500 samples of Devanagari characters. Four feature extraction methodologies such as zoning, diagonal, transition and peak extent-based features and classification methods such as  $k$ -NN and linear SVM are used with identification accuracy of 91.53% when using zoning, transition and peak extent-based features with a linear SVM classifier.

**Keywords** Forensic record examination · Writer identification · Devanagari character recognition · Feature extraction ·  $k$ -NN · SVM

## 1 Introduction

Writer identification system is one of the important applications of document analysis and recognition. Writer identification is the process of identifying the author or the writer based on the handwritten text, and the nature of the script may be Indic or non-Indic. Each script has its own specific characteristics and uniqueness that can be explored to devise a unique approach for the target such as the extraction of specific features from of the text samples

written by a person, implementation of classification algorithms and so on. Automatic identification of the writer for the online or offline digitized handwriting sample images has been widely used and important for many applications such as crime investigation, forensic analysis, theft cases and personal identification for wills. It is the part of document analysis and identification which comes under the field of pattern recognition and the machine learning research community. The proposed experiment was implemented on the Devanagari script for the development of the writer identification system for achieving high accuracy rates as compared to the state-of-the-art work. Devanagari is the national language of India and is the third popular language of the world. It has been believed that the Devanagari script was originally developed from the Brahmi script by applying numerous transformations. As there is a close linkage between handwriting recognition and writer identification, it is therefore believed that writing by hand can also be an important source for solving many important issues such as the name of the writer, age, handedness (left or right) and

---

Communicated by V. Loia.

---

✉ Munish Kumar  
munishcse@gmail.com

<sup>1</sup> Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

<sup>2</sup> Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, India

<sup>3</sup> Department of Professional Security Studies, Cyber Security, New Jersey City University, Jersey City, USA

even prediction of the state and the country to which person belongs.

The relationship between the handwriting recognition and the writer identification is helpful for many problems such as:

- Handwriting recognition deals with the ignoring the writer-dependent variations in writing and extracting set of features that can be identified in the different allographs of the same character.
- Writer identification system analyzes the differences of an individual handwriting style with another and takes the benefits of the variations in characters and words for the identification of the writer from a given sample image.

### 1.1 Offline and online modes

Writer identification can be of two types, namely offline writer identification and online writer identification. Offline writer identification is the process of identifying the writer based on a text which was already submitted, scanned and transformed into a computer as a sample image. Offline writer identification is difficult as compared to online because of the number of factors like variations in handwriting style, quality of paper, preprocessing methods and so on. Online writer identification is the process of identifying the writer online, while writing using the electronic devices such as stylus pen and tablet. Working with online data is easy as compared to offline data. A sample of offline and online handwriting is shown in Fig. 1a, b. Here, in the present paper, the authors are presenting a system for offline Devanagari writer identification.

### 1.2 Text-dependent and text-independent systems

Writer identification system is further divided into two categories called text dependent and text independent. In

text-dependent methodology, a writer is required to compose the same text as previously given to the recognition, but in text-independent methodology, any text can be utilized for the identification. And the recognition rate or accuracy can be measured by using a few factors such as number of writers, the type of handwriting, the division of training and testing dataset.

### 1.3 Applications of writer identification system

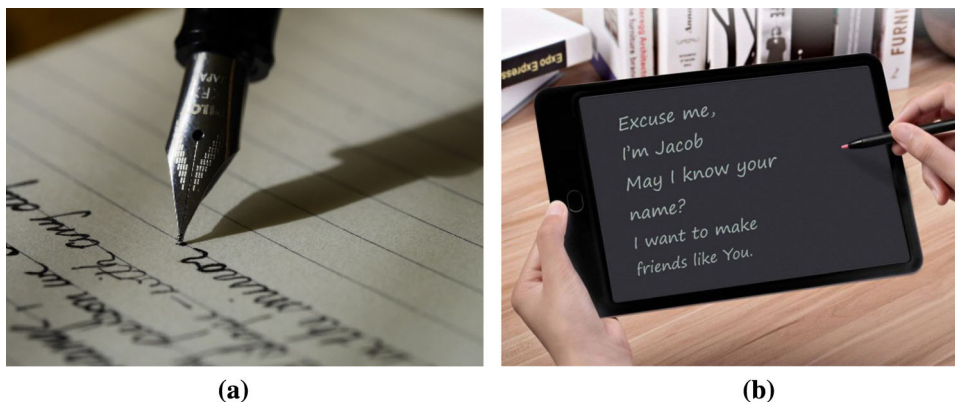
A variety of application areas of writer identification system include:

- Biometric recognition
- Forensic record analysis and investigation
- Authentication of customers in the banking system
- Crime investigations
- Analysis of Wills
- Mobile bank transactions
- Signature verification
- Resolving any kind of suspects.

### 1.4 Layout of the paper

The paper is divided into eight sections. The brief introduction of the sections is shown here, and the detail introductory part of the writer identification system is presented in Sect. 1. Section 2 illustrates the state-of-the-art work on the writer identification system mainly focused on the Indic scripts. Motivations and the various challenges of the proposed work are discussed in Sect. 3. Section 4 consists of fundamentals, characteristics and introduction of Devanagari script. Section 5 portrays the proposed framework of the writer identification system for Devanagari script. Section 6 presents the experimental results and discussions based on the proposed work. Section 7 shows the comparative study of the findings on the writer identification, their feature extraction methods, classification methods and the accuracy achieved with the proposed

**Fig. 1** Handwritten text: **a** offline text, **b** online text



system. Finally, limitations, conclusion and future directions are presented in Sect. 8.

## 2 State-of-the-art work

Sethi and Chatterjee (1977) have presented a handwritten Devanagari character recognition system in 1977 with simple primitives. Presence, absence and positional relationships of these primitives were used to make decisions and reaching conclusions. For training of their model, they considered multilayer perceptron (MLP) and the radial basis function (RBF) networks and error back-propagation algorithms. Pal et al. (2010) proposed a novel method for the text recognition of multi-script recognition in Bangla and Devanagari. By considering the cavity regions and background information, water reservoir and convex hull approaches were applied. From the foreground part, invariant features were extracted for the character. Circular features and convex hull features are integrated with support vector machine as the classifier and the experiment successfully attained the accuracy of 99.18% when tested with 7515 Devanagari characters. For the text document in Bangla script, 98.86% accuracy has been achieved when they collected 7874 Bangla characters. Shaw and Puri (2010) proposed a framework for offline handwritten Devanagari word recognition system. For feature extraction process, they used vertical and horizontal stroke-based features. Hidden Markov model was used for the classification process and the team realized an identification accuracy rate of 81.63%. Siddiqi and Vincent (2010) used the concept of graphemes for characterization of writer's identification. Chain code-based technique was proposed to extract features from the handwriting contours. These features were integrated and validated with the IAM database. They attained an identification accuracy of 77.0% with the IAM database and 79.0% with distribution of chain code. Khanale and Chitnis (2011) proposed a system for character recognition of Devanagari script by inculcating a two-layer feed-forward neural network. For the training of the network, with back-propagation model, they achieved a recognition accuracy of 96.0%. Agnihotri (2012) presented a new method of generation for the chromosome functions and fitness function for offline handwritten Devanagari character recognition. He used diagonal-based method for the feature extraction and genetic algorithm for classification. The corpus consisted of 1000 samples and achieved accuracy of 85.78%. Singh and Lehri (2014) proposed an offline character recognition system for Devanagari script. They used one-dimensional vector arrangements of size  $49 \times 1$  with neural network for the classification, and they successfully got 93.0% accuracy rate.

Halder et al. (2015) proposed a novel technique for writer identification with Devanagari script, with five copies of the character set of Devanagari script by 50 different writers. They computed 64-dimensional feature set which was constructed from the gradient of images. For the classification phase, they used SVM classifier and achieved accuracy of 99.12%. Panda and Tripathy (2015) proposed a new method for offline Odia character recognition. They employed template matching and Unicode mapping for the writer identification process. By using these techniques, they reported the identification accuracy of 97.0%. Sagar and Pandey (2015) proposed an intelligent method for the writer identification system based on the Devanagari script with artificial neural network (ANN). By employing slant estimation techniques, textural features, random, Hough transform and finally with Zernike moments the experiment attained better results. Alwzawy et al. (2016) proposed a deep learning method called CNN for the Arabic handwritten digit recognition. They performed their study on 45,000 samples. Deep CNN was used for the classification and proved to give more accurate results, i.e., 95.7% for the Arabic handwritten digit recognition. Xing and Qiao (2016) proposed a novel framework for writer identification system by employing multi-stream deep CNN. They used local handwritten patches followed by training with soft-max classification loss. They also used multi-stream structure and data augmentation learning for improving the performance. They developed Deep Writer, a deep multi-stream CNN, which was used to perform a deep representation for identifying writers. For the experiment, they employed IAM and HWDB datasets and attained the accuracy of 99.0%. Roy et al. (2017) presented a novel approach with deep belief networks for the compressed delineation of the data. Hidden Markov model (HMM) was used for the word recognition process. For the dataset, RIMES and IFN/ENIT datasets for the Latin and Arabic languages, respectively, are selected and they also conducted the experimental results on Devanagari script. The experimental work proved that the proposed method was performing better than the MLP-HMMs approaches. Andrew et al. (2017) presented a novel technique for writer identification for Telugu script with 150 writers. They extracted directional features and descriptive convolution-based features for recognition. For classifier, they considered nearest neighbor and SVM with linear and RBF kernels. The experiment realized the identification accuracy of 71.0%.

Adak and Chaudhuri (2015) presented a novel method for writer identification which was specially designed for Bangla numerals and characters. They used 193 orthosyllabic numerals and characters of NewISldb:HwC and got improved results for identification as compared to previous studies. Adak et al. (2017) presented writer

identification system on two scripts, namely Bangla and English. Their techniques assumed that writer sometimes has struck-out and cross-out inappropriate words in the document. A corpus of 29,341 words was collected and maintained with the handcrafted feature set. For classification, they used support vector machine (SVM) with radial basis function (RBF) and achieved the accuracy of 79.07% for identification. Dhandra and Vijaylaxmi (2015) presented a novel way to explore recognition of writer of the Kannada script. They used to feature vector that consisted of multi-resolution spatial and directional features based on discrete cosine transformation, random transforms and structural features. Here, features from two or more words were combined, and for the classification,  $k$ -nearest neighbor and fivefold cross-validation technique was used and got satisfactory results. Thendral et al. (2013) proposed a text-dependent writer identification system with the Tamil script and their approach is based upon the local and global features. With tenfold cross-validation method and decision tree for the classification, they attained an identification accuracy of 98.6%. Deng et al. (2017a) presented a novel multi-objective optimization model based on the minimum walking distance of the passengers, idle time variations of gates, number of flights and utilization of gates. This novel method can effectively help in providing a reference for assigning the gates in hub airport and balance the utilization rates of gates by reducing walking distance, improving the services of airport and most important is the satisfaction level of passengers. Deng et al. (2017b) proposed a new collaborative optimization algorithm to improve the local search ability of the genetic algorithms and improving the convergence speed in ant colony optimization (ACO) algorithm that uses multi-population strategy is used for the information exchange and to overcome long search time. The collaborative strategy is used to dynamically balance the global ability and local search ability and to improve the convergence speed. Dargan and Kumar (2018) have presented a state-of-the-artwork for various Indic and non-Indic scripts. This state-of-the-art work gives the cognizance and beneficial assistance to the novice researchers in this field by providing in a nut shell the studies of various feature extraction methods and classification techniques required for writer identification on both Indic and non-Indic scripts. They have also presented a comprehensive survey on deep learning and its application in the different research areas (Dargan and Kumar (2019)). Kumar et al. (2011) have recognized offline handwritten Gurmukhi characters using  $k$ -NN classifier. They have also presented a hierarchical technique for recognition of offline handwritten Gurmukhi characters. Using hierarchical technique, they have reported a recognition accuracy of 91.8% (Kumar et al. 2014). Kumar et al. (2018) have collected 31,500 samples of

Gurmukhi characters from 90 different writers for 35 class problem. They have achieved an identification accuracy of 89.85% and 94.76% by using a combination of zoning, transition and peak extent-based features with linear SVM classifier with partitioning strategy and tenfold cross-validation technique, respectively. Deng et al. (2019) proposed a novel improved ant colony optimization ICMACO method in which they employed co-evolution, multi-population, pheromone updating strategy and pheromone diffusion approach is also used. The goal is to maintain the balance between the solution diversity and the convergence speed. Here, the problem is decomposed into several sub-problems and the ants in the population are divided into elite ants and common ants. By employing traveling salesman problem and actual gate assignment problem, they got better assignment results and improved optimization value. Zhao et al. (2018) proposed a novel identification high-order morphology gradient spectrum entropy method for describing the fault damage degree of bearing and to testify the vibration signal of motors under no-load and load states. The experiment shows improved accuracy rates of fault damage degree and fault prediction of rotating machinery. Zhao et al. (2019) presented a PABSFD method, i.e., a fault diagnostic method based on the principal component analysis (PCA) and broad learning system (BLS). This method can help in eliminating the feature correlation and dimensional reduction in the feature matrix and accurately achieved the fault diagnosis results.

### 3 Motivations and challenges

The motivation for the proposed system, i.e., the development of writer identification framework, originates from the usefulness and continuous need of forensic record analysis, records verification in the banking system, biometric recognition and so on. As handwriting is really an art which is so flexible and dynamic, Fig. 2 shows the variations in handwriting style in Devanagari script with the name of the writer.

There are a variety of challenges that emphasize more interest and excitement of the researchers for the development of proposed system based on the handwritten text. The various challenges are like complex to distinguish the different handwriting styles of individuals, diverse shapes and size of alphabets, the quality of the document, constrained handwriting styles and unconstrained handwritings, etc. The quality of the paper, flexibilities and variations in the font styles is the major challenges that one can face for the identification process in Indic scripts. Working on Indic scripts causes more difficulties as compared to the non-Indic scripts, e.g., poor-quality document consisting of holes, noise, spots, broken strokes, larger



**Fig. 2** Variations in handwriting styles of different writers

1) मेरा नाम श्वेता है ।	Shweta .
2) मेरा नाम संदीप कौर है ।	Sandeep Kaur
3) मेरा नाम अबमोलप्रीत कौर है ।	Akash
4) मेरा नाम गायत्री है	Gayatri
5) मेरा नाम अर्शदीप है।	Arshdeep ka
6) मेरा नाम अर्शदीप है।	Arashdeep kaur
7) मेरा नाम अमनदीप है।	Amndeep
8) मेरा नाम रमनदीप कौर है।	Ramdeep I .
9) मेरा नाम प्रभजोत कौर है।	Prabhjot kaur
10) मेरा नाम कमलजीत कौर है।	Kamajeet kaur
11) मेरा नाम रमनप्रीत कौर है।	Ramanpreet Kaur
12) मेरा नाम गुरप्रीत कौर है।	Gurpreet kaur
13) मेरा नाम गरीमा बेदी है।	Garima Bedi
14) मेरा नाम जसप्रीत कौर है।	Jaspreet kaur
15) मेरा नाम हरविन्द कौर है।	Harvi
16) मेरा नाम रुपिंदर जीत सिंह है।	Rupinder
17) मेरा नाम शीनम है।	Sheenam
18) मेरा नाम संजीव है	Sanjiv
19) मेरा नाम दीपक कुमार है।	Deepak

character set, modifiers and the absence of standard test databases.

#### 4 Devanagari script and data collection

Devanagari script is also known as Nagari, God, Deva which is used to write the Hindi, Sanskrit, Marathi and Nepali language. It was developed from the North India monumental script which was called Gupta. It is written from the left to right and has a strong preference in the symmetrical rounded shapes of characters. The script includes a total family of 47 primary characters out of which, 14 are vowels and 33 are consonants. A character set of Devanagari script is depicted in Fig. 3.

The Devanagari script has been playing a marvelous role for developing the literature. Hindi, Marathi, Nepali and Sanskrit are the languages that mainly use the Devanagari script for writing the text in these languages. This script is written from left to right with no distinct letter cases and is recognizable with the horizontal line along the top of the full letter. In this study, 49 characters of Devanagari script have been considered. Samples of these characters are collected from various public places such as school and college which were taken from the 100 different writers with five samples of each character written by each writer. Hence, a corpus of 24,500 samples for 49-class Devanagari problem has been created.

**Independent vowel signs, anusvāra and visarga**

अ a	आ ā	इ i	ई ī	उ u	ऊ ū
ऋ ṛ	ॠ ṝ	ऌ ḷ	ॡ ḹ		
ए e	ऐ ai	ओ o	औ au		
म् m	अं an̄	ः ḥ	अः aḥ		

**Virama and dependent vowel signs**

क् k	का kā	कि ki	की kī	कु ku	कू kū
कृ kṛ	कृ kāṛ	क्लृ klṛ			
के ke	कै kai	को ko	कौ kau		
कं kaṁ	कः kaḥ				

**Basic consonant signs**

क ka	ख kha	ग ga	घ gha	ङ ṅa
च ca	छ cha	ज ja	झ jha	ञ ña
ट ṭa	ठ ṭha	ड ḍa	ढ ḍha	ण ṇa
त ta	थ tha	द da	ध dha	न na
प pa	फ pha	ब ba	भ bha	म ma
य ya	र ra	ल la	व va	
श śa	ष ṣa	स sa	ह ha	

Fig. 3 Character set of Devanagari script

### 5 Proposed writer identification system

We are presenting here a neoteric system for the text-dependent writer identification that relies on the isolated handwritten Devanagari characters. We are going to develop a robust system with efficient feature extraction phase that performs the extraction of discriminative features followed by the classification phase. The proposed framework consists of various stages such as digitization, preprocessing, feature extraction and classification. These phases are briefly discussed in the following subsections.

#### 5.1 Digitization and preprocessing

Digitization phase deals with the conversion of handwritten character into a digital image. This is done using a scanner with 300 dpi, which is considered as a standard value. While performing writer identification, the first step is to convert the multilevel images into black and white images. In the preprocessing stage, various operations will be applied to a character image. Firstly, the character image is normalized into 64 × 64 size by using the nearest neighbor interpolation method and then images are converted into a bitmap image as [0, 1]. Then, the bitmap image is transformed into a thinned image by using a parallel thinning algorithm (Zhang and Suen 1984).

#### 5.2 Feature extraction

Feature extraction phase is the most important phase of writer identification framework as this phase deals with the extraction of features from the preprocessed characters. It

extracts important features, characteristics and relevant properties about the character and finally classifies the writer with the least measure of time and calculations in the classification phase. Here, four features such as transition features, diagonal features, zoning features and peak extent-based features are used for the development of the writer identification system. These feature extraction methods are briefly discussed in the next subsections.

#### 5.2.1 Zoning features ( $F_1$ )

In this method, decomposition of the thinned image of a character into  $n$  ( $= 100$ ) number of equivalent estimated zones was done. Presently, the number of front area pixels in each zone is determined. The numbers  $p_1, p_2, \dots, p_n$  acquired for all  $n$  zones, are standardized to [0, 1] coming about into a list of capabilities of  $n$  components. So, using this process, 100 features have been collected for an image of a handwritten Devanagari character.

#### 5.2.2 Diagonal features ( $F_2$ )

This method deals with the division of the original thinned image of a character into  $n$  ( $= 100$ ) number of equivalent estimated zones. These features are extracted from the pixels of each zone by moving along its diagonals. Each zone has  $2n - 1$  diagonal and foreground pixels which are present along each diagonal. Then, these are summed up and finally produced a single sub-feature. Using this algorithm, the authors obtained  $n$  features corresponding to each sample.

#### 5.2.3 Transition features ( $F_3$ )

Extraction of transition feature relies on the estimations and area of transitions from background to foreground pixels in the vertical and the horizontal directions. For this, the image is examined from left to right and top to bottom. This approach, as depicted in Fig. 4, this process gives

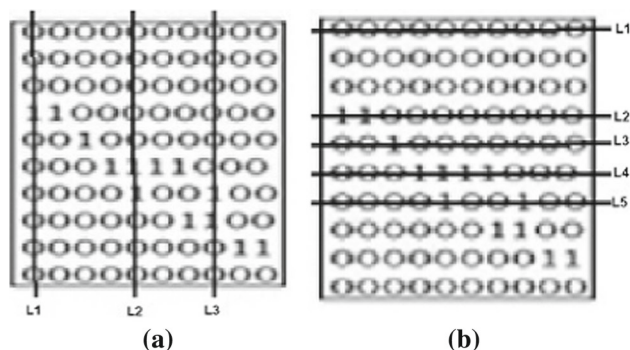


Fig. 4 Transition feature extraction: a transitions in horizontal direction, b transitions in vertical direction

2n features for a character image. Using this process, 200 features have been collected for an image of a handwritten Devanagari character.

### 5.2.4 Peak extent-based features (F<sub>4</sub>)

In this section, authors are implemented presented peak extent-based features on the dataset. They have used these features for offline handwritten Gurmukhi character recognition, and using these features, they have achieved the best recognition accuracy for character recognition of Gurmukhi script. In this technique, features are extracted by considering the sum of the peak extent that fits successive black pixels along each zone. Peak extent-based features can be extracted horizontally and vertically. In the horizontal peak extent features, they considered the sum of the peak extents that fit successive black pixels horizontally in each row of a zone as shown in Fig. 5b, and in vertical peak extent features, they considered the sum of the peak extents that fit successive black pixels vertically in each column of a zone as shown in Fig. 5c.

Using this process, again 200 features (100 features in horizontal peak extent and 100 features in vertical peak extent) have been collected for each sample of handwritten Devanagari character. In our current study, the feature extraction techniques are applied individually, and various combinations of these techniques are also used for identifying the writers based on their Devanagari handwriting styles. Different combinations of these feature extraction vectors are used to improve the identification accuracy as discussed in Sect. 6.

### 5.3 Classification

Classification is the final phase of the writer identification system. Classification phase is a decision-making phase, which is used to identify the writer in view of the features extracted in the previous phase. For classification, the authors have considered two classification techniques, namely *k*-NN and SVM. The authors have computed

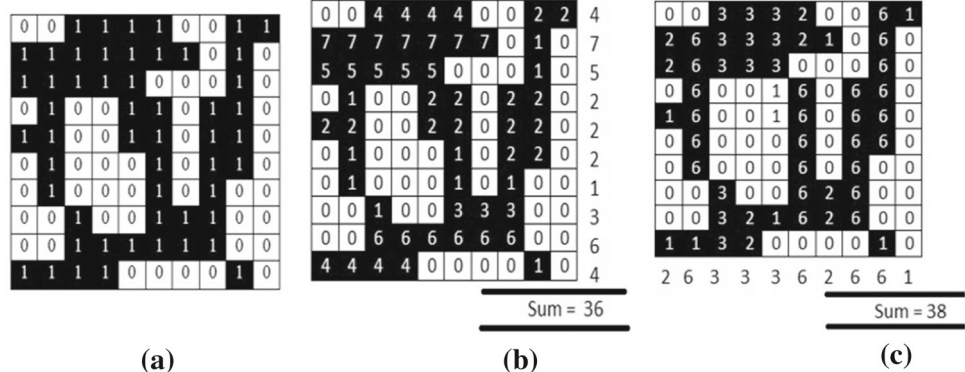
experimental results with neural network and decision tree classifiers also. But the combination of *k*-NN and linear SVM is performing better than the other classifiers, so the authors have presented the experimental results of *k*-NN and linear SVM in this article.

- *k*-nearest neighbor (*k*-NN) is a method for classifying unknown samples in view of neighboring samples in the training feature space. Locations and labels of the training samples are used to divide the space into regions. Usually, Euclidean distance is used to calculate the distance between stored feature vector and candidate feature vector in *k*-nearest neighbor algorithm. In the present work, the value of *k* is considered as *k* = 1.
- SVM is an exceptionally valuable method for data classification. SVM is a learning machine, which has been broadly applied in pattern recognition. It depends upon a measurable learning hypothesis that uses supervised learning. In supervised learning, a machine is trained instead of programmed to perform a given task on a few inputs/outputs' pairs. SVM classifier has also been considered with a linear kernel.

## 6 Experimental results and discussion

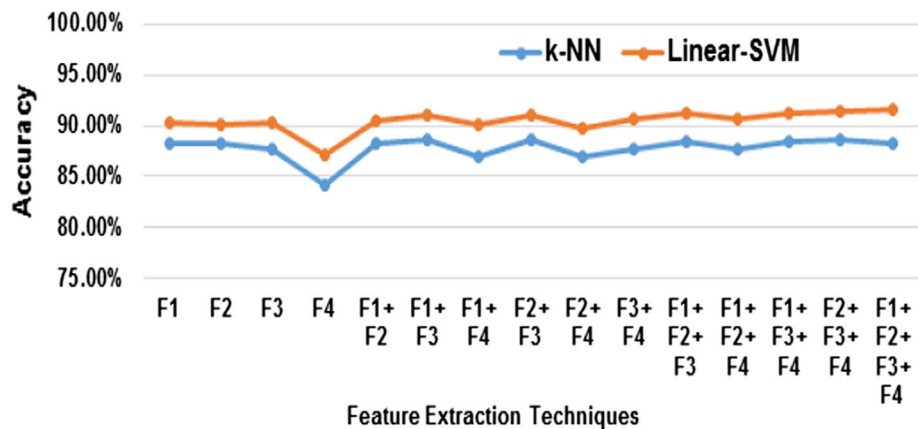
In this section, the authors have presented the experimental results of the proposed writer identification framework in view of the Devanagari characters. The results are produced by using different feature extraction techniques and classification techniques. For experimental results, authors have collected a corpus consisting of total 24,500 samples of segmented characters of Devanagari script collected from 100 different writers. For training dataset, the authors have taken 70% of the total collected data, i.e., 17,150 samples, and for the testing phase, they have taken dataset which is 30% of the total data, i.e., 7350 samples. In feature extraction phase, four feature sets, namely zoning features, diagonal features, transition features and peak extent-based features, tested individually and different

**Fig. 5** Peak extent-based features: **a** zone of bitmap image, **b** horizontally peak extent-based features, **c** vertically peak extent-based features



**Table 1** Feature-wise experimental results using  $k$ -NN and linear SVM classifier

Feature extraction technique	$k$ -NN classifier			Linear SVM classifier		
	Accuracy (%)	TPR (%)	FPR (%)	Accuracy (%)	TPR (%)	FPR (%)
$F_1$	88.24	87.09	0.4	90.23	89.12	0.4
$F_2$	88.20	87.05	0.4	90.15	89.97	0.3
$F_3$	87.65	86.12	0.5	90.23	89.05	0.6
$F_4$	84.09	83.02	0.2	87.13	86.24	0.3
$F_1 + F_2$	88.25	87.19	0.6	90.44	89.26	0.5
$F_1 + F_3$	88.55	86.12	0.4	91.09	89.90	0.4
$F_1 + F_4$	87.01	86.45	0.4	90.03	88.85	0.5
$F_2 + F_3$	88.55	87.42	0.7	91.05	89.86	0.4
$F_2 + F_4$	86.92	85.70	0.3	89.81	88.64	0.7
$F_3 + F_4$	87.76	86.61	0.4	90.74	89.56	0.3
$F_1 + F_2 + F_3$	88.43	87.28	0.8	91.19	90.24	0.2
$F_1 + F_2 + F_4$	87.68	86.94	0.5	90.58	89.28	0.2
$F_1 + F_3 + F_4$	88.45	87.12	0.6	91.23	90.04	0.3
$F_2 + F_3 + F_4$	88.58	87.42	0.4	91.39	90.20	0.2
$F_1 + F_2 + F_3 + F_4$	88.33	87.18	0.7	91.53	90.34	0.4

**Fig. 6** Feature-wise writer identification accuracy using linear SVM classifier

combinations of all these feature sets are considered to improve the accuracy of the system. By using this proposed framework, promising and acceptable results of 91.53% have been achieved by using a combination of zoning, diagonal, transitions and peak extent-based features with a linear SVM classifier, as depicted in Table 1. In Table 1, the authors have presented accuracy, TPR (true positive rate) and FAR (false positive rate) that are computed using Eqs. (1)–(3), respectively. These experimental results are

also shown graphically in Fig. 6. We have also presented writer-wise accuracy for analyzing the handwriting quality of each writer based on their handwriting flow and shapes of characters. Writer-wise recognition accuracy with a combination of zoning, diagonal, transitions and peak extent-based features with linear SVM classifier is depicted in Table 2. Using individual feature set maximum accuracy of 90.23% has been achieved by using zoning features or transition features.

$$\text{Accuracy} = \frac{\text{True Positive (TP)} + \text{True Negative (TN)}}{\text{True Positive (TP)} + \text{True Negative (TN)} + \text{False Positive (FP)} + \text{False Negative (FN)}} \quad (1)$$



**Table 2** Writer-wise recognition accuracy using proposed framework using linear SVM classifier

Writer (W) number	Recognition accuracy (%)	Writer (W) number	Recognition accuracy (%)
W <sub>1</sub>	98.50	W <sub>51</sub>	85.63
W <sub>2</sub>	95.68	W <sub>52</sub>	62.62
W <sub>3</sub>	96.82	W <sub>53</sub>	68.95
W <sub>4</sub>	53.75	W <sub>54</sub>	72.26
W <sub>5</sub>	98.50	W <sub>55</sub>	98.50
W <sub>6</sub>	85.13	W <sub>56</sub>	97.38
W <sub>7</sub>	84.82	W <sub>57</sub>	95.68
W <sub>8</sub>	98.50	W <sub>58</sub>	93.51
W <sub>9</sub>	75.28	W <sub>59</sub>	68.66
W <sub>10</sub>	98.50	W <sub>60</sub>	97.56
W <sub>11</sub>	63.49	W <sub>61</sub>	71.48
W <sub>12</sub>	76.79	W <sub>62</sub>	90.07
W <sub>13</sub>	98.50	W <sub>63</sub>	85.38
W <sub>14</sub>	75.28	W <sub>64</sub>	59.08
W <sub>15</sub>	98.50	W <sub>65</sub>	81.62
W <sub>16</sub>	45.45	W <sub>66</sub>	83.37
W <sub>17</sub>	81.63	W <sub>67</sub>	90.76
W <sub>18</sub>	98.50	W <sub>68</sub>	82.57
W <sub>19</sub>	98.50	W <sub>69</sub>	97.94
W <sub>20</sub>	98.50	W <sub>70</sub>	98.50
W <sub>21</sub>	69.24	W <sub>71</sub>	95.68
W <sub>22</sub>	82.57	W <sub>72</sub>	93.51
W <sub>23</sub>	89.24	W <sub>73</sub>	68.66
W <sub>24</sub>	83.50	W <sub>74</sub>	68.95
W <sub>25</sub>	98.50	W <sub>75</sub>	72.26
W <sub>26</sub>	75.07	W <sub>76</sub>	72.78
W <sub>27</sub>	98.50	W <sub>77</sub>	84.01
W <sub>28</sub>	60.88	W <sub>78</sub>	78.10
W <sub>29</sub>	98.50	W <sub>79</sub>	85.13
W <sub>30</sub>	95.68	W <sub>80</sub>	84.82
W <sub>31</sub>	76.01	W <sub>81</sub>	98.50
W <sub>32</sub>	98.50	W <sub>82</sub>	59.08
W <sub>33</sub>	82.32	W <sub>83</sub>	81.62
W <sub>34</sub>	79.50	W <sub>84</sub>	83.37
W <sub>35</sub>	72.60	W <sub>85</sub>	90.76
W <sub>36</sub>	98.50	W <sub>86</sub>	81.73
W <sub>37</sub>	75.98	W <sub>87</sub>	82.32
W <sub>38</sub>	95.68	W <sub>88</sub>	81.36
W <sub>39</sub>	72.78	W <sub>89</sub>	82.28
W <sub>40</sub>	84.01	W <sub>90</sub>	81.15
W <sub>41</sub>	78.10	W <sub>91</sub>	82.00
W <sub>42</sub>	75.98	W <sub>92</sub>	82.41
W <sub>43</sub>	63.78	W <sub>93</sub>	79.75
W <sub>44</sub>	98.50	W <sub>94</sub>	80.02
W <sub>45</sub>	73.74	W <sub>95</sub>	78.63
W <sub>46</sub>	79.76	W <sub>96</sub>	80.02
W <sub>47</sub>	98.50	W <sub>97</sub>	78.54

**Table 2** (continued)

Writer (W) number	Recognition accuracy (%)	Writer (W) number	Recognition accuracy (%)
W <sub>48</sub>	62.39	W <sub>98</sub>	79.30
W <sub>49</sub>	98.50	W <sub>99</sub>	79.91
W <sub>50</sub>	98.50	W <sub>100</sub>	79.23

$$TPR = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \quad (2)$$

$$FPR = \frac{\text{False Positive (FP)}}{\text{False Positive (FP)} + \text{True Negative (TN)}} \quad (3)$$

### 7 Comparative study with the state-of-the-art work

In this section, the authors have presented a comparative study of proposed work the state-of-the-art work, after getting the output from the proposed study with the huge collection of corpora. The comparative study is shown in Table 3 which presents the parameters such as author name, accuracy rates, feature extraction method and classifier used. Table 3 also depicts that the proposed system performs better than already existing work for writer identification of handwritten characters of Devanagari script.

### 8 Conclusion and future scope

Writer identification system based on the handwritten text is a novel, changing and very useful approach in the field of biometric recognition system. Development of writer identification system for Devanagari script based on handwritten text is such a system in which one can identify the author of a given text by using handwriting as input. The paper comprehensively and systematically presented state-of-the-art work on the writer identification system based on the Devanagari script. Framework consisting of phases such as digitization and preprocessing, feature extraction and then classification techniques are deeply explained. An accuracy rate of 91.53% has been obtained by using the proposed framework with the blending of zoning, diagonal, transitions and peak extent-based features and linear SVM classifier. The advantage of this proposed system is to extend the designed framework for the other Indian scripts like Bengali, Gujarati, Tamil, Gurmukhi and so on. Based on the handwritten text either online or offline, gender classification, handedness (left or

**Table 3** Comparative study of proposed work versus state-of-the-art work

References	Script/ scripts	Number of writers/datasets	Feature extraction methods	Classification methods	Accuracy achieved
Jayanthi and Rajalakshmi (2011)	Tamil	70 writers	Texture analysis, gray-level co-occurrence matrix	Maximum correlation coefficient	82.8%
Akbari et al. (2012)	Persian	50 writers	Wavelet transform and co-occurrence matrix	$k$ -NN	93.3%
Biswas and Das (2012)	Bangla	55 writers	Random transformation	Distance matrix, histogram	92.72%
Chanda et al. (2012)	Oriya	100 writers	Directional chain code and curvature feature	Support vector machine	99.0%
Choksi and Thakkar (2012)	Gujarati	The test set consists of 30 samples of each character with a variety of regular, bold and italic (online)	Structural, geometric and wavelet	$k$ -Nearest neighbor, general regression neural network, fuzzy $k$ -NN	$k$ -Nearest neighbor 67%, general regression neural network 97%, fuzzy $k$ -NN 100%
Halder and Roy (2013)	Bangla	90 writers	400 dimensional	MLP	99.1%
Newell and Griffin (2014)	Roman	300 writers	Oriented basic image feature (OBIF) columns, delta encoding	Euclidean distance	99.0%
Adak and Chaudhuri (2015)	Bangla	212,300 characters	Local binary pattern, pseudo-structural descriptor, index of dissimilarity	Multiple SVM classifiers With RBF kernel	86.6%
Desai (2015)	Gujarati	Trained for 610 digits and tested for 2650 digits	Horizontal, vertical and two diagonal features	Multilayered feed-forward neural network	82.0%
Maadeed et al. (2016)	Arabic	100 writers	Edge-based directional probability distributions	$k$ -Nearest neighbor	90.0%
Yang et al. (2016)	Chinese and Roman	187 writers Chinese 134 writers Roman	Deep writer ID, drop segment	Convolutional neural network	95.72% for Chinese 98.51% for Roman
Adak et al. (2017)	Bengali	29,341 words	Handcrafted feature	Support vector machine with radial basis function	79.1%
Fukue et al. (2017)	Thai	Set of Thai characters	Individual change control processing	Euclidean distance	99.5%
Kalra and Rani (2017)	Gurmukhi	30 writers	Zoning features, open end point and intersection	$k$ -NN and multilayer perceptron model	53.0%
Proposed work (2019)	Devanagari	100 writers (24,500 samples)	Zoning, diagonal, transitions and peak extent-based features	Linear SVM	91.53%

right) prediction, age estimation and verification and nationality prediction can be developed. Various challenges in developing the proposed system are the presence of touching characters in the dataset, overlapping characters, unconstrained databases of writers, poor scanning,

setting of a threshold, noisy data, etc. Deep learning, autoencoders, hybridization of classifiers and use of multiple feature extraction methods are good solutions to be used for the future directions.

## Compliance with ethical standards

**Conflict of interest** The authors have no conflict of interest.

## References

- Adak C, Chaudhuri B (2015). Writer identification from offline isolated Bangla characters and numerals. In: Proceedings of the international conference on document analysis and recognition (ICDAR), vol 15, pp 486–490
- Adak C, Chaudhuri BB, Blumenstein M (2017). Impact of struck-out text on writer identification. In: Proceedings of the international joint conference on neural networks, pp 1465–1473
- Agnihotri VP (2012). Offline handwritten Devanagari script recognition. In: MEC, pp 37–42
- Akbari M, Eslami R, Kashani MH (2012) Offline Persian writer identification based on wavelet analysis. In: Proceedings of the 4th international conference on bioinformatics and biomedical technology, vol 29, pp 180–186
- Alwzawy HA, Albehadili HA, Alwan YS, Islam NE (2016) Handwritten digit recognition using convolutional neural networks. *Int J Innov Res Comput Commun Eng* 4(2):2971–2976
- Andrew C, Reddy S, Pulabaigari V, Pal U (2017) Text independent writer identification for Telugu script using directional filter based features. In: Proceedings of the 14th IAPR international conference on document analysis and recognition, pp 65–70
- Biswas S, Das AK (2012) Writer identification of Bangla handwritings by Radon Transform Projection Profile. In: Proceedings of the 10th IAPR international workshop on document analysis systems, pp 215–219
- Chanda S, Franke K, Pal U (2012) Text independent writer identification for Oriya script. In: Proceedings of the 10th IAPR international workshop on document analysis systems, pp 369–373
- Choksi AH, Thakkar SP (2012) Recognition of similar appearing gujarati characters using fuzzy-KNN algorithm. *Int J Comput Appl* 55(6):12–17
- Dargan S, Kumar M (2018) Writer identification System for Indic and non-Indic scripts: state-of-the-art survey. *Arch Comput Methods Eng* 26:1283–1311
- Dargan S, Kumar M (2019) A survey on deep learning and its applications: a new paradigm to machine learning. *Arch Comput Methods Eng* 9:9. <https://doi.org/10.1007/s11831-019-09344-w>
- Deng W, Zhao H, Yang X, Xiong J, Sun M, Li B (2017a) Study on an improved adaptive PSO algorithm for solving multi-objective gate assignment. *Appl Soft Comput* 59:288–302. <https://doi.org/10.1016/j.asoc.2017.06.004>
- Deng W, Zhao H, Li Zou, Li G, Yang X, Wu D (2017b) A novel collaborative optimization algorithm in solving complex optimization problems. *Soft Comput* 21(15):4387–4398
- Deng W, Xu J, Zhao H (2019) An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem. *IEEE Access* 7:20281–20292
- Desai AA (2015) Support vector machine for identification of handwritten Gujarati alphabets using hybrid feature space. *CSI Trans ICT* 2(4):235–241
- Dhendra BV, Vijaylaxmi MV (2015) A novel approach to text dependent writer identification to Kannada handwriting. In: Proceedings of the 4th international conference on advances in computing, communication and control, vol 49, pp 33–41
- Fukue K, Tomokiyo D, Tangtisanon P (2017) Offline handwriting identification system for Thai characters using individual change control processing. In: Fourteen international conference on electrical engineering/electronics, computer, telecommunications and information technology, pp 123–126
- Halder C, Roy K (2013) Individuality of isolated Bangla numerals. *J Netw Innov Comput* 1:33–42
- Halder C, Thakur K, Phadikar S, Roy K (2015) Writer identification from handwritten Devanagari script. In: Mandal J, Satapathy S, Kumar Sanyal M, Sarkar P, Mukhopadhyay A (eds) *Information systems design and intelligent applications Advances in intelligent systems and computing*, vol 340. Springer, New Delhi, pp 497–505
- Jayanthi SK, Rajalakshmi D (2011) Writer identification for offline Tamil handwriting based on gray-level co-occurrence matrices. In: Proceedings of the international conference on advanced computing (ICoAC), pp 187–192
- Kalra K, Rani S (2017) Writer identification from offline isolated handwritten Gurmukhi characters. *Adv Comput Sci Technol* 10(5):903–914
- Khanale PB, Chitmis SD (2011) Handwritten Devanagari character recognition using artificial neural network. *J Artif Intell* 4:55–62
- Kumar M, Jindal MK, Sharma RK (2011) *k*-nearest neighbor based offline handwritten Gurmukhi character recognition. In: Proceedings of international conference on image information processing, Jaypee University of Information Technology, Waknaghat (Shimla), pp 1–4
- Kumar M, Jindal MK, Sharma RK (2014) A novel hierarchical techniques for offline handwritten Gurmukhi character recognition. *Natl Acad Sci Lett* 37(6):567–572
- Kumar M, Jindal MK, Sharma RK (2018) A novel framework for writer identification based on pre-segmented Gurmukhi characters. *Sadhana* 43(197):1–9
- Maadeed AI, Hassaine A, Bouridane A, Tahir MA (2016) Novel Geometric features for off-line writer identification. *Pattern Anal Appl* 19(3):699–708
- Newell AJ, Griffin LD (2014) Writer identification using oriented basic image features and the delta encoding. *Pattern Recognit* 47(6):2255–2265
- Pal U, Roy PP, Tripathy N, Lladós J (2010) Multi oriented Bangla and Devanagari multi script character recognition. *Pattern Recognit* 43:4124–4136
- Panda SK, Tripathy J (2015) Odia offline typewritten character recognition using template matching with unicode mapping. In: Proceedings of the 2015 international symposium on advanced computing and communication (ISACC), pp 1–7. <https://doi.org/10.1109/isacc.2015.7377325>
- Roy PP, Zhong G, Cheriet M (2017) Tandem hidden Markov models using deep belief networks for offline handwritten recognition. *Frontiers Inf Technol Electr Eng* 18(7):978–988
- Sagar P, Pandey A (2015) An intelligent system for writer identification of Devanagari handwritten documents using ANN. *Int J Res Appl Sci Eng Technol (IJRASET)* 3(12):194–201
- Sethi IK, Chatterjee B (1977) Machine recognition of constrained hand printed Devanagari. *Pattern Recognit* 9(2):69–75
- Shaw B, Parui SK (2010) A two stage recognition scheme for offline handwritten Devanagari words. *Mach Interpret Patterns* 145–165. [https://doi.org/10.1142/9789814299190\\_0008](https://doi.org/10.1142/9789814299190_0008)
- Siddiqi I, Vincent N (2010) Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. *Pattern Recognit* 43(11):3853–3865
- Singh G, Lehri S (2014) Recognition of handwritten Hindi characters using backpropagation neural network. *Int J Comput Sci Inf Technol* 4(2):4892–4895
- Thendral T, Vijaya MS, Karpagavalli S (2013) Supervised learning approach for Tamil writer identity prediction using global and local features. *Int J Res Eng Technol (IJRET)* 2(5):204–208
- Xing L, Qiao Y (2016) DeepWriter: a multi-stream deep CNN for textindependent writer identification. In: Proceedings of the international conference on Frontiers in handwriting recognition, pp 584–589

- Yang W, Jin L, Liu M (2016) DeepWriterID: an end-to-end online text-independent writer identification system. *IEEE Intell Syst* 31:45–53
- Zhang TY, Suen CY (1984) A fast parallel algorithm for thinning digital patterns. *Commun ACM* 27(3):236–239
- Zhao H, Yao R, Xu L, Li G, Deng W (2018) Study on a novel fault damage degree identification method using high-order differential mathematical morphology gradient spectrum entropy. *Entropy* 20(9):682
- Zhao H, Zheng J, Xu J, Deng W (2019) Fault diagnosis method based on principal component analysis and broad learning system. *IEEE Access* 7:99263–99272

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